Comprehensive Review of Deep Unfolding Techniques for Next-Generation Wireless Communication Systems

Sukanya Deka, Kuntal Deka, Nhan Thanh Nguyen, Sanjeev Sharma, Vimal Bhatia, and Nandana Rajatheva

arXiv:2502.05952v3 [eess.SP] 12 Jul 2025

***Abstract*—The massive surge in device connectivity demands higher data rates, increased capacity with low latency and high throughput. Hence, to provide ultra-reliable, low-latency communication with ubiquitous connectivity for Internet-of-Things (IoT) devices, next-generation wireless communication leverages the incorporation of machine learning tools. However, standard data-driven models often need large datasets and lack interpretability. To overcome this, model-driven deep learning (DL) approaches combine domain knowledge with learning to improve accuracy and efficiency. Deep unfolding is a model-driven method that turns iterative algorithms into deep neural network (DNN) layers. It keeps the structure of traditional algorithms while allowing end-to-end learning. This makes deep unfolding both interpretable and effective for solving complex signal processing problems in wireless systems. We first present a brief overview of the general architecture of deep unfolding to provide a solid foundation. We also provide an example to outline the steps involved in unfolding a conventional iterative algorithm. We then explore the application of deep unfolding in key areas, including signal detection, channel estimation, beamforming design, decoding for error-correcting codes, integrated sensing and communication, power allocation, and physical layer security. Each section focuses on a specific task, highlighting its significance in emerging 6G technologies and reviewing recent advancements in deep unfolding-based solutions. Finally, we discuss the challenges associated with developing deep unfolding techniques and propose potential improvements to enhance their applicability across diverse wireless communication scenarios.**

***Index Terms*—Deep learning, Deep unfolding, Model-driven machine learning, Wireless communication signal processing, Integrated sensing and communication (ISAC), Intelligent reflecting surfaces (IRS), Terahertz, IoT.**

I. INTRODUCTION

THE future mobile generations are expected to experience a surge in resource demand due to the exponential growth of connected devices. With the advent of 5G, billions of devices require seamless connectivity, intensifying competition for efficient spectrum allocation [1–3]. Each network has unique requirements and constraints, making resource allocation a complex challenge. Given the inherent limitations of available resources, problem formulation must prioritize flexibility, adaptability, and efficient utilization. Significant advancements have taken place in wireless communication due to various exciting physical layer technologies, including massive multiple-input multiple-output (MIMO), Internet-of-Things (IoT), integrated sensing and communication (ISAC), intelligent reflecting surfaces (IRS), millimeter-wave (mmWave) communication, physical layer security, and next-generation multiple access such as nonorthogonal multiple access (NOMA). Additionally, orthogonal time-frequency space (OTFS) modulation is emerging as a promising alternative to orthogonal frequency-division multiplexing (OFDM), further improving spectral efficiency (SE) and resilience in next-generation networks.

Deep learning (DL), a machine learning subclass, maps the relationship between input and output without explicitly modeling it in the architecture [4]. It involves the use of deep neural network (DNN) with multiple hidden layers (hence "deep") to analyze and interpret data. Numerous architectures have been proposed such as multilayer perceptron (MLP), convolutional neural network (CNN), deep reinforcement learning (DRL), generative adversarial networks (GANs), etc. DNNs have been applied to various components of wireless communication, including detection, channel estimation, beamforming designs, and power allocation [5–9]. However, DL architectures exhibit several inherent limitations. These include complex training procedures and a black-box nature of algorithms. The training process for these models requires extensive datasets, considerable time, and intricate fine-tuning. The complexity of implementing DNN architectures often constrains their applicability in real-world systems. Therefore, DL models are highly sensitive to the specific configurations of wireless systems, such as the number of antennas, users, and operating frequencies, which require retraining or reimplementation whenever these parameters change. These challenges are not typically encountered in conventional optimization techniques, underscoring the potential advantages of using classical algorithms [10, 11].

Communication problems are inherently influenced by dynamic physical constraints that evolve in real-time. These constraints stem from factors such as the physical environment, hardware limitations, and the specific operational requirements of communication systems. Different communication scenarios impose varying demands on bandwidth, power consumption, and latency, which fluctuate with user density, data rates, and service types. Interpreting and addressing these constraints using DNNs presents significant challenges. Data-driven neural network (NN) architectures were developed primarily to process language and vision data, rendering them suboptimal for wireless communication tasks that involve channels, signals, and outputs governed by complex constraints. Although certain constraints can be handled through the application of specific activation functions, for example, employing the $tanh$ function to enforce a constant modulus constraint in analog beamforming, standard activation functions generally lack the versatility to manage the intricate constraints characteristic of wireless communication systems. To mitigate these challenges, model-driven DL approaches facilitate the development of problem-specific network architectures and customized activation functions, optimized for the unique signal processing requirements of wireless communication [12].

Compared to the conventional fully-connected DL architectures, model-driven DL have fewer trainable parameters, implying less training time and computational complexity. Deep

Sukanya Deka and Kuntal Deka are with the Department of EEE, Indian Institute of Technology Guwahati, India. Email: {sukanya.deka, kuntaldeka}@iitg.ac.in.

Nhan Thanh Nguyen and Nandana Rajatheva are with Centre for Wireless Communications, University of Oulu, Finland. Email: {nhan.nguyen, nandana.rajatheva}@oulu.fi.

Sanjeev Sharma is with the Department of Electronics Engineering, Indian Institute of Technology (BHU) Varanasi, India. Email: sanjeev.ece@iitbhu.ac.in.

Vimal Bhatia is with Department of EE, Indian Institute of Technology Indore, India. Email: vbhatia@iiti.ac.in.

unfolding is an approach that falls into the category of model-driven DL [13]. There are several types of model-driven DL approaches, including deep unfolding [14],[15], physics informed NN [16, 17], learned optimizers [18, 19], etc. A deep unfolded network can be trained in an easier and faster fashion with smaller datasets than DNNs. This is because deep unfolding incorporates domain knowledge from iterative algorithms into layers of DNNs, offering improved performance and interpretability [20].

A. Related Works

The growing adoption of deep unfolding in communication systems has led to several surveys covering its diverse applications. Wang *et al.* [21] reviewed DL and deep unfolding methods for physical layer tasks such as MIMO detection, channel estimation, and decoding. The authors of [22] discussed its use in transceiver design and channel state information (CSI) estimation. In [23], deep learning methods including DNNs, GANs, and recurrent NN (RNN) were reviewed, with deep unfolding applied to Bayesian estimation. The authors of [15] focused on MIMO detection, multiuser precoding, and belief propagation. A broader comparison of model-based, DL, and hybrid methods is found in [24]. Monga *et al.* [14] focused on algorithm unrolling for signal/image processing. Additional works [13, 25] explored signal processing and optimization. Eldar *et al.* [12] and Shi *et al.* [26] addressed DL-based communication frameworks. In [27], the authors highlight the advantages of deep unfolding for physical receiver design, with particular emphasis on its applications in signal detection and channel estimation. A comparison between data-driven and unfolding-based beamforming appears in [28] and Table II.

B. Contributions

Existing surveys lack coverage of recent studies addressing critical 6G technologies, such as mmWave spectrum, new air interface that includes massive MIMO, IRS, next-generation multiple access, and are often limited to specific application areas. Consequently, a comprehensive survey is needed to encompass the broad applications of deep unfolding techniques across various aspects of the physical layer in wireless communication. This paper provides an in-depth review of deep unfolding methods applied to the physical layer of wireless communication. The contributions are summarized below:

1) This paper presents a systematic, step-by-step approach for developing deep unfolding algorithms, with a detailed tutorial on the alternating direction method of multipliers (ADMM)-based MIMO detection. The tutorial offers insights into the ADMM unfolding process, identifies trainable parameters, and demonstrates the design of the resulting unfolded network.
2) We survey the application of deep unfolding across various domains, including signal detection, channel estimation, precoder design, ISAC, decoding, power allocation, and physical layer security. Each section highlights the integration of deep unfolding with key 6G technologies such as IRS, mmWave, and IoT, providing readers with guidance on designing frameworks adaptable to future wireless systems.
3) In addition, each section concludes with a comparative analysis table that evaluates prominent unfolded networks across key metrics such as computational complexity, generalization capability, and convergence behavior. These tables highlight the practical trade-offs and deployment boundaries associated with each architecture. Furthermore, we discuss the key takeaways to provide readers with deeper insights into the practical applicability and design considerations of deep unfolding in physical layer communication.
4) This paper concludes the survey by outlining a set of challenges in developing deep unfolding models and propose future directions that could guide advancements in deep

TABLE I: Acronyms and corresponding full names.

Abbreviation	Description
AO	Alternate optimization
AP	Access point
AMP	Approximate message passing
ADC	Analog-to-digital converters
ADMM	Alternating direction method of multipliers
BP	Belief propagation
BS	Base station
BER	Bit error rate
CD	Coordinate descent
CNN	Convolutional neural network
CSI	Channel state information
CGD	Conjugate gradient descent
DL	Deep learning
DNN	Deep neural network
DRL	Deep reinforcement learning
EM	Expectation maximization
EP	Expectation propagation
FDD	Frequency division duplexing
GAN	Generative adversarial network
GNN	Graph neural network
HBF	Hybrid beamforming
IRS	Intelligent reflecting surface
IoT	Internet-of-things
ISI	Inter-symbol interference
ISAC	Integrated sensing and communication
ISTA	Iterative shrinkage-thresholding algorithm
LR	Low-resolution
LS	Least squares
LLR	Log-likelihood ratio
LDPC	Low-density parity check
LMMSE	Linear minimum mean square error
MF	Matched filter
ML	Maximum likelihood
MO	Manifold optimization
MU	Multi-user
MAP	Maximum *a posteriori*
MPD	Message passing detector
MSE	Mean square error
MMSE	Minimum mean square error
MIMO	Multiple input multiple output
MISO	Multiple input single output
MSER	Minimum symbol error rate
mMTC	Massive machine type communication
mmWave	millimeter-wave
NN	Neural network
NOMA	Non-orthogonal multiple access
NMSE	Normalized mean square error
OAMP	Orthogonal approximate message passing
OFDM	Orthogonal frequency division multiplexing
OTFS	Orthogonal time frequency space
PGA	Projected gradient ascent
PGD	Projected gradient descent
QAM	Quadrature amplitude modulation
QoS	Quality of service
QPSK	Quadrature phase shift keying
RF	Radio frequency
RNN	Recurrent neural network
RSMA	Rate splitting multiple access
SD	Sphere decoder
SE	Spectral efficiency
SER	Symbol error rate
SCA	Successive convex approximation
SGD	Stochastic gradient descent
SIC	Successive interference cancellation
SNR	Signal-to-noise ratio
SSCA	Stochastic successive convex approximation
SINR	Signal-to-interference and noise ratio
THz	Terahertz
TPGD	Trainable projected gradient descent
WMMSE	Weighted minimum mean square error
ZF	Zero forcing

TABLE II: Existing survey papers on deep unfolding for communication.

SNo.	Reference	Year	Detection	Channel Estimation	Precoder Design	Sensing and Communication	Decoding in Error Correcting Codes	Power Allocation	Security
1	Deep Learning for Wireless Physical Layer: Opportunities and Challenges [21]	2017	✓	✓			✓		
2	Model-Driven Deep Learning for Physical Layer Communications [22]	2019	✓	✓					
3	Deep Learning in Physical Layer Communications [23]	2019	✓	✓					
4	Deep Unfolding for Communications Systems: A Survey and Some New Directions [15]	2019	✓	✓	✓		✓		
5	Wireless Networks Design in the Era of Deep Learning: Model-Based, AI-Based, or Both? [24]	2019	✓	✓					
6	Algorithm Unrolling: Interpretable, Efficient Deep Learning for Signal and Image Processing [14]	2021	✓	✓					
7	Intelligent Radio Signal Processing: A Survey [29]	2021	✓	✓	✓				
8	Redefining Wireless Communication for 6G: Signal Processing Meets Deep Learning With Deep Unfolding [25]	2021	✓	✓			✓		
9	Model-Based Deep Learning: Key Approaches and Design Guidelines [30]	2021	✓						
10	Model-Based Deep Learning: On the Intersection of Deep Learning and Optimization [13]	2022	✓						
11	Machine Learning and Wireless Communications [12]	2022	✓	✓			✓		
12	Machine Learning for Large-Scale Optimization in 6G Wireless Networks [26]	2023	✓	✓	✓			✓	
13	Artificial Intelligence-Empowered Hybrid Multiple-Input/Multiple-Output Beamforming: Learning to Optimize for High-Throughput Scalable MIMO [28]	2024			✓				
14	Model and Data Dual-Driven: A Promising Receiver Design for 6G Wireless Communications [27]	2025	✓	✓					
15	This Survey		✓	✓	✓	✓	✓	✓	✓

unfolding techniques to enhance their performance.

C. Paper Outline

The skeleton of this paper is depicted in Fig. 1. This survey

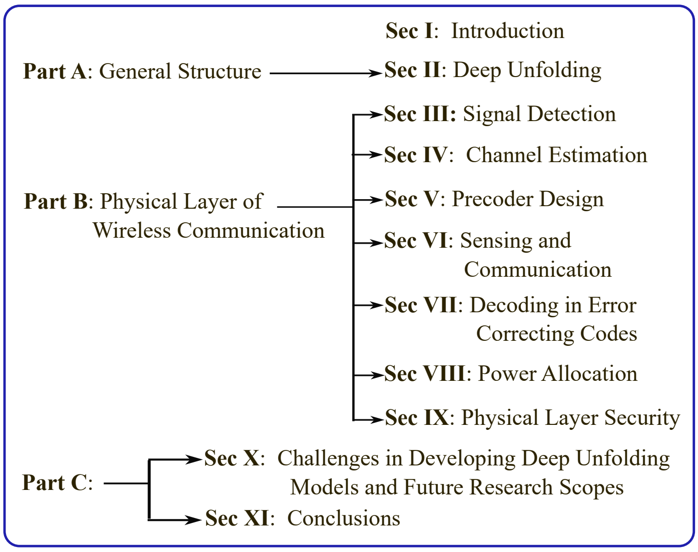

Fig. 1: Outline of the paper.

paper is organized as follows:

- **Part A**: Section II presents the deep unfolding framework, which combines classical iterative algorithms with DNNs. The understanding of deep unfolding is exemplified via ADMM based MIMO detection. A step-by-step approach is presented for developing the unfolding of ADMM specifically for MIMO detection.
- **Part B**: This part explores deep unfolding applications in physical layer tasks: signal detection (Section III), channel estimation (Section IV), precoder design (Section V), ISAC (Section VI), decoding for error correcting code (Section VII), power allocation (Section VIII), and security (Section IX).
- **Part C** : Section X discusses the challenges in adopting the deep unfolding approach in wireless communication and the future scope of research. Section XI concludes the paper.

D. Notation and Acronyms

Lower case, bold-face lower case, and bold-face upper case letters denote scalars, vectors, and matrices respectively. $(\cdot)^T$ and $(\cdot)^H$ denote transpose and Hermitian transpose, respectively. $\|\cdot\|$ denotes the Euclidean norm of a vector. $\mathbb{R}^n$ and $\mathbb{C}^n$ denote n-dimensional real and complex vector spaces, respectively. $\mathcal{N}(\mu,\Sigma)$ denotes the multivariate Gaussian distribution with mean vector μ and covariance matrix Σ, while $\mathcal{CN}(\mu,\Sigma)$ denotes the complex multi-variate Gaussian distribution. Table I lists the main acronyms used in this paper.

II. Deep Unfolding: General Architecture and Fundamentals

Deep unfolding, or algorithm unrolling [14], bridges model-based iterative algorithms and data-driven DNNs [10], [22], [31], [32], [33], [34], [35], [36]. It maps each iteration of an iterative algorithm to a DNN layer, enabling end-to-end learning while retaining interpretability. In deep unfolding, the learnable parameters include model-specific variables and regularization coefficients, optimized layer by layer. Fig. 2 illustrates the process of unfolding an iterative algorithm. In general, for any iterative algorithm, the update equation for each iteration is given as:

$$\mathbf{s}^{l+1} = h(\mathbf{s}^l; \boldsymbol{\theta}^l) \tag{1}$$

where $\{\mathbf{s}^l\}_{l=1}^{L} \in \mathbb{R}^n$ denotes the sequence of variable vector updated across L layers of the algorithm. At each layer, the update is performed by executing $h(\cdot;\cdot)$. Repeated application of this update rule over L layers, yields the unfolded network architecture, $h(\cdot;\boldsymbol{\theta}^l)_{l=1}^{L}$. The trainable parameters, $\left\{\boldsymbol{\theta}^l\right\}_{l=1}^{L}$ (shown in red fonts in Fig. 2) includes the model parameters and the step-size for the regularization coefficients of the algorithm

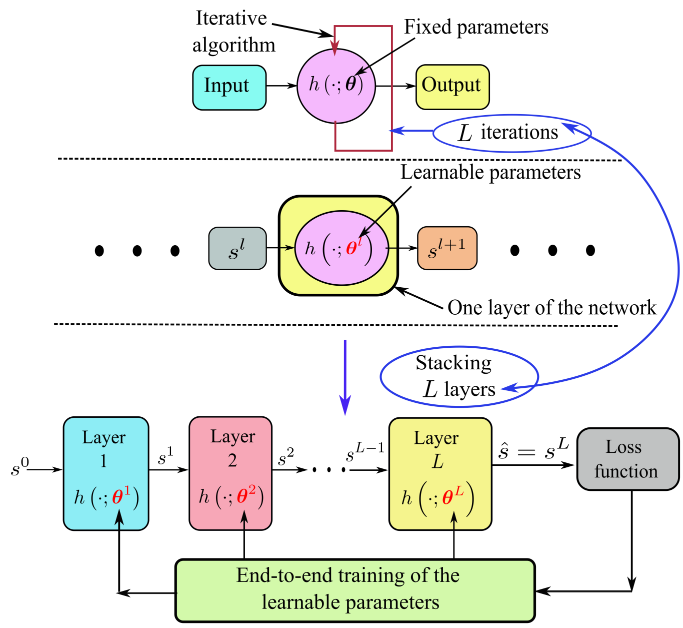

Fig. 2: General framework of training for deep unfolding. The learnable parameters are shown in red color.

[26]. To learn the trainable parameters $\{\boldsymbol{\theta}\}_{l=1}^{L}$ in an end-to-end approach, the loss $\ell(\mathbf{s}^L, \mathbf{s})$ is computed, where $\mathbf{s}^L$ is the output of the network and $\mathbf{s}$ is the ground truth target. The key idea behind end-to-end learning is that all parameters across L layers are simultaneously adjusted through backpropagation, where the loss is minimized using gradient-based optimization. Feeding the data forward through the unfolded network with trainable parameters in the testing stage is similar to executing the parameter-optimized iterative algorithm for a finite number of iterations. Thus, the number of layers in deep unfolding models is fixed, resulting in a predefined computational complexity. Within these fixed layers, the model is trained to optimize the trainable parameters, enabling the problem to be addressed using a structured algorithm with parameters tailored for optimal performance.

Example 1. To develop an understanding on deep unfolding, we take ADMM as an example. ADMM decomposes a complex optimization problem into simpler subproblems, which are solved iteratively. The algorithm combines the principles of dual ascent and the method of multipliers. Each iteration involves solving the subproblems for the primal variables in an alternating manner, followed by updating the dual variables (Lagrange multipliers) to enforce consistency with the original problem's constraints [37]. This iterative process ensures that the subproblem solutions converge to the global solution of the main optimization problem. ADMM in its unfolded version has been widely used in MIMO signal detection [25], [38] and decoding in error correction codes [39, 40]. Here, we show the unfolding of ADMM for MIMO detection as an illustrative example.

Consider a MIMO system with M transmit antennas and N receive antennas. The received signal vector at the receiver can be given by

$$\bar{\mathbf{y}} = \bar{\mathbf{H}}\bar{\mathbf{x}} + \bar{\mathbf{n}}$$

where $\bar{\mathbf{H}} \in \mathbb{C}^{N\times M}$ represents the channel matrix, $\bar{\mathbf{x}} \in \bar{\mathcal{X}}^M$ is the transmitted symbol vector, and $\bar{\mathbf{n}} \in \mathbb{C}^N$ denotes additive white Gaussian noise (AWGN) vector at the receiver. Here, $\bar{\mathcal{X}}$ represents the complex constellation set. In this example, we consider the QPSK modulation scheme, and thus, $\left(\bar{\mathcal{X}} = \{1+i, -1+i, -1-i, 1-i\}\right)$. The maximum likelihood (ML) detection rule is given as

$$\min_{\bar{\mathbf{x}}\in\bar{\mathcal{X}}^{\mathbf{M}}} \left\|\bar{\mathbf{y}} - \bar{\mathbf{H}}\bar{\mathbf{x}}\right\|_2^2 \tag{2}$$

Denote the following real-valued vectors and matrix:

$$\mathbf{y} \triangleq \begin{bmatrix}\Re(\bar{\mathbf{y}})\\ \Im(\bar{\mathbf{y}})\end{bmatrix}, \quad \mathbf{x} \triangleq \begin{bmatrix}\Re(\bar{\mathbf{x}})\\ \Im(\bar{\mathbf{x}})\end{bmatrix}, \quad \mathbf{n} \triangleq \begin{bmatrix}\Re(\bar{\mathbf{n}})\\ \Im(\bar{\mathbf{n}})\end{bmatrix},$$
$$\mathbf{H} \triangleq \begin{bmatrix}\Re(\bar{\mathbf{H}}) & -\Im(\bar{\mathbf{H}})\\ \Im(\bar{\mathbf{H}}) & \Re(\bar{\mathbf{H}})\end{bmatrix},$$

where $\Re(\cdot)$ and $\Im(\cdot)$ denote the real and imaginary part of a complex vector or matrix, respectively. With this transformation, each element of $\mathbf{x}$ is drawn from the discrete space {-1, 1} . We then relax this space to the continuous domain $\mathcal{X} = [-1, 1]$. Consequently, (2) can be relaxed to the following problem in the real domain: $\min_{\mathbf{x}\in\mathcal{X}^{2M}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2$. We further introduce an auxiliary variable $\mathbf{z}$ to reformulate this problem as:

$$\begin{aligned} &\min_{\mathbf{x}\in\mathcal{X}^{2M}} && \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 \\ &\text{subject to} && \mathbf{x} = \mathbf{z}, \mathbf{z} \in \mathcal{X}^{2M}. \end{aligned}$$

The augmented Lagrangian is given as:

$$\mathcal{L}_\rho(\mathbf{x}, \mathbf{z}, \boldsymbol{\lambda}) = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \boldsymbol{\lambda}^T(\mathbf{x} - \mathbf{z}) + \frac{\rho}{2}\|\mathbf{x} - \mathbf{z}\|_2^2 \tag{4}$$

where $\boldsymbol{\lambda} \in \mathbb{R}^{2M}$ is the dual variable (or Lagrange multiplier) and $\rho > 0$ is the penalty parameter. The iterative steps of the ADMM involve alternating minimization of the augmented Lagrangian with respect to the primal variables $\mathbf{x}$ and $\mathbf{z}$, followed by an update of the dual variable, $\boldsymbol{\lambda}$ [41, 42]. The closed-form solutions for $\mathbf{x}$ and $\mathbf{z}$ can be obtained as:

$$\begin{aligned}
\mathbf{x}^{l+1} &= \arg\min_{\mathbf{x}} \mathcal{L}_\rho(\mathbf{x}, \mathbf{z}^l, \boldsymbol{\lambda}^l) \\
&= \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + (\boldsymbol{\lambda}^l)^T(\mathbf{x} - \mathbf{z}^l) + \frac{\rho}{2}\left\|\mathbf{x} - \mathbf{z}^l\right\|_2^2 \\
&= \left(\mathbf{H}^T\mathbf{H} + \rho\mathbf{I}\right)^{-1}\left(\mathbf{H}^T\mathbf{y} + \rho\mathbf{z}^l - \boldsymbol{\lambda}^l\right), && (5) \\
\mathbf{z}^{l+1} &= \arg\min_{\mathbf{z}} \mathcal{L}_\rho(\mathbf{x}^{l+1}, \mathbf{z}, \boldsymbol{\lambda}^l) \\
&= \arg\min_{\mathbf{z}} -(\boldsymbol{\lambda}^l)^T\mathbf{z} + \frac{\rho}{2}\left\|\mathbf{x}^{l+1} - \mathbf{z}\right\|_2^2 \\
&= \arg\min_{\mathbf{z}} \frac{\rho}{2}\left\|\mathbf{z} - (\mathbf{x}^{l+1} + \frac{\boldsymbol{\lambda}^l}{\rho})\right\|_2^2 \\
&= \Pi_{[-1,1]}\left(\mathbf{x}^{l+1} + \frac{\boldsymbol{\lambda}^l}{\rho}\right), \quad \text{and} && (6) \\
\boldsymbol{\lambda}^{l+1} &= \boldsymbol{\lambda}^l + \rho\left(\mathbf{x}^{l+1} - \mathbf{z}^{l+1}\right) \quad \forall l && (7)
\end{aligned}$$

where $\Pi_{[-1,1]}(\boldsymbol{a})$ denotes an element-wise projection of $\boldsymbol{a}$ onto the set $[-1, 1]$. More specifically, if $\boldsymbol{b} = \Pi_{[-1,1]}(\boldsymbol{a})$, we have

$$b_i = \begin{cases} +1, & \text{if } a_i > 1 \\ -1, & \text{if } a_i < 1 \\ a_i, & \text{otherwise.} \end{cases}$$

The detailed procedure for the unfolded ADMM approach is provided in Algorithm 1. In deep unfolding, to ensure gradient flow during backpropagation in the training stage, all operations must be differentiable. To enable differentiable approximation, a soft projection is used that allows gradient to flow while keeping outputs near the desired constellation points. Consequently, (6) becomes:

$$\mathbf{z}^{l+1} = \Pi_{\text{soft}}\left(\mathbf{x}^{l+1} + \frac{\boldsymbol{\lambda}^l}{\rho}\right). \tag{8}$$

Furthermore, the layer-wise unfolding of ADMM for MIMO detection is illustrated in Fig. 3. The received symbol vector

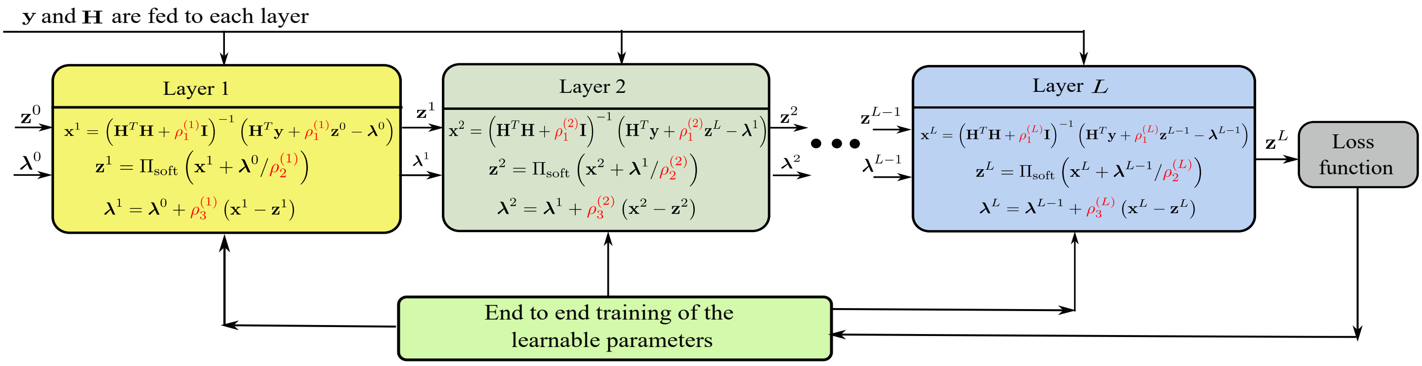

Fig. 3: Deep unfolding training framework for ADMM-based MIMO detection. The learnable parameters are shown in red color.

$\mathbf{y}$ and the channel matrix $\mathbf{H}$ are treated as fixed inputs and are provided to every layer of the network. Each layer of the unfolded network corresponds to one iteration of the ADMM procedure, specifically the update steps in equations (5), (7), and (8). Corresponding to each of these equations, a specific ρ is considered to enable learning. These parameters form the trainable set $\boldsymbol{\theta} = \left\{\rho_1^{(l)}, \rho_2^{(l)}, \rho_3^{(l)}\right\}_{l=1}^{L}$, where each layer l has its own set of penalty coefficients. In the final layer L, the network produces the estimate $\mathbf{z}^L = \hat{\mathbf{z}}$. For $n = 1, \ldots, N$, data samples, the output $\hat{\mathbf{z}}_n(\mathbf{y}_n, \mathbf{H}; \rho_1^{(l)}, \rho_2^{(l)}, \rho_3^{(l)})$ is retrieved from the network as the predicted transmitted symbol for $\mathbf{y}$. The network training loss function formed between the predicted output $\hat{\mathbf{z}}$, and ground truth transmitted symbol $\mathbf{z}_n$ is given as:

$$\ell(\rho_1^{(l)}, \rho_2^{(l)}, \rho_3^{(l)}) = \frac{1}{N}\sum_{n=1}^{N}\left\|\hat{\mathbf{z}}_n(\mathbf{y}_n, \mathbf{H}; \rho_1^{(l)}, \rho_2^{(l)}, \rho_3^{(l)}) - \mathbf{z}_n\right\|_2^2. \tag{9}$$

The network is trained by minimizing the loss with gradient-based learning techniques such as stochastic gradient descent (SGD) to learn $\rho_1^{(l)}$, $\rho_2^{(l)}$ and $\rho_3^{(l)}$.

Algorithm 1 Unfolded ADMM for MIMO detection.

Input: Initial values $\mathbf{z}^{(0)}$, $\boldsymbol{\lambda}^{(0)}$
Given: $\mathbf{y}$ and $\mathbf{H}$
Learnable parameters $\boldsymbol{\theta} = \left\{\rho_1^{(l)}, \rho_2^{(l)}, \rho_3^{(l)}\right\}_{l=1}^{L}$
for $l = 1$ to L **do**
  $\mathbf{x}^{l+1} = \left(\mathbf{H}^T\mathbf{H} + \rho_1^{(l)}\mathbf{I}\right)^{-1}\left(\mathbf{H}^T\mathbf{y} + \rho_1^{(l)}\mathbf{z}^{l-1} - \boldsymbol{\lambda}^l\right)$
  $\mathbf{z}^{l+1} = \Pi_{\text{soft}}\left(\mathbf{x}^{l+1} + \frac{\boldsymbol{\lambda}^{l-1}}{\rho_2^{(l)}}\right)$
  $\boldsymbol{\lambda}^{l+1} = \boldsymbol{\lambda}^l + \rho_3^{(l)}\left(\mathbf{x}^{l+1} - \mathbf{z}^{l+1}\right)$
end for
Output: Final estimate $\hat{\mathbf{z}} = \mathbf{z}^L$

Fig. 4 presents the bit error rate (BER) performance comparison of the unfolded ADMM detector with zero forcing (ZF), minimum mean squared error (MMSE), ADMM and sphere decoder (SD), under QPSK modulation. We consider a 16×16 MIMO system with Rayleigh fading channels. The results demonstrate that the unfolded ADMM detector outperforms ZF, MMSE, and standard ADMM detectors, except for SD. The unfolded ADMM detector performs better due to the use of deep unfolding, where learnable parameters are adaptively trained. This example has been included to provide the readers with an illustrative understanding of the deep unfolding process. The comparative analysis highlights the advantage of the unfolded method over classical approaches. Additional performance improvements can be achieved by optimizing or introducing trainable parameters in the unfolded network.

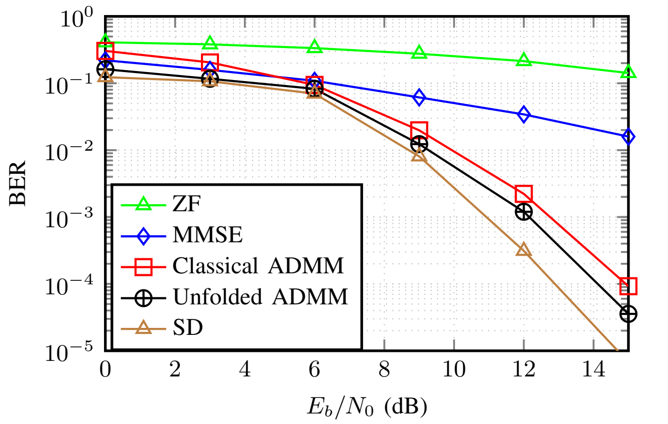

Fig. 4: BER performance comparison for 16×16 MIMO system. For soft projection, *tanh* is used.

In the following sections, we provide a comprehensive survey of deep unfolding algorithms for a variety of critical wireless communication tasks.

III. Signal Detection

This section discusses the application of deep unfolding techniques for signal detection. We segregate these signal detection techniques into various key technologies, as described in the following.

A. Massive MIMO

Massive MIMO uses numerous antennas at both the transmitter and receiver, improving spectral and energy efficiency, but making symbol detection more complex due to inter-symbol interference (ISI) [3]. Designing low-complexity, high-performance detectors is still challenging. MIMO detection can be linear or non-linear. Linear detectors, like matched filter (MF), ZF, and MMSE, involve matrix inversion and can be enhanced with lattice reduction [43]. Non-linear detectors, such as successive interference cancellation (SIC) and parallel interference cancellation (PIC), remove interference in stages [3], [44]. SIC schemes (e.g., MF-SIC, ZF-SIC, MMSE-SIC) use linear detectors to detect symbols, estimate interference, and remove it before detecting the next symbol. SD, a non-linear method, approximates ML detection by searching a bounded hypersphere around the received lattice point, offering a more efficient search than exhaustive methods [45], [46], [47].

The projected gradient descent (PGD) method is a popular optimization-driven approach for MIMO detection. In each iteration, PGD performs gradient descent to minimize detection error, then projects the solution onto a feasible, valid constellation set [48]. One of the first papers to use deep unfolding in detection is the DetNet architecture [49]. To achieve ML optimization, the PGD algorithm is unfolded where the gradient's step size is learned across layers. The authors in [49] claimed that the computational complexity is reduced compared to the SD. DetNet nearly matches SD's performance (within 0.5 order of BER at 18 dB). The computational complexity is linear ($\mathcal{O}(L.K^2)$), where L and K are the number of layers and constellation size, respectively. On the other hand, the SD exhibits ($\mathcal{O}(2^K)$) for worst-case complexity. However, the proposed network is not flexible to changes in the size of the constellation or the number of users. To address MIMO detection under imperfect CSI, He *et al.* introduced model-driven DL networks, orthogonal approximate message passing (OAMP) network, OAMP-Net [10], and its enhanced variant OAMP-Net2 [11]. OAMP-Net2 outperforms OAMP-Net by adapting to varying channel conditions. The OAMP algorithm is unfolded, and the linear MMSE (LMMSE) channel estimator is used to design the receiver in [11]. Learnable parameters, such as the step-size for updating the prior mean and variance in the MMSE estimator, are trained using SGD. These parameters in OAMP-Net2 help compensate for channel estimation errors, improving performance compared to DetNet [49].

As the number of transmit antennas exceeds the number of receive antennas, low complexity detectors are required for such overloaded massive MIMO systems. In [50], the trainable PGD (TPGD) algorithm was introduced for such systems. TPGD uses a DL-aided iterative PGD method with trainable parameters, such as step-size, which are optimized via backpropagation and SGD. TPGD achieves lower computational cost with fewer trainable parameters compared to DetNet [49], OAMP-Net2 [10]. Yun *et al.* in [51] enhanced the unfolded PGD design by adding momentum components to the iterative hidden layers. Momentum helps guide parameter updates in the right direction and reduces oscillations during PGD optimization [52]. This modification improves convergence speed and overall performance, outperforming DetNet [49] and OAMP-Net2 [11]. Furthermore, a trainable detector based on the Hubbard–Stratonovich transformation (to avoid matrix inversion) is discussed in [53]. Compared to the TPGD detector [50] and OAMP-Net2 [11], the network in [53] achieves faster training and lower execution cost. In [54], the authors introduced the MIMO detection network (MMNet), an unfolded detector designed for ill-conditioned channels and non i.i.d. scenarios. MMNet is based on an iterative shrinkage-thresholding algorithm (ISTA), alternating between a linear detector and a nonlinear denoising step. The architectural flexibility and computational complexity are balanced, outperforming DetNet [49] and OAMP-Net [10]. The MMNet outperforms the classical LMMSE detector with a performance gain of approximately $4-6$ dB at a symbol error rate (SER) of 10^{-3} for different channel conditions.

In [55], the CONCRETE MAP method is proposed, based on a CONtinuous relaxation of the probability mass function for disCRETE random variables, as derived in [56]. This approach solves the maximum *a posteriori* (MAP) detection problem using continuous optimization instead of exhaustive search, approximating the MAP solution for any probabilistic nonlinear model. By unfolding the gradient descent algorithm, the detector is able to generate soft outputs, making it suitable for integration with soft-input decoders. The network in [55] outperforms DetNet [49], OAMP-Net2 [11], and MMNet [54], in terms of BER under correlated channels. In [57], conjugate gradient descent (CGD) iterations are unfolded to design the network, LcgNet. The authors also proposed a quantized version, QLcgNet, which minimizes memory usage with negligible performance degradation. LcgNet outperforms DetNet [49] in terms of SER, computational complexity, and memory cost, making it a robust solution for massive MIMO systems. To enhance LcgNet, Olutayo *et al.* [58] incorporated a preconditioner during training, designing a preconditioned LcgNet. Preconditioning improves the condition number of a matrix [59], optimizing the receiver filter's spectrum. This enhancement significantly increases the convergence speed compared to the original LcgNet in [57]. Additionally, the dynamic conjugate gradient network utilizes self-supervised learning with forward error correction codes, effectively adapting to non-stationary, time-varying channels.

Physical layer signal processing involves continuous-time analog waveforms, which must be quantized for hardware implementation. Low-resolution (LR) one-bit analog-to-digital converters (ADC) are commonly used at radio frequency (RF) ends due to their high sampling rates, low cost, and low energy consumption [60]. To recover high-dimensional signals from noisy measurements, Khobahi *et al.* [61] developed an LR detection network (LoRDNet). This network extracts signals from noisy one-bit measurements with unknown CSI. It is based on unfolding the PGD method for symbol detection using the ML estimator. The channel effects are embedded in the training data, so LoRDNet implicitly learns a representation of the channel behavior, eliminating the need for CSI. The network outperforms classical ML estimators [62] and DL networks like DeepSIC [63], even with small datasets of about 500 samples. In [64], the authors proposed a generalized PGD algorithm. The generalized PGD algorithm executes multiple gradient descent steps before applying the projection, allowing for a better balance between convergence speed and detection performance. The authors introduced the self-correcting auto detector (SAD) by unfolding the generalized PGD algorithm. The architecture has two modules: the self-correcting module that executes multiple gradient descent steps to refine the detection estimate, and a denoising autoencoder, a learnable projection module. The numerical results demonstrate that SAD [64] achieves a better trade-off between detection accuracy and complexity compared to DetNet [49]. Although DetNet requires 15 projection steps, SAD only performs 4, significantly reducing detection latency and computational cost.

The channel matrix for a MIMO fading channel can be modeled as a graph. Graph models, such as Bayesian belief networks, Markov random fields, and factor graphs, combine the posterior probability of transmitted data variables and graph theory, establishing a relationship between different nodes for proper detection [3]. The performance of these methods largely depends on the underlying graph structure and convergence properties. Iterative algorithms, like message-passing detectors (MPD) with belief propagation (BP), approximate message passing (AMP), and channel hardening-exploiting message passing (CHEMP), further reduce computational complexity. Expectation propagation (EP), as an extension of BP, enables efficient detection in higher-dimensional MIMO systems and large constellations. EP approximates the posterior distribution of transmitted symbols iteratively, enabling efficient detection in polynomial time [65, 66]. Tan *et al.* [67] unfolded MPD iterations to update messages at symbol nodes using prior and posterior log-likelihood ratio (LLR). Learnable parameters, such as prior probabilities and correction factors (damping, scaling factor, LLR rescaling), are optimized via mini-batch SGD. The authors unfolded the existing classical MPDs: BP and CHEMP, where the correction factors are learned instead of heuristic tuning. A greedy search determines the number of layers. The unfolded networks adapt better to correlated channels and perform consistently in higher-order modulation than classical MPDs. In [68], a low-complexity AMP-graph NN (GNN) model (AMP-GNN) is proposed, which outperforms OAMP-Net2 [11] by avoiding matrix inversion.

Due to the high computational complexity, message-passing algorithms are generally unsuitable for single-carrier frequency division multiple access (FDMA) systems, where a low peak-to-average power ratio is a key design objective for efficient transmission. To address this, the unfolded network in [69] simplifies generalized AMP iterations as discussed in [70]. The proposed network simplified generalized AMP iterations by approximating per-subcarrier channel statistics through statistical averaging. The unfolded network [69] maintains high performance even with channel mismatches and achieves 0.83 dB at BER$= 10^{-3}$ gain over classical generalized AMP [70]. To enhance EP-based detection, the authors in [71] proposed a network, a modified EP MIMO detector (MEPD). In the network, parameter selection is optimized using a modified moment-matching scheme and shows superior performance in high modulation cases compared to OAMP-Net2 [11]. In [72], an approximate EP inspired by [73] is unfolded, achieving better detection performance and robust convergence.

To address high computational complexity of SD, Nguyen *et al.* [74] proposed a fast DL-aided SD algorithm inspired by [75]. The DNN generates reliable initial candidates for SD, prioritizing candidates closer to the DNN's output for testing. This approach accelerates sphere shrinkage, reducing computational cost. The sparse network architecture eliminates matrix inversion, relying only on element-wise matrix multiplications. The proposed method performs better in highly correlated channels than DetNet [49], OAMP-Net2 [11], MMNet [54], and LcgNet [57]. A MIMO detector based on annealed Langevin dynamics is introduced in [76]. It replaces the traditional two-step iterative framework (continuous-space optimization followed by projection) with a single-step approximation of the MAP estimator by sampling from Langevin dynamics. This approach shows superior performance in correlated channels compared to DetNet [49], OAMP-Net2 [11], and MMNet [54]. To mitigate the computational costs of matrix inversion in classical detectors like ZF and MMSE, Berra *et al.* [77] proposed unfolding accelerated Chebyshev successive over-relaxation algorithm. This network requires fewer trainable parameters and features a fast, stable training scheme. The unfolded detector achieves better performance compared to DetNet [49], OAMP-Net2 [11], TPGD-detector [50], MEPD [71], EP-Net [34], and other methods like [32].

To avoid the computational burden incurred due to matrix inversion, the authors in [78] have proposed a sparse refinement architecture for symbol detection in uplink massive MIMO. The authors incorporated deep unfolding for components in matrix inversion, to reduce the detection complexity. The sparse weight matrix associated with the Jacobian iteration is trained across the layers of the network. Liao *et al.* [32] incorporated interference cancellation into DNN layers. Inspired by [79], their sequential detector trains residual error coefficients with mean square error (MSE) loss, reducing complexity by avoiding matrix inversion and relying on matrix addition and multiplication. Compared to DetNet [49], which requires 90 layers for convergence, their model achieves convergence with just 8 layers, offering significant computational simplicity. The authors in [38] developed an unfolded ADMM-based network for MIMO detection. They employed an inexact ADMM update as in [37, 80] to reduce complexity while achieving near-optimal performance. Compared to DetNet [49], the proposed network [38] requires fewer trainable parameters, resulting in improved performance in terms of symbol error. Since, in [38], perfect CSI was assumed for MIMO detection, the authors in [81], proposed an ADMM-based unfolded algorithm with channel estimation errors. A set of penalty and relaxation parameters are learned layer wise. Moreover, the complexity of the proposed network is reduced as compared to OAMP-Net2 [11].

B. Modern Multicarrier Modulation

A cellular network must simultaneously support a growing number of active subscribers within a limited time-frequency spectrum [82]. The current trend in wireless communication is to increase the signal bandwidth to achieve higher transmission rates. However, increasing bandwidth results in shorter symbol periods due to the inverse relationship between bandwidth and symbol period. Consequently, in a multi-path fading channel, the reduced symbol period exacerbates ISI. Multi-carrier modulation techniques like OFDM provide an effective solution to mitigate ISI by dividing a wideband signal into a set of orthogonal narrowband subcarriers [83]. MIMO-OFDM detectors face many challenges due to the complexity introduced by high dimensional signal space. A low complexity receiver for a cyclic prefix-free MIMO-OFDM system is explored in [84]. The receiver unfolds the OAMP algorithm and scalable parameters are introduced for optimizing the channel estimate matrix. The network was tested in over the air experiments and it performed reliably with low synchronization errors under imperfect CSI. Similarly, detection of coarsely quantized or LR signals in one-bit OFDM-based MIMO systems is a challenge. Hence, to address this challenge Shao *et al.* [85] have proposed an unfolded expectation maximization (EM) algorithm. One-bit ADCs severely quantize received signals, discarding amplitude information, and hence, ML detection becomes intractable due to nonlinear quantization effects. Hence, EM iteratively estimates unquantized symbols by computing its conditional expectation and optimizes the parameters in the maximization step. At high SNR, unfolded EM [85] performs better than ZF for one bit MIMO-OFDM detection. Furthermore, by unfolding the likelihood ascent search algorithm, Ullah *et al.* [86] developed an unfolded network to jointly detect and estimate the soft output in MIMO-OFDM systems. The likelihood ascent search algorithm is a neighborhood search-based detector for the MIMO system. The algorithm begins with an initial symbol estimate from linear detectors (ZF or MMSE) and iteratively refines the solution by flipping bits in the detected symbols to minimize the likelihood cost function. Near-optimal detection is achieved with the unfolded network, 1.2 dB gap with 16-QAM as compared to SD at BER of 10^{-5} at polynomial complexity. Zhang *et al.* developed an unfolding-based framework for MIMO-OFDM detectors in [87]. The EP is unfolded with layer-specific damping factors that are trained to adjust effectively to varying channel conditions. The proposed network achieved near-optimal detection across a wide range of channels and noise levels and demonstrated robustness in over-the-air experiment tests. Underwater acoustic communication is widely used for long-range transmissions, reaching up to tens of kilometers [88–90]. Sound waves are used to transmit data through water since radio and light waves attenuate rapidly underwater. Channel estimation in underwater acoustic communication is particularly challenging due to doubly-selective channels that exhibit both severe time variation (from Doppler effects) and frequency selectivity (from long multipath delays). Hence, for OFDM systems in complex doubly-selective underwater acoustic channels, Zhao *et al.* in [35] proposed an unfolded equalizer (UDNet). The unfolded algorithm is based on unfolding the classical MMSE equalizer. In terms of performance, UDNet achieved better generalization and maintained lower computational complexity than DetNet [49]. Li *et al.* [34] proposed an unfolded EP network (EP-Net) for OTFS detection. The trainable parameters include the step-size,

TABLE III: Summary of typical deep unfolding models for signal detection and their advantages and disadvantages.

Unfolding Model	Main Novelties and Performance Gains	Advantages and Disadvantages
DetNet [49]	• First deep unfolding model for MIMO signal detection based on unrolling PGD method. • In a 60×30 MIMO system, DetNet achieves 3 dB SNR gain over ZF for BER $= 10^{-3}$, performs comparable to SDR and AMP, and approaches SD.	• **Advantages:** Explainable (follows PGD optimization flow), performs well when $N_r \gg N_t$. • **Disadvantages:** Complex architecture (high computational cost), poor generalization under correlated channels.
LcgNet [57]	• Unfolded CGD for massive MIMO detection. • Achieves 1.52 dB SNR gain over LMMSE in 32×128 MIMO at BER $= 10^{-4}$. Outperforms DetNet with lower complexity and faster training.	• **Advantages:** Fast training (practical for low-end CPUs). • **Disadvantages:** Trained on Rayleigh fading, suffers significant BER degradation when tested on a spatially correlated channel.
OAMP-Net2 [11]	• Lightweight unfolding of OAMP with soft input/output. • Outperforms DetNet and AMP in 4×4 and 8×8 MIMO system, robust under imperfect CSI.	• **Advantages:** Fast training (4 parameters). Supports turbo equalization (soft I/O). • **Disadvantages:** Underperforms in 3GPP channels (assumes unitarily-invariant matrix).
MMNet [54]	• Online-trained unfolded ISTA for 3GPP channels. • Achieves 8 dB SNR gain over MMSE for SER $= 10^{-3}$ (and 2.5 dB over OAMP-Net2 [11] under imperfect CSI).	• **Advantages:** Robust to correlated channels, high-order QAM, and estimation errors. • **Disadvantages:** Online training adds latency (unsuitable for fast-varying channels).
ADMM-Net [38]	• Inexact ADMM-based unfolding with soft $tanh$ demodulation (supports $16/64$ QAM). • Achieves SER within 1.2 dB of SD in 30×60 MIMO, outperforms DetNet with lower complexity.	• **Advantages:** Robust to channel correlation and imperfect CSI. • **Disadvantages:** More layers (5–20) improve accuracy but increase computation, hence a tradeoff between performance and runtime.
LoRDNet [61]	• Unfolding PGD for 1-bit ADC receivers without CSI. • Outperforms classical ML detectors with just 500 samples.	• **Advantages:** Works with low-resolution ADCs; data-efficient. • **Disadvantages:** Periodic retraining required for block fading channels (reduced SE due to pilot overhead) .

damping factor, and offset values for the mean and variance of the posterior conditional probability distribution. To enhance convergence and prevent the network from being trapped in local optima, the Adam optimizer with warm restarts is employed. In [91], an unfolded OAMP-based detector for OTFS is designed to handle channel estimation errors. This detector is trained using imperfect CSI and optimizes three trainable parameters: step-sizes for updating the linear estimate, nonlinear estimate, and noise variance. Numerical analysis demonstrates that at SNR of 30 dB, the proposed detector achieves a BER of 2.4×10^{-4} while EP-Net [34] achieves BER of 1.2×10^{-3}.

C. Next Generation Multiple Access

Multiple access techniques allow multiple users to share the same communication channel and resources, such as time, frequency, or code, to transmit data efficiently. Usually multiple access schemes achieve this by dividing signaling dimensions orthogonally across time, frequency, or code [92]. This orthogonality enables low-complexity multi-user detection at the receiver. However, orthogonal multiple access cannot achieve the desired sum-rate capacity for a multiuser wireless system. To address this, NOMA has been introduced, allowing more users to be served than the number of orthogonal resource units by leveraging non-orthogonal resource sharing [82]. NOMA is categorized into power-domain NOMA and code-domain NOMA [93]. Power domain NOMA multiplexes users by assigning different power levels on the same resource, such as time, frequency, or code, and relies on SIC at the receiver. Whereas, code domain NOMA multiplexes users by assigning different non-orthogonal spreading codes on the same resource block.

A deep unfolded detector, named the learned preconditioned CGD network with SIC, is proposed for downlink MIMO-NOMA systems in [94]. The algorithm incorporates a data transfer learning strategy to achieve a robust model with low data set requirements. This approach efficiently adapts the model to variations in environmental factors, such as channel models, modulation schemes, or power allocation, ensuring consistent performance in diverse scenarios. Tanner-graph-based transmission schemes for NOMA are discussed in [95]. To enhance detection efficiency, a message-passing detector is unfolded, allowing the detector to improve as more information about transmitted data is learned during training epochs.

In sparse code division multiple access (CDMA), each user transmits using sparse signature codes to reduce interference and enable efficient multiuser access [96, 97]. Takabe *et al.* proposed the complex sparse TPGD (C-STPG) detector in [98], achieving superior detection over LMMSE with low complexity and minimal training due to few trainable parameters. They also introduced gradual sparsification, a deep unfolding method that jointly learns sparse signature matrices by minimizing MSE, tuning step size and a softness parameter.

D. Discussions and Takeaways

Deep unfolding has proven to be a compelling solution for signal detection in complex and high-dimensional problems, such as those in massive MIMO, low-resolution ADC, and time-varying systems. By transforming iterative algorithms such as gradient descent variants, ISTA, ADMM, and AMP into structured networks, unfolding approaches bridge interpretability and performance. Table III summarizes the novelties, performance gains, advantages and disadvantages of the most popular deep unfolding models for signal detection.

While pioneering unfolding architectures such as DetNet [49] and LcgNet [57] have demonstrated the advantages of data-driven signal detection with explainable DNN structures, they exhibit limited generalization in challenging scenarios, such as correlated channels. More recent unfolding-based approaches, including MMNet [54] and ADMM-Net [38], address these limitations by offering robustness to both channel correlation and imperfect CSI. However, methods like MMNet [54] and LoRD-Net [61] introduce additional latency and training complexity due to their reliance on online training and periodic retraining, making them less suitable for fast-varying channel environments.

The research community has actively tested deep unfolded detectors in real-world scenarios. Over-the-air test platforms [35, 84, 87] have been employed to validate the performance of unfolded detector networks under practical conditions. These real-world experiments confirm the feasibility of model-driven deep unfolding networks that are trained offline and subsequently deployed on software-defined radios or software-based receivers for real-time inference. The results demonstrate that deep unfolding techniques in signal detection can achieve robust BER performance in practical indoor environments, outperforming both traditional and purely data-driven methods.

IV. Channel Estimation

Channel estimation is essential for signal detection because it allows the receiver to compensate for distortions introduced by the transmission medium such as fading, delay, and phase shifts. In the following, we discuss deep unfolding aided channel estimation methods.

A. Massive MIMO

MMSE-based channel estimation is commonly used in MIMO systems to estimate the channel response in noisy environments. In massive MIMO systems with frequency division duplexing (FDD), reducing CSI feedback overhead is challenging. A method in [99] combines DL and superposition coding to address this. Downlink CSI is encoded, superimposed on uplink user data, and sent to the BS. A multi-task NN unfolds the MMSE process to reduce interference and reconstruct the CSI efficiently.

ISTA is commonly used for sparse channel estimation, enabling efficient CSI feedback. Each ISTA iteration includes a linear operation followed by a nonlinear soft-thresholding step, similar to the rectified linear unit (ReLU) activation function [20]. In [100], Chen *et al.* proposed a trainable DL architecture for sparse recovery problems using an adaptive depth approach. Instead of a fixed number of layers, the method dynamically adjusts the depth based on the problem's complexity. Highly sparse channels use fewer layers, while more complex channels require additional layers. This adaptive approach improves accuracy and lowers computational costs compared to classical methods and fixed-depth networks. Results in [100] demonstrate that the method reduces reconstruction error and speeds up channel estimation, making it more efficient. A scenario-adaptive CSI compression framework is proposed in [101], requiring minimal channel measurements for training. The authors introduced a CSI feedback architecture designed for efficient encoder updates in dynamic wireless environments, which is essential for downlink precoding in massive MIMO systems. The encoding network focuses on improving the robustness and efficiency of CSI compression. It employs a deep unfolding-based feedback network to iteratively refine the feedback process and enhance feature extraction. Similarly, the decoding network uses a deep unfolding structure inspired by ISTA. This iterative approach improves CSI recovery accuracy, making the process more efficient and adaptable to changing conditions.

In massive MIMO systems, where the channel matrix is often sparse, the majorization-minimization (MM) approach is effective for addressing sparse channel estimation [102]. In [103], the authors address the challenge of downlink channel estimation in massive MIMO. A two-timescale joint uplink/downlink dictionary learning and channel estimation algorithm is proposed. The short-term subproblems focuses on estimating sparse channel vectors for each timeslot. The MM algorithm is unfolded, and the sparse channel vectors are treated as trainable parameters. In contrast, the long-term subproblem involves updating the dictionary to adapt to dynamic environments for which a constrained stochastic successive convex approximation (SSCA) framework is employed without deep unfolding. Simulations results in [103] demonstrate that the proposed approach reduces pilot overhead and outperforms existing methods in both static and dynamic environments, particularly in environments with mobile users.

The matching pursuit algorithm is a greedy iterative method often applied in sparse channel scenarios like massive MIMO. It identifies channel components (basis functions) by matching them to the residual, which is the difference between the observed signal and the signal generated by the current channel estimate. Yassine *et al.* [104] introduced a deep unfolded algorithm, called mpNet, based on the classical matching pursuit algorithm. This network estimates the multiuser massive MIMO channel in an unsupervised manner, initializing with a dictionary of imperfect steering vectors. The system operates in time division duplex mode to leverage channel reciprocity. The network dynamically adjusts the number of layers based on incoming data, ensuring efficient performance in realistic environments.

The basis pursuit algorithm is a widely used method for sparse signal recovery, particularly in compressed sensing. It identifies the sparsest solution to an underdetermined system by minimizing the ℓ_1-norm of the signal. This approach is crucial in scenarios where data is acquired under constraints, and the signal is assumed to be sparse in certain domains. In [105], the authors observed that existing sparse channel estimation schemes for downlink CSI often deliver suboptimal reconstruction due to the use of random measurement matrices. To address this, they proposed deep unfolding-based autoencoders, that optimize the measurement matrix by unfolding the basis pursuit algorithm. This customized design enhances reconstruction performance in sparse channel estimation tasks.

Sparse Bayesian learning is another effective approach for channel recovery [106]. It estimates the non-zero elements of the channel matrix while utilizing prior knowledge of sparsity. Hu *et al.* [107] introduced a deep unfolding framework called deep deterministic policy gradient (DDPG) for MU-MIMO systems. DDPG, a variant of deep Q-networks, is designed to handle problems in continuous action spaces [108]. Unlike conventional deep unfolding NNs with fixed depth, the authors leveraged DDPG to dynamically adjust the network depth based on the complexity of the problem. This adaptive design enhances the efficiency and performance of sparse channel recovery in varying scenarios. The channel estimation problem is formulated using an off-grid sparse Bayesian learning model. The sparse Bayesian learning algorithm is unfolded with trainable parameters grid points and off-grid variables. Simulation results show that the proposed method significantly outperforms fixed-depth models, achieving a 30% reduction in the number of layers while improving normalized MSE (NMSE) performance.

B. mmWave Massive MIMO

mmWave massive MIMO operates in the high-frequency range of 30 GHz to 300 GHz, offering abundant spectral resources to alleviate the bandwidth crunch at sub-6 GHz frequencies [82]. However, mmWave signals suffer from severe propagation and penetration losses. To compensate for these losses, large-scale antenna arrays with tens to hundreds of elements are required to achieve high power gains. The short wavelength (1–10 mm) of mmWave signals allows compact integration of these arrays into small transceivers. These features introduce unique challenges in designing physical-layer transmission algorithms [109]. The BS, equipped with numerous antennas, must acquire downlink CSI. However, the requirement of large amount of feedback creates significant overhead, posing a challenge in channel estimation in mmWave massive MIMO systems.

Compressive sensing techniques are widely used to reduce feedback overhead by exploiting channel sparsity [110]. The limited scattering properties of mmWave propagation and the power-focusing capability of lens antenna arrays transform the spatial domain channel into a sparse beamspace representation. Consequently, the channel estimation problem becomes a sparse signal recovery task, often addressed through compressed sensing methods. However, performing beamspace channel estimation in massive MIMO systems with limited RF chains remains a challenging task.

Compressive sensing aided deep unfolding is explored in [111], where the authors apply Onsager correction for AMP as explained in [112]. The architecture of the unfolded model, learned AMP (LAMP) [111] is similar to unfolded ISTA in [20]. A learned denoising-based AMP network (LDAMP) is proposed in [113] to learn beamspace channel estimates. The authors demonstrate that the iterative sparse signal recovery algorithm with denoising achieves robustness in channel estimation particularly in scenarios with large antenna arrays and limited RF chains. Yi *et al.* in [114] introduced an unfolded network called learned AMP with deep residual learning (LampResNet). In this model, the AMP algorithm is unfolded to estimate the channel, and deep residual learning is employed to mitigate channel noise. NMSE analysis shows that the proposed network in [114] outperforms LAMP [111] and achieves performance comparable to LDAMP [113] in terms of estimation accuracy. However, exploiting channel sparsity does not always yield optimal estimation performance. To address this, the authors of [115] investigated the sparsity of the mmWave channel in the beamspace. They formulate the estimation problem as a sparse signal recovery task and improve convergence speed by unfolding the iterative PGD algorithm. In this approach, the step-size of the descent step is treated as a trainable parameter. In [116], an iterative trimmed-ridge regression algorithm is unfolded to improve beamspace channel estimation. The proposed algorithm involves parallel computation to achieve negligible latency. These algorithms were applied to both narrowband and wideband massive MIMO systems, to ensure versatility in channel estimation.

C. Modern Multicarrier Modulation

Yiyun *et al.* [117] proposed a CSI recovery network for FDD downlink massive MIMO-OFDM systems. ISTA is unfolded and the network consists of compression and reconstruction modules for CSI compression and feedback. Building on ISTA, Hu *et al.* [118] introduced an unfolded algorithm called the learnable optimization and regularization algorithm (LORA) to enhance the precision of CSI feedback. This approach also unfolds ISTA, and the regularization term is learned through end-to-end training. The model captures CSI characteristics and adapts to different levels of channel estimation error. Furthermore, to improve CSI feedback accuracy and efficiency, Yangyang *et al.* [119] proposed an unfolded ISTA network. Their study showed that the real and imaginary components of the channel differential terms, i.e, the difference between CSIs at adjacent time slots, exhibit similar correlation patterns in both the angular and delay domains. This property enables a shared model to process both components, thereby reducing network complexity and enhancing the efficiency of the CSI encoder. Each of the networks proposed in [117–119] enhance CSI feedback performance for FDD massive MIMO-OFDM system. In [33], Guo *et al.* proposed a two-stage low-rank CSI feedback scheme for a low-rank OFDM mmWave channel. In massive MIMO systems operating in FDD mode, CSI is estimated by unfolding the fast ISTA. Fast ISTA is an improved version of the classical ISTA due to the introduction of momentum term that accelerates the convergence rate. The network, FISTA-Net, includes trainable step-size and thresholds to achieve accurate reconstruction performance. In [120], the authors explore quantized compressed CSI at the user equipment before feedback for massive MIMO-OFDM system. They propose a deep unfolding-based bit-level CSI feedback network, TiLISTA, to mitigate the effects of accumulated quantization errors. Building upon ISTA, they used tied learned ISTA as discussed in [121]. Since the trainable parameters are shared across all layers, the number of learnable parameters are reduced, offering a faster and more accurate solution to the signal reduction problem. FISTA-Net [33] focuses on faster convergence while TiLISTA [120] targets robustness to quantization errors.

In [122], an unfolded sparse Bayesian learning algorithm is derived to capture channel sparsity in a mmWave massive MIMO-OFDM system with hybrid beamforming (HBF). The proposed network achieved a lower NMSE of 0.033 outperforming the classical sparse Bayesian learning algorithm which recorded an NMSE of 0.1203. Jianqiao *et al.* [123] proposed an unfolded inverse-free variational Bayesian learning framework to address the high computational complexity of matrix inversion in mmWave MIMO-OFDM systems. The sparse channel recovery performance is enhanced by learning the a priori parameters of the sparsely represented dictionary for channel estimation. In [124], based on approximate Bayesian inference and an extension of expectation-consistent approximation, the authors proposed a generalized expectation consistent algorithm, an iterative message-passing approach for sparse channel estimation. In a lens-based beamspace massive MIMO-OFDM system, this approach compensates for large channel attenuation in the mmWave band. The authors developed a deep unfolded unsupervised learning network called the learned denoising-based generalized expectation consistent signal recovery network. The generalized expectation consistent signal recovery algorithm as discussed in [125] is unfolded. Both the networks in [123] and [124] provide accurate channel recovery estimates for mmWave channel with few RF chains. In [126], the authors leveraged angle-domain channel sparsity to reduce uplink pilot overhead during channel estimation for mmWave MIMO-OFDM systems. They propose an unfolded multiple-measurement-vectors-based AMP network. The phase shift network and channel estimator are jointly trained, thus integrating the channel sparsity for improved performance.

Gao *et al.* [127] proposed a network, ComNet, that unfolds an OFDM receiver into two subnetworks: channel estimation and signal detection. The LMMSE and ZF are unfolded for channel estimation and signal detection, respectively. The weights of the LMMSE and ZF matrices are trained layer-wise. The simulation results show that the BER performance is better compared to the classical LMMSE estimation. Furthermore, Wang *et al.* in [128] proposed an unfolded least squares (LS) estimation to enhance channel estimation for OTFS MIMO system under fractional Doppler condition. The network, CENet, is structured as parallel sub-networks, each responsible for estimating one row (delay tap) of the delay-time channel matrix. At BER= 3×10^{-3}, CENet achieved ~ 5 dB SNR gain over LS estimation algorithm.

TABLE IV: Summary of typical deep unfolding models for channel estimation and their advantages and disadvantages.

Unfolding Model	Main Novelties and Performance Gains	Advantages and Disadvantages
LDAMP [113]	• Denoising-based unfolded AMP model for sparse-beamspace channel estimation in mmWave massive MIMO. • Outperforms classical denoising AMP method with small number of RF chains at the receiver.	• **Advantages:** Simple architecture and fast convergence (within 5 layers), require a few RF chains. • **Disadvantages:** Fixed parameters post-training restrict the model's adaptability, performance may degrade in varying scenarios.
LampResNet [114]	• Combines unfolded AMP with deep residual learning to enhance robustness against noise. • Achieves close performance to LDAMP [113] at high SNR regime (25 dB).	• **Advantages:** Resilient to noise, consistent NMSE improvement under SNR variations. • **Disadvantages:** Rely on prior information of beamspace channel, not consistent with realistic channel models, and network may fail to converge.
FISTA-Net [33]	• Fast ISTA-based deep unfolding for low-rank CSI feedback in mmWave OFDM systems. • Outperforms the classical FISTA with lower complexity.	• **Advantages:** Fast convergence (20 layers). • **Disadvantages:** Requires retraining for major environment variations.
mpNet [104]	• Deep unfold model based on the matching pursuit method for multiuser massive MIMO systems. • Outperforms classical LS estimator in terms of NMSE under varying SNR.	• **Advantages:** Avoids ground-truth channel data by using unsupervised online learning on noisy LS estimates. • **Disadvantages:** Need prior knowledge of physical wave propagation model, as small uncertainty of system parameters lead to high estimation performance loss.
LORA [118]	• The deep unfolding model based on ISTA introduces a learnable regularization module to adaptively capture the CSI structure. • Improved CSI reconstruction accuracy by >2 dB NMSE over TiLISTA [117].	• **Advantages:** LORA exhibits strong denoising capability under channel estimation error. • **Disadvantages:** The number of layers is critical: using fewer than four layers degrades reconstruction accuracy, while increasing the number beyond four leads to higher computational complexity with only marginal improvements in NMSE.

D. IRS

To mitigate the blockage effect and extend the connectivity range for mmWave communication, IRS is a key technology proposed to enhance both energy and SE. IRS consists of many reflecting elements that steer the propagation of electromagnetic waves by adjusting their reflection coefficients. However, to unlock the full potential of IRS, accurate channel estimation is required for both direct and cascaded links. This is a challenging task because IRS-based channel estimation involves estimating multiple channels simultaneously: the direct channels between the BS and each user, the channel between the BS and IRS elements, and the channel between IRS elements and users.

In [129], Jiguang *et al.* proposed a method for estimating the rank-deficient cascaded channel in an IRS-assisted single-input multiple-output system. They developed a deep unfolding network based on a gradient descent algorithm. The model structure is inspired by DetNet [49]. In massive MIMO, increasing the number of IRS elements and BS antennas enhances the system's capacity. However, this leads to increased training overhead to obtain accurate channel estimates within the channel coherence time. The proposed network addresses this challenge by compensating for the computational complexity.

Awais *et al.* [130] developed two networks based on the MAP estimation criterion to address the challenges in IRS cascaded channel estimation. The first network is a deep denoising network designed for noise-dominated SNR regions, providing support at various noise levels. The second network is a deep unfolded half-quadratic splitting network that alternates between optimizing extrapolation and denoising for interference-dominated regions. As the SNR increases, the proposed network shows significant improvement in NMSE. Under severe channel variations, the network delivers a performance gain of 20% in the high SNR range.

Tsai *et al.* [131] proposed a two-stage learned AMP network with row compression to address the joint estimation problem of direct and cascaded channels in IRS-aided mmWave systems. By utilizing the deep unfolding technique, the network combines the advantages of compressive sensing and DL. This approach enables reliable channel estimation with reduced training, which is crucial for efficiently reconfiguring wireless environments under IRS reflection control.

E. IoT devices

Joint activity detection and channel estimation are critical in grant-free random access for massive machine type communication (mMTC), where numerous IoT devices access the network sporadically without scheduling. To enhance detection under imperfect CSI, Qiang *et al.* [31] proposed AMP-Net, an unfolded AMP based network. Each AMP iteration alternates between MMSE-based denoising and residual updates to minimize the MSE. The network uses four trainable scalar parameters to estimate activity probability, channel variance, regulate the denoiser, and control the state evolution. AMP-Net achieved accurate detection of the active device by extracting the CSI from the estimated device state matrix with a 4 dB gain over classical AMP at SNR$=15$ dB. The authors in [132] proposed a grant-free access protocol for mmWave cell-free massive MIMO using a hybrid structure based on learned vector AMP, reducing reliance on prior channel knowledge and improving convergence over classical AMP. In [133], Ma *et al.* introduced a one-phase non-coherent scheme for mMTC. The data bits are embedded in pilot sequences to enable joint activity and data detection without channel estimation. They used unfolded AMP to leverage pilot correlation. In [134], an unfolded AMP with backpropagation was applied to an asynchronous random access scheme with data length diversity, addressing activity detection, channel estimation, and data recovery in mMTC.

The authors in [135] proposed an ADMM-based deep unfolding network for joint channel estimation and active user detection in massive IoT systems. The trainable parameters include an auxiliary variable for matrix inversion, a regularization term balancing sparsity and accuracy, and a shrinkage threshold controlling output sparsity per iteration. Similarly, in [136], the ADMM algorithm is unfolded to improve convergence rate and recovery accuracy in grant-free vehicular networks, particularly for BS with a limited number of antennas. In [137], two unfolded networks were introduced for sporadic user detection and channel estimation in mMTC: ADMM-Net and VAMP-ISTA-Net. ADMM-Net unfolds a linearized ADMM [80] to address constrained convex problems, using a single iteration of the proximal gradient method to avoid solving subproblems explicitly. VAMP-ISTA-Net unfolds a vector AMP enhanced with an ISTA-based denoiser. Both architectures required only 5 layers, significantly reducing the iteration count from 300 in classical algorithms and achieving faster convergence. Overall, the approaches in [135–137] tackle the multiple measurement vector sparse recovery problem using unfolded ADMM frameworks for improved reconstruction performance.

Shi *et al.* [138] considered grant-free massive access in an IoT network. A large number of single-antenna IoT devices sporadically transmit non-orthogonal sequences to a multi-antenna BS. An unfolded ISTA with trainable weight matrices is designed. The network trains faster due to less number of trainable parameters. Zou *et al.* in [139] formulated the joint activity detection and channel estimation problem as a group-row-sparse matrix recovery problem. A proximal gradient-based method is unfolded and it achieves better convergence than the network in [138].

F. Discussions and Takeaways

Channel estimation in modern wireless systems, especially in massive MIMO and mmWave settings, poses distinct challenges due to high dimensionality, feedback overhead, and sparse propagation characteristics. Deep unfolding addresses these issues by embedding structural priors into learning pipelines, as demonstrated by models like LDAMP [113] and LampResNet [114]. Table IV presents notable unfolded networks for channel estimation with their performance gains, advantages, and disadvantages.

Nonetheless, deep unfolding for channel estimation remains sensitive to model assumptions (e.g., sparsity or statistical independence). Performance degradation in dense or correlated environments, retraining overhead, and reliance on fixed unfolding depth are key limitations. Moreover, model-agnostic unfolding and joint learning of estimation and feedback (especially in IRS and wideband settings) will be crucial for making deep unfolding a mainstream tool in physical layer design. Unlike deep unfolding models for signal detection, the practical deployment of such models for channel estimation remains limited. Moreover, their performance has yet to be thoroughly validated through real-world over the air experiments.

V. Precoder Design

This section reviews the applications of deep unfolding for precoder design used in key 6G technologies.

A. Massive MIMO

Beamforming in MIMO systems enhances signal transmission by steering signals toward the intended receivers using multiple antennas, thereby improving received SNR [140, 141]. Beamforming techniques can be broadly divided into three types: analog, digital, and hybrid [140]. Digital beamforming uses a digital precoding matrix with one RF chain per antenna, leading to high power consumption and complexity in large arrays. Analog beamforming employs phase shifters but struggles with inter-user interference. HBF [142, 143] addresses these issues by combining digital and analog methods, reducing RF chain count while maintaining performance and efficiency which is ideal for mmWave and massive MIMO in 5G and beyond.

Shlezinger *et al.* [28] discuss HBF optimization using classical, data-driven, and deep unfolding methods, highlighting deep unfolding's efficiency with limited RF chains. For precoder design, the weighted MMSE (WMMSE) algorithm [144] maximizes the weighted sum rate via iterative updates involving matrix inversions and eigendecompositions. Given the nonconvexity of the power-constrained problem [145], a distributed inexact cyclic coordinate descent (CD) method is proposed for efficient locally optimal solutions.

A deep unfolded architecture, iterative algorithm induced deep unfolding network (IAIDNN) is proposed for precoder design in MU-MIMO systems in [146]. The iterative WMMSE algorithm is unfolded. To improve efficiency, WMMSE is reformulated to approximate matrix inversions. A generalized chain rule in matrix form helps visualize the gradient recurrence between adjacent layers during backpropagation. Using this rule, the gradients of trainable parameters are computed effectively. To improve convergence while avoiding matrix inversion, a downlink beamforming method using an unfolded PGD-WMMSE algorithm for multi-user multiple-input single-output (MISO) system is proposed in [147]. The authors replace matrix inversions with PGD and compute gradients in parallel. Unlike IAIDNN [146], this approach eliminates matrix inversion, making hardware implementation more practical. Additionally, it achieves a high weighted sum rate in fully loaded scenarios, where the number of users matches the number of transmit antennas at the BS.

In [146], learnable parameters are used to approximate matrix multiplication and inversion in WMMSE update rules. However, the network fails to exploit the inherent graph structure of wireless networks. This limitation is addressed in [148]. An unfolded WMMSE algorithm for transceiver design in multicell multi-user MIMO (MU-MIMO) interference channels with local CSI is proposed in [148]. The authors use graph convolutional networks (a variant of GNN) to unfold WMMSE, enabling coordination in multicell MU-MIMO systems. Graph filters and convolutional networks approximate WMMSE-related variables, such as weighted beamformer matrices, by leveraging graph learning techniques. However, the approach in [148] lacks flexibility in learning the exact transformation of WMMSE variables. To overcome this, [149] introduces a hybrid unfolded WMMSE algorithm for transmit beamforming in MU-MIMO ad-hoc networks. This method leverages GNNs to learn functional transformations of WMMSE parameters. Compared to the classical WMMSE algorithm, the proposed network reduces the number of layers, leading to faster convergence of the optimization problem. In [150], the joint optimization of precoders for multicell MU-MIMO systems is studied. The WMMSE algorithm is unfolded to enable distributed and synchronized precoding across cooperative BSs. The methods in [146, 147] rely on centralized optimization, leading to high complexity and backhaul overhead. In contrast, due to the distributed precoding, the proposed network in [150] reduces computational cost while maintaining efficiency. An unfolded WMMSE-based NN for massive MU-MIMO under imperfect CSI is studied in [151]. The proposed network enhances performance as compared to [144], with low complexity by leveraging CSI statistics for near-optimal precoder solutions, thus accelerating convergence.

Shi *et al.* [152] designed a hybrid analog-digital transceiver using iterative gradient descent to obtain minimum SER (MSER) in massive MIMO systems. They unfolded the iterative gradient descent algorithm into a DNN, where step-sizes for the gradients are the learnable parameters. The authors claim that their proposed network matches the performance of the MSER-based iterative gradient descent algorithm while significantly reducing complexity.

Gradient-based methods have been successfully applied to beamforming optimization with reduced complexity. However, these approaches can be further improved using advanced optimization frameworks like PGD or projected gradient ascent (PGA). A robust frequency-selective precoding algorithm that maximizes the expected weighted sum rate while considering per-antenna power constraints is developed for wideband MIMO-OFDM system in [153]. This frequency-selective precoding algorithm is then unfolded into a deep network, with the aid of PGD. The trainable parameters include step sizes and projections in each PGD layer. The proposed network achieves near-optimal sum rate with lower computational complexity. Zhu *et al.* [154] proposed a beamforming learning architecture by unfolding a parallel gradient projection algorithm for MISO heterogeneous systems. The trainable set includes the step-size for the gradient ascent update and weighted sum rate coefficients used to construct the beamforming gradient direction. The proposed network generalizes well across different numbers of antennas, different numbers of BSs, and different network sizes especially without retraining. PGA is a variant of the gradient ascent algorithm. It updates the solution in the gradient direction to maximize the objective function and then projects it back onto the feasible set. Lavi *et al.* [155] unfolded the PGA algorithm to optimize hybrid precoding for multiuser downlink MIMO, assuming perfect CSI availability. To handle noisy CSI, they implemented the projected conceptual mirror prox algorithm [156, 157]. The step-sizes for updating the analog and digital precoder matrices are tuned using mini-batch SGD, a data-driven approach. The solution is computed within a fixed number of iterations, defined as network layers, ensuring a balance between interpretability and efficiency. In [158], Bilbao *et al.* proposed an efficient analog beamforming design to maximize the achievable rate of massive MIMO in-band full-duplex systems. The underlying beamforming problem is non-convex, primarily due to constant modulus constraints imposed by the use of analog phase shifters. To overcome this challenge, they developed an unfolded algorithm based on PGA combined with ZF projection onto convex sets. The convex feasible set arises due to the requirements of the ZF constraint and unit norm beamforming vectors. This approach leverages the structure of the zero-forcing conditions and the constant modulus constraints, achieving significantly faster convergence and improved performance compared to classical iterative methods. Besides PGA and PGD, CGD is also an iterative algorithm that falls under the gradient-based optimization approach. In [159], the CGD algorithm is unfolded for constant envelope precoding to reduce computational overhead. The Riemannian manifold optimization (MO)-based CGD is implemented across network layers. Simulations show that this approach effectively suppresses multi-user interference, improving overall performance.

A fast and robust precoding network for MU-MIMO systems with imperfect CSI and per-antenna power constraints is proposed in [160]. The objective function is a geometric program that minimizes the MSE between the transmitted and recovered symbol vectors at the user. The iterative optimization framework is derived from the MSE uplink-downlink duality-based algorithm, which is unfolded to design a robust MSE duality network (RMSED-Net). RMSED-Net is robust to channel estimation errors and achieves sum rate close to the upper bound of its benchmark.

In [161], the interior point method is unfolded for precoding design. It leverages interference in multi-user MISO systems to minimize power while meeting signal-to-interference and noise ratio (SINR) constraints by training the proximal operators and Lagrangian multipliers. The authors show that the execution time is reduced by $50 - 70\%$ compared to the classical method. However, the network is limited to phase shift keying (PSK) modulation, requiring reformulation for higher order and QAM modulations.

A constructive interference-based symbol-level precoding for multi-user MISO downlink transmission is discussed in [162]. The proximal Jacobian ADMM is unfolded. The unfolded network reduces average transmit power and execution time by 98% compared to proximal Jacobian ADMM. The ADMM algorithm struggles with nonlinear discrete optimization due to biased alternating updates. To address this, iterative discrete estimation (IDE) is introduced in [163], ensuring unbiased updates and better convergence. Wang *et al.* in [163] proposed an algorithm for downlink massive MU-MIMO systems with finite-alphabet precoding to minimize inter-user interference. To reduce hardware costs and power consumption in multi-user massive MIMO, He *et al.* [164] developed an unfolded network using the IDE-based algorithm in [163]. The unfolded network with learnable step-size and damping factor achieves lower BER across a range of SNR.

In [165] a learning-based channel semantic acquisition and beamforming solution is proposed for cell-free massive MIMO. The beamformer is constructed by unfolding successive over-relaxation-based linear beamforming algorithm which reduces estimation errors in HBF. The proposed approach reduces computational complexity and enhances robustness to imperfect CSI.

B. mmWave MIMO

mmWave MIMO systems rely on beamforming to mitigate multi-user interference while enhancing diversity and spectrum efficiency. Typically, the beam selection and digital precoding matrices are designed separately, leading to suboptimal solutions. To address this, Hu *et al.* [166] proposed a joint design approach combining DL and model-driven techniques. DRL optimizes beam selection, while a deep unfolding NN, based on iterative WMMSE, optimizes the digital precoding matrix. The goal is to maximize the system sum rate while satisfying beam selection constraints. The DRL and deep unfolding networks are trained sequentially, with one network's output serving as the input to the other. Backpropagation and SGD are used for training and parameter updates. The authors verified the proposed framework outperforms benchmarks in sum-rate, complexity, and robustness while maintaining interpretability and scalability.

Lin *et al.* [167] studied a LMMSE transceiver design for uplink massive MU-MIMO with one-bit ADCs. To minimize the system MSE by optimizing the precoder at each user equipment, the PGD algorithm is unfolded. The authors focused on training only the step-size in each PGD iteration, significantly reducing the training overhead. In [168], the authors combined deep generative models with unfolding techniques to achieve near-optimal HBF while minimizing feedback and computational complexity. The receiver estimates the channel using GAN, enabling efficient optimization of digital and analog precoders. The HBF problem is formulated as a bilevel optimization. The algorithm unfolds an iterative bi-level optimization problem into the layers of the CNN. There are two levels are upper and lower layer.

TABLE V: Summary of typical deep unfolding models for precoder design and their advantages and disadvantages.

Unfolding Model	Main Novelties and Performance Gains	Advantages and Disadvantages
Unfolded WMMSE [166]	• Pioneering deep unfolding model based on the WMMSE algorithm for downlink precoding in MU-MIMO systems. • Achieves 10% improvement in sum rate metric compared with the conventional WMMSE at SNR = 45 dB.	• **Advantages:** The network has a few trainable parameters, leading to fast convergence. • **Disadvantages:** Performance loss with ill-conditioned channel matrices because the matrix inversion in WMMSE is replaced by a trainable approximation.
IAIDNN [146]	• Unfolded WMMSE model for MU-MIMO precoding. • Converges after 2.94 seconds compared to 291.01 seconds required in classical WMMSE and reduces WMMSE computational cost by up to 30%.	• **Advantages:** IAIDNN generalizes to varying system sizes (numbers of users and antennas) with minor adjustments. • **Disadvantages:** High training computational complexity due to the increased number of trainable parameters.
PGA-Net [155]	• Unfolds PGA with trainable step sizes for hybrid precoding. • Proposed network achieves better sum-rate than classical PGA with fewer iterations.	• **Advantages:** Fast convergence and generalize well with varying channel size, effective for wideband channels. • **Disadvantages:** Lacks inter-user interference cancellation and involves high gradient computation complexity.
ManNet [181]	• Deep unfolding models for HBF (with both fully- and sub-connected architectures) based on Riemannian MO method. • Over 600 times faster and 6 times complexity reduction compared to MO-AltMin algorithm [182] ,has lower complexity than PGA-Net [155].	• **Advantages:** Sparsely connected network architecture, and thus, has small number of training parameters. • **Disadvantages:** The dynamic RF chain-antenna connections has not been incorporated into the unfolding models.

In the upper layer digital precoder is optimized given a fixed analog precoder while in the lower layer the analog precoder is optimized given a fixed digital precoder. This alternating iterative steps are then unfolded. The trainable parameter set include the weights of the CNN that maps digital precoder to the analog precoder. The unfolded bilevel optimization algorithm achieved near optimal HBF with low feedback and computational cost.

C. Next Generation Multiple Access

Rate-splitting multiple access (RSMA) is a flexible multiple access technology that merges the strengths of space division multiple access (SDMA) and NOMA [169]. It uses linearly precoded rate-splitting at the transmitter and SIC at the receivers. RSMA balances interference management by partially decoding some interference while treating the rest as noise, bridging the gap between SDMA (which treats all interference as noise) and NOMA (which fully decodes interference). In [170], the authors applied fractional programming with hyperplane fixed-point iteration to design the RSMA beamformer. The algorithm is unfolded into an NN, replacing its inner loop with a small-scale DNN to predict Lagrangian dual variables. Simulations show that the proposed approach achieves high performance while reducing computational complexity.

D. IRS

With the aid of IRS, Chen *et al.* [171] addressed the joint design of the HBF matrix and the phase-shifting matrix of the IRS elements. The authors proposed the deep unfolded WMMSE-MO network (DU-WMMSE-MO), for IRS-assisted mmWave MIMO-OFDM systems. The trainable parameters include the weights for the precoder, phase configuration of the IRS elements and the combiner, and biases for the precoder and phase-shift matrix of the IRS. The algorithm showed improved performance in IRS phase-shift design, utilizing two loops to jointly update the step-size of the beamforming and IRS matrices. Compared to the Armijo backtracking line search method used in classical WMMSE-MO [172], DU-WMMSE-MO [171] provides more flexibility but with a higher runtime.

An IRS-assisted multiuser MIMO full-duplex system is explored in [173]. The active beamforming matrices at the BS and uplink users, along with the passive beamforming matrix at the IRS, are jointly optimized to maximize the weighted sum rate of the system. A deep unfolding approach is proposed to approximate the stochastic SSCA-based beamforming algorithm, reducing the complexity of matrix inversion while maintaining near-optimal performance.

In [174], Min *et al.* introduced a two-timescale transmission approach to reduce the signal processing complexity and channel training overhead for the IRS-aided multiuser MISO system. To minimize the average transmit power at the analog precoder while ensuring the individual quality of service (QoS) constraints for the users, a primal-dual decomposition-based two-timescale HBF method is proposed and unfolded. This approach allows for efficient beamforming optimization by separating the optimization into two timescales, improving both computational efficiency and performance.

In [175], a low-complexity IRS-assisted downlink MISO system is designed using an unfolded block CD algorithm to address the joint optimization problem of transmit beamforming and IRS phase shifts. The multi-ratio fractional programming structure of the objective function is handled by decoupling the problem into two subproblems: one for optimizing the beamforming at the BS and another for optimizing the phase shifts of the IRS. This is achieved through the use of Lagrange dual transformation, effectively simplifying the optimization process. The numerical analysis validated that the achieved weighted sum rate is comparable to that of [176] but with reduced computational complexity. Furthermore, the authors in [177], have developed an unfolded network to address the joint beamforming task for an IRS-assisted MISO system. The WMMSE and power iteration algorithm are used as an alternating optimization (AO) strategy to derive the local optimal solution for the active and passive beamforming. The proposed network reduce the computational runtime to approximately 3% of the block CD algorithm in [176].

In [178], an IRS-assisted MISO covert symbiotic radio communication system is explored, and a deep unfolding algorithm based on gradient descent is proposed for beamforming design.

The goal is to jointly optimize active and passive beamforming under covert communication constraints. The step-size for updating the active and passive beamformers, and Lagrange multipliers form the trainable set. In contrast, Gangyong *et al.* in [179] focused purely on passive beamforming and unfolded the Riemannian gradient descent algorithm. They have designed the IRS phase shift matrix by maximizing the sum path gain of the channel. In this network, the only trainable parameter is the step-size associated with each Riemannian gradient descent iteration. The network in [179] is simpler compared to that of [178], due to less number of parameters to train but is specific to optimizing the phase shift of the IRS.

E. Terahertz

6G wireless networks are expected to achieve terabit-per-second (Tbps) data rates to support ultra-reliable low-latency communication. To enhance spectral and energy efficiency, terahertz (THz)-enabled massive MIMO is a promising solution, leveraging the abundant bandwidth at THz frequencies to significantly boost data rates [180].

Nguyen *et al.* [181] proposed a deep unfolding framework that unfolds the iterations of Riemannian MO (specifically the MO-AltMin algorithm) [182]. Instead of directly solving the original SE maximization or WMMSE minimization problems, which are highly nonconvex and computationally intensive, they reformulated the HBF design as an approximate matrix factorization problem. This factorization leads to LS formulation for analog and digital beamforming enabling a more efficient solution. By leveraging this structure, their proposed ManNet-based approach [181] requires significantly fewer layers to estimate the analog precoder, resulting in faster training and lower computational complexity as compared to the approach in [155].

Minghui *et al.* [183] proposed an unfolded network to enhance RSMA precoding in IRS-aided THz MU-MIMO systems under CSI imperfections. They unfold an approximate WMMSE algorithm to optimize the digital active precoding layers based on the estimated equivalent channels. The trainable set includes the transformer weights associated with the private and common streams and a regularization parameter. Compared to the original approximate WMMSE algorithm, the proposed unfolding method achieves superior precoding performance with substantially lower computational complexity.

F. Discussions and Takeaways

Precoding design in massive MIMO and HBF systems involves solving non-convex optimization problems under hardware constraints and channel uncertainty. Popular unfolded networks for precoding design are listed in Table V. For instance, IAIDNN [146] unfolds the WMMSE algorithm and replaces complex operations like matrix inversion with a learnable block. This results in a significant reduction in computational complexity while achieving performance close to traditional WMMSE solutions. Deep unfolding enables the incorporation of domain-specific constraints such as the constant modulus constraint in analog beamformers through tailored activation functions. This approach significantly reduces computational complexity by avoiding matrix inversion and accelerates convergence, making it well-suited for real-time hybrid precoding under practical hardware limitations.

Despite their benefits, current unfolded precoding models often assume fixed system dimensions and ideal CSI, which limits generalization and applications in practical systems. Future work should focus on scalable, adaptive unfolding frameworks that can operate reliably in time-varying, multi-user, and wideband settings. Integration with channel estimation modules and robust designs under CSI uncertainty will be key for practical deployment in next-generation wireless networks. Moreover, applying such unfolded frameworks for precoders in real systems and analyzing the performance in real experiments will be very beneficial.

VI. Sensing and Communication

The independent operation of sensing and communication faces challenges in the 6G era due to spectrum congestion in high-frequency bands like mmWave and THz [184]. This has driven the development of ISAC systems, which unify both functions using shared architectures and processing techniques [185], [186], [187]. ISAC includes active (mono-static) and passive (bi-static) sensing methods [188]. While ISAC enhances SE and system performance, understanding its fundamental limits is crucial. Key sensing parameter estimations have been thoroughly analyzed in [189], revealing theoretical and practical boundaries.

A. MIMO ISAC

The integration of deep unfolding in ISAC systems has shown promising advances in optimizing waveform design, beamforming, and communication-sensing trade-offs. In [190], the challenge of designing constant-modulus waveforms for MU-MIMO ISAC systems was explored. Since constant-modulus waveforms help avoid signal distortion in non-linear power amplifiers, their design is crucial. However, the non-convex nature of the constant-modulus constraint makes optimization challenging. To address this, the authors employed a deep unfolded network based on PGD algorithm, ensuring efficient waveform design under these constraints.

HBF in MIMO ISAC systems poses additional complexity due to balancing communication rates and sensing accuracy. The authors in [191] tackled this challenge by unfolding the iterative PGA algorithm. Their deep unfolding-based method accelerated convergence and enhanced overall system performance, achieving an optimal balance between communication and sensing objectives. In addition, joint optimization of radar communication beamforming in MIMO system was addressed in [192]. The authors incorporated symbol-level precoding with constructive interference constraints to improve energy efficiency by exploiting multi-user interference. Since, the objective problem is computationally complex due to non-convexity and higher dimensional matrix inversions, a custom iterative algorithm is developed. The iterative algorithm is based on block CD and primal-dual optimization (augmented Lagrangian method). Therefore, the trainable parameter set for training the deep unfolded network include step-sizes to update the primal, dual and auxiliary variable, as well as a correction matrix to approximate matrix inversion. The proposed network reduces computational complexity due to small number of trainable parameters, improve scalability, and support real-time implementation.

The large-scale antenna arrays in mmWave MIMO systems enable sub-meter user localization by providing high angular resolution. However, these methods often incur significant computational complexity, particularly in high dimensional mmWave MIMO settings. To improve this, Fan *et al.* proposed, a fast and accurate direct localization method in [193]. An ADMM solver is developed to solve the optimization problem with low complexity. The ADMM iterations are unfolded to learn the trainable parameters that include penalty and proximal parameters, and weight coefficient for channel gain between the BS and user. The unfolded network converges faster than classical ADMM and computational time is reduced.

TABLE VI: Summary of typical deep unfolding models for ISAC and their advantages and disadvantages.

Unfolding Model	Main Novelties and Performance Gains	Advantages and Disadvantages
Unfolded PGA network [191]	• Unfolds PGA to maximize the system sum rate while minimizing the sensing beampattern error. • Achieves 33.5% higher sum rate, 2.5 dB lower radar beam pattern error compared with SCA, and 65% reduction in runtime and complexity compared with the conventional PGA.	• **Advantages:** Ensure convergence in the challenging ISAC HBF designs, perform well in both single use and MU systems. • **Disadvantages:** The convergence speed is relatively slow, high complexity to compute the gradients.
ISAC-Net [194]	• Unfolds the denoising AMP method for communication and DFT-based sensing in narrowband OFDM ISAC. • Achieve communication performance close to OAMP-Net2 [11] and better sensing performance than classical 2D-DFT algorithm.	• **Advantages:** Leverages shared resources to overcome challenges in signal processing without prior knowledge of transmitted signals, enhanced accuracy and security. • **Disadvantages:** Beam squint effect is not considered leading to discrepancies extracting sensing parameters (angle of arrival or angle of departure) across various subcarriers in OFDM.
ADMM-Net for OTFS ISAC [196]	• Unfolds ADMM to estimate delay-Doppler targets and demodulate OTFS symbols under high mobility. • With 2 targets, ADMM-Net achieve 4 dB gain in NMSE compared with the classical ADMM and ISTA at SNR=14 dB.	• **Advantages:** Suitable for vehicular/high-mobility scenarios and robust to delay-Doppler spread. • **Disadvantages:** Real time CSI estimation is not addressed.
Double loop unfolded network [197]	• Unfolds gradient descent to jointly optimize IRS phase shifts in ISAC system. • Minimizes worst case CRB for sensing targets.	• **Advantages:** Has low computational complexity and fast convergence (within 15 layers). • **Disadvantages:** Offline training (1000 channel realization, 20 epochs) may not be feasible for frequently changing environments.

B. Modern Multicarrier Modulation for ISAC

A joint passive sensing and communication demodulation for OFDM is addressed in [194]. The ISAC-Net is designed by unfolding two algorithms: a denoising-based AMP algorithm for the communication module and the 2D discrete Fourier transform (DFT) for the sensing module. The step-size to update the channel and DFT matrix form the trainable set to balance the communication demodulation and sensing accuracy. Furthermore, Jingcai *et al.* proposed an unfolded network (FISTA-Net) for joint target parameter estimation and communication in MIMO-OFDM systems [195]. The fast ISTA is unfolded, with learnable parameters such as step-size for weight matrix, enabling high-precision solutions with low computational complexity. The proposed network converges with 8 layers as compared to 2000 iteration in fast ISTA. Also, the network achieves an NMSE of -20.12 dB while the classical complex ISTA achieved -16.8 dB. The problem of estimating parameters of moving targets in high-speed scenarios using OTFS modulation for ISAC systems is explored in [196]. The ADMM iterations are unfolded (ADMM-Net) to accelerate convergence and enhance sparse recovery of target position. Regarding parameter estimation, ADMM-Net [196] and FISTA-Net [195] reduce NMSE significantly. Also, ISAC-Net [194] and FISTA-Net [195] show faster convergence than iterative baselines.

C. IRS-Assisted ISAC

Liu *et al.* addressed a robust beamforming design for an IRS-assisted ISAC system under imperfect CSI and target direction estimates in [197]. The objective is to minimize the sum of worst-case Cramér-Rao bounds (CRB) for all sensing targets. They formulated a bi-level optimization structure that consists of two levels: the lower-level problem finds the worst-case CSI and worst-case target sensing angles. The upper-level problem optimizes the active and passive beamforming matrices. A double-loop deep unfolding NN is designed to tackle the bi-level optimization problem. The CSI and angle uncertainties are updated via PGD in the lower-level loop. The beamforming variables are updated in the upper-level loop using PGD and Riemannian MO optimization. The proposed network effectively achieves near-optimal performance by handling statistical CSI errors and bounded target direction errors.

D. Discussions and Takeaways

Deep unfolding offers a structured and interpretable framework for tackling challenging optimization problems in ISAC systems. By unrolling classical algorithms like PGD and WMMSE into NN layers, unfolding enables the joint design of beamformers and waveforms that balance sensing accuracy and communication performance. These approaches reduce computational complexity and provide faster convergence compared to traditional iterative solutions. We summarize typical deep unfolding models for ISAC in Table VI. It is observed that deep unfolding exhibits advantages in constant modulus waveform design, optimization of dual-functional beamformers, and power allocation. Indeed, the deep unfolding technique can effectively handle non-convex constraints or objective functions like SINR, sum rate, and CRB, which are infeasible with black-box models.

In [198], the authors have provided techniques such as gradient calculation for estimating position that can be employed to facilitate practical deployment. However, the application of deep unfolding in ISAC is still underexplored. Indeed, most existing unfolding solutions are developed to optimize the transmit waveform in monostatic ISAC scenarios. In sensing functions, the processing of the received echo signal plays an important role in translating the observation into a decision on target detection and localization. Well-designed deep unfolding models can perform these tasks with low complexity and high accuracy, which is important for real-time sensing applications.

There is limited literature addressing the practical deployment of unfolded networks in ISAC systems.

VII. Decoding in Error Correcting Codes

Mobile communication has evolved across generations. 1G used analog transmission (without any coding), and 2G introduced digital voice coding with convolutional and block codes. Turbo codes, adopted in 3G and 4G, offer near-Shannon limit performance with moderate complexity [199]. Low-density parity check (LDPC) codes, rediscovered in the 1960s, are now widely used in 4G and 5G [200], [201], [202]. Polar codes, introduced in 2009, achieve the capacity of symmetric binary-input discrete memoryless channels [203], approaching the Shannon limit as the block length increases. They have been adopted in 5G new radio for both uplink and downlink control channels [204]. BP operates on a bipartite graph known as Tanner graph, introduced by Michael Tanner [205] The min-sum and max-sum algorithms are derived from the generic sum-product decoding [206]. Polar codes use successive cancellation decoding, where the received bits are decoded sequentially by leveraging frozen bits and previously decoded bits. More advanced techniques, such as successive list decoding assisted by cyclic redundancy check (CRC), further improve the performance of polar codes [207].

Wei *et al.* in [39] proposed an unfolded ADMM-based penalized algorithm for binary linear codes to enhance the decoding performance. The learnable ADMM decoding network (LADN) trains the ADMM penalty parameters. Additionally, decoding irregular binary LDPC codes via unfolded ADMM is explored in [40]. The LADN [39] outperforms the original ADMM-penalized decoder in block error rate for LDPC codes, while the network in [40] outperforms the original ADMM-penalized decoder in block error rate for irregular LDPC codes.

In [36], the authors proposed the TurboNet architecture, to improve turbo decoding by unfolding the max-log MAP algorithm. Trainable parameters were inserted into the max-log-MAP decoding flow like weights for prior LLRs. Numerical analysis demonstarted that TurboNet outperforms classical turbo decoding based on the max-log-MAP algorithm, particularly at moderate-to-high SNRs. Furthermore, TurboNet has been validated through over the air experiments confirming its robustness to SNR mismatches. Moreover, the authors have extended the work by including joint detection and decoding in [208]. A CGD-based OAMP detector (CG-OAMPNet) for the MIMO-OFDM system is proposed by unfolding EP. The robustness of CG-OAMPNet was further validated through over-the-air experiments, demonstrating practical resilience to environmental variations. In correlated channel scenario, CG-OAMPNet outperforms OAMP-Net2 [11] by performance gain of ~ 2 dB at BER$= 10^{-3}$. The authors in [209] tackle the decoding challenges of marker codes in insertion and deletion channels without perfect CSI. They proposed a network called FBNet to improve detection performance; the forward-backward algorithm is unfolded for marker codes. The trainable weights are the CSI probabilities for insertion/deletion channels. FBNet performs quite close to the classical forward-backward algorithm but is robust to CSI mismatch. FBNet [209] is based on RNN structure and target synchronization errors (insertions/deletions) without accurate CSI, while TurboNet [36] is a deep feedforward DNN structure and targets substitution errors in turbo codes under AWGN channel.

Gao *et al.* [210] have unfolded the polar BP algorithm to improve polar code decoding performance under one-bit quantization. The network with trainable weights for the factor graph edges outperforms standard BP with two bit quantizer by 0.2 dB. In [211], a double-loop iterative decoding algorithm for LDPC codes is unfolded within a penalty dual decomposition framework. The network alternately updates primal and dual variables in a double loop structure. The penalty coefficients, penalty parameters and over relaxation factor form the trainable parameter set. The network has lower computational complexity than ADMM based decoders. Additionally, by unfolding the iterative decoding process between check nodes and variable nodes for LDPC codes, a deep unfolded normalized min-sum LDPC decoder is implemented in [212]. The unfolded network outperforms the classical normalized min-sum LDPC decoder in BER and is robust across code lengths, rates and different channel conditions. The networks in [211, 212] focus on LDPC codes while the network in [210] consider polar codes. Moreover, the networks in [211, 212] use multi-SNR training for robustness.

In [213], theoretical results for the unfolded BP decoder for Tanner codes are presented. The authors define the generalization gap of a decoder as the difference between the empirical and expected BER. They provide a set of theoretical results that bound this gap and examine its dependence on decoder complexity, which is influenced by code parameters (such as block length, message length, and variable/check node degrees), the number of decoding iterations, and the size of the training dataset.

Traditionally, channel estimation, signal detection, and channel decoding in a MIMO receiver are performed sequentially and independently. This disjoint approach leads to suboptimal performance. To address this, Sun *et al.* [214] propose an unfolded ADMM-based receiver for joint detection and decoding. The detection and LDPC decoding components are jointly unfolded from ADMM iterations. Simulations show that the proposed receiver outperforms classical turbo receivers. In [215], Zhang *et al.* proposed a joint signal detection and channel decoding for MIMO systems. Signal detection is achieved by unfolding the EP algorithm. The channel decoding is based on an unfolded turbo code decoding network, TurboNet, as proposed in [36]. The damping factors in EP and decoder parameters in turbo decoding are learned in the training stage. The networks showed fast adaptation and strong BER gains in over-the-air experiments.

A. Discussions and Takeaways

In Table VII, we summarize the most popular unfolded network for the decoding of channel codes. Deep unfolding spans their applications in modern error-correcting codes, such as LDPC, turbo, and polar codes. The soft probabilistic estimates are progressively refined through learnable decoding layers in deep unfolding while enforcing parity constraints, enabling near-optimal hard decisions without the need for manual hyperparameter tuning. Hence, it merges model based optimization, for structured soft-output refinement and codeword validity with data-driven learning, for adaptive penalty tuning, outperforming traditional decoders in accuracy and efficiency.

Despite these advantages, there are several limitations in existing deep unfolding architectures for decoding and channel coding. For example, the unfolded decoders in [39] is tied to specific code structures, requiring retraining when parameters such as code length or interleaving patterns change. Some methods demand significant training effort or hyperparameter tuning to maintain generalization.

There are several works [36, 208, 215] that developed a prototyping system to experimentally validate the effectiveness and practical feasibility of unfolded networks in real-world channel conditions. The experimental results indicated that the unfolded network demonstrated strong robustness to environmental variations and operate effectively without the need for frequent retraining.

TABLE VII: Summary of typical deep unfolding models for decoding and their advantages and disadvantages.

Unfolding Model	Main Novelties and Performance Gains	Advantages and Disadvantages
LADN [39]	• Introduces learnable ADMM-based decoder with trainable penalties for binary LDPC decoding. • Achieves a block error rate reduction of $\sim 10\times$ at SNR $= 2$ dB over the baseline ADMM decoder.	• **Advantages:** Has small number of learning parameters and fast convergence. • **Disadvantages:** Limited by code length. Inherent codeword structure not exploited leading to high computational complexity.
TurboNet [36]	• Unfolds max-log MAP decoding steps for turbo codes with learnable weights. • Outperforms the conventional max-log MAP scheme by $0.25 - 0.5$ dB BER.	• **Advantages:** Compatible with 3GPP turbo codes and validated through practical experiment. • **Disadvantages:** Training requires fixed interleaver and careful SNR selection.
Quantized Polar BP [210]	• Learns BP decoding under 1-bit quantization for polar codes over an AWGN channel. • Outperforms standard BP with one-bit quantization by $1.2 - 2$ dB (block error rate).	• **Advantages:** Hardware-friendly and applicable to IoT systems with coarse ADCs. • **Disadvantages:** Less scalable. Require large number of trainable weights assigned to each edge of factor graph for decoding.
ADMM-Net [40]	• Extends LADN [39] to irregular LDPC codes. Unfolded ADMM, learn different penalty weights for each variable node in decoding. • Achieve better BER performance than BP and ADMM penalized decoders for irregular LDPC codes.	• **Advantages:** Suitable for irregular code graphs. • **Disadvantages:** The initialization strategy for the trainable parameters assumes prior knowledge of the code structure, which makes it less generalizable.
FBNet [209]	• Introduce a RNN-based unfolding model based on forward-backward decoding for Marker codes over insertion-deletion channels. • Outperforms the classical forward-backward algorithm under imperfect CSI.	• **Advantages:** Robust to synchronization errors with flexible sequence handling. • **Disadvantages:** Degraded performance under burst insertion-deletion channels.

Future research should aim to create more adaptable unfolding architectures that can handle a variety of code structures, channel models, and SNR conditions. Joint optimization with modulation and detection, as well as hardware-aware training, may further improve the viability of deep unfolding for end-to-end error correction in next-generation wireless communication systems.

VIII. Power Allocation

Power allocation in wireless communication refers to the process of distributing available transmit power across different communication channels or users to optimize system performance. It aims to maximize throughput, minimize interference, and ensure fairness, while considering factors like channel conditions, user requirements, and energy efficiency for improved overall network efficiency. Several classical approaches, such as the Dinkelbach algorithm, Lagrange dual decomposition, interference pricing, successive convex approximation (SCA), WMMSE, and matching theory, have been employed for power allocation [144, 216–220]. However, a major drawback of these methods is their reliance on iterative algorithms, which are computationally intensive and result in long runtimes, ultimately degrading system energy efficiency. The deep unfolding approach mitigates this issue by reducing the number of iterations required to achieve a near-optimal solution, thereby enhancing computational efficiency and energy savings.

Chowdhury *et al.* in [221] unfolded WMMSE into GNN based architecture to learn a power allocation policy from the channel matrix. The key feature of their architecture is that graph topology is leveraged with GNN for efficient rate utility maximization. The unfolded WMMSE method integrates domain-specific elements with trainable components, which are parameterized by generated adaptive weigh of the GNN. Hu *et al.* [222] propose an unsupervised power allocation optimization approach for adhoc networks by integrating attentive graph representation with deep unfolded WMMSE. The key advantage of attentive graph representation is its ability to allow network nodes to selectively focus on the most relevant connections during learning, rather than treating all links equally. The power allocation framework is built upon this attentive graph representation, as described in [223, 224]. The model [222] achieves improved interpretability and accelerated training over [221] while eliminating the need for supervised data. In [225], Li *et al.* unroll the SCA algorithm into NN layers for energy-efficient power allocation in wireless interference networks. The proposed unfolded algorithm is highly adaptable, as it generalizes across varying network sizes and channel distributions. The power allocation policy is formulated based on the CSI matrix and the total power constraint, within a layered architecture incorporating trainable weights. Each of the networks discussed in [221, 222] and [225] leverages the GNN structure in the deep unfolding approach and achieve near optimal performance of their benchmark iterative algorithm i.e, WMMSE and SCA respectively.

IX. Physical Layer Security

Physical layer security aims to ensure secure and reliable communication while preventing eavesdropping and impersonation attacks. Wireless communication's broadcast nature exposes it to adversaries like eavesdroppers, who intercept transmissions, and jammers, who degrade signal quality. The primary objective is to maximize the secrecy rate, ensuring high transmission rates for the intended user while minimizing information leakage to the eavesdropper. The secrecy capacity represents the maximum rate at which secure communication can be achieved, balancing efficiency and confidentiality. It is defined as the difference between mutual information between transmit and receive signals and information leakage to the eavesdropper [226].

Jamming attacks pose a serious threat to the reliable operation of wireless communication systems [226, 227]. A jamming attacker deliberately transmits interference or noise on the same frequencies as legitimate users. Marti *et al.* in [228] explored solutions to combat smart jamming in massive multiuser MIMO systems. The authors proposed an optimization-based approach for joint jammer estimation, channel estimation, and data detection, using the BS multiantenna configuration. This method does not require prior knowledge of the jammer's behavior, making communication more resilient even during sporadic or bursty jamming. The authors have unfolded an iterative algorithm, called as mitigation estimation and detection (MAED), that uses forward-backward splitting optimization to jointly estimate the jammer subspace, user equipment's channels, and user equipment's data within a coherence block. The soft-output MAED (SO-MAED), is the deep unfolded network. The trainable set include error precision (to represent noise in symbol estimation), gradient descent step-sizes, scaling factor and momentum weights. SO-MAED improves detection accuracy, especially in high-interference environments, and does not require knowledge of the attack type.

There is limited literature on the application of deep unfolding techniques in physical layer security. One possible reason is that physical layer security strategies are often dynamic and must adapt to adversarial attacks. In contrast, deep unfolding methods typically rely on a structured iterative algorithm which are seldom available in scenarios such as jamming attacks, where the system behavior must adjust rapidly to unpredictable threats.

X. Challenges in Developing Deep Unfolding Models and Future Research Scopes

The extensive review of recently emerging techniques and applications of deep unfolding have demonstrated its potential in addressing complex signal processing problems with reduced complexity and runtime. Despite these advantages, the adoption of deep unfolding in practical scenarios remains limited. In this section, we outline the main challenges in developing deep unfolding models, explore the reasons behind these challenges, and propose potential improvements to make deep unfolding a more powerful and widely applicable technique in wireless communications.

A. Limitations of Deep Unfolding Techniques

Although deep unfolding has found applications in various design problems, it has limitations. Most of these limitations arise from the fact that deep unfolding is not an independent design methodology; it must be built upon existing conventional optimization frameworks. Consequently, it faces the following challenges in the development and real-world applications:

(C1) **Methodological challenges:** Deep unfolding methods are inspired by the iterative procedures of conventional optimization algorithms, where each iteration is typically interpreted as a layer in an NN. In traditional wireless transmitter-receiver design and signal processing, optimization problems are often non-convex and highly coupled due to factors such as fractional SINR, constant-modulus constraints (e.g., for analog beamforming), or mixed-integer variables (e.g., in symbol detection or antenna selection). They are typically solved using a two-layer approach. In the first layer, AO or block coordinate ascent/descent is used to decouple variables, breaking the original problem into smaller, more manageable sub-problems. In the second layer, convex optimization techniques such as SCA or the ADMM are employed to solve each sub-problem. While this hierarchical structure enables the use of convex solvers (e.g., CVX), it often introduces auxiliary variables and results in nested loops that must converge at each layer, significantly increasing algorithmic complexity.

Deep unfolding models, on the other hand, aim to transform these iterative procedures into end-to-end trainable architectures. Instead of solving sub-problems sequentially, the unfolding framework typically approximates the entire optimization process in a more unified and parallelizable manner. This shift in design philosophy, from explicitly solving decoupled sub-problems to learning a unified mapping, poses non-trivial challenges, especially when attempting to translate complex multi-stage optimizers with auxiliary variables into a compact, learnable structure. Therefore, while deep unfolding builds upon conventional iterative methods, the methodological differences in structure and implementation represent a major challenge in developing effective unfolding models.

(C2) **Unfolding principle algorithms:** While deep unfolding typically enhances conventional iterative algorithms, not all can be effectively unfolded. This limitation arises from the core principle of deep unfolding, which structures the iterations of an iterative algorithm as layers within a DNN. Complex operations like matrix decomposition or the use of convex optimization tools such as CVX or YALMIP (with solvers like MOSEK, SeDuMi, etc.) are difficult to implement in a DNN framework. In large-scale systems such as massive MIMO, even matrix multiplications result in high complexity for unfolding models.

Another challenge is parameter generalization, particularly with respect to the number of layers in the unfolding model. In iterative algorithms with nested loops (discussed in ***C1***), both inner and outer loops run until convergence, but deep unfolding models require the number of layers to be predetermined before deployment. This is typically done through empirical tuning, which may not provide optimal values for all channel realizations. As a result, convergence is not guaranteed across different scenarios, potentially leading to performance degradation. Proper tuning is essential to achieve a favorable performance-complexity tradeoff in deep unfolding models.

(C3) **Handling design constraints and objectives:** One of the biggest challenges in developing deep unfolding models lies in managing the design constraints formulated in the problem. These constraints require the DNN layers' outputs to meet specific requirements, such as total power, QoS, or other element-wise constraints. Simple constraints, like power or constant modulus, can often be enforced through normalization [181]. For instance, the $tanh$ activation function can be employed to impose constant-modulus constraints on analog beamforming coefficients or sensing waveforms, similar to black-box DL architectures. However, QoS constraints are significantly more complex. While they resemble activation functions, none of the existing activation functions can adequately manage such intricate requirements. One approach is to incorporate these constraints into the design objective. However, this technique often results in an nonequivalent problem with a more complicated objective function [191].

Not only do complex constraints pose challenges for developing deep unfolding models, but intricate objective functions also present difficulties in training the models. This is because their loss functions are generally formulated based on the objective function. For instance, unfolding architectures like DetNet [49], FSNet [75], and ScNet [229] for symbol detection are trained to minimize the Euclidean distance between the model's output (symbol vector) and the true transmitted symbol vector. In particular, when labels are unavailable, as in beamforming problems, unfolding networks use the objective function as the loss function for minimization or its negative for maximization problems. Typical deep unfolding techniques [155, 191] require calculating the gradient of the objective function. However, this becomes challenging when the objective function is a complex, high-order, or fractional function of the design variables, especially when those variables are vectors or matrices.

(C4) **Sophisticated mappings:** While deep unfolding is powerful, certain inference tasks still require black-box DNN models, especially when there is no clear relationship between the input data, the output solution, and performance metrics. For instance, in multi-modal sensing-assisted beam tracking, tasks may involve complex mappings from environmental data, such as sensory information from sensors and cameras, to the optimal beam [230, 231]. In such cases, there is no explicit relationship between environmental images and the optimal beam, and standard communication metrics like throughput or SINR cannot be directly expressed as functions of these images. These problems cannot be formulated as optimization tasks solvable with principle algorithms, making it difficult to develop unfolding models that map environmental information to beam patterns or blockage states.

***(C5)* Real-world applications:** Deep unfolding architectures, while promising in mimicking iterative algorithms through trainable layers, often entail considerable depth and parameter complexity [191]. This imposes severe challenges for real-time inference under strict latency and throughput constraints, especially in systems with multiple and large dimensions like massive MIMO-OFDM. From a hardware perspective, deep unfolding designs are not naturally aligned with the requirements for area-efficient and parallelizable architectures needed in scalable OFDM deployments. For example, the non-uniform and sequential computations common in unfolding detectors can result in inefficient silicon utilization and high power consumption, limiting the ability to instantiate multiple parallel detection cores across subcarriers [232].

Furthermore, in real-world scenarios, the data distribution is highly dynamic because of user mobility, varying propagation environments, and hardware impairments. These variations often violate the assumptions underlying many unfolding-based transmitter and receiver designs, which are typically developed under idealized conditions. However, this assumption breaks down in practical deployments, where channels often exhibit strong line-of-sight components and spatial correlation. As a result, detectors relying on such assumptions can suffer significant performance degradation. To maintain robustness in these conditions, deep unfolding models must be redesigned or trained using data that captures the complexity and variability of realistic, deployment-specific channels.

The above five challenges (***C1–C5***) also outline the scenarios wherein deep unfolding models cannot be built or does not offer much gain in performance-complexity tradeoff. These include when there is lack of suitable conventional iterative algorithms (***C1***, ***C2***, ***C4***), when the design problems are highly complicated with constraints and objective function (***C3***), or when the exist imperfection in real-world propagation environments and hardware (***C5***). It can be seen that deep unfolding is not a game changer that will completely replace conventional principle algorithms or black-box ML models. However, the applicability of deep unfolding can be enhanced by overcoming those challenges, as will be discussed next.

B. Development of Deep Unfolding Techniques

1) Coexistence of Deep Unfolding Models, Conventional Solvers, and Black-Box Models: Deep unfolding doesn't need to fully replace conventional algorithms or black-box ML models. Instead, these approaches can coexist, leveraging their combined strengths. In hybrid designs, deep unfolding can (i) reduce the number of sub-problems, (ii) lower the complexity of solving each sub-problem, and (iii) enhance the learning process in black-box models. Benefits (i) and (ii) correspond to challenges ***C1*** and ***C2***, while (iii) relates to ***C4***.

Benefit (i) arises because deep unfolding can rely on techniques like PGA/PGD, which avoids transforming the original problem into convex sub-problems. Benefit (ii) comes from the fact that unfolded models typically exhibit lower complexity compared to methods like SCA, and their runtime is usually shorter than that of solvers like CVX. For example, in the AO framework (discussed in ***C1***), the most challenging sub-problem often dominates the overall complexity. By applying an efficient deep unfolding model to this sub-problem, the overall efficiency improves, while simpler sub-problems can be handled by conventional solvers like CVX. Benefit (iii) is frequently seen when deep unfolding is used to preprocess input data for black-box models, as in multimodal sensing applications. The combination of deep unfolding structures with conventional solvers and black-box models leads to simpler yet more efficient designs.

2) Deep Unfolding for Multi-Objective Learning: Most existing deep unfolding models are developed for single-objective designs, such as maximizing throughput, SINR, or minimizing BER. However, in many scenarios, system designs involve multiple objectives. Traditional approaches often formulate these as constrained problems, such as maximizing energy efficiency under a QoS constraint (e.g., throughput or SINR). In ISAC designs, weighted sums of communication and sensing performance metrics are used when the goal is to balance both functions without prioritizing one over the other. The complexity grows in multi-static ISAC systems, where distributed sensing receivers have different performance metrics depending on their functions, such as signal clutter noise ratio (SCNR), mutual information, CRB, or probability of detection.

These designs are often difficult to solve using traditional constrained formulations, as the problem can easily become infeasible if constraints are not properly tuned. However, by incorporating all design metrics into the objective function with appropriate weights, to form a multi-objective design, the feasibility of the problem can be significantly improved [191, 233]. This multiobjective formulation also makes it easier to develop unfolding models, optimizing multiple objectives simultaneously, and improving performance trade-offs. However, developing unfolding models for multi-objective designs presents several challenges. First, as the number of terms in the objective function increases, training the unfolding DNN becomes more difficult. This is because the loss function is typically set to mirror the objective function (or its negative in maximization problems). Furthermore, balancing the changes of these objectives with respect to the design variables across layers is crucial [191, 234], necessitating careful design of the unfolding models.

3) Distributed Deep Unfolding Architectures: While deep unfolding models provide advantages in terms of complexity due to faster convergence compared to conventional iterative algorithms, they can still face high complexity issues due to challenge ***C2***. This limits their applicability on platforms with constrained computational resources, such as edge devices or distributed access points (APs). Future wireless networks will advance beyond simple communications, evolving into large-scale, distributed systems such as cell-free massive MIMO that integrate advanced sensing, communications, and computing capabilities. However, most existing AI models, both data-driven and model-driven, were designed for centralized scenarios, where they are deployed at a central node with substantial computing power. These models are not well-suited for distributed scenarios, highlighting the need for appropriate distributed unfolding architectures.

One potential solution is to develop distributed deep unfolding architectures. In this approach, rather than executing all the unfolded layers at every distributed node or edge device, the layers can be distributed across multiple platforms for parallel execution, thereby reducing latency. For example, to design L beamformers for L APs for their downlink transmission in a distributed system, a CPU with high computing power and hardware resources can handle the majority of the layers, generating high-quality beamformers. These beamformers are then sent to the APs, serving as initial solutions and inputs for smaller sub-unfolding models at the distributed APs. This ensures that reliable beamformers can be generated at the APs with only a few layers, minimizing complexity and runtime to meet the computational constraints of the APs. Conversely, in the uplink, a distributed unfolding detection model can be deployed at the APs to generate initial detected symbols, which are then forwarded to the CPU for further processing and final decision-making by the large unfolding detector. By such an unfolding framework, highly reliable symbols are generated with low complexity at distributed APs.

4) Deep Unfolding to Enhance Black-box Solution: As its name suggests, a black-box ML model operates with a black-box nature, making it unexplainable and resistant to re-structuring beyond tuning parameters like the number of layers or nodes. The performance and efficiency of such models are heavily dependent on the quality of input data. Deep unfolding can contribute by performing pre-processing steps to improve the input data quality. One example is in multi-modal sensing-assisted communication tasks, where data from distributed cameras, LiDAR, and radar sensors are used to enhance communication performance, such as in beam tracking and blockage prediction. However, significant challenges arise in collecting, transferring, and processing large sensory data. Specifically, high-resolution camera images generate considerable traffic overhead and latency, impacting real-time communication processing. Deep unfolding models can be particularly useful in pre-processing such high resolution images. In image processing, deep unfolding has been successfully applied to tasks such as denoising, super-resolution, and image reconstruction, allowing for high-quality outcomes with fewer iterations than conventional algorithms. For example, unfolded algorithms based on ISTA for sparse coding have been shown to significantly accelerate image reconstruction while maintaining high-quality results [235–237]. This makes deep unfolding a valuable tool for handling large-scale sensory data in real-time communication systems.

XI. Conclusions

The integration of classical algorithmic principles with DL techniques has led to the development of deep unfolding, a method that overcomes the limitations associated with both classical algorithms and standalone DL models in wireless communication. This article offers a comprehensive survey of deep unfolding techniques, examining their application across various facets of the physical layer in wireless communication systems. Key applications, such as signal detection, channel estimation, precoder design, ISAC, decoding for error-correcting codes, power allocation, and physical layer security, are explored, with an emphasis on technologies anticipated for beyond-5G networks. The comparisons of the unfolded networks have also been given, which reflect the performance gains over the benchmark iterative algorithms. Also, we have analyzed the advantages and limitations of the design of various unfolding models. Moreover, we have explored the practical feasibility of deep unfolding networks across key application areas of wireless communications. The survey highlights the strengths of deep unfolding in addressing complex challenges, particularly in achieving low-latency, high-accuracy performance in dynamic environments. We conclude by discussing the key challenges in developing deep unfolding models, including issues related to scalability, computational efficiency, and robustness across diverse scenarios. Finally, this paper presents potential advancements to enhance the adaptability and resilience of deep unfolding, with the goal of solidifying its role as a robust and versatile technique for next-generation wireless communications.

References

[1] M. Agiwal, A. Roy, and N. Saxena, "Next Generation 5G Wireless Networks: A Comprehensive Survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1617–1655, 2016.

[2] W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, "The Road Towards 6G: A Comprehensive Survey," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 334–366, 2021.

[3] A. Chockalingam and B. S. Rajan, *Large MIMO Systems*. Cambridge University Press, 2014.

[4] I. J. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[5] H. Huang, S. Guo, G. Gui, Z. Yang, J. Zhang, H. Sari, and F. Adachi, "Deep Learning for Physical-Layer 5G Wireless Techniques: Opportunities, Challenges and Solutions," *IEEE Wirel. Commun.*, vol. 27, no. 1, pp. 214–222, 2020.

[6] H. Ye, G. Y. Li, and B.-H. Juang, "Power of Deep Learning for Channel Estimation and Signal Detection in OFDM Systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 114–117, 2018.

[7] M. B. Mashhadi and D. Gündüz, "Pruning the Pilots: Deep Learning-Based Pilot Design and Channel Estimation for MIMO-OFDM Systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 10, pp. 6315–6328, 2021.

[8] H. Huang, Y. Peng, J. Yang, W. Xia, and G. Gui, "Fast Beamforming Design via Deep Learning," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 1065–1069, 2020.

[9] Q. Bai, J. Wang, Y. Zhang, and J. Song, "Deep Learning-Based Channel Estimation Algorithm Over Time Selective Fading Channels," *IEEE Trans. Cogn. Commun. Netw.*, vol. 6, no. 1, pp. 125–134, 2020.

[10] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "A Model-Driven Deep Learning Network for MIMO Detection," in *2018 IEEE GlobalSIP*, 2018, pp. 584–588.

[11] ——, "Model-Driven Deep Learning for MIMO Detection," *IEEE Trans. Signal Process.*, vol. 68, pp. 1702–1715, 2020.

[12] A. G. D. G. Yonina C. Eldar and H. V. Poor, *Machine Learning and Wireless Communications*. Cambridge University Press, 2022.

[13] 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. 115 384–115 398, 2022.

[14] 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, 2021.

[15] A. Balatsoukas-Stimming and C. Studer, "Deep Unfolding for Communications Systems: A Survey and Some New Directions," in *2019 IEEE SiPS*, 2019, pp. 266–271.

[16] Q. Han, H. Zhang, Y. Li, D. Yuan, and X. Ma, "Physics Informed Digital Twin for RIS-Assisted Wireless Communication System," *IEEE Wirel. Commun.*, vol. 32, no. 3, pp. 106–112, 2025.

[17] J. Shi, Y. Liu, Z. Luo, Z. Li, C. Shen, J. Zhang, G. Wang, and N. Chi, "Simplified Neural Network With Physics-Informed Module in MIMO Visible Light Communication Systems," *J. Light. Technol.*, vol. 42, no. 1, pp. 57–68, 2024.

[18] A. Agrawal, S. Barratt, and S. Boyd, "Learning convex optimization models," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 8, pp. 1355–1364, 2021.

[19] A. Agrawal, S. Barratt, S. Boyd, and B. Stellato, "Learning Convex Optimization Control Policies," in *Proceedings of the 2nd Conf. on Learning for Dynamics and Control*, ser. Proceedings of Machine Learning Research, vol. 120. PMLR, 2020, pp. 361–373. [Online]. Available: https://proceedings.mlr.press/v120/agrawal20a.html

[20] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *27th ICML*, ser. ICML'10. Madison, WI, USA: Omnipress, 2010, p. 399–406.

[21] T. Wang, C.-K. Wen, H. Wang, F. Gao, T. Jiang, and S. Jin, "Deep Learning for Wireless Physical Layer: Opportunities and Challenges," *China Communications*, vol. 14, pp. 92–111, 10 2017.

[22] H. He, S. Jin, C.-K. Wen, F. Gao, G. Y. Li, and Z. Xu, "Model-Driven Deep Learning for Physical Layer Communications," *IEEE Wirel. Commun.*, vol. 26, no. 5, pp. 77–83, 2019.

[23] Z. Qin, H. Ye, G. Y. Li, and B.-H. F. Juang, "Deep Learning in Physical Layer Communications," *IEEE Wirel. Commun.*, vol. 26, no. 2, pp. 93–99, 2019.

[24] 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. Commun.*, vol. 67, no. 10, pp. 7331–7376, 2019.

[25] 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, 2021.

[26] Y. Shi, L. Lian, Y. Shi, Z. Wang, Y. Zhou, L. Fu, L. Bai, J. Zhang, and W. Zhang, "Machine Learning for Large-Scale Optimization in 6G Wireless Networks," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 4, pp. 2088–2132, 2023.

[27] S. Wu, S. Zheng, H. Jia, Z. Ji, A. Xiao, and C. Jiang, "Model and Data Dual-Driven: A Promising Receiver Design for 6G Wireless Communications," *IEEE Commun. Mag.*, vol. 63, no. 6, pp. 42–48, 2025.

[28] N. Shlezinger, M. Ma, O. Lavi, N. T. Nguyen, Y. C. Eldar, and M. Juntti, "Artificial Intelligence-Empowered Hybrid Multiple-Input/Multiple-Output Beamforming: Learning to Optimize for High-Throughput Scalable MIMO," *IEEE Veh. Technol. Mag.*, pp. 2–11, 2024.

[29] Q.-V. Pham, N. T. Nguyen, T. Huynh-The, L. Bao Le, K. Lee, and W.-J. Hwang, "Intelligent Radio Signal Processing: A Survey," *IEEE Access*, vol. 9, pp. 83 818–83 850, 2021.

[30] N. Shlezinger, J. Whang, Y. C. Eldar, and A. G. Dimakis, "Model-Based Deep Learning: Key Approaches and Design Guidelines," in *2021 IEEE DSLW*, 2021, pp. 1–6.

[31] Y. Qiang, X. Shao, and X. Chen, "A Model-Driven Deep Learning Algorithm for Joint Activity Detection and Channel Estimation," *IEEE Commun. Lett.*, vol. 24, no. 11, pp. 2508–2512, 2020.

[32] J. Liao, J. Zhao, F. Gao, and G. Y. Li, "A Model-Driven Deep Learning Method for Massive MIMO Detection," *IEEE Commun. Lett.*, vol. 24, no. 8, pp. 1724–1728, 2020.

[33] J. Guo, L. Wang, F. Li, and J. Xue, "CSI Feedback With Model-Driven Deep Learning of Massive MIMO Systems," *IEEE Commun. Lett.*, vol. 26, no. 3, pp. 547–551, 2022.

[34] S. Li, C. Ding, L. Xiao, X. Zhang, G. Liu, and T. Jiang, "Expectation Propagation Aided Model Driven Learning for OTFS Signal Detection," *IEEE Trans. Veh. Technol.*, vol. 72, no. 9, pp. 12 407–12 412, 2023.

[35] H. Zhao, C. Yang, Y. Xu, F. Ji, M. Wen, and Y. Chen, "Model-Driven Based Deep Unfolding Equalizer for Underwater Acoustic OFDM Communications," *IEEE Trans. Veh. Technol.*, vol. 72, no. 5, pp. 6056–6067, 2023.

[36] Y. He, J. Zhang, S. Jin, C.-K. Wen, and G. Y. Li, "Model-Driven DNN Decoder for Turbo Codes: Design, Simulation, and Experimental Results," *IEEE Trans. Commun.*, vol. 68, no. 10, pp. 6127–6140, 2020.

[37] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011. [Online]. Available: http://dx.doi.org/10.1561/2200000016

[38] M. Kim and D. Park, "Learnable MIMO Detection Networks Based on Inexact ADMM," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 565–576, 2021.

[39] Y. Wei, M.-M. Zhao, M.-J. Zhao, and M. Lei, "ADMM-Based Decoder for Binary Linear Codes Aided by Deep Learning," *IEEE Commun. Lett.*, vol. 24, no. 5, pp. 1028–1032, 2020.

[40] X. Guo, T.-H. Chang, and Y. Wang, "Model-Driven Deep Learning ADMM Decoder for Irregular Binary LDPC Codes," *IEEE Commun. Lett.*, vol. 27, no. 2, pp. 571–575, 2023.

[41] Q. Zhang, J. Wang, and Y. Wang, "Efficient QAM Signal Detector for Massive MIMO Systems via PS/DPS-ADMM Approaches," *IEEE Trans. Wireless Commun.*, vol. 21, no. 10, pp. 8859–8871, 2022.

[42] X. Li, C. Shan, H. Zhao, W. Yuan, and R. Zhang, "Iterative Detection for Orthogonal Time Frequency Space Modulation with ADMM," *IEEE Commun. Lett.*, pp. 1–1, 2025.

[43] M. A. Albreem, M. Juntti, and S. Shahabuddin, "Massive MIMO Detection Techniques: A Survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3109–3132, 2019.

[44] S. Yang and L. Hanzo, "Fifty Years of MIMO Detection: The Road to Large-Scale MIMOs," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 1941–1988, 2015.

[45] S. T. McCormick, B. Peis, R. Scheidweiler, and F. Vallentin, "A Polynomial Time Algorithm for Solving the Closest Vector Problem in Zonotopal Lattices," *SIAM Journal on Discrete Mathematics*, vol. 35, no. 4, pp. 2345–2356, 2021.

[46] B. Hassibi and H. Vikalo, "On the sphere decoding algorithm I. Expected complexity," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 2806–2818, 2005.

[47] ——, "On the sphere decoding algorithm II. Generalizations, second-order statistics, and applications to communications," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 2819–2834, 2005.

[48] S. Takabe, M. Imanishi, T. Wadayama, and K. Hayashi, "Deep Learning-Aided Projected Gradient Detector for Massive Overloaded MIMO Channels," in *ICC 2019*, 2019, pp. 1–6.

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

[50] S. Takabe, M. Imanishi, T. Wadayama, R. Hayakawa, and K. Hayashi, "Trainable Projected Gradient Detector for Massive Overloaded MIMO Channels: Data-Driven Tuning Approach," *IEEE Access*, vol. 7, pp. 93 326–93 338, 2019.

[51] S. Yun, S. Moon, Y.-S. Jeon, and Y. Lee, "Intelligent MIMO Detection With Momentum-Induced Unfolded Layers," *IEEE Wireless Commun. Lett.*, vol. 13, no. 3, pp. 879–883, 2024.

[52] N. Qian, "On the momentum term in gradient descent learning algorithms," *Neural Networks*, vol. 12, no. 1, pp. 145–151, 1999.

[53] S. Takabe and T. Abe, "Hubbard–Stratonovich Detector for Simple Trainable MIMO Signal Detection," *IEEE Wireless Commun. Lett.*, vol. 13, no. 3, pp. 701–705, 2024.

[54] M. Khani, M. Alizadeh, J. Hoydis, and P. Fleming, "Adaptive Neural Signal Detection for Massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5635–5648, 2020.

[55] E. Beck, C. Bockelmann, and A. Dekorsy, "CMDNet: Learning a Probabilistic Relaxation of Discrete Variables for Soft Detection With Low Complexity," *IEEE Trans. Commun.*, vol. 69, no. 12, pp. 8214–8227, 2021.

[56] C. J. Maddison, A. Mnih, and Y. W. Teh, "The concrete distribution: A continuous relaxation of discrete random variables," *arXiv preprint arXiv:1611.00712*, 2016.

[57] Y. Wei, M.-M. Zhao, M. Hong, M.-J. Zhao, and M. Lei, "Learned Conjugate Gradient Descent Network for Massive MIMO Detection," *IEEE Trans. Signal Process.*, vol. 68, pp. 6336–6349, 2020.

[58] T. Olutayo and B. Champagne, "Dynamic Conjugate Gradient Unfolding for Symbol Detection in Time-Varying Massive MIMO," *IEEE Open J. Veh. Technol.*, vol. 5, pp. 792–806, 2024.

[59] K. Chen, *Matrix Preconditioning Techniques and Applications*. Cambridge University Press, 2005.

[60] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What Will 5G Be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, 2014.

[61] S. Khobahi, N. Shlezinger, M. Soltanalian, and Y. C. Eldar, "LoRD-Net: Unfolded Deep Detection Network With Low-Resolution Receivers," *IEEE Trans. Signal Process.*, vol. 69, pp. 5651–5664, 2021.

[62] J. Choi, J. Mo, and R. W. Heath, "Near Maximum-Likelihood Detector and Channel Estimator for Uplink Multiuser Massive MIMO Systems With One-Bit ADCs," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 2005–2018, 2016.

[63] N. Shlezinger, R. Fu, and Y. C. Eldar, "DeepSIC: Deep Soft Interference Cancellation for Multiuser MIMO Detection," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1349–1362, 2021.

[64] L. He, Z. Wang, S. Yang, T. Liu, and Y. Huang, "Generalizing Projected Gradient Descent for Deep-Learning-Aided Massive MIMO Detection," *IEEE Trans. Wireless Commun.*, vol. 23, no. 3, pp. 1827–1839, 2024.

[65] J. Céspedes, P. M. Olmos, M. Sánchez-Fernández, and F. Perez-Cruz, "Expectation Propagation Detection for High-Order High-Dimensional MIMO Systems," *IEEE Trans. Commun.*, vol. 62, no. 8, pp. 2840–2849, 2014.

[66] I. Santos, J. J. Murillo-Fuentes, R. Boloix-Tortosa, E. Arias-de Reyna, and P. M. Olmos, "Expectation Propagation as Turbo Equalizer in ISI Channels," *IEEE Trans. Commun.*, vol. 65, no. 1, pp. 360–370, 2017.

[67] X. Tan, W. Xu, K. Sun, Y. Xu, Y. Be'ery, X. You, and C. Zhang, "Improving Massive MIMO Message Passing Detectors With Deep Neural Network," *IEEE Trans. Veh. Technol.*, vol. 69, no. 2, pp. 1267–1280, 2020.

[68] H. He, X. Yu, J. Zhang, S. Song, and K. B. Letaief, "Message Passing Meets Graph Neural Networks: A New Paradigm for Massive MIMO Systems," *IEEE Trans. Wireless Commun.*, vol. 23, no. 5, pp. 4709–4723, 2024.

[69] Y. Zeng, Y. Ge, X. Tan, Z. Ji, Z. Zhang, X. You, and C. Zhang, "A Deep-Learning-Aided Message Passing Detector for MIMO SC-FDMA," *IEEE Trans. Veh. Technol.*, vol. 73, no. 7, pp. 10 767–10 771, 2024.

[70] Y. Zhao, Y. Xiao, P. Yang, B. Dong, R. Shi, and K. Deng, "Generalized Approximate Message Passing Aided Frequency Domain Turbo Equalizer for Single-Carrier Spatial Modulation," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3630–3634, 2018.

[71] H. Chen, G. Yao, and J. Hu, "Algorithm Parameters Selection Method With Deep Learning for EP MIMO Detector," *IEEE Trans. Veh. Technol.*, vol. 70, no. 10, pp. 10 146–10 156, 2021.

[72] Y. Ge, X. Tan, Z. Ji, Z. Zhang, X. You, and C. Zhang, "Improving Approximate Expectation Propagation Massive MIMO Detector With Deep Learning," *IEEE Wireless Commun. Lett.*, vol. 10, no. 10, pp. 2145–2149, 2021.

[73] X. Tan, Y.-L. Ueng, Z. Zhang, X. You, and C. Zhang, "A Low-Complexity Massive MIMO Detection Based on Approximate Expectation Propagation," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 7260–7272, 2019.

[74] N. T. Nguyen, K. Lee, and H. DaiIEEE, "Application of Deep Learning to Sphere Decoding for Large MIMO Systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 10, pp. 6787–6803, 2021.

[75] 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, 2020.

[76] N. Zilberstein, C. Dick, R. Doost-Mohammady, A. Sabharwal, and S. Segarra, "Annealed Langevin Dynamics for Massive MIMO Detection," *IEEE Trans. Wireless Commun.*, vol. 22, no. 6, pp. 3762–3776, 2023.

[77] S. Berra, S. Chakraborty, R. Dinis, and S. Shahabuddin, "Deep Unfolding of Chebyshev Accelerated Iterative Method for Massive MIMO Detection," *IEEE Access*, vol. 11, pp. 52 555–52 569, 2023.

[78] A. Datta, A. Nema, and V. Bhatia, "Deep Unfolded Sparse Refinement Network Based Detection in Uplink Massive MIMO," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6825–6830, 2022.

[79] M. Mandloi and V. Bhatia, "Low-Complexity Near-Optimal Iterative Sequential Detection for Uplink Massive MIMO Systems," *IEEE Commun. Lett.*, vol. 21, no. 3, pp. 568–571, 2017.

[80] N. Parikh and S. Boyd, *Proximal Algorithms*. Now Foundations and Trends, 2014.
[81] Y. Sun, H. Shen, W. Xu, N. Hu, and C. Zhao, "Robust MIMO Detection With Imperfect CSI: A Neural Network Solution," *IEEE Trans. Commun.*, vol. 71, no. 10, pp. 5877–5892, 2023.
[82] W. Jiang and F. L. Luo, *6G Key Technologies: A Comprehensive Guide*. Wiley-IEEE Press, 2023.
[83] S. Weinstein and P. Ebert, "Data Transmission by Frequency-Division Multiplexing Using the Discrete Fourier Transform," *IEEE Trans. Commun.*, vol. 19, no. 5, pp. 628–634, 1971.
[84] J. Zhang, C.-K. Wen, S. Jin, and G. Y. Li, "Artificial Intelligence-Aided Receiver for a CP-Free OFDM System: Design, Simulation, and Experimental Test," *IEEE Access*, vol. 7, pp. 58 901–58 914, 2019.
[85] M. Shao, W.-K. Ma, J. Liu, and Z. Huang, "Accelerated and Deep Expectation Maximization for One-Bit MIMO-OFDM Detection," *IEEE Trans. Signal Process.*, vol. 72, pp. 1094–1113, 2024.
[86] A. Ullah, W. Choi, T. M. Berhane, Y. Sambo, and M. A. Imran, "Soft-Output Deep LAS Detection for Coded MIMO Systems: A Learning-Aided LLR Approximation," *IEEE Trans. Veh. Technol.*, pp. 1–15, 2024.
[87] J. Zhang, C.-K. Wen, and S. Jin, "Adaptive MIMO Detector Based on Hypernetwork: Design, Simulation, and Experimental Test," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 1, pp. 65–81, 2022.
[88] E. Sozer, M. Stojanovic, and J. Proakis, "Underwater acoustic networks," *IEEE J. Ocean. Eng.*, vol. 25, no. 1, pp. 72–83, 2000.
[89] S. Jiang, "On Reliable Data Transfer in Underwater Acoustic Networks: A Survey From Networking Perspective," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 2, pp. 1036–1055, 2018.
[90] H. Lin, T. Guo, G. Peng, Q. Zeng, and X. Zhong, "DNN-VTRM Channel Estimation for OFDM Underwater Acoustic Communication System," *IEEE Access*, vol. 12, pp. 194 946–194 961, 2024.
[91] Y. Yue, J. Shi, Z. Li, J. Hu, and Z. Tie, "Model-Driven Deep Learning Assisted Detector for OTFS With Channel Estimation Error," *IEEE Commun. Lett.*, vol. 28, no. 4, pp. 842–846, 2024.
[92] B. Clerckx, Y. Mao, Z. Yang, M. Chen, A. Alkhateeb, L. Liu, M. Qiu, J. Yuan, V. W. S. Wong, and J. Montojo, "Multiple Access Techniques for Intelligent and Multifunctional 6G: Tutorial, Survey, and Outlook," *Proc. IEEE*, vol. 112, no. 7, pp. 832–879, 2024.
[93] S. M. R. Islam, N. Avazov, O. A. Dobre, and K.-s. Kwak, "Power-Domain Non-Orthogonal Multiple Access (NOMA) in 5G Systems: Potentials and Challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 721–742, 2017.
[94] X. Wang, D. Zhang, B. Chen, T. Liu, Y. Xin, and Y. Xu, "Deep Transfer Learning for Model-Driven Signal Detection in MIMO-NOMA Systems," *IEEE Trans. Veh. Technol.*, vol. 72, no. 10, pp. 13 039–13 054, 2023.
[95] J. Liu and X. Wang, "Tanner-Graph-Based Massive Multiple Access—Transmission and Decoding Schemes," *IEEE Trans. Wireless Commun.*, vol. 22, no. 11, pp. 8468–8482, 2023.
[96] M. Yoshida and T. Tanaka, "Analysis of Sparsely-Spread CDMA via Statistical Mechanics," in *2006 IEEE ISIT*, 2006, pp. 2378–2382.
[97] J. Choi, "Low density spreading for multicarrier systems," in *Eighth IEEE ISSSTA*, 2004, pp. 575–578.
[98] S. Takabe, Y. Yamauchi, and T. Wadayama, "Deep-Unfolded Sparse CDMA: Multiuser Detector and Sparse Signature Design," *IEEE Access*, vol. 9, pp. 40 027–40 038, 2021.
[99] C. Qing, B. Cai, Q. Yang, J. Wang, and C. Huang, "Deep Learning for CSI Feedback Based on Superimposed Coding," *IEEE Access*, vol. 7, pp. 93 723–93 733, 2019.
[100] W. Chen, B. Zhang, S. Jin, B. Ai, and Z. Zhong, "Solving Sparse Linear Inverse Problems in Communication Systems: A Deep Learning Approach With Adaptive Depth," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 4–17, 2021.
[101] Z. Liu, L. Wang, L. Xu, and Z. Ding, "Deep Learning for Efficient CSI Feedback in Massive MIMO: Adapting to New Environments and Small Datasets," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2024.
[102] F. Liu, X. Shang, and H. Zhu, "Efficient Majorization-Minimization-Based Channel Estimation for One-Bit Massive MIMO Systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3444–3457, 2021.
[103] Y. Zhao, Y. Teng, A. Liu, X. Wang, and V. K. N. Lau, "Joint UL/DL Dictionary Learning and Channel Estimation via Two-Timescale Optimization in Massive MIMO Systems," *IEEE Trans. Wireless Commun.*, vol. 23, no. 3, pp. 2369–2382, 2024.
[104] T. Yassine and L. Le Magoarou, "mpNet: Variable Depth Unfolded Neural Network for Massive MIMO Channel Estimation," *IEEE Trans. Wireless Commun.*, vol. 21, no. 7, pp. 5703–5714, 2022.
[105] P. Wu and J. Cheng, "Deep Unfolding Basis Pursuit: Improving Sparse Channel Reconstruction via Data-Driven Measurement Matrices," *IEEE Trans. Wireless Commun.*, vol. 21, no. 10, pp. 8090–8105, 2022.
[106] S. Ji, Y. Xue, and L. Carin, "Bayesian Compressive Sensing," *Trans. Sig. Proc.*, vol. 56, no. 6, p. 2346–2356, jun 2008. [Online]. Available: https://doi.org/10.1109/TSP.2007.914345
[107] Q. Hu, S. Shi, Y. Cai, and G. Yu, "DDPG-Driven Deep-Unfolding With Adaptive Depth for Channel Estimation With Sparse Bayesian Learning," *IEEE Trans. Signal Process.*, vol. 70, pp. 4665–4680, 2022.
[108] K. K. Nguyen, T. Q. Duong, N. A. Vien, N.-A. Le-Khac, and L. D. Nguyen, "Distributed Deep Deterministic Policy Gradient for Power Allocation Control in D2D-Based V2V Communications," *IEEE Access*, vol. 7, pp. 164 533–164 543, 2019.
[109] T. Kebede, Y. Wondie, J. Steinbrunn, H. B. Kassa, and K. T. Kornegay, "Precoding and Beamforming Techniques in mmWave-Massive MIMO: Performance Assessment," *IEEE Access*, vol. 10, pp. 16 365–16 387, 2022.
[110] K. Hassan, M. Masarra, M. Zwingelstein, and I. Dayoub, "Channel Estimation Techniques for Millimeter-Wave Communication Systems: Achievements and Challenges," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 1336–1363, 2020.
[111] M. Borgerding and P. Schniter, "Onsager-corrected deep learning for sparse linear inverse problems," in *2016 IEEE GlobalSIP*, 2016, pp. 227–231.
[112] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: I. motivation and construction," in *2010 IEEE ITW*, 2010, pp. 1–5.
[113] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Deep Learning-Based Channel Estimation for Beamspace mmWave Massive MIMO Systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 852–855, 2018.
[114] Y. Wei, M.-M. Zhao, M. Zhao, M. Lei, and Q. Yu, "An AMP-Based Network With Deep Residual Learning for mmWave Beamspace Channel Estimation," *IEEE Wireless Commun. Lett.*, vol. 8, no. 4, pp. 1289–1292, 2019.
[115] P. Zheng, X. Lyu, and Y. Gong, "Trainable Proximal Gradient Descent-Based Channel Estimation for mmWave Massive MIMO Systems," *IEEE Wireless Commun. Lett.*, vol. 12, no. 10, pp. 1781–1785, 2023.
[116] P. Wu, J. Cheng, Y. C. Eldar, and J. M. Cioffi, "Learned Trimmed-Ridge Regression for Channel Estimation in Millimeter-Wave Massive MIMO," *IEEE Trans. Commun.*, pp. 1–1, 2024.
[117] Y. Wang, X. Chen, H. Yin, and W. Wang, "Learnable Sparse Transformation-Based Massive MIMO CSI Recovery Network," *IEEE Commun. Lett.*, vol. 24, no. 7, pp. 1468–1471, 2020.
[118] Z. Hu, G. Liu, Q. Xie, J. Xue, D. Meng, and D. Gündüz, "A Learnable Optimization and Regularization Approach to Massive MIMO CSI Feedback," *IEEE Trans. Wireless Commun.*, vol. 23, no. 1, pp. 104–116, 2024.
[119] Y. Zhang, X. Zhang, and Y. Liu, "Lightweight Differential Frameworks for CSI Feedback in Time-Varying Massive MIMO Systems," *IEEE Trans. Veh. Technol.*, vol. 73, no. 5, pp. 6878–6893, 2024.
[120] Z. Cao, J. Guo, C.-K. Wen, and S. Jin, "Deep-Unfolding-Based Bit-Level CSI Feedback in Massive MIMO Systems," *IEEE Wireless Commun. Lett.*, vol. 12, no. 2, pp. 371–375, 2023.
[121] J. Liu, X. Chen, Z. Wang, and W. Yin, "ALISTA: Analytic Weights Are As Good As Learned Weights in LISTA," in *International Conference on Learning Representations*, 2019. [Online]. Available: https://openreview.net/forum?id=B1lnzn0ctQ
[122] J. Gao, C. Zhong, G. Y. Li, J. B. Soriaga, and A. Behboodi, "Deep Learning-Based Channel Estimation for Wideband Hybrid MmWave Massive MIMO," *IEEE Trans. Commun.*, vol. 71, no. 6, pp. 3679–3693, 2023.
[123] J. Chen, F. Meng, N. Ma, X. Xu, and P. Zhang, "Model-Driven Deep Learning-Based Sparse Channel Representation and Recovery for Wideband mmWave Massive MIMO Systems," *IEEE Trans. Veh. Technol.*, vol. 72, no. 12, pp. 16 788–16 793, 2023.
[124] H. He, R. Wang, W. Jin, S. Jin, C.-K. Wen, and G. Y. Li, "Beamspace Channel Estimation for Wideband Millimeter-Wave MIMO: A Model-Driven Unsupervised Learning Approach," *IEEE Trans. Wireless Commun.*, vol. 22, no. 3, pp. 1808–1822, 2023.
[125] H. He, C.-K. Wen, and S. Jin, "Generalized Expectation Consistent Signal Recovery for Nonlinear Measurements," in *2017 IEEE ISIT*, 2017, pp. 2333–2337.
[126] X. Ma, Z. Gao, F. Gao, and M. Di Renzo, "Model-Driven Deep Learning Based Channel Estimation and Feedback for Millimeter-Wave Massive Hybrid MIMO Systems," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 8, pp. 2388–2406, 2021.
[127] X. Gao, S. Jin, C.-K. Wen, and G. Y. Li, "ComNet: Combination of Deep Learning and Expert Knowledge in OFDM Receivers," *IEEE Commun. Lett.*, vol. 22, no. 12, pp. 2627–2630, 2018.
[128] Q. Wang, X. Chen, Q. Tao, L. P. Qian, P.-Y. Kam, and Y. Wu, "Model-Driven Channel Estimation Network for Orthogonal Time-Frequency Space Systems," *IEEE Trans. Veh. Technol.*, pp. 1–5, 2025.
[129] J. He, H. Wymeersch, M. Di Renzo, and M. Juntti, "Learning to Estimate RIS-Aided mmWave Channels," *IEEE Wireless Commun. Lett.*, vol. 11, no. 4, pp. 841–845, 2022.
[130] M. Awais, M. A. Khan, and Y. H. Kim, "Deep Denoising and Unfolding for IRS Cascaded Channel Estimation With Element-Grouping," *IEEE Wireless Commun. Lett.*, vol. 12, no. 10, pp. 1726–1730, 2023.
[131] W.-C. Tsai, C.-W. Chen, and A.-Y. Wu, "Deep Unfolding-Based Channel Estimation for IRS-Aided mmWave Systems via Two-Stage LAMP Network with Row Compression," *IEEE Trans. Veh. Technol.*, pp. 1–14, 2024.
[132] T. Li, Y. Jiang, Y. Huang, P. Zhu, F.-C. Zheng, and D. Wang, "Model-Based Deep Learning for Massive Access in mmWave Cell-Free Massive MIMO System," in *2024 IEEE ICC Workshops*, 2024, pp. 828–833.
[133] Z. Ma, W. Wu, F. Gao, and X. Shen, "Model-Driven Deep Learning for Non-Coherent Massive Machine-Type Communications," *IEEE Trans. Wireless Commun.*, vol. 23, no. 3, pp. 2197–2211, 2024.
[134] Y. Bai, W. Chen, B. Ai, and P. Popovski, "Deep Learning for Asyn-

chronous Massive Access With Data Frame Length Diversity," *IEEE Trans. Wireless Commun.*, vol. 23, no. 6, pp. 5529–5540, 2024.

[135] Z. Mao, X. Liu, M. Peng, Z. Chen, and G. Wei, "Joint Channel Estimation and Active-User Detection for Massive Access in Internet of Things—A Deep Learning Approach," *IEEE Internet Things J.*, vol. 9, no. 4, pp. 2870–2881, 2022.

[136] X. Dang, W. Xiang, L. Yuan, Y. Yang, E. Wang, and T. Huang, "Deep Unfolding Scheme for Grant-Free Massive-Access Vehicular Networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 12, pp. 14 443–14 452, 2023.

[137] J. Johnston and X. Wang, "Model-Based Deep Learning for Joint Activity Detection and Channel Estimation in Massive and Sporadic Connectivity," *IEEE Trans. Wireless Commun.*, vol. 21, no. 11, pp. 9806–9817, 2022.

[138] Y. Shi, H. Choi, Y. Shi, and Y. Zhou, "Algorithm Unrolling for Massive Access via Deep Neural Networks With Theoretical Guarantee," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 945–959, 2022.

[139] Y. Zou, Y. Zhou, X. Chen, and Y. C. Eldar, "Proximal Gradient-Based Unfolding for Massive Random Access in IoT Networks," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2024.

[140] I. Ahmed, H. Khammari, A. Shahid, A. Musa, K. S. Kim, E. De Poorter, and I. Moerman, "A Survey on Hybrid Beamforming Techniques in 5G: Architecture and System Model Perspectives," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3060–3097, 2018.

[141] A. F. Molisch, V. V. Ratnam, S. Han, Z. Li, S. L. H. Nguyen, L. Li, and K. Haneda, "Hybrid Beamforming for Massive MIMO: A Survey," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 134–141, 2017.

[142] R. Heath, S. Sandhu, and A. Paulraj, "Antenna Selection for Spatial Multiplexing Systems with :Linear Receivers," *IEEE Commun. Lett.*, vol. 5, no. 4, pp. 142–144, 2001.

[143] 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, 2016.

[144] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An Iteratively Weighted MMSE Approach to Distributed Sum-Utility Maximization for a MIMO Interfering Broadcast Channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, 2011.

[145] Y.-F. Liu, Y.-H. Dai, and Z.-Q. Luo, "Coordinated Beamforming for MISO Interference Channel: Complexity Analysis and Efficient Algorithms," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 1142–1157, 2011.

[146] Q. Hu, Y. Cai, Q. Shi, K. Xu, G. Yu, and Z. Ding, "Iterative Algorithm Induced Deep-Unfolding Neural Networks: Precoding Design for Multiuser MIMO Systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1394–1410, 2021.

[147] L. Pellaco, M. Bengtsson, and J. Jaldén, "Matrix-Inverse-Free Deep Unfolding of the Weighted MMSE Beamforming Algorithm," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 65–81, 2022.

[148] L. Schynol and M. Pesavento, "Coordinated Sum-Rate Maximization in Multicell MU-MIMO With Deep Unrolling," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 4, pp. 1120–1134, 2023.

[149] A. Chowdhury, G. Verma, A. Swami, and S. Segarra, "Deep Graph Unfolding for Beamforming in MU-MIMO Interference Networks," *IEEE Trans. Wireless Commun.*, vol. 23, no. 5, pp. 4889–4903, 2024.

[150] C. Yi, W. Xu, Y. Sun, D. W. K. Ng, and Z. Wang, "Distributed Multiuser Precoding via Unfolding WMMSE for Multicell MIMO Downlinks," *IEEE Trans. Cogn. Commun. Netw.*, pp. 1–1, 2024.

[151] K. Wang and A. Liu, "Robust WMMSE-Based Precoder With Practice-Oriented Design for Massive MU-MIMO," *IEEE Wireless Commun. Lett.*, vol. 13, no. 7, pp. 1858–1862, 2024.

[152] 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, 2023.

[153] Q. Xu, J. Sun, and Z. Xu, "Robust Frequency Selective Precoding for Downlink Massive MIMO in 5G Broadband System," *IEEE Trans. Veh. Technol.*, vol. 72, no. 12, pp. 15 941–15 952, 2023.

[154] M. Zhu, T.-H. Chang, and M. Hong, "Learning to Beamform in Heterogeneous Massive MIMO Networks," *IEEE Trans. Wireless Commun.*, vol. 22, no. 7, pp. 4901–4915, 2023.

[155] O. Lavi and N. Shlezinger, "Learn to Rapidly and Robustly Optimize Hybrid Precoding," *IEEE Trans. Commun.*, vol. 71, no. 10, pp. 5814–5830, 2023.

[156] A. Juditsky, A. S. Nemirovskii, and C. Tauvel, "Solving variational inequalities with Stochastic Mirror-Prox algorithm," 2011. [Online]. Available: https://arxiv.org/abs/0809.0815

[157] K. K. Thekumprampil, P. Jain, P. Netrapalli, and S. Oh, *Efficient algorithms for smooth minimax optimization*. Curran Associates Inc., 2019.

[158] I. Bilbao, E. Iradier, J. Montalbán, P. Angueira, N. Thanh Nguyen, and M. Juntti, "Deep Unfolding-Powered Analog Beamforming for In-Band Full-Duplex," *IEEE Open J. Commun. Soc.*, vol. 5, pp. 3753–3761, 2024.

[159] Y. He, H. He, C.-K. Wen, and S. Jin, "Model-Driven Deep Learning for Massive Multiuser MIMO Constant Envelope Precoding," *IEEE Wireless Commun. Lett.*, vol. 9, no. 11, pp. 1835–1839, 2020.

[160] J. Xu, C. Kang, J. Xue, and Y. Zhang, "A Fast Deep Unfolding Learning Framework for Robust MU-MIMO Downlink Precoding," *IEEE Trans. Cogn. Commun. Netw.*, vol. 9, no. 2, pp. 359–372, 2023.

[161] A. Mohammad, C. Masouros, and Y. Andreopoulos, "An Unsupervised Deep Unfolding Framework for Robust Symbol-Level Precoding," *IEEE Open J. Commun. Soc.*, vol. 4, pp. 1075–1090, 2023.

[162] J. Yang, A. Li, X. Liao, and C. Masouros, "ADMM-SLPNet: A Model-Driven Deep Learning Framework for Symbol-Level Precoding," *IEEE Trans. Veh. Technol.*, vol. 73, no. 1, pp. 1376–1381, 2024.

[163] C.-J. Wang, C.-K. Wen, S. Jin, and S.-H. Tsai, "Finite-Alphabet Precoding for Massive MU-MIMO With Low-Resolution DACs," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4706–4720, 2018.

[164] H. He, M. Zhang, S. Jin, C.-K. Wen, and G. Y. Li, "Model-Driven Deep Learning for Massive MU-MIMO With Finite-Alphabet Precoding," *IEEE Commun. Lett.*, vol. 24, no. 10, pp. 2216–2220, 2020.

[165] Z. Gao, S. Liu, Y. Su, Z. Li, and D. Zheng, "Hybrid Knowledge-Data Driven Channel Semantic Acquisition and Beamforming for Cell-Free Massive MIMO," *IEEE J. Sel. Topics Signal Process.*, vol. 17, no. 5, pp. 964–979, 2023.

[166] 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, 2021.

[167] Q. Lin, H. Shen, and C. Zhao, "Learning Linear MMSE Precoder for Uplink Massive MIMO Systems With One-Bit ADCs," *IEEE Wireless Commun. Lett.*, vol. 11, no. 10, pp. 2235–2239, 2022.

[168] 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, 2021.

[169] Y. Mao, B. Clerckx, and V. O. K. Li, "Rate-splitting Multiple Access for Downlink Communication Systems: Bridging, Generalizing, and Outperforming SDMA and NOMA," *EURASIP J Wirel Commun Netw*, vol. 1, p. 133, 2018.

[170] Y. Wang, Y. Mao, and S. Ji, "RS-BNN: A Deep Learning Framework for the Optimal Beamforming Design of Rate-Splitting Multiple Access," *IEEE Trans. Veh. Technol.*, pp. 1–6, 2024.

[171] K.-M. Chen, H.-Y. Chang, R. Y. Chang, and W.-H. Chung, "Deep Unfolded Hybrid Beamforming in Reconfigurable Intelligent Surface Aided mmWave MIMO-OFDM Systems," *IEEE Wireless Commun. Lett.*, vol. 13, no. 4, pp. 1118–1122, 2024.

[172] X. Zhao, T. Lin, Y. Zhu, and J. Zhang, "Partially-Connected Hybrid Beamforming for Spectral Efficiency Maximization via a Weighted MMSE Equivalence," *IEEE Trans. Wireless Commun.*, vol. 20, no. 12, pp. 8218–8232, 2021.

[173] Y. Liu, Q. Hu, Y. Cai, G. Yu, and G. Y. Li, "Deep-Unfolding Beamforming for Intelligent Reflecting Surface Assisted Full-Duplex Systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 7, pp. 4784–4800, 2022.

[174] M.-M. Zhao, A. Liu, Y. Wan, and R. Zhang, "Two-Timescale Beamforming Optimization for Intelligent Reflecting Surface Aided Multiuser Communication With QoS Constraints," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 6179–6194, 2021.

[175] W. Xia, Y. Jiang, B. Zhao, H. Zhao, and H. Zhu, "Deep Unfolded Fractional Programming-Based Beamforming in RIS-Aided MISO Systems," *IEEE Wireless Commun. Lett.*, vol. 13, no. 2, pp. 515–519, 2024.

[176] H. Guo, Y.-C. Liang, J. Chen, and E. G. Larsson, "Weighted Sum-Rate Maximization for Reconfigurable Intelligent Surface Aided Wireless Networks," *IEEE Trans. Wirel. Commun.*, vol. 19, no. 5, pp. 3064–3076, 2020.

[177] W. Jin, J. Zhang, C.-K. Wen, S. Jin, X. Li, and S. Han, "Low-Complexity Joint Beamforming for RIS-Assisted MU-MISO Systems Based on Model-Driven Deep Learning," *IEEE Trans. Wireless Commun.*, vol. 23, no. 7, pp. 6968–6982, 2024.

[178] X. He, H. Xu, J. Wang, W. Xie, X. Li, and A. Nallanathan, "Joint Active and Passive Beamforming in RIS-Assisted Covert Symbiotic Radio Based on Deep Unfolding," *IEEE Trans. Veh. Technol.*, pp. 1–6, 2024.

[179] G. Zhu, J. Hu, K. Zhong, X. Cheng, and Z. Song, "Sum-Path-Gain Maximization for IRS-Aided MIMO Communication System via Riemannian Gradient Descent Network," *IEEE Signal Process. Lett.*, vol. 31, pp. 51–55, 2024.

[180] 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, 2022.

[181] N. T. Nguyen, M. Ma, O. Lavi, N. Shlezinger, Y. C. Eldar, A. L. Swindlehurst, and M. Juntti, "Deep Unfolding Hybrid Beamforming Designs for THz Massive MIMO Systems," *IEEE Trans. Signal Process.*, vol. 71, pp. 3788–3804, 2023.

[182] 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, 2016.

[183] M. Wu, Z. Gao, Y. Huang, Z. Xiao, D. W. K. Ng, and Z. Zhang, "Deep Learning-Based Rate-Splitting Multiple Access for Reconfigurable Intelligent Surface-Aided Tera-Hertz Massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 5, pp. 1431–1451, 2023.

[184] F. Liu, L. Zheng, Y. Cui, C. Masouros, A. P. Petropulu, H. Griffiths, and Y. C. Eldar, "Seventy Years of Radar and Communications: The road from separation to integration," *IEEE Signal Process. Mag.*, vol. 40, no. 5, pp. 106–121, 2023.

[185] W. Zhou, R. Zhang, G. Chen, and W. Wu, "Integrated Sensing and Communication Waveform Design: A Survey," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 1930–1949, 2022.

[186] Z. Feng, Z. Fang, Z. Wei, X. Chen, Z. Quan, and D. Ji, "Joint radar and

communication: A survey," *China Communications*, vol. 17, no. 1, pp. 1–27, 2020.

[187] F. Liu, C. Masouros, A. P. Petropulu, H. Griffiths, and L. Hanzo, "Joint Radar and Communication Design: Applications, State-of-the-Art, and the Road Ahead," *IEEE Trans. Commun.*, vol. 68, no. 6, pp. 3834–3862, 2020.

[188] J. A. Zhang, M. L. Rahman, K. Wu, X. Huang, Y. J. Guo, S. Chen, and J. Yuan, "Enabling Joint Communication and Radar Sensing in Mobile Networks—A Survey," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 1, pp. 306–345, 2022.

[189] A. Liu, Z. Huang, M. Li, Y. Wan, W. Li, T. X. Han, C. Liu, R. Du, D. K. P. Tan, J. Lu, Y. Shen, F. Colone, and K. Chetty, "A Survey on Fundamental Limits of Integrated Sensing and Communication," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 2, pp. 994–1034, 2022.

[190] P. Krishnananthalingam, N. T. Nguyen, and M. Juntti, "Deep Unfolding Enabled Constant Modulus Waveform Design for Joint Communications and Sensing," 06 2023. [Online]. Available: https://arxiv.org/abs/2306.14702

[191] N. T. Nguyen, L. Nguyen, N. Shlezinger, Y. Eldar, A. Swindlehurst, and M. Juntti, "Joint Communications and Sensing Hybrid Beamforming Design via Deep Unfolding," *IEEE J. Sel. Top. Signal Proces.*, 07 2024.

[192] J. Zhang, C. Masouros, F. Liu, Y. Huang, and A. L. Swindlehurst, "Low-Complexity Joint Radar-Communication Beamforming: From Optimization to Deep Unfolding," *IEEE J. Sel. Topics Signal Process.*, pp. 1–16, 2025.

[193] W. Fan, S. Liu, C. Li, and Y. Huang, "Fast Direct Localization for Millimeter Wave MIMO Systems via Deep ADMM Unfolding," *IEEE Wireless Commun. Lett.*, vol. 12, no. 4, pp. 748–752, 2023.

[194] W. Jiang, D. Ma, Z. Wei, Z. Feng, P. Zhang, and J. Peng, "ISAC-NET: Model-driven Deep Learning for Integrated Passive Sensing and Communication," *IEEE Trans. Commun.*, pp. 1–1, 2024.

[195] J. Lin, H. Zheng, and X. Feng, "Target Parameter Estimation with Deep Unfolding Networks for MIMO-OFDM Based Integrated Sensing and Communication Systems," in *2023 WCSP*, 2023, pp. 371–376.

[196] W. Lin, H. Zheng, X. Feng, and Y. Chen, "Deep Unfolding Network for Target Parameter Estimation in OTFS-based ISAC Systems," in *2024 IEEE WCNC*, 2024, pp. 1–6.

[197] W. Liu, H. Xu, X. He, Y. Ye, and A. Zhou, "Bi-Level Deep Unfolding Based Robust Beamforming Design for IRS-Assisted ISAC System," *IEEE Access*, vol. 12, pp. 76 663–76 672, 2024.

[198] Z. Hu, A. Liu, W. Xu, T. Q. S. Quek, and M. Zhao, "A Stochastic Particle Variational Bayesian Inference Inspired Deep-Unfolding Network for Sensing Over Wireless Networks," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 10, pp. 2832–2846, 2024.

[199] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes. 1," in *Proc. of ICC '93*, vol. 2, 1993, pp. 1064–1070 vol.2.

[200] R. Gallager, "Low-density parity-check codes," *IRE trans. inf. theory*, vol. 8, no. 1, pp. 21–28, 1962.

[201] R. G. Gallager, "Low-Density Parity-Check Codes," Ph.D. Thesis, MIT, Cambridge, MA, 1962. [Online]. Available: https://dspace.mit.edu/handle/1721.1/13799

[202] D. MacKay, "Good Error-Correcting Codes Based on Very Sparse Matrices," in *IEEE Int. Symp. Inf. Theory Proc.*, vol. 42, no. 2, 1997, pp. 399–431.

[203] E. Arikan, "Channel Polarization: A Method for Constructing Capacity-Achieving Codes for Symmetric Binary-Input Memoryless Channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, 2009.

[204] Z. B. Kaykac Egilmez, L. Xiang, R. G. Maunder, and L. Hanzo, "The Development, Operation and Performance of the 5G Polar Codes," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 1, pp. 96–122, 2020.

[205] R. Tanner, "A recursive approach to low complexity codes," *IEEE Trans. Inf. Theory*, vol. 27, no. 5, pp. 533–547, 1981.

[206] F. Kschischang, B. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, 2001.

[207] I. Tal and A. Vardy, "List Decoding of Polar Codes," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2213–2226, 2015.

[208] X. Zhou, J. Zhang, C.-W. Syu, C.-K. Wen, J. Zhang, and S. Jin, "Model-Driven Deep Learning-Based MIMO-OFDM Detector: Design, Simulation, and Experimental Results," *IEEE Trans. Commun.*, vol. 70, no. 8, pp. 5193–5207, 2022.

[209] G. Ma, X. Jiao, J. Mu, H. Han, and Y. Yang, "Deep Learning-Based Detection for Marker Codes over Insertion and Deletion Channels," *IEEE Trans. Commun.*, pp. 1–1, 2024.

[210] J. Gao, K. Niu, and C. Dong, "Learning to Decode Polar Codes With One-Bit Quantizer," *IEEE Access*, vol. 8, pp. 27 210–27 217, 2020.

[211] Y. Liu, M.-M. Zhao, C. Wang, M. Lei, and M.-J. Zhao, "PDD-Based Decoder for LDPC Codes With Model-Driven Neural Networks," *IEEE Commun. Lett.*, vol. 26, no. 11, pp. 2532–2536, 2022.

[212] Q. Wang, Q. Liu, S. Wang, L. Chen, H. Fang, L. Chen, Y. Guo, and Z. Wu, "Normalized Min-Sum Neural Network for LDPC Decoding," *IEEE Trans. Cogn. Commun. Netw.*, vol. 9, no. 1, pp. 70–81, 2023.

[213] S. Adiga, X. Xiao, R. Tandon, B. Vasić, and T. Bose, "Generalization Bounds for Neural Belief Propagation Decoders," *IEEE Trans. Inf. Theory*, vol. 70, no. 6, pp. 4280–4296, 2024.

[214] Y. Sun, H. Shen, B. Li, W. Xu, P. Zhu, N. Hu, and C. Zhao, "Trainable Joint Channel Estimation, Detection and Decoding for MIMO URLLC Systems," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2024.

[215] J. Zhang, Y. He, Y.-W. Li, C.-K. Wen, and S. Jin, "Meta Learning-Based MIMO Detectors: Design, Simulation, and Experimental Test," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1122–1137, 2021.

[216] K. T. K. Cheung, S. Yang, and L. Hanzo, "Achieving Maximum Energy-Efficiency in Multi-Relay OFDMA Cellular Networks: A Fractional Programming Approach," *IEEE Trans. Commun.*, vol. 61, no. 7, pp. 2746–2757, 2013.

[217] D. W. K. Ng, E. S. Lo, and R. Schober, "Energy-Efficient Resource Allocation in OFDMA Systems with Hybrid Energy Harvesting Base Station," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3412–3427, 2013.

[218] Z. Ding, S. M. Perlaza, I. Esnaola, and H. V. Poor, "Power Allocation Strategies in Energy Harvesting Wireless Cooperative Networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 846–860, 2014.

[219] Z. Yang, M. Chen, W. Saad, W. Xu, M. Shikh-Bahaei, H. V. Poor, and S. Cui, "Energy-Efficient Wireless Communications With Distributed Reconfigurable Intelligent Surfaces," *IEEE Trans. Wireless Commun.*, vol. 21, no. 1, pp. 665–679, 2022.

[220] W. Liang, Z. Ding, Y. Li, and L. Song, "User Pairing for Downlink Non-Orthogonal Multiple Access Networks Using Matching Algorithm," *IEEE Trans. Commun.*, vol. 65, no. 12, pp. 5319–5332, 2017.

[221] A. Chowdhury, G. Verma, C. Rao, A. Swami, and S. Segarra, "Unfolding WMMSE Using Graph Neural Networks for Efficient Power Allocation," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 6004–6017, 2021.

[222] H. Hu, Z. Xie, H. Shi, B. Liu, H. Zhao, and G. Gui, "Unsupervised Power Allocation Based on Combination of Edge Aggregated Graph Attention Network with Deep Unfolded WMMSE," *IEEE Trans. Veh. Technol.*, pp. 1–15, 2024.

[223] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in *6th ICLR*, 2018. [Online]. Available: https://arxiv.org/abs/1710.10903

[224] "EGRET: Edge Aggregated Graph Attention Networks and Transfer Learning Improve Protein–Protein Interaction Site Prediction, author=Mahbub, Sazan and Shamsuzzoha Bayzid, Md, journal=Briefings in Bioinformatics, volume=23, number=2, pages=bbab578, year=2022, publisher=Oxford University Press, doi=10.1093/bib/bbab578."

[225] B. Li, G. Verma, and S. Segarra, "Graph-Based Algorithm Unfolding for Energy-Aware Power Allocation in Wireless Networks," *IEEE Trans. Wireless Commun.*, vol. 22, no. 2, pp. 1359–1373, 2023.

[226] Y. Huo, Y. Tian, L. Ma, X. Cheng, and T. Jing, "Jamming Strategies for Physical Layer Security," *IEEE Wirel. Commun.*, vol. 25, no. 1, pp. 148–153, 2018.

[227] K. N. Vaishnavi, S. D. Khorvi, R. Kishore, and S. Gurugopinath, "A Survey on Jamming Techniques in Physical Layer Security and Anti-Jamming Strategies for 6G," in *2021 ICT*, 2021, pp. 174–179.

[228] G. Marti, T. Kölle, and C. Studer, "Mitigating Smart Jammers in Multi-User MIMO," *IEEE Trans. Signal Process.*, vol. 71, pp. 756–771, 2023.

[229] G. Gao, C. Dong, and K. Niu, "Sparsely Connected Neural Network for Massive MIMO Detection," in *2018 IEEE ICCC*, 2018, pp. 397–402.

[230] A. Ali, N. Gonzalez-Prelcic, R. W. Heath, and A. Ghosh, "Leveraging sensing at the infrastructure for mmWave communication," *IEEE Commun. Mag.*, vol. 58, no. 7, pp. 84–89, 2020.

[231] A. Klautau, N. González-Prelcic, and R. W. Heath, "LIDAR Data for Deep Learning-Based mmWave Beam-Selection," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 909–912, 2019.

[232] Z. Li, S. H. Mirfarshbafan, O. Castañeda, and C. Studer, "A deep-unfolding-optimized coordinate-descent data-detector asic for mmwave massive mimo," *IEEE J. Sel. Areas Commun.*, vol. 43, no. 4, pp. 1323–1338, 2025.

[233] F. Liu, C. Masouros, A. Li, H. Sun, and L. Hanzo, "MU-MIMO Communications With MIMO Radar: From Co-Existence to Joint Transmission," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2755–2770, 2018.

[234] M. Ma, T. Fang, N. Shlezinger, A. Swindlehurst, M. Juntti, and N. Nguyen, "Model-Based Machine Learning for Max-Min Fairness Beamforming Design in JCAS Systems," *arXiv preprint arXiv:2409.17644*, 2024.

[235] X. Deng and P. L. Dragotti, "Deep Coupled ISTA Network for Multi-Modal Image Super-Resolution," *IEEE Trans. Image Process.*, vol. 29, pp. 1683–1698, 2020.

[236] Z. Zheng, W. Dai, D. Xue, C. Li, J. Zou, and H. Xiong, "Hybrid ISTA: Unfolding ISTA With Convergence Guarantees Using Free-Form Deep Neural Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 3, pp. 3226–3244, 2023.

[237] J. Zhang, Z. Zhang, J. Xie, and Y. Zhang, "High-Throughput Deep Unfolding Network for Compressive Sensing MRI," *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 4, pp. 750–761, 2022.