Deep Learning for Estimation and Pilot Signal Design in Few-Bit Massive MIMO Systems

Ly V. Nguyen, *Member, IEEE*, Duy H. N. Nguyen, *Senior Member, IEEE*, and A. Lee Swindlehurst, *Fellow, IEEE*

***Abstract*— Estimation in few-bit MIMO systems is challenging, since the received signals are nonlinearly distorted by the low-resolution ADCs. In this paper, we propose a deep learning framework for channel estimation, data detection, and pilot signal design to address the nonlinearity in such systems. The proposed channel estimation and data detection networks are model-driven and have special structures that take advantage of domain knowledge in the few-bit quantization process. While the first data detection network, B-DetNet, is based on a linearized model obtained from the Bussgang decomposition, the channel estimation network and the second data detection network, FBM-CENet and FBM-DetNet respectively, rely on the original quantized system model. To develop FBM-CENet and FBM-DetNet, the maximum-likelihood channel estimation and data detection problems are reformulated to overcome the indeterminant gradient issue. An important feature of the proposed FBM-CENet structure is that the pilot matrix is integrated into the weight matrices of its channel estimator. Thus, training the proposed FBM-CENet enables a joint optimization of both the channel estimator at the base station and the pilot signal transmitted from the users. Simulation results show significant performance gains in estimation accuracy by the proposed deep learning framework.**

***Index Terms*— Deep learning, deep neural network, massive MIMO, low-resolution ADCs, channel estimation, data detection.**

I. Introduction

ONE practical solution for reducing hardware cost and power consumption in massive MIMO systems is to use low-resolution (e.g., 1–3 bits) analog-to-digital converters (ADCs), due to their simple structure and very low power consumption. In particular, the number of comparators in a b-bit ADC grows exponentially with b, which means both the hardware complexity and the power consumption of an ADC scales exponentially with the resolution [1]. Therefore, the cost and power consumption of low-resolution ADCs are substantially lower than for high-resolution ADCs. Furthermore, the hardware structure of other components in an RF chain can also be simplified or removed when low-resolution ADCs are used. For example, the simplest architecture involving one-bit ADCs does not require an automatic gain control (AGC) since only the *sign* of the real and imaginary parts of the received signals is retained. The stringent linearity requirement of the low-noise amplifier (LNA) can be relaxed and a simpler low-cost amplifier can be used instead. These benefits on the hardware side make it possible to deploy low-resolution ADCs in practical massive MIMO systems. However, the lower-complexity and lower-power-consumption hardware necessitates special care in the subsequent signal processing. More specifically, the nonlinearities introduced by the low-resolution ADCs makes signal processing tasks such as channel estimation and data detection in few-bit MIMO systems much more challenging compared to those in unquantized systems. Therefore, it is crucial that efficient signal processing methods for channel estimation and data detection be developed for such systems so that they can be transitioned to commercial systems.

Channel estimation for massive MIMO systems with low-resolution ADCs has attracted significant research interest and has been studied intensively. The majority of the proposed approaches focus on one-bit systems in different scenarios, e.g., [2]–[17]. For example, a one-bit maximum-likelihood (ML) channel estimator was proposed in [2]. The work in [3] exploits the Bussgang decomposition to form a one-bit Bussgang-based minimum mean-squared error (BMMSE) channel estimator. Another BMMSE channel estimator was also proposed in [4] but for one-bit spatial sigma-delta ADCs in a spatially oversampled array. Channel estimation with temporally oversampled one-bit ADCs is studied in [5] and [6]. It has been shown that one-bit ADCs with spatial and temporal oversampling can help improve the channel estimation accuracy but more resources and computation are required due to the oversampling process. Angular-domain channel estimation for one-bit massive MIMO systems was studied in [7]–[9]. Spatially/temporally correlated channels and multi-cell processing with pilot contamination were investigated in [10] and [11], respectively. For sparse millimeter-wave MIMO channels, ML and maximum a posteriori (MAP) channel estimation were examined in [12] and [13], respectively. Taking into account the sparsity of such channels, the one-bit ADC channel estimation problem has been formulated as

Manuscript received 9 August 2021; revised 25 April 2022 and 8 July 2022; accepted 9 July 2022. Date of publication 2 August 2022; date of current version 9 January 2023. This work was supported in part by the U.S. National Science Foundation under Grant CCF-2225575 and Grant CCF-2225576. The associate editor coordinating the review of this article and approving it for publication was C. Brinton. *(Corresponding author: Duy H. N. Nguyen.)*

Ly V. Nguyen is with the Computational Science Research Center, San Diego State University, San Diego, CA 92182 USA (e-mail: vnguyen6@sdsu.edu).

Duy H. N. Nguyen is with the Department of Electrical and Computer Engineering, San Diego State University, San Diego, CA 92182 USA (e-mail: duy.nguyen@sdsu.edu).

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

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TWC.2022.3193885.

Digital Object Identifier 10.1109/TWC.2022.3193885

a compressed sensing problem in [14]–[16]. Performance bounds on the channel estimation of mmWave one-bit massive MIMO channels were reported in [17]. The Bussgang decomposition was exploited in [18] to derive two linear channel estimators for few-bit ADCs including an extension of the BMMSE approach as well as a Bussgang-based weighted zero-forcing (BWZF) algorithm.

Recently, machine learning techniques have been studied to address massive MIMO channel estimation for low-resolution ADCs [19]–[25]. For example, the work in [19] shows that support-vector machine (SVM) models can be applied to estimate massive MIMO channels with one-bit observations. A DNN-based joint pilot signal and channel estimator design is proposed in [23]. The work in [24], [25] studied mixed-resolution channel estimation where low-resolution ADCs were used for only some of the receive antennas, while the rest are equipped with conventional ADCs.

Data detection for low-resolution massive MIMO systems has also been studied intensively in the literature. Again, most of the results have been reported for the case of one-bit ADCs, e.g., [2], [26]–[36]. In particular, a one-bit ML detector and a one-bit sphere decoding (OSD) technique were proposed in [2] and [26], respectively. The very high computational complexity of the ML and OSD methods nevertheless make them impractical for large-scale systems. A near-ML (nML) data detection method for large-scale MIMO systems was proposed in [2]. However, the nML method is not robust at high signal-to-noise ratios (SNRs) when the channel state information (CSI) is not perfectly known. The learning-based method in [27] is a blind detection method for which CSI is not required, but it is only applicable to MIMO systems with a small number of transmit antennas and only low-dimensional constellations. Various one-bit linear detectors were introduced in [28], [29]. These linear detectors are applicable for large-scale systems but often suffer from high detection-error floors. The authors in [31] proposed a one-bit detection method based on the alternating direction method of multipliers (ADMM) algorithm that takes hardware impairments into account. The SVM-based and DNN-based one-bit detectors proposed in [19] and [29] were shown to be robust, applicable to large-scale systems, and also to outperform other existing one-bit detectors. Another DNN-based one-bit detector was proposed in [30] but it requires online training since the network has to be retrained whenever the channel changes. This significantly increases the computational complexity and resources as well as the pilot overhead. Several other one-bit data detection approaches can be found in [33]–[36], but they are only applicable in systems where either a cyclic redundancy check (CRC) [33]–[35] or an error correcting code such as a low-density parity-check (LDPC) code [36] is available.

Data detection in few-bit massive MIMO systems has been studied in recent papers [18], [37]–[40]. Generalized approximate message passing (GAMP) and Bayes inference were exploited in [37], while the work in [38] employed variational Bayesian (VB) inference and belief propagation (BP) for soft symbol decoding. However, the resulting methods are complicated and expensive to implement. Unlike the blind detection method in [27] which was developed for one-bit systems, the learning-based blind detection methods in [39], [40] are applicable for few-bit systems, but they are also restricted to MIMO systems with a small number of transmit antennas and only low-dimensional constellations. The BMMSE and BWZF detection methods in [18] are linear detectors and thus simple and applicable for large-scale MIMO systems, but their performance is limited.

In this paper, we develop a deep learning framework for channel estimation and data detection for massive MIMO systems with low-resolution ADCs. Using deep unfolding of the first-order optimization iterations, we propose a channel estimator and two data detectors that are applicable for both one-bit and few-bit ADCs as well as large-scale systems without the need for CRC or error correcting codes. The proposed channel estimation and data detection networks are model-driven and have special structures that can take advantage of domain knowledge in few-bit MIMO systems.

We first reformulate the ML channel estimation problem by exploiting the approximation of the cumulative distribution function (cdf) of a normal random variable as a Sigmoid activation function. Unlike the original problem, the reformulated channel estimation approach does not lead to occasionally indeterminant gradients. Based on the reformulated problem and a deep unfolding technique, we propose a Few-Bit massive MIMO Channel Estimation Network, referred to as FBM-CENet. An interesting feature of the proposed FBM-CENet is that the pilot signal matrix is directly integrated in the weight matrices of the estimation network. When the pilot matrix is not given, it can be treated as additional trainable parameters and therefore training FBM-CENet is equivalent to *jointly optimizing both the channel estimator at the base station and the pilot signal transmitted from the users*. This is a significant advantage of the proposed FBM-CENet structure since existing channel estimators are often designed only for a known pilot matrix. The proposed DNN is based on a novel reformulation of the network layers, and is shown via several simulation results to significantly outperform the conventional DNN architecture in [23] as well as other existing channel estimation methods.

For data detection, we first propose a Bussgang-based few-bit massive MIMO Data Detection Network, referred to as B-DetNet, that is based on a linearized system model obtained through the Bussgang decomposition. Then we propose a Few-Bit massive MIMO Data Detection Network, referred to as FBM-DetNet. Previously presented in [41], FBM-DetNet is developed based on the original quantized system model. The special structure of FBM-DetNet is also obtained through a reformulated ML data detection problem that parallels the reformulated channel estimation problem. We stress that the proposed B-DetNet and FBM-DetNet are highly adaptive to the channel since their weight matrices and the bias vectors are defined by the channel matrix and the received signal vector, respectively. The proposed detection networks have relatively few parameters and are thus easier to train. Simulation results also show that they significantly outperform existing methods.

The rest of this paper is organized as follows: Section II introduces the assumed system model. Channel estimation is considered in Section III, where the FBM-CENet estimator is proposed. The two proposed data detection networks B-DetNet and FBM-DetNet are presented in Section IV. Numerical results are given in Section V. Finally, Section VI concludes the paper.

Notation: Upper-case and lower-case boldface letters denote matrices and column vectors, respectively. $\mathbb{E}[\cdot]$ represents expectation. The operator $|\cdot|$ denotes the absolute value of a number and the operator $\|\cdot\|$ denotes the ℓ_2-norm of a vector. The transpose is denoted by $[\cdot]^T$. The notation $\Re\{\cdot\}$ and $\Im\{\cdot\}$ respectively indicates the real and imaginary parts of the complex argument. If $\Re\{\cdot\}$ and $\Im\{\cdot\}$ are applied to a matrix or vector, they are applied separately to every element of that matrix or vector. The operator $\mathrm{vec}(\mathbf{A})$ vectorizes $\mathbf{A}$ by stacking its columns on top of one another. The operator $\otimes$ denotes the Kronecker product, $\mathbb{R}$ and $\mathbb{C}$ denote the set of real and complex numbers, respectively, and j is the unit imaginary number satisfying $j^2 = -1$. The symbols $\mathcal{N}(\cdot,\cdot)$ and $\mathcal{CN}(\cdot,\cdot)$ represent the real and the circular complex normal distributions respectively, where the first argument is the mean and the second argument is the variance or the covariance matrix. The functions $\Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{\tau^2}{2}} d\tau$ and $\phi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}$ are the cdf and pdf of the standard normal random variable.

II. System Model

We consider an uplink massive MIMO system with K single-antenna users and an N-antenna base station (BS), where it is assumed that $N \geq K$. Let $\bar{\mathbf{x}} = [\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_K]^T \in \mathbb{C}^K$ denote the transmitted signal vector, where $\bar{x}_k$ is the signal transmitted from the k^{th} user under the power constraint $\mathbb{E}[|\bar{x}_k|^2] = p_k^{\text{tx}}$. The signal $\bar{x}_k$ is drawn from a constellation $\bar{\mathcal{M}}$. Let $\bar{\mathbf{H}} \in \mathbb{C}^{N\times K}$ denote the channel, which is assumed to be block flat fading. Let $\bar{\mathbf{r}} = [\bar{r}_1, \bar{r}_2, \ldots, \bar{r}_N]^T \in \mathbb{C}^N$ be the unquantized received signal vector at the base station, which is given as

$$\bar{\mathbf{r}} = \bar{\mathbf{H}}\bar{\mathbf{x}} + \bar{\mathbf{z}} \tag{1}$$

where $\bar{\mathbf{z}} = [\bar{z}_1, \bar{z}_2, \ldots, \bar{z}_N]^T \in \mathbb{C}^N$ is a noise vector whose elements are assumed to be independent and identically distributed (i.i.d.) as $\mathcal{CN}(0, N_0)$ with noise power N_0. Each received analog signal is then quantized by a pair of b-bit ADCs to produce the quantized received signal:

$$\bar{\mathbf{y}} = \mathcal{Q}_b(\bar{\mathbf{r}}) = \mathcal{Q}_b(\Re\{\bar{\mathbf{r}}\}) + j\mathcal{Q}_b(\Im\{\bar{\mathbf{r}}\}). \tag{2}$$

For vector or matrix arguments, the operator $\mathcal{Q}_b(\cdot)$ is applied separately to every element.

It is assumed that the ADCs perform b-bit uniform scalar quantization. The b-bit ADC model is characterized by a set of $2^b - 1$ thresholds denoted as $\{\tau_1, \ldots, \tau_{2^b-1}\}$. Without loss of generality, we assume $-\infty = \tau_0 < \tau_1 < \cdots < \tau_{2^b-1} < \tau_{2^b} = \infty$. Let Δ be the step size, so the thresholds of the uniform quantizer are given as

$$\tau_l = (-2^{b-1} + l)\Delta, \text{ for } l \in \mathcal{L} = \{1, \ldots, 2^b - 1\}. \tag{3}$$

The quantized output is then defined as

$$\mathcal{Q}_b(r) = q_l = \begin{cases} \tau_l - \frac{\Delta}{2} & \text{if } r \in (\tau_{l-1}, \tau_l] \text{ with } l \in \mathcal{L} \\ (2^b - 1)\frac{\Delta}{2} & \text{if } r \in (\tau_{2^b-1}, \tau_{2^b}]. \end{cases} \tag{4}$$

III. Channel Estimation in Few-Bit MIMO Systems

In order to estimate the channel, a pilot sequence $\bar{\mathbf{X}}_{\text{t}} \in \mathbb{C}^{K\times T_{\text{t}}}$ of length T_{t} is used to generate the training data

$$\bar{\mathbf{Y}}_{\text{t}} = \mathcal{Q}_b\left(\bar{\mathbf{H}}\bar{\mathbf{X}}_{\text{t}} + \bar{\mathbf{Z}}_{\text{t}}\right) \tag{5}$$

where $\bar{\mathbf{Y}}_{\text{t}} \in \mathbb{C}^{N\times T_{\text{t}}}$ and $\bar{\mathbf{Z}}_{\text{t}} \in \mathbb{C}^{N\times T_{\text{t}}}$. The subscript 't' in this paper indicates the training phase where channel estimation is performed. We vectorize the received signal in (5) to obtain

$$\bar{\mathbf{y}}_{\text{t}} = \mathcal{Q}_b(\bar{\mathbf{P}}\bar{\mathbf{h}} + \bar{\mathbf{z}}_{\text{t}}), \tag{6}$$

where $\bar{\mathbf{y}}_{\text{t}} = \mathrm{vec}(\bar{\mathbf{Y}}_{\text{t}}) \in \mathbb{C}^{NT_{\text{t}}\times 1}$, $\bar{\mathbf{P}} = \bar{\mathbf{X}}_{\text{t}}^T \otimes \mathbf{I}_N \in \mathbb{C}^{NT_{\text{t}}\times NK}$, $\bar{\mathbf{h}} = \mathrm{vec}(\bar{\mathbf{H}}) \in \mathbb{C}^{NK\times 1}$, and $\bar{\mathbf{z}}_{\text{t}} = \mathrm{vec}(\bar{\mathbf{Z}}_{\text{t}}) \in \mathbb{C}^{NT_{\text{t}}\times 1}$. For convenience in later derivations, we convert the notation in (6) into the real domain as

$$\mathbf{y}_{\text{t}} = \mathcal{Q}_b(\mathbf{P}\mathbf{h} + \mathbf{z}_{\text{t}}) \tag{7}$$

where

$$\mathbf{y}_{\text{t}} = \begin{bmatrix} \Re\{\bar{\mathbf{y}}_{\text{t}}\} \\ \Im\{\bar{\mathbf{y}}_{\text{t}}\} \end{bmatrix}, \ \mathbf{h} = \begin{bmatrix} \Re\{\bar{\mathbf{h}}\} \\ \Im\{\bar{\mathbf{h}}\} \end{bmatrix}, \text{ and } \mathbf{P} = \begin{bmatrix} \Re\{\bar{\mathbf{P}}\} & -\Im\{\bar{\mathbf{P}}\} \\ \Im\{\bar{\mathbf{P}}\} & \Re\{\bar{\mathbf{P}}\} \end{bmatrix}.$$

Note that $\mathbf{y}_{\text{t}} \in \mathbb{R}^{2NT_{\text{t}}\times 1}$, $\mathbf{h} \in \mathbb{R}^{2NK\times 1}$, $\mathbf{P} \in \mathbb{R}^{2NT_{\text{t}}\times 2NK}$, and $\mathbf{z}_{\text{t}} \in \mathbb{R}^{2NT_{\text{t}}\times 1}$.

A. Bussgang-Based Linear Channel Estimators

We first revisit the Bussgang-based linear channel estimators including BMMSE and BWZF for low-resolution massive MIMO systems [3], [18]. The system model in (7) can be linearized by the Bussgang decomposition as follows:

$$\begin{aligned} \mathbf{y}_{\text{t}} &= \mathbf{V}_{\text{t}}\mathbf{P}\mathbf{h} + \mathbf{V}_{\text{t}}\mathbf{z}_{\text{t}} + \mathbf{d}_{\text{t}} \\ &= \mathbf{A}_{\text{t}}\mathbf{h} + \mathbf{n}_{\text{t}} \end{aligned} \tag{8}$$

where $\mathbf{A}_{\text{t}} \equiv \mathbf{V}_{\text{t}}\mathbf{P} \in \mathbb{R}^{2NT_{\text{t}}\times 2NK}$, $\mathbf{n}_{\text{t}} \equiv \mathbf{V}_{\text{t}}\mathbf{z}_{\text{t}} + \mathbf{d}_{\text{t}}$ combines the receiver and equivalent quantization noise, and $\mathbf{V}_{\text{t}} \in \mathbb{R}^{2NT_{\text{t}}\times 2NT_{\text{t}}}$ is a diagonal matrix and given as [18]

$$\begin{aligned} \mathbf{V}_{\text{t}} = \frac{\Delta}{\sqrt{2\pi}} &\,\mathrm{diag}(\mathbf{\Sigma}_{\mathbf{r}_{\text{t}}})^{-\frac{1}{2}} \\ &\times \sum_{i=1}^{2^b-1} \exp\Big\{ -\frac{1}{2}\Delta^2(i - 2^{b-1})^2 \,\mathrm{diag}(\mathbf{\Sigma}_{\mathbf{r}_{\text{t}}})^{-1} \Big\} \end{aligned}$$

where $\mathbf{\Sigma}_{\mathbf{r}_{\text{t}}} = \mathbf{P}\mathbf{\Sigma}_{\mathbf{h}}\mathbf{P}^T + \frac{N_0}{2}\mathbf{I} \in \mathbb{R}^{2NT_{\text{t}}\times 2NT_{\text{t}}}$ is the covariance matrix of $\mathbf{r}_{\text{t}} = \mathbf{P}\mathbf{h} + \mathbf{z}_{\text{t}}$. For the case of one-bit ADCs with $\Delta = \sqrt{2}$, $\mathbf{V}_{\text{t}}$ reduces to the form reported in [3, Eq. (10)].

The BMMSE channel estimator is given as [3], [18]

$$\begin{aligned} \hat{\mathbf{h}}_{\texttt{BMMSE}} &= \mathbf{\Sigma}_{\mathbf{h}\mathbf{y}_{\text{t}}}\mathbf{\Sigma}_{\mathbf{y}_{\text{t}}}^{-1}\mathbf{y}_{\text{t}} \\ &= \mathbf{\Sigma}_{\mathbf{h}}\mathbf{A}_{\text{t}}^T\mathbf{\Sigma}_{\mathbf{y}_{\text{t}}}^{-1}\mathbf{y}_{\text{t}} \end{aligned} \tag{9}$$

where $\mathbf{\Sigma}_{\mathbf{h}\mathbf{y}_{\text{t}}} \in \mathbb{R}^{2NK\times 2NT_{\text{t}}}$ is the cross-covariance matrix between $\mathbf{h}$ and $\mathbf{y}_{\text{t}}$, and $\mathbf{\Sigma}_{\mathbf{y}_{\text{t}}} \in \mathbb{R}^{2NT_{\text{t}}\times 2NT_{\text{t}}}$ is the covariance matrix of $\mathbf{y}_{\text{t}}$. For the case of one-bit ADCs, $\mathbf{\Sigma}_{\mathbf{y}_{\text{t}}}$ is given as [3]

$$\mathbf{\Sigma}_{\mathbf{y}_{\text{t}}} = \frac{\Delta^2}{2\pi} \arcsin\Big(\mathrm{diag}(\mathbf{\Sigma}_{\mathbf{r}_{\text{t}}})^{-\frac{1}{2}} \mathbf{\Sigma}_{\mathbf{r}_{\text{t}}} \,\mathrm{diag}(\mathbf{\Sigma}_{\mathbf{r}_{\text{t}}})^{-\frac{1}{2}} \Big). \tag{10}$$

TABLE I
OPTIMUM UNIFORM QUANTIZER FOR $\mathcal{N}(0,1)$ GAUSSIAN INPUTS [42]

Resolution b	1-bit	2-bit	3-bit	4-bit
Step size Δ_b	$\sqrt{8/\pi}$	0.996	0.586	0.335
Distortion η_b	$1-2/\pi$	0.1188	0.0374	0.0115

For the case of two-bit or higher resolution ADCs, $\mathbf{\Sigma}_{\mathbf{y}_\mathrm{t}}$ is given as [18]

$$\mathbf{\Sigma}_{\mathbf{y}_\mathrm{t}} = \mathbf{V}_\mathrm{t}\mathbf{\Sigma}_{\mathbf{r}_\mathrm{t}}\mathbf{V}_\mathrm{t}^T + \mathbf{\Sigma}_{\mathbf{d}_\mathrm{t}}, \tag{11}$$

where $\mathbf{\Sigma}_{\mathbf{d}_\mathrm{t}} \in \mathbb{R}^{2NT_\mathrm{t}\times 2NT_\mathrm{t}}$ is the covariance matrix of $\mathbf{d}_\mathrm{t}$ and can be approximated as $\mathbf{\Sigma}_{\mathbf{d}_\mathrm{t}} \approx \eta_b \,\mathrm{diag}(\mathbf{\Sigma}_{\mathbf{r}_\mathrm{t}})$. The distortion factor η_b depending on the number of quantization bits b is given in Table I.

A BWZF channel estimator was also proposed in [18] as follows:

$$\hat{\mathbf{h}}_{\mathtt{BWZF}} = \left(\mathbf{A}_\mathrm{t}^T \,\mathrm{diag}(\mathbf{w})\mathbf{A}_\mathrm{t}\right)^{-1}\mathbf{A}_\mathrm{t}^T \,\mathrm{diag}(\mathbf{w})\mathbf{y}_\mathrm{t} \tag{12}$$

where $\mathrm{diag}(\mathbf{w})$ is a diagonal matrix with $\mathbf{w} = [w_1, w_2, \ldots, w_{2NT_\mathrm{t}}]$ on the diagonal, and

$$w_i = \frac{1}{\mathbb{E}[z_{\mathrm{t},i}^2] + \mathbb{E}[d_{\mathrm{t},i}^2|y_{\mathrm{t},i}]}, \quad i = 1, \ldots, 2NT_\mathrm{t}.$$

Here, $y_{\mathrm{t},i}$, $z_{\mathrm{t},i}$, and $d_{\mathrm{t},i}$ are the i-th elements of $\mathbf{y}_\mathrm{t}$, $\mathbf{z}_\mathrm{t}$, and $\mathbf{d}_\mathrm{t}$, respectively. The key idea of BWZF is that given an observed quantized signal vector $\mathbf{y}_\mathrm{t}$, the elements of $\mathbf{r}_\mathrm{t}$ have different variances. Exploiting this fact, the BWZF estimator sets the signals with lower variances to have higher weights.

B. Proposed FBM-CENet

1) Maximum-Likelihood Channel Estimation Problem: Let $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_{2NT_\mathrm{t}}]^T$, $\mathbf{y}_\mathrm{t} = [y_{\mathrm{t},1}, y_{\mathrm{t},2}, \ldots, y_{\mathrm{t},2NT_\mathrm{t}}]^T$, and $\mathbf{z}_\mathrm{t} = [z_{\mathrm{t},1}, \ldots, z_{\mathrm{t},2NT_\mathrm{t}}]^T$, then we have

$$y_{\mathrm{t},i} = \mathcal{Q}_b\left(\mathbf{p}_i^T\mathbf{h} + z_{\mathrm{t},i}\right), \quad i = 1, 2, \ldots, 2NT_\mathrm{t}. \tag{13}$$

Let $s_{\mathrm{t},i}^{\mathrm{up}} = \sqrt{2\rho}(q_{\mathrm{t},i}^{\mathrm{up}} - \mathbf{p}_i^T\mathbf{h})$ and $s_{\mathrm{t},i}^{\mathrm{low}} = \sqrt{2\rho}(q_{\mathrm{t},i}^{\mathrm{low}} - \mathbf{p}_i^T\mathbf{h})$, where $\rho = 1/N_0$, and

$$q_{\mathrm{t},i}^{\mathrm{up}} = \begin{cases} y_{\mathrm{t},i} + \frac{\Delta}{2} & \text{if } y_{\mathrm{t},i} < \tau_{2^b-1} \\ \infty & \text{otherwise,} \end{cases}$$

$$q_{\mathrm{t},i}^{\mathrm{low}} = \begin{cases} y_{\mathrm{t},i} - \frac{\Delta}{2} & \text{if } y_{\mathrm{t},i} > \tau_1 \\ -\infty & \text{otherwise.} \end{cases}$$

Hence, $q_{\mathrm{t},i}^{\mathrm{up}}$ and $q_{\mathrm{t},i}^{\mathrm{low}}$ are the upper and lower quantization thresholds of the bin to which $y_{\mathrm{t},i}$ belongs.

The ML channel estimator is given as follows:

$$\begin{aligned}\hat{\mathbf{h}}_{\mathtt{ML}} &= \arg\max_{\mathbf{h}} \; f(\mathbf{y}_\mathrm{t}\,|\,\mathbf{h}) \\ &= \arg\max_{\mathbf{h}} \; \sum_{i=1}^{2NT_\mathrm{t}} \log\left[\Phi\left(s_{\mathrm{t},i}^{\mathrm{up}}\right) - \Phi\left(s_{\mathrm{t},i}^{\mathrm{low}}\right)\right].\end{aligned} \tag{14}$$

Let $\mathcal{P}_\mathrm{t}(\mathbf{h})$ be the objective function of (14). Since $\mathcal{P}_\mathrm{t}(\mathbf{h})$ is a concave function [43], the unconstrained optimization problem (14) is convex, and therefore an iterative gradient ascent method can be used to solve it. However, the gradient of $\mathcal{P}_\mathrm{t}(\mathbf{h})$, given by

$$\nabla\mathcal{P}_\mathrm{t}(\mathbf{h}) = \sum_{i=1}^{2NT_\mathrm{t}} \frac{-\sqrt{2\rho}\mathbf{p}_i\left(\phi\left(s_{\mathrm{t},i}^{\mathrm{up}}\right) - \phi\left(s_{\mathrm{t},i}^{\mathrm{low}}\right)\right)}{\Phi\left(s_{\mathrm{t},i}^{\mathrm{up}}\right) - \Phi\left(s_{\mathrm{t},i}^{\mathrm{low}}\right)}, \tag{15}$$

is undefined at certain points, since the function $\Phi(\cdot)$ very rapidly approaches zero or one. Specifically, as the iterative gradient descent method sequentially updates the estimated channel $\hat{\mathbf{h}}$, there exist instances of $\hat{\mathbf{h}}$ that make both $\Phi\left(s_{\mathrm{t},i}^{\mathrm{up}}\right)$ and $\Phi\left(s_{\mathrm{t},i}^{\mathrm{low}}\right)$ equal to zero or one. Thus, the denominator in (15) can be zero for some $\hat{\mathbf{h}}$ causing the gradient to become unbounded. In addition, a lack of a closed-form expression for $\Phi(\cdot)$ complicates the evaluation in (14).

These observations motivate us to reformulate the ML channel estimation problem (14) to address the indeterminant gradient issue as well as the complicated evaluation of the objective function in (14). We exploit a result in [44], which shows that the function $\Phi(t)$ can be accurately approximated by the Sigmoid function $\sigma(t) = 1/(1+e^{-t})$ as follows:

$$\Phi(t) \approx \sigma(ct) = \frac{1}{1+e^{-ct}} \tag{16}$$

where $c = 1.702$ is a constant. It was shown in [44] that $|\Phi(t) - \sigma(ct)| \le 0.0095$, $\forall t \in \mathbb{R}$. Using this approximation, the objective function $\mathcal{P}_\mathrm{t}(\mathbf{h})$ can be re-written as

$$\mathcal{P}_\mathrm{t}(\mathbf{h}) \approx \tilde{\mathcal{P}}_\mathrm{t}(\mathbf{h}) = \textstyle\sum_{i=1}^{2NT_\mathrm{t}} \log\left[\frac{1}{1+e^{-cs_{\mathrm{t},i}^{\mathrm{up}}}} - \frac{1}{1+e^{-cs_{\mathrm{t},i}^{\mathrm{low}}}}\right] \tag{17}$$

and the reformulated ML channel estimation problem is

$$\hat{\mathbf{h}} = \arg\max_{\mathbf{h}} \; \tilde{\mathcal{P}}_\mathrm{t}(\mathbf{h}). \tag{18}$$

The gradient of $\tilde{\mathcal{P}}_\mathrm{t}(\mathbf{h})$ is

$$\begin{aligned}\nabla\tilde{\mathcal{P}}_\mathrm{t}(\mathbf{h}) &= \sum_{i=1}^{2NT_\mathrm{t}} c\sqrt{2\rho}\,\mathbf{p}_i\left(1 - \frac{1}{1+e^{cs_{\mathrm{t},i}^{\mathrm{up}}}} - \frac{1}{1+e^{cs_{\mathrm{t},i}^{\mathrm{low}}}}\right) \\ &= c\sqrt{2\rho}\,\mathbf{P}^T\Big[\mathbf{1} - \sigma\left(c\sqrt{2\rho}\,(\mathbf{P}\mathbf{h} - \mathbf{q}_\mathrm{t}^{\mathrm{up}})\right) \\ &\quad - \sigma\left(c\sqrt{2\rho}\left(\mathbf{P}\mathbf{h} - \mathbf{q}_\mathrm{t}^{\mathrm{low}}\right)\right)\Big]\end{aligned} \tag{19}$$

in which $\mathbf{q}_\mathrm{t}^{\mathrm{up}} = [q_{\mathrm{t},1}^{\mathrm{up}}, \ldots, q_{\mathrm{t},2NT_\mathrm{t}}^{\mathrm{up}}]^T$ and $\mathbf{q}_\mathrm{t}^{\mathrm{low}} = [q_{\mathrm{t},1}^{\mathrm{low}}, \ldots, q_{\mathrm{t},2NT_\mathrm{t}}^{\mathrm{low}}]^T$. Here, it should be noted that, for a matrix or vector argument, $\sigma(\cdot)$ is applied separately to every element.

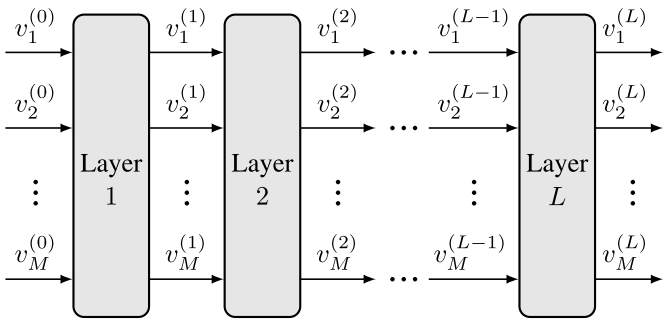
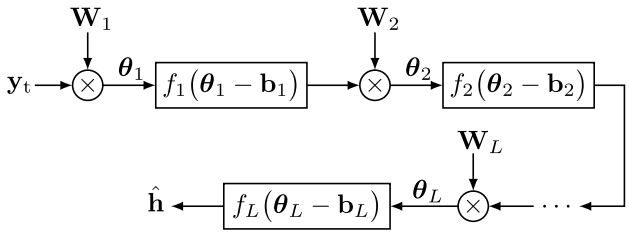
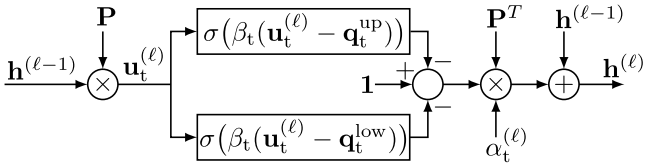


Fig. 1. Overall structure of the proposed FBM-CENet, FBM-DetNet, and B-DetNet. For FBM-CENet, v plays the role of h and $M = 2NK$. For FBM-DetNet and B-DetNet, v plays the role of x and $M = 2K$.



(a) Conventional DNN channel estimation structure. Each layer ℓ contains a trainable weight matrix $\mathbf{W}_\ell$, a trainable bias vector $\mathbf{b}_\ell$, and an activation function $f_\ell(\cdot)$.



(b) Specific structure of layer ℓ of the proposed FBM-CENet.

Fig. 2. Conventional versus proposed DNN structure for channel estimation.

Unlike (19), it can be seen that the gradient of $\tilde{\mathcal{P}}_\mathrm{t}(\mathbf{h})$ does not suffer from the divide-by-zero issue. Thus, an iterative gradient descent method for solving (18) can be written as

$$\mathbf{h}^{(\ell)} = \mathbf{h}^{(\ell-1)} + \alpha_\mathrm{t}^{(\ell)} \nabla \tilde{\mathcal{P}}_\mathrm{t}\big(\mathbf{h}^{(\ell-1)}\big) \tag{20}$$

where ℓ is the iteration index and $\alpha_\mathrm{t}^{(\ell)}$ is the step size.

2) Structure of the Proposed FBM-CENet: We employ the deep unfolding technique [45] to unfold each iteration in (20) as a layer of a deep neural network. The overall structure of the proposed FBM-CENet estimator is illustrated in Fig. 1, where each of the L layers takes a vector of $2NK$ elements as the input and generates an output vector of the same size.

The specific structure for each layer ℓ of the proposed FBM-CENet is illustrated in Fig. 2b. The proposed layer structure is unique due to the use of the approximation in (16) and the structure of the reformulated gradient in (19). Specifically, each layer of the proposed FBM-CENet consists of two weight matrices and two bias vectors where the pilot matrix $\mathbf{P}$ plays the role of the weight matrices and the received signals $\mathbf{q}_\mathrm{t}^{\mathrm{up}}$ and $\mathbf{q}_\mathrm{t}^{\mathrm{low}}$ play the role of the bias vectors. By contrast, each layer ℓ of a conventional DNN-based channel estimator as illustrated in Fig. 2a contains one weight matrix $\mathbf{W}_\ell$ and one bias vector $\mathbf{b}_\ell$. Such a conventional DNN structure has been employed in several existing works, e.g., [23]–[25]. An interesting feature of the proposed network is the Sigmoid activation function $\sigma(\cdot)$, which is not arbitrary but results from the use of the approximation in (16). This is unlike conventional DNN structures where the activation functions $\{f_\ell(\cdot)\}$ are often chosen heuristically by experiments.

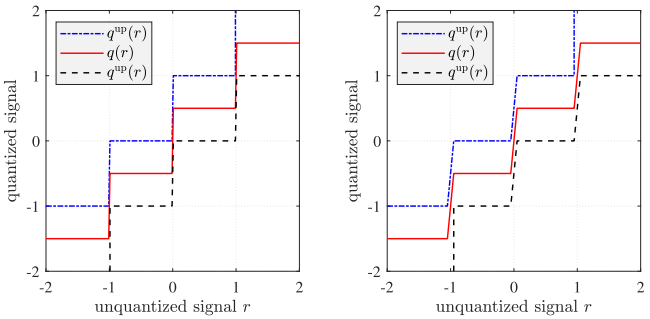
It should be noted that the proposed FBM-CENet structure in Fig. 2b is free of the constant $c\sqrt{2\rho}$ since it is absorbed into the trainable parameters $\alpha_\mathrm{t}^{(\ell)}$ and β_t. If the constant is kept, each layer ℓ will contain only one trainable parameter, which is the step size $\alpha_\mathrm{t}^{(\ell)}$. Training $\alpha_\mathrm{t}^{(\ell)}$ can be interpreted as moving along the gradient directions and optimizing the step size at each layer. We refer this network structure to as purely gradient-based FBM-CENet (PG-FBM-CENet). Since different values of β_t result in different directions in the vicinity of the gradient, training FBM-CENet can be interpreted as jointly learning the optimal directions and the associated optimal step sizes. This helps FBM-CENet improve the performance compared to PG-FBM-CENet. The reason is that always moving along the gradient direction may not be optimal. FBM-CENet learns an optimal path that makes the network output (the channel estimate) closer to the true channel vector. We will numerically show that FBM-CENet outperforms PG-FBM-CENet later.

3) Trainable Parameters: For a given pilot matrix $\mathbf{P}$, the trainable parameters in the proposed FBM-CENet are the step sizes $\{\alpha_\mathrm{t}^{(\ell)}\}$ and the scaling parameter β_t inside the Sigmoid function. Note that as mentioned above, the coefficient $c\sqrt{2\rho}$ is omitted in the proposed network since it is absorbed in the trainable parameters $\{\alpha_\mathrm{t}^{(\ell)}\}$ and β_t.

It is important to note that the pilot matrix $\mathbf{P}$ directly plays the role of the weight matrices. Therefore, when the pilot matrix $\mathbf{P}$ is not given, it too can be treated as a trainable parameter. In this case, training the proposed FBM-CENet is equivalent to *jointly optimizing both the channel estimator at the base station and the pilot signal transmitted from the users*. This is a significant advantage of the proposed network structure since the conventional DNN-based channel estimator is often trained or optimized for a given pilot matrix, and thus is unable to convey information about the optimal pilot signal. The approach proposed in [23] also jointly optimized the pilot signal and the channel estimator for massive MIMO systems with low-resolution ADCs, but it employs the conventional DNN structure illustrated in Fig. 2a. We will later show that the proposed FBM-CENet estimator significantly outperforms the method in [23].

4) Training Strategy: Here we present the strategy for training the proposed FBM-CENet estimator. Let $\hat{\mathbf{h}}$ denote the channel estimate, which is set to be the output of the last layer of FBM-CENet, i.e., $\hat{\mathbf{h}} = \mathbf{h}^{(L)}$. The cost function to be minimized is $\|\hat{\mathbf{h}} - \mathbf{h}\|^2$. We choose this cost function instead of the objective function in (17) because the value of (17) is undefined when the argument of the logarithm approaches zero. In our investigation, training using the cost function (17) was not successful due to this issue. When the pilot matrix $\mathbf{P}$ is given, a training sample for FBM-CENet consists of the given matrix $\mathbf{P}$, a channel vector realization $\mathbf{h}$ and a Gaussian noise vector $\mathbf{z}$, which can be randomly generated. When the pilot matrix $\mathbf{P}$ is not given and is trainable, a training sample only consists of $\mathbf{h}$ and a Gaussian noise vector $\mathbf{z}$. Note that $\mathbf{h}$ is randomly generated according to a particular channel model.

It is important to note that the received signals $\mathbf{q}_\mathrm{t}^{\mathrm{up}}$ and $\mathbf{q}_\mathrm{t}^{\mathrm{low}}$ depend on the pilot matrix $\mathbf{P}$. Therefore, when the pilot matrix $\mathbf{P}$ is trainable, gradient back-propagation during the training process should also go through $\mathbf{q}_\mathrm{t}^{\mathrm{up}}$ and $\mathbf{q}_\mathrm{t}^{\mathrm{low}}$. However, the low-resolution ADCs are discontinuous functions, which make gradient back-propagation through $\mathbf{q}_\mathrm{t}^{\mathrm{up}}$ and $\mathbf{q}_\mathrm{t}^{\mathrm{low}}$ infeasible. To overcome this issue, we employ a soft quantizer model based on the Rectified Linear Unit (ReLU) function



(a) $c_1 = 0.01$. (b) $c_1 = 0.05$.

Fig. 3. Two-bit ReLU-based soft quantizer with $\Delta = 1$.

$f_{\text{relu}}(r) = \max(0, r)$ for the training process as follows:

$$q^{\text{up}}(r) = q(r) + \frac{\Delta}{2} + c_2\big[f_{\text{relu}}(r - B\Delta + c_1) - f_{\text{relu}}(r - B\Delta - c_1)\big] \tag{21}$$

$$q^{\text{low}}(r) = q(r) - \frac{\Delta}{2} - c_2\big[f_{\text{relu}}(-r - B\Delta + c_1) - f_{\text{relu}}(-r - B\Delta - c_1)\big] \tag{22}$$

where $B = 2^{b-1} - 1$, c_1 and c_2 are positive constants, and

$$q(r) = -(2^b - 1)\frac{\Delta}{2} + \frac{\Delta}{2c_1}\sum_{i=-B}^{B}\big[f_{\text{relu}}(r + i\Delta + c_1) - f_{\text{relu}}(r + i\Delta - c_1)\big]. \tag{23}$$

The resulting ReLU-based function is continuous and therefore back-propagation is feasible. The effect of c_1 is illustrated in Fig. 3. It can be seen that smaller values of c_1 make the soft quantizer sharper, or in other words closer to the actual hard quantizer. The constant c_2 accounts for the two thresholds $\tau_0 = -\infty$ and $\tau_{2^b} = \infty$, and hence should be large. The constants $\{c_1, c_2\}$ should not be treated as trainable parameters because it is necessary for the soft quantizer to be a reasonable approximation of the hard quantizer. Allowing these constants to be trained may produce a large deviation between the soft and hard quantizers.

Note that the soft quantizer could also be modeled using the tanh function as follows:

$$q^{\text{up}}(r) = q(r) + \frac{\Delta}{2} + c_4 f_{\tanh}(c_3(r - B\Delta)) \tag{24}$$

$$q^{\text{low}}(r) = q(r) - \frac{\Delta}{2} - c_4 f_{\tanh}(c_3(-r - B\Delta)) \tag{25}$$

where $f_{\tanh}(r) = (\tanh(r) + 1)/2$ and

$$q(r) = \frac{\Delta}{2}\big[f_{\tanh}(c_3 r) - f_{\tanh}(-c_3 r)\big] + \Delta\sum_{i=1}^{B} f_{\tanh}(c_3(r - i\Delta)) - f_{\tanh}(c_3(-r - i\Delta)). \tag{26}$$

Larger values of c_3 make the soft quantizer sharper. The constant c_4 accounts for the two thresholds τ_0 and τ_{2^b}, and hence should also be large. Although we implement our networks with the ReLU-based model, in the simulations we will show that both the tanh- and ReLU-based soft quantizers yield essentially the same performance.

IV. Data Detection in Few-Bit MIMO Systems

In this section, we propose two DNN-based detectors, referred to as B-DetNet and FBM-DetNet, for low-resolution massive MIMO systems. For convenience in later derivations, we convert (1) and (2) into the real domain as follows:

$$\mathbf{y} = \mathcal{Q}_b\left(\mathbf{H}\mathbf{x} + \mathbf{z}\right), \tag{27}$$

where

$$\mathbf{y} = \begin{bmatrix}\Re\{\bar{\mathbf{y}}\}\\ \Im\{\bar{\mathbf{y}}\}\end{bmatrix},\ \mathbf{x} = \begin{bmatrix}\Re\{\bar{\mathbf{x}}\}\\ \Im\{\bar{\mathbf{x}}\}\end{bmatrix},\ \mathbf{z} = \begin{bmatrix}\Re\{\bar{\mathbf{z}}\}\\ \Im\{\bar{\mathbf{z}}\}\end{bmatrix},\ \text{and}\ \mathbf{H} = \begin{bmatrix}\Re\{\bar{\mathbf{H}}\} & -\Im\{\bar{\mathbf{H}}\}\\ \Im\{\bar{\mathbf{H}}\} & \Re\{\bar{\mathbf{H}}\}\end{bmatrix}.$$

Note that $\mathbf{y} \in \mathbb{R}^{2N}$, $\mathbf{x} \in \mathbb{R}^{2K}$, $\mathbf{z} \in \mathbb{R}^{2N}$, and $\mathbf{H} \in \mathbb{R}^{2N\times 2K}$. We also denote $\mathbf{y} = [y_1, \ldots, y_{2N}]^T$ and $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_{2N}]^T$.

A. Proposed B-DetNet

Applying the Bussgang decomposition to (27), we obtain

$$\begin{aligned}\mathbf{y} &= \mathbf{V}\mathbf{H}\mathbf{x} + \mathbf{V}\mathbf{z} + \mathbf{d}\\ &= \mathbf{A}\mathbf{x} + \mathbf{n}\end{aligned} \tag{28}$$

where $\mathbf{V} \in \mathbb{R}^{2N\times 2N}$ is a diagonal matrix and given as

$$\mathbf{V} = \frac{\Delta}{\sqrt{2\pi}}\operatorname{diag}(\mathbf{\Sigma_r})^{-\frac{1}{2}} \times \sum_{i=1}^{2^b-1}\exp\left\{-\frac{1}{2}\Delta^2(i - 2^{b-1})^2\operatorname{diag}(\mathbf{\Sigma_r})^{-1}\right\}$$

and $\mathbf{\Sigma_r} = \mathbf{H}\mathbf{\Sigma_x}\mathbf{H}^T + \frac{N_0}{2}\mathbf{I} \in \mathbb{R}^{2N\times 2N}$. For the case of 1-bit ADCs, the covariance of $\mathbf{n}$ is given in closed form as [46]

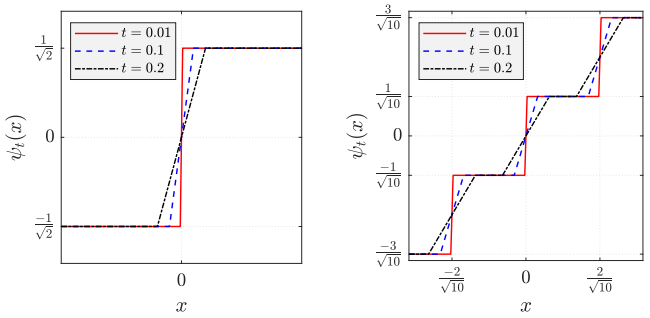
$$\mathbf{\Sigma_n} = \frac{\Delta^2}{2\pi}\Big[\arcsin\Big(\operatorname{diag}(\mathbf{\Sigma_r})^{-\frac{1}{2}}\mathbf{\Sigma_r}\operatorname{diag}(\mathbf{\Sigma_r})^{-\frac{1}{2}}\Big) - \operatorname{diag}(\mathbf{\Sigma_r})^{-\frac{1}{2}}\mathbf{\Sigma_r}\operatorname{diag}(\mathbf{\Sigma_r})^{-\frac{1}{2}} + \frac{N_0}{2}\operatorname{diag}(\mathbf{\Sigma_r})^{-1}\Big]. \tag{29}$$

For few-bit ADCs, the covariance of $\mathbf{n}$ can be approximated as $\mathbf{\Sigma_n} \approx \frac{N_0}{2}\mathbf{V}\mathbf{V}^T + \eta_b\operatorname{diag}(\mathbf{\Sigma_r})$. The effective noise $\mathbf{n}$ is often modeled as $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma_n})$. Based on this linearized model, different linear detectors such as BZF, BMMSE, and BWZF were introduced in [18], [28], [29].

Here, we propose the data detection network B-DetNet based on the linearized system model in (28). Since the effective noise $\mathbf{n}$ is assumed to be Gaussian, the Bussgang-based maximum likelihood detection problem is given as

$$\hat{\mathbf{x}}_{\text{BML}} = \arg\min_{\bar{\mathbf{x}}\in\bar{\mathcal{M}}^K}(\mathbf{y} - \mathbf{A}\mathbf{x})^T\mathbf{\Sigma_n}^{-1}(\mathbf{y} - \mathbf{A}\mathbf{x}). \tag{30}$$

Let $P_{\text{B}}(\mathbf{x})$ be the objective function of (30). Note that $P_{\text{B}}(\mathbf{x})$ is a quadratic function of $\mathbf{x}$ and thus convex, but the optimization problem is not convex due to the discrete feasibility constraint $\bar{\mathbf{x}} \in \bar{\mathcal{M}}^K$. An optimal solution to (30) therefore requires an



(a) QPSK signaling. (b) 16QAM signaling.

Fig. 4. Projector function $\psi_t(\cdot)$ with different values of t.

exhaustive search, which is very expensive for large scale systems. Instead, an iterative projected gradient descent method

$$\mathbf{x}^{(\ell)} = \psi_{t_\ell}\left(\mathbf{x}^{(\ell-1)} - \alpha^{(\ell)}\nabla P_{\mathrm{B}}(\mathbf{x}^{(\ell-1)})\right) \tag{31}$$

can be applied to search for the optimal solution. Herein, the gradient of $P_{\mathrm{B}}(\mathbf{x})$ evaluated at $\mathbf{x}^{(\ell-1)}$ is given by

$$\nabla P_{\mathrm{B}}(\mathbf{x}^{(\ell-1)}) = -2\mathbf{A}^T\boldsymbol{\Sigma}_{\mathbf{n}}^{-1}\left(\mathbf{y} - \mathbf{A}\mathbf{x}^{(\ell-1)}\right) \tag{32}$$

and $\psi_{t_\ell}(\cdot)$, characterized by the positive parameter t_ℓ, is a non-linear projector that forces the signal to the nearest constellation point. Based on the ReLU activation function $q(r)$ in (23), $\psi_{t_\ell}(\cdot)$ can be written as

$$\psi_{t_\ell}(x) = -(2^{b'}-1)\frac{\Delta'}{2} + \frac{\Delta'}{2t_\ell}\sum_{i=-B'}^{B'}\left[f_{\mathrm{relu}}(r+i\Delta+t_\ell) - f_{\mathrm{relu}}(r+i\Delta-t_\ell)\right] \tag{33}$$

where $B' = 2^{b'-1}-1$. For QPSK signalling, $\{b', \Delta'\} = \{1, \frac{2}{\sqrt{2}}\}$ and for 16-QAM signalling, $\{b', \Delta'\} = \{2, \frac{2}{\sqrt{10}}\}$. The effect of t_ℓ on $\psi_t(\cdot)$ is illustrated in Fig. 4. It can be seen that a smaller t_ℓ also makes the projector sharper. Such a projection function was used in [47], which studied deep learning-based detection for unquantized MIMO systems.

Our proposed B-DetNet approach is realized by unfolding the projected gradient descent in (31). The overall structure of B-DetNet is illustrated in Fig. 1. Each layer takes an input vector of size $2K$ and generates an output vector of the same size. The specific structure of each B-DetNet layer is given in Fig. 5 where $\hat{\mathbf{A}}$ and $\hat{\mathbf{A}}^T\hat{\boldsymbol{\Sigma}}_{\mathbf{n}}^{-1}$ play the role of weight matrices. Note that $\hat{\mathbf{A}}$ and $\hat{\boldsymbol{\Sigma}}_{\mathbf{n}}^{-1}$ are obtained using a channel estimate $\hat{\mathbf{H}}$ from, e.g., FBM-CENet. The received signal vector $\mathbf{y}$ is seen to be the bias vector. Hence, B-DetNet is highly adaptive to the channel. The only trainable parameters in layer ℓ of B-DetNet are the step size $\alpha^{(\ell)}$ and the scaling parameter t_ℓ in the projector function $\psi_{t_\ell}(\cdot)$.

We note that similar structures for data detection in full-resolution systems have been developed in [47], [48]. However, the received signal in full-resolution systems is given as $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}$, and therefore the gradient of interest becomes $-2\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{x})$. For low-resolution systems, we have a new effective channel $\mathbf{A}$ and a new noise covariance matrix $\boldsymbol{\Sigma}_{\mathbf{n}}$, resulting in a new form of the gradient as in (32).

B. Proposed FBM-DetNet

1) Maximum-Likelihood Data Detection Problem: Let $s_i^{\mathrm{up}} = \sqrt{2\rho}(q_i^{\mathrm{up}} - \mathbf{h}_i^T\mathbf{x})$ and $s_i^{\mathrm{low}} = \sqrt{2\rho}(q_i^{\mathrm{low}} - \mathbf{h}_i^T\mathbf{x})$, where

$$q_i^{\mathrm{up}} = \begin{cases} y_i + \frac{\Delta}{2} & \text{if } y_i < \tau_{2^b-1} \\ \infty & \text{otherwise,} \end{cases}$$

$$q_i^{\mathrm{low}} = \begin{cases} y_i - \frac{\Delta}{2} & \text{if } y_i > \tau_1 \\ -\infty & \text{otherwise.} \end{cases}$$

Hence, q_i^{up} and q_i^{low} are the upper and lower quantization thresholds of the bin to which y_i belongs. The ML detection problem based on the log-likelihood function for the model in (27) is defined as follows [49]:

$$\hat{\mathbf{x}}_{\mathrm{ML}} = \arg\max_{\bar{\mathbf{x}}\in\bar{\mathcal{M}}^K}\sum_{i=1}^{2N}\log\left[\Phi\left(s_i^{\mathrm{up}}\right) - \Phi\left(s_i^{\mathrm{low}}\right)\right]. \tag{34}$$

Let $\mathcal{P}(\mathbf{x})$ denote the objective function of (34), which is a concave function of $\mathbf{x}$. However, the optimization problem (34) is not convex since the feasible set is discrete. Therefore, an optimal solution for ML detection in (34) also requires an exhaustive search over $\bar{\mathcal{M}}^K$, which is prohibitively complex for large-scale systems. One can relax the constraint on the feasible set from $\bar{\mathbf{x}} \in \bar{\mathcal{M}}^K$ to $\bar{\mathbf{x}} \in \mathbb{C}^K$ in order to obtain a convex optimization problem and allow an iterative gradient descent method to be used. Unfortunately, such an approach also suffers from the indeterminant gradient issue discussed earlier for the channel estimation problem. In addition, there is no closed-form expression for $\Phi(\cdot)$, which complicates the evaluation in (34). As before, we exploit the approximation in (16) to obtain an approximate version of the function $\mathcal{P}(\mathbf{x})$ as follows:

$$\mathcal{P}(\mathbf{x}) \approx \tilde{\mathcal{P}}(\mathbf{x}) = \textstyle\sum_{i=1}^{2N}\log\left[\frac{1}{1+e^{-cs_i^{\mathrm{up}}}} - \frac{1}{1+e^{-cs_i^{\mathrm{low}}}}\right] \tag{35}$$

The reformulated ML detection problem is thus

$$\hat{\mathbf{x}}_{\mathrm{ML}} = \arg\max_{\bar{\mathbf{x}}\in\bar{\mathcal{M}}^K}\tilde{\mathcal{P}}(\mathbf{x}), \tag{36}$$

and the gradient of $\tilde{\mathcal{P}}(\mathbf{x})$ is

$$\nabla\tilde{\mathcal{P}}(\mathbf{x}) = \sum_{i=1}^{2N} c\sqrt{2\rho}\,\mathbf{h}_i\left(1 - \frac{1}{1+e^{cs_i^{\mathrm{up}}}} - \frac{1}{1+e^{cs_i^{\mathrm{low}}}}\right) \tag{37}$$

$$= c\sqrt{2\rho}\,\mathbf{H}^T\left[\mathbf{1} - \sigma\left(c\sqrt{2\rho}\left(\mathbf{H}\mathbf{x} - \mathbf{q}^{\mathrm{up}}\right)\right) - \sigma\left(c\sqrt{2\rho}\left(\mathbf{H}\mathbf{x} - \mathbf{q}^{\mathrm{low}}\right)\right)\right] \tag{38}$$

where $\mathbf{q}^{\mathrm{up}} = [q_1^{\mathrm{up}}, \ldots, q_{2N}^{\mathrm{up}}]^T$ and $\mathbf{q}^{\mathrm{low}} = [q_1^{\mathrm{low}}, \ldots, q_{2N}^{\mathrm{low}}]^T$. Thus, an iterative projected gradient decent method for solving (36) can be written as

$$\mathbf{x}^{(\ell)} = \psi_{t_\ell}\left(\mathbf{x}^{(\ell-1)} + \alpha^{(\ell)}\nabla\tilde{\mathcal{P}}(\mathbf{x}^{(\ell-1)})\right) \tag{39}$$

where ℓ is the iteration index, $\alpha^{(\ell)}$ is a step size, and $\psi_{t_\ell}(\cdot)$ is also a projector as defined in (33).

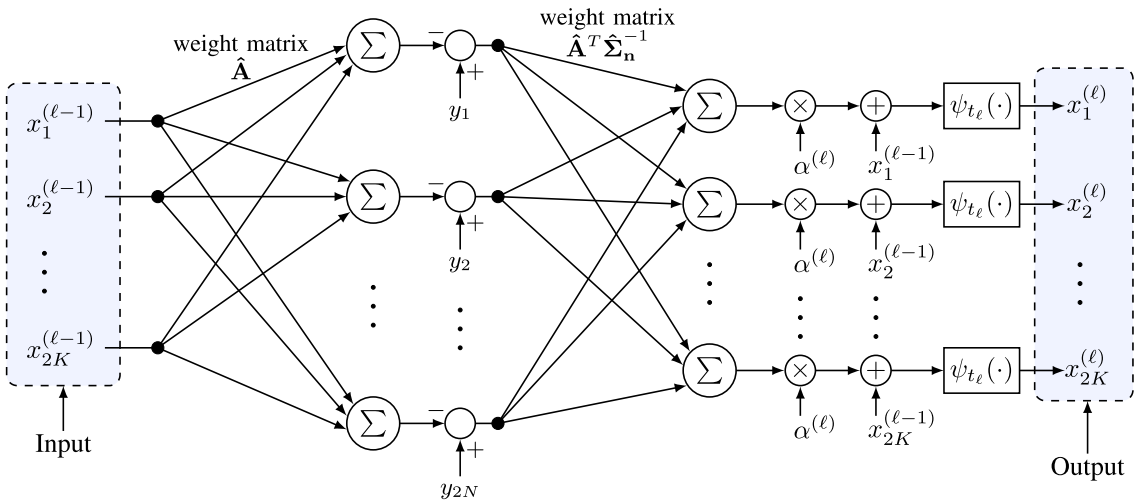


Fig. 5. Specific structure of layer ℓ of the proposed B-DetNet.

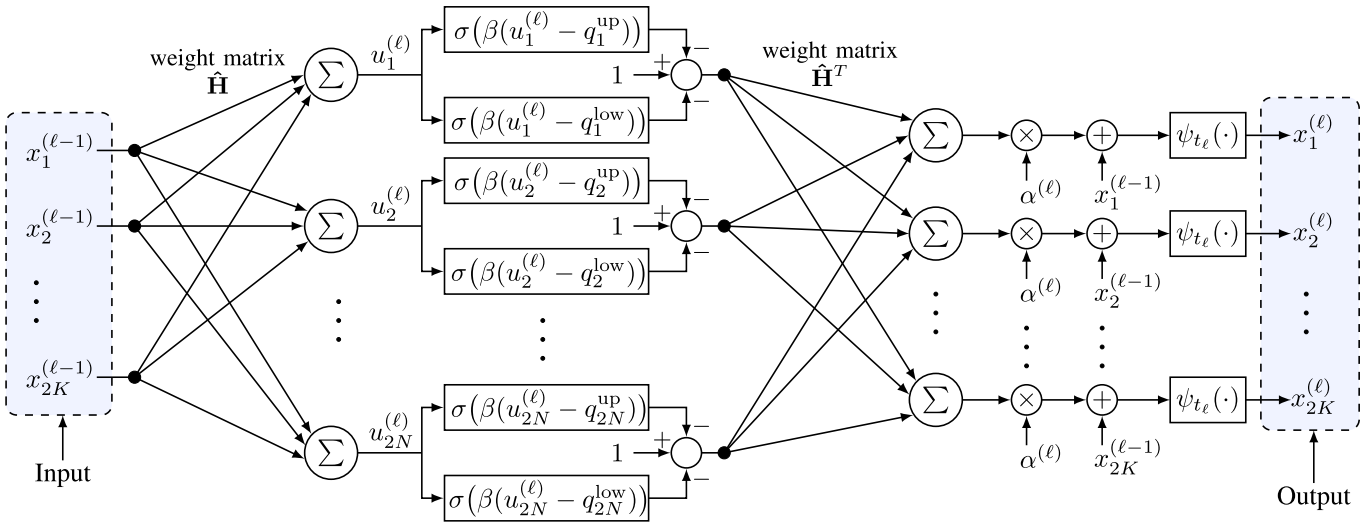


Fig. 6. Specific structure of layer ℓ of FBM-DetNet. The weight matrices and the bias vectors are defined by the channel and the received signal, respectively.

2) Structure of the Proposed FBM-DetNet: In order to optimize the step sizes $\{\alpha^{(\ell)}\}$ and scaling parameters $\{t_\ell\}$ of the projection function, we also unfold each iteration in (39) as a separate DNN layer. The overall structure of the proposed detector FBM-DetNet is also illustrated in Fig. 1, and is similar to that of B-DetNet as each layer of both networks takes a vector of $2K$ elements as the input and generates an output vector of the same size.

The specific structure for each layer ℓ of FBM-DetNet is illustrated in Fig. 6. Each layer of FBM-DetNet has two weight matrices $\mathbf{H}$ and $\mathbf{H}^T$ and two bias vectors $\mathbf{q}^{\mathrm{up}}$ and $\mathbf{q}^{\mathrm{low}}$ defined by the channel and the received signal, respectively. The activation function is the Sigmoid function $\sigma(\cdot)$ due to the use of the approximation in (16). Since $\mathbf{H} \in \mathbb{R}^{2N\times 2K}$, the learning process for each layer of FBM-DetNet can be interpreted as first up-converting the signal $\mathbf{x}^{(\ell-1)}$ from dimension $2K$ to dimension $2N$ using the weight matrix $\mathbf{H}$, then applying the nonlinear activation function $\sigma(\cdot)$ before down-converting the signal back to dimension $2K$ using the weight matrix $\mathbf{H}^T$. Finally, the function $\psi_{t_\ell}(\cdot)$ is implemented to project $\mathbf{x}^{(\ell-1)}$ onto the discrete set $\tilde{\mathcal{M}}^K$.

The layers of FBM-DetNet are similar to those of FBM-CENet in Fig. 2b. However, while the weight matrices of FBM-CENet are defined by the pilot matrix $\mathbf{P}$ which is trainable, the weight matrices of FBM-DetNet are defined by the channel matrix $\mathbf{H}$ which is not. Thus, FBM-DetNet is highly adaptive to the channel. The trainable parameters of FBM-DetNet are the step sizes $\{\alpha^{(\ell)}\}$, scaling parameters $\{t_\ell\}$ for the projector, and a scaling parameter β for the Sigmoid function. Note that the coefficient $c\sqrt{2\rho}$ is also omitted in FBM-DetNet for the same reason as in FBM-CENet.

C. Training Strategy

A training sample for the two proposed data detection networks, B-DetNet and FBM-DetNet, can be obtained by first randomly generating a channel matrix $\mathbf{H}$ according to a particular channel model, then obtaining an estimate $\hat{\mathbf{H}}$ of

TABLE II

COMPUTATIONAL COMPLEXITY COMPARISON OF CHANNEL ESTIMATION METHODS, WHERE $f_{\mathtt{s1}}(\cdot)$ REPRESENTS A SUPER-LINEAR FUNCTION

Method	Complexity
BMMSE	i.i.d. channels: $\mathcal{O}(KN^2T_{\mathrm{t}})$
	correlated channels: $\mathcal{O}(KN^3T_{\mathrm{t}}^2)$
BWZF	$\mathcal{O}\big(K^2N^3T_{\mathrm{t}}\big)$
SVM-based	$\mathcal{O}\big(KNT_{\mathrm{t}}f_{\mathtt{s1}}(T_{\mathrm{t}})\big)$
Proposed FBM-CENet	$\mathcal{O}\big(KN^2LT_{\mathrm{t}}\big)$

TABLE III

COMPUTATIONAL COMPLEXITY COMPARISON OF DATA DETECTION METHODS

Method	Preprocessing	Detection Stage
BMMSE	$\mathcal{O}(N^3)$	$\mathcal{O}(KNT_{\mathrm{d}})$
BWZF	$\mathcal{O}\big(KN\big)$	$\mathcal{O}\big(K^2NT_{\mathrm{d}}\big)$
SVM-based	–	$\mathcal{O}\big(KNf_{\mathtt{s1}}(N)T_{\mathrm{d}}\big)$
Proposed B-DetNet	$\mathcal{O}(N^3)$	$\mathcal{O}\big(KNLT_{\mathrm{d}}\big)$
Proposed FBM-DetNet	–	$\mathcal{O}\big(KNLT_{\mathrm{d}}\big)$

$\mathbf{H}$ by using, e.g., FBM-CENet. Next, the transmit signal $\mathbf{x}$ can be randomly drawn from the signal constellation and a Gaussian noise vector $\mathbf{z}$ can be chosen to obtain a representative received signal vector $\mathbf{y} = \mathcal{Q}_b(\mathbf{H}\mathbf{x} + \mathbf{z})$. Similar to the channel estimation network, the cost function to be minimized is $\|\mathbf{x}^{(L)} - \mathbf{x}\|^2$, where $\mathbf{x}$ is the target (transmitted) signal. For training the proposed data detection networks, we do not need to use the soft quantization model because the trainable parameters do not appear in the received signals $\mathbf{y}$ or $\mathbf{q}^{\mathrm{up}}$ and $\mathbf{q}^{\mathrm{low}}$, and therefore the exact hard quantizer can be used.

V. COMPUTATIONAL COMPLEXITY ANALYSIS AND NUMERICAL RESULTS

A. Computational Complexity Analysis

Here we present a Big-$\mathcal{O}$ computational complexity analysis for the considered channel estimation and data detection methods. The presented complexities only account for the online processing phase. Offline computations are excluded. Table II compares the complexity of different channel estimation methods. It can be seen that the complexity of BMMSE is the lowest and highest when the channels are i.i.d. and correlated, respectively. This is because the BMMSE estimation matrix can be computed offline for i.i.d. channels, but online for correlated channels. The complexity of BWZF is higher than that of the SVM method and the proposed FBM-CENet because the BWZF estimation matrix must also be computed online as it depends on the received signal. The complexity order of the proposed FBM-CENet is higher than BMMSE with i.i.d. channels, but scales more slowly than the SVM method since the SVM complexity is a super-linear function of T_{t}.

The complexity comparison of different data detection methods is given Table III. Note that while the detection methods SVM and FBM-DetNet do not require preprocessing, BMMSE, BWZF, and B-DetNet require a preprocessing step due to the linearization process of the Bussgang decomposition. In the detection stage, BMMSE has the lowest complexity since it requires only one matrix-vector multiplication for each time slot. The complexity of BWZF is higher than BMMSE since the demultiplexing matrix of BWZF has to be re-computed in each time slot. The detection complexities of the proposed B-DetNet and FBM-DetNet are higher than the complexity of BMMSE, but lower than that of the SVM-based method.

B. Numerical Results

Here we present numerical results that illustrate the superior performance of the proposed channel estimation and data detection networks. For training the networks, we use TensorFlow [50] and the Adam optimizer [51] with a learning rate that starts at 0.002 and decays at a rate of 0.97 after every 100 training epochs. The size of each training batch is set to 1000. The input of the first layer is set to a zero vector. When the pilot matrix $\mathbf{P}$ is trainable, we use the soft quantization model in (21) and (22) for the training phase and set $c_1 = 0.01$ and $c_2 = c_3 = c_4 = 1000$. For the channel estimation phase, we set the training length to be five times the number of users, i.e., $T_{\mathrm{t}} = 5K$.

We consider the following channel model:

$$\bar{\mathbf{H}} = \bar{\mathbf{H}}^{\mathtt{LoS}}\mathbf{\Xi}^{\mathtt{LoS}} + \bar{\mathbf{H}}^{\mathtt{NLoS}}\mathbf{\Xi}^{\mathtt{NLoS}}, \tag{40}$$

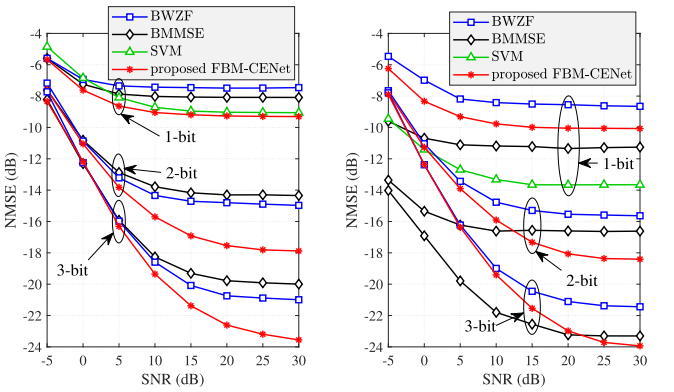
where $\bar{\mathbf{H}}^{\mathtt{LoS}}\mathbf{\Xi}^{\mathtt{LoS}}$ and $\bar{\mathbf{H}}^{\mathtt{NLoS}}\mathbf{\Xi}^{\mathtt{NLoS}}$ account for the Line-of-Sight (LoS) and Non-Line-of-Sight (NLoS) channels, respectively. The matrices $\mathbf{\Xi}^{\mathtt{LoS}}$ and $\mathbf{\Xi}^{\mathtt{NLoS}}$ are diagonal and defined as $\mathbf{\Xi}^{\mathtt{LoS}} = \mathrm{diag}(\xi_1^{\mathtt{LoS}}, \ldots, \xi_K^{\mathtt{LoS}})$ and $\mathbf{\Xi}^{\mathtt{NLoS}} = \mathrm{diag}(\xi_1^{\mathtt{NLoS}}, \ldots, \xi_K^{\mathtt{NLoS}})$ where

$$\xi_k^{\mathtt{LoS}} = \sqrt{\frac{\kappa_k}{\kappa_k + 1}} \text{ and } \xi_k^{\mathtt{NLoS}} = \sqrt{\frac{1}{\kappa_k + 1}}, \tag{41}$$

and κ_k is the Rician factor of the channel from the k-th user to the BS. If $\kappa_k = 0$, there is no LoS component between user-k and the BS. Let $\bar{\mathbf{H}}^{\mathtt{LoS}} = [\bar{\mathbf{h}}_1^{\mathtt{LoS}}, \ldots, \bar{\mathbf{h}}_K^{\mathtt{LoS}}]$ and $\bar{\mathbf{H}}^{\mathtt{NLoS}} = [\bar{\mathbf{h}}_1^{\mathtt{NLoS}}, \ldots, \bar{\mathbf{h}}_K^{\mathtt{NLoS}}]$. The LoS channel vector $\bar{\mathbf{h}}_k^{\mathtt{LoS}}$ is given as [52]

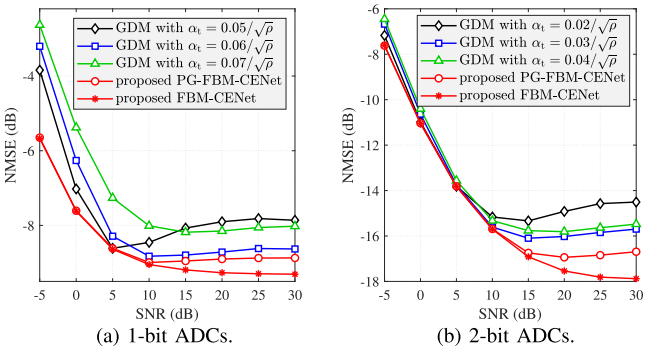
$$\bar{\mathbf{h}}_k^{\mathtt{LoS}} = \sqrt{\gamma_k}e^{j\varphi_k}[1, e^{j2\pi d_{\mathtt{A}}\sin(\theta_k)}, \ldots, e^{j2\pi d_{\mathtt{A}}(N-1)\sin(\theta_k)}]^T \tag{42}$$

where γ_k is the large-scale fading coefficient, $\varphi_k \in [0, 2\pi]$ is a random phase shift, and $d_{\mathtt{A}}$ is the antenna spacing parameter (in fractions of a wavelength), and $-\pi/3 \leq \theta_k \leq \pi/3$ is the angle-of-arrival seen at the BS for user-k. The NLoS channel is given as $\bar{\mathbf{h}}_k^{\mathtt{NLoS}} \sim \mathcal{CN}(\mathbf{0}, \bar{\mathbf{C}}_k)$ where $\mathrm{tr}(\bar{\mathbf{C}}_k)/N = \gamma_k$. Note that the channels between a user and different receive antennas can be correlated, but the channels between the users and the BS are uncorrelated. The large-scale fading coefficient is modeled (in dB) as [53] $\gamma_k = -30.18 - 26\log_{10}(d_k) + F_k$ where d_k is the distance between user-k and the BS, and $F_k \sim \mathcal{N}(0, \sigma_{\mathtt{sh}}^2)$ is the shadow fading coefficient with $\sigma_{\mathtt{sh}} = 4$.



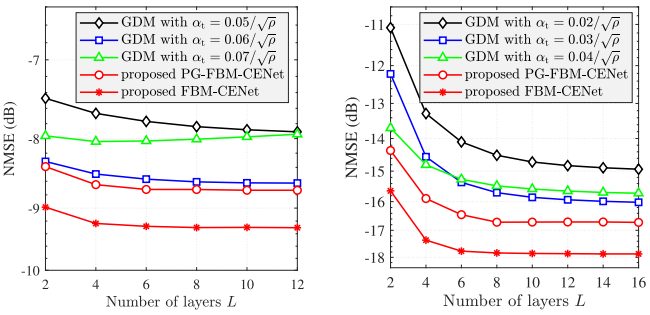
(a) Uncorrelated NLoS channels. (b) Spatially correlated and mixed LoS-NLoS channels.

Fig. 7. Channel estimation performance comparison for a given pilot matrix with $K = 4$, $L = 8$, and $N = 32$.



(a) 1-bit ADCs. (b) 2-bit ADCs.

Fig. 8. Channel estimation performance of GDM versus the proposed FBM-CENet with $K = 4$, $L = 8$, and $N = 32$.



(a) 1-bit ADCs. (b) 2-bit ADCs.

Fig. 9. Channel estimation performance of GDM versus the proposed FBM-CENet with different values of L, $K = 4$, $N = 32$, and SNR $= 30$ dB.

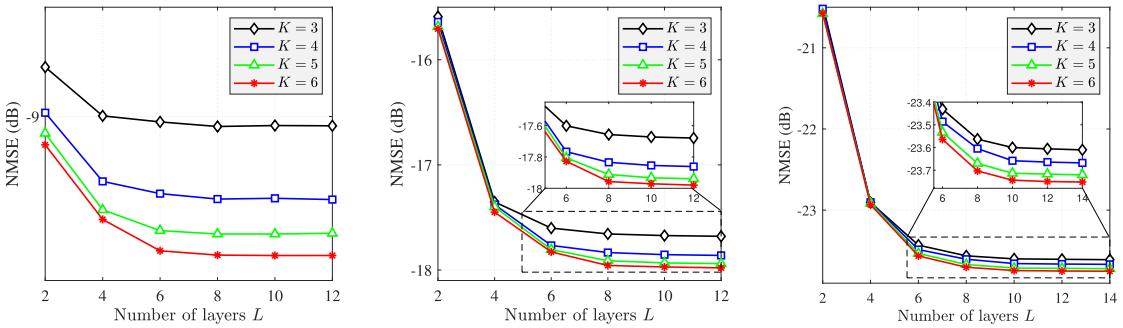
The Rician factor is given as $\kappa_k = 13 - 0.03 d_k$ (dB) [53]. We assume perfect transmit power control at the users so that the received signal powers of different users are the same and equal to one. The SNR is defined as $\rho = 1/N_0$.

We compare the channel estimation performance of different methods in terms of normalized mean squared error (NMSE), defined here as $\text{NMSE} = \mathbb{E}[\|\hat{\mathbf{H}} - \bar{\mathbf{H}}\|_{\text{F}}^2]/\mathbb{E}[\|\bar{\mathbf{H}}\|_{\text{F}}^2]$, where $\hat{\mathbf{H}}$ is an estimate of the channel $\bar{\mathbf{H}}$. When the pilot matrix is prespecified, it is assumed to contain K columns of a $T_{\text{t}} \times T_{\text{t}}$ discrete Fourier transform (DFT) matrix. In particular, the k^{th} row of the pilot matrix $\bar{\mathbf{X}}_{\text{t}}$ is column $(k+1)$ of the DFT matrix.

Fig. 7 presents a performance comparison of different channel estimation methods for the given DFT-based pilot matrix and considering both uncorrelated NLoS and spatially correlated mixed LoS-NLoS channels. Numerical results for the uncorrelated NLoS scenario are given in Fig. 7a where we set $\kappa_k = 0$ for all k, and $\bar{\mathbf{h}}^{\text{NLoS}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$. It can be seen from Fig. 7a that the proposed FBM-CENet significantly outperforms existing methods. For the case of one-bit ADCs, it is observed that the proposed FBM-CENet slightly outperforms the SVM-based method in [19] at medium-to-high SNRs. However, at low SNRs, the performance gap between FBM-CENet and the SVM method is larger. For few-bit ADCs, it is clear that FBM-CENet significantly outperforms other existing channel estimation methods. Note that the SVM-based method in [19] was specifically designed for one-bit ADCs, and thus SVM results for other ADC resolutions are not available. The BWZF method does not perform well for one-bit ADCs since it gives a higher weight to signals with lower variance. However, for one-bit ADCs, there is only one bin on each side of the quantization threshold, and thus the weighting has no impact in this case. On the other hand, higher resolution ADCs result in more quantization bins and thus different weights come into play, and thus we see that BWZF performs better with few-bit quantization.

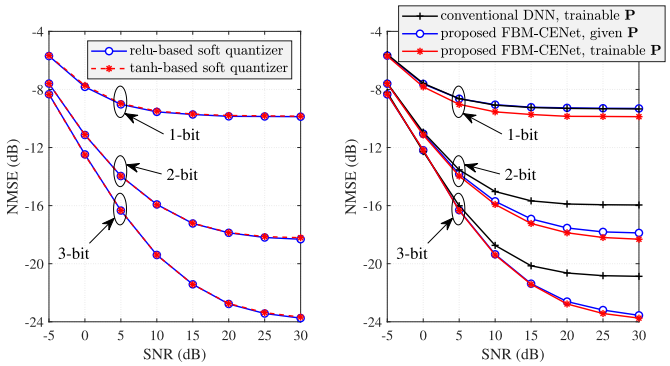
Numerical results for spatially correlated mixed LoS-NLoS channels are provided in Fig. 7b. For this scenario, we use the typical urban correlation model as in [3] and set $10 \leq d_k \leq 1000$. For the case of one-bit ADCs, the SVM-based method gives the best performance, while the proposed FBM-CENet performs worse than the SVM-based and BMMSE methods, but better than BWZF. The reason for this is because both BMMSE and SVM exploit knowledge of the channel correlation matrix, which is not used by FBM-CENet. Only received signals and pilot matrix are used by FBM-CENet. However, for few-bit ADCs, the proposed FBM-CENet approach outperforms both BMMSE and BWZF at higher SNRs even without using information about the channel statistics (recall that the SVM method only applies in the one-bit case). BMMSE gives the best performance at low SNRs for an additional reason beyond its use of the channel correlation information, namely that its use of the Bussgang decomposition is more accurate when the received signal is Gaussian, which becomes a better approximation as the power of the Gaussian noise increases. However, BMMSE uses an approximation for the Bussgang decomposition with few-bit ADCs that limits its performance at higher SNRs where FBM-CENet is superior. Note that BMMSE and SVM were implemented assuming perfect knowledge of the channel correlation, which may not be available in practice. FBM-CENet still provides very good performance without any relying on any correlation information, but extending the network to be able to exploit this information is an interesting area for future work.

In the following comparisons and evaluations, from Fig. 8 to Fig. 11, we present results for uncorrelated NLoS channels since we found that the results were similar for spatially correlated mixed LoS-NLoS channels. Fig. 8 compares the proposed FBM-CENet with PG-FBM-CENet as well as the conventional gradient descent method (GDM) in (20) using a constant step size α_{t} for all iterations. The step size α_{t} used



(a) 1-bit ADCs. (b) 2-bit ADCs. (c) 3-bit ADCs.

Fig. 10. Channel estimation performance of the proposed FBM-CENet with various values of K and L, $N = 32$, and SNR $= 30$ dB.



(a) Proposed FBM-CENet with relu-based and tanh-based soft quantizers. (b) Proposed FBM-CENet versus conventional DNN.

Fig. 11. Channel estimation performance comparison with trainable pilot matrix, $K = 4$, $L = 8$, and $N = 32$.

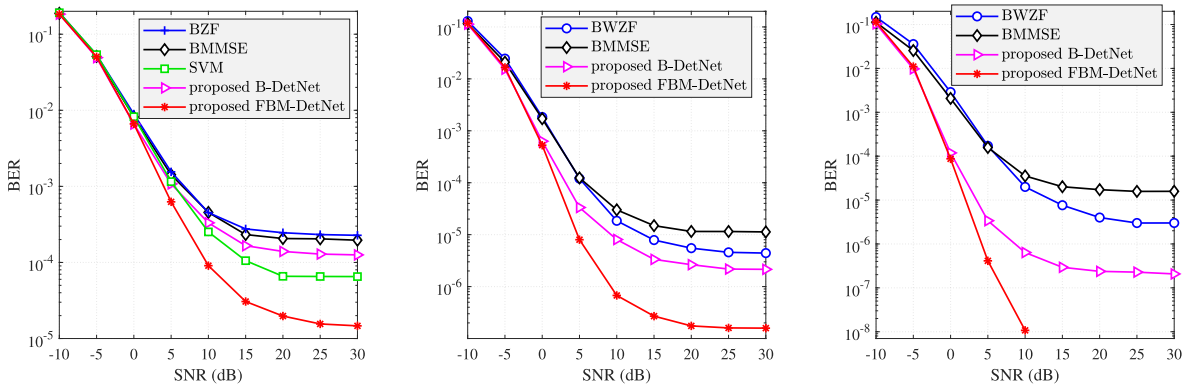
in GDM is normalized by the SNR as we found this gives stable performance for different SNR regimes. Note that FBM-CENet, PG-FBM-CENet, and GDM use the same number of layers (iterations) so that they have the same complexity. Simulation results show that the proposed FBM-CENet significantly outperforms PG-FBM-CENet and the conventional GDM. This results because, while GDM uses a common step size in all iterations and PG-FBM-CENet only optimizes the step sizes, the proposed FBM-CENet learns an optimal path by jointly optimizing the optimal step sizes $\{\alpha_{\mathrm{t}}^{(\ell)}\}$ as well as the optimal scaling parameter β_{t}. The value of these optimal parameters will depend on the specific channel and system model. If an analytical connection between the system model and the optimal parameters $\{\alpha_{\mathrm{t}}^{(\ell)}\}$ and β_{t} could be found, then such a connection could be used to infer $\{\alpha_{\mathrm{t}}^{(\ell)}\}$ and β_{t} for different channel and system models without the need for a training process. However, our simulations show that the trained values of $\{\alpha_{\mathrm{t}}^{(\ell)}\}$ and β_{t} do not appear to have a clearly interpretable connection to the model parameters. We leave study of this interesting problem for future work.

In practice, the step size α_{t} can also be tuned by, for example, the backtracking line search method. However, this method requires an inner search loop in each iteration and therefore significantly increases the computational complexity compared to using fixed step sizes. Note that GDM, PG-FBM-CENet, and FBM-CENet presented above use fixed step sizes. For PG-FBM-CENet and FBM-CENet, the step sizes are obtained by the training process. Thus, GDM, PG-FBM-CENet, and FBM-CENet have significantly lower complexity. In addition, the inner search loop in each iteration requires the calculation of the objective function (17), which is undefined when the argument of the logarithm approaches zero. In our investigation, this issue occurs frequently. Note however that although the value of the objective function (17) can become undefined, its gradient (19) is robust against this issue.

In Fig. 9, we evaluate GDM, PG-FBM-CENet, and FBM-CENet for different numbers of layers L. It is observed that the proposed FBM-CENet performs better than both PG-FBM-CENet and GDM for different values of L and also requires fewer layers for convergence.

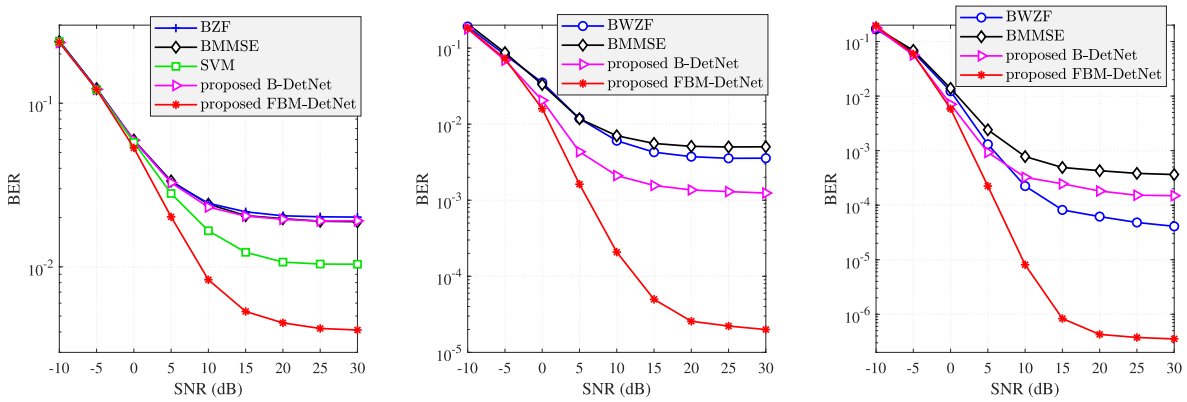
We investigate the performance of FBM-CENet as K, L, and b vary in Fig. 10. We see that for a given bit resolution b, the number of layers L need not be increased as the number of users K increases. However, as the bit resolution increases, improved performance can be achieved with an increased number of layers. Specifically, with one-bit ADCs, we can fix the number of layers to 6 as K increases from 3 to 6. However, as the bit resolution increases to 2 and 3, it is best to increase the number of layers to 8 and 10, respectively.

In Fig. 11, we consider the case where the pilot matrix is trained concurrently with the channel estimator. As mentioned earlier, when the pilot matrix is not given, we need to use a soft quantizer, based on either the ReLU or tanh function. In Fig. 11a, it is seen that the ReLU- and tanh-based soft quantizers give essentially identical performance. This is due to the fact that the parameters of the soft quantizers should be chosen so that they act similar to a hard quantizer. In Fig. 11b, the proposed FBM-CENet is compared with the existing conventional DNN-based method in [23] which also jointly optimizes the pilot matrix and the channel estimator. Note that we use the network structure and training method proposed in [23] to obtain the performance of the conventional DNN-based method. FBM-CENet significantly outperforms the channel estimator in [23] since the method of [23] uses the conventional data-driven DNN structure in Fig. 2a. On the other hand, the structure of FBM-CENet takes advantage of domain knowledge in the ML estimation framework. In Fig. 11b, we also include the channel estimation performance of FBM-CENet for a given orthogonal DFT-based pilot matrix in order to show that jointly optimizing the pilot matrix and the estimator can improve the estimation



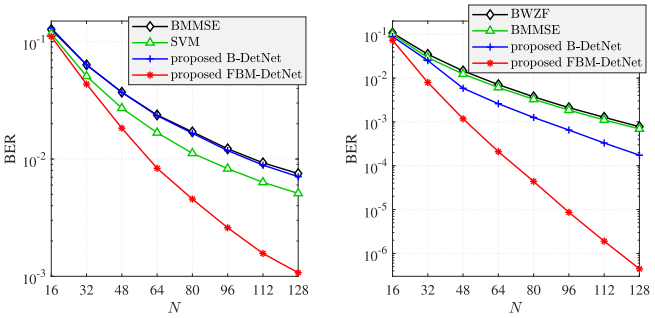
(a) $b = 1$ bit, $K = 4$, and $L = 8$. (b) $b = 2$ bit, $K = 8$, and $L = 16$. (c) $b = 3$ bit, $K = 16$, and $L = 24$.

Fig. 12. Performance comparison for data detection methods with QPSK signalling and $N = 32$.



(a) $b = 1$ bit, $K = 4$, and $L = 8$. (b) $b = 2$ bit, $K = 8$, and $L = 16$. (c) $b = 3$ bit, $K = 16$, and $L = 24$.

Fig. 13. Performance comparison for data detection methods with 16-QAM signalling and $N = 64$.



(a) 1-bit ADCs, $K = 4$, $L = 8$ (b) 2-bit ADCs, $K = 8$, $L = 16$

Fig. 14. Data detection performance comparison for various values of N at 10-dB SNR and 16-QAM signalling.

accuracy. This improvement is obtained since orthogonal pilot data is known to be sub-optimal in low-resolution quantized systems [23]. When the pilot matrix is not given and treated as trainable, the training process of FBM-CENet produces a non-orthogonal pilot matrix that yields better performance than orthogonal pilots.

In the following, we present performance comparisons for data detection. Unless otherwise stated, uncorrelated NLoS channels are considered and the estimated CSI is obtained by FBM-CENet with a trainable pilot matrix. Comparisons given in Fig. 12 and Fig. 13 are for QPSK and 16-QAM signaling, respectively. The results show that FBM-DetNet significantly outperforms other data detection methods. FBM-DetNet outperforms B-DetNet because FBM-DetNet is developed based on the original quantized system model whereas B-DetNet relies on the linearized system model in (28) whose effective noise $\mathbf{n}$ is approximated as Gaussian. Furthermore, the distortion covariance matrix $\mathbf{\Sigma_n}$ assumed by B-DetNet for the case of few-bit ADCs is approximate since a closed-form expression for $\mathbf{\Sigma_n}$ is intractable. For the case of 3-bit ADCs and 16-QAM signaling, B-DetNet performs worse than the BWZF method. As mentioned earlier, this is because BWZF performs better when there are more quantization bins (i.e., few-bit quantization), and also because B-DetNet is developed by unfolding the gradient descent of a linearized system, similar to the methodology applied in FS-Net [47] and DetNet [54], whose performance tends to degrade with higher dimensional constellations [55]. Note that FS-Net was developed for unquantized systems while B-DetNet is for the low-resolution quantized case.

In Fig. 14, we present a detection performance comparison for various values of N at 10-dB SNR and with 16-QAM signalling. It can be seen that the performance improvement of the proposed detection networks is maintained as the number of receive antennas increases. Since our derivations and methods assume no constraint on N, the proposed networks can work with an arbitrary number of receive antennas.

We provide a data detection performance comparison for spatially correlated mixed LoS-NLoS channels in Fig. 15 where the estimated CSI is obtained by the BMMSE method.

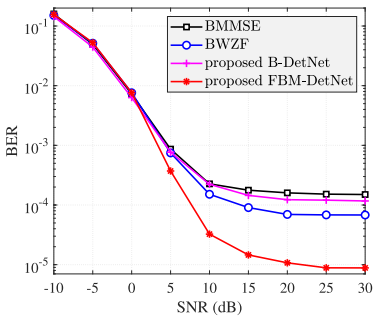


Fig. 15. Data detection performance comparison with spatially correlated mixed LoS-NLoS channels, $b = 2$, $K = 4$, $N = 64$, $L = 8$, 16-QAM signalling, and BMMSE-based estimated CSI.

It is still observed that the proposed FBM-DetNet gives the best performance. This shows that the proposed detection networks can work well with the estimated CSI given by not only FBM-CENet but also other channel estimation methods.

VI. Conclusion

In this paper, we have developed a channel estimation network (FBM-CENet) and two data detection networks (B-DetNet and FBM-DetNet) for massive MIMO systems with low-resolution ADCs. The proposed networks are model-driven and have special structures that can take advantage of domain-knowledge to efficiently address the severe non-linearity caused by the low-resolution ADCs. An interesting feature of the proposed FBM-CENet is that the pilot matrix directly plays the role of the weight matrices in the network structure, which makes it possible to jointly optimize the estimation network and the pilot signal by simply treating the pilot matrix as trainable parameters. The proposed detection networks are highly adaptive to the channel and easy to train since they have a small number of trainable parameters. Simulation results show that the proposed networks significantly outperform existing methods.

References

[1] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 4, pp. 539–550, Apr. 1999.

[2] 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, Mar. 2016.

[3] Y. Li, C. Tao, G. Seco-Granados, A. Mezghani, A. L. Swindlehurst, and L. Liu, "Channel estimation and performance analysis of one-bit massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4075–4089, Apr. 2017.

[4] S. Rao, A. L. Swindlehurst, and H. Pirzadeh, "Massive MIMO channel estimation with 1-bit spatial sigma-delta ADCS," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Brighton, U.K., May 2019, pp. 4484–4488.

[5] Z. Shao, L. T. N. Landau, and R. C. D. Lamare, "Oversampling based channel estimation for 1-bit large-scale multiple-antenna systems," in *Proc. 23rd Int. ITG Workshop Smart Antennas*, Vienna, Austria, Apr. 2019, pp. 1–5.

[6] Z. Shao, L. T. N. Landau, and R. C. de Lamare, "Channel estimation using 1-bit quantization and oversampling for large-scale multiple-antenna systems," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Brighton, U.K., May 2019, pp. 4669–4673.

[7] F. Liu *et al.*, "Angular-domain channel estimation for one-bit massive MIMO systems: Performance bounds and algorithms," *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 2928–2942, Mar. 2020.

[8] I.-S. Kim, N. Lee, and J. Choi, "Dominant channel estimation via MIPS for large-scale antenna systems with one-bit ADCs," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Abu Dhabi, United Arab Emirates, Dec. 2018, pp. 1–6.

[9] H. Kim and J. Choi, "Channel AoA estimation for massive MIMO systems using one-bit ADCs," *J. Commun. Netw.*, vol. 20, no. 4, pp. 374–382, Aug. 2018.

[10] H. Kim and J. Choi, "Channel estimation for spatially/temporally correlated massive MIMO systems with one-bit ADCs," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, no. 1, p. 267, Dec. 2019.

[11] B. Srinivas, K. Mawatwal, D. Sen, and S. Chakrabarti, "An iterative semi-blind channel estimation scheme and uplink spectral efficiency of pilot contaminated one-bit massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 7854–7868, Aug. 2019.

[12] A. Mezghani and A. L. Swindlehurst, "Blind estimation of sparse broadband massive MIMO channels with ideal and one-bit ADCs," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2972–2983, Jun. 2018.

[13] I.-S. Kim and J. Choi, "Channel estimation via gradient pursuit for mmWave massive MIMO systems with one-bit ADCs," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, no. 1, p. 289, Dec. 2019.

[14] J. Mo, P. Schniter, and R. W. Heath, Jr., "Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1141–1154, Mar. 2018.

[15] J. Rodríguez-Fernández, N. González-Prelcic, and R. W. Heath, "Channel estimation in mixed hybrid-low resolution MIMO architectures for mmWave communication," in *Proc. 50th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2016, pp. 768–773.

[16] C. Rusu, R. Mendez-Rial, N. Gonzalez-Prelcic, and R. W. Heath, "Adaptive one-bit compressive sensing with application to low-precision receivers at mmWave," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, San Diego, CA, USA, Dec. 2015, pp. 1–6.

[17] S. Rao, A. Mezghani, and A. L. Swindlehurst, "Channel estimation in one-bit massive MIMO systems: Angular versus unstructured models," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 5, pp. 1017–1031, Sep. 2019.

[18] N. Kolomvakis, T. Eriksson, M. Coldrey, and M. Viberg, "Quantized uplink massive MIMO systems with linear receivers," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Dublin, Ireland, Jun. 2020, pp. 1–6.

[19] L. V. Nguyen, A. L. Swindlehurst, and D. H. N. Nguyen, "SVM-based channel estimation and data detection for one-bit massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 69, pp. 2086–2099, 2021.

[20] E. Balevi and J. G. Andrews, "Two-stage learning for uplink channel estimation in one-bit massive MIMO," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Nov. 2019, pp. 1764–1768.

[21] Y. Dong, H. Wang, and Y.-D. Yao, "Channel estimation for one-bit multiuser massive MIMO using conditional GAN," *IEEE Commun. Lett.*, vol. 25, no. 3, pp. 854–858, Mar. 2021.

[22] Y. Zhang, M. Alrabeiah, and A. Alkhateeb, "Deep learning for massive MIMO with 1-bit ADCs: When more antennas need fewer pilots," *IEEE Wireless Commun. Lett.*, vol. 9, no. 8, pp. 1273–1277, Aug. 2020.

[23] D. H. N. Nguyen, "Neural network-optimized channel estimator and training signal design for MIMO systems with few-bit ADCs," *IEEE Signal Process. Lett.*, vol. 27, pp. 1370–1374, 2020.

[24] S. Gao, P. Dong, Z. Pan, and G. Y. Li, "Deep learning based channel estimation for massive MIMO with mixed-resolution ADCs," *IEEE Commun. Lett.*, vol. 23, no. 11, pp. 1989–1993, Nov. 2019.

[25] J. Zicheng, G. Shen, L. Nan, P. Zhiwen, and Y. Xiaohu, "Deep learning-based channel estimation for massive-MIMO with mixed-resolution ADCs and low-resolution information utilization," *IEEE Access*, vol. 9, pp. 54938–54950, 2021.

[26] Y.-S. Jeon, N. Lee, S.-N. Hong, and R. W. Heath, "One-bit sphere decoding for uplink massive MIMO systems with one-bit ADCs," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4509–4521, Jul. 2018.

[27] S. Kim, J. Chae, and S.-N. Hong, "Machine learning detectors for MU-MIMO systems with one-bit ADCs," *IEEE Access*, vol. 8, pp. 86608–86616, 2020.

[28] A. S. Lan, M. Chiang, and C. Studer, "Linearized binary regression," in *Proc. 52nd Annu. Conf. Inf. Sci. Syst. (CISS)*, Princeton, NJ, USA, Mar. 2018, pp. 1–6.

[29] L. V. Nguyen, A. L. Swindlehurst, and D. H. N. Nguyen, "Linear and deep neural network-based receivers for massive MIMO systems with one-bit ADCs," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7333–7345, Nov. 2021.

[30] 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.
[31] O. T. Demir and E. Bjornson, "ADMM-based one-bit quantized signal detection for massive MIMO systems with hardware impairments," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Barcelona, Spain, May 2020, pp. 9120–9124.
[32] S. H. Mirfarshbafan, M. Shabany, S. A. Nezamalhosseini, and C. Studer, "Algorithm and VLSI design for 1-bit data detection in massive MIMO-OFDM," *IEEE Open J. Circuits Syst.*, vol. 1, pp. 170–184, 2020.
[33] Y.-S. Jeon, N. Lee, and H. V. Poor, "Robust data detection for MIMO systems with one-bit ADCs: A reinforcement learning approach," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1663–1676, Mar. 2020.
[34] S. H. Song, S.-C. Lim, G. Kwon, and H. Park, "CRC-aided soft-output detection for uplink multi-user MIMO systems with one-bit ADCs," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Marrakesh, Morocco, Apr. 2019, pp. 1–5.
[35] Y. Cho and S.-N. Hong, "One-bit successive-cancellation soft-output (OSS) detector for uplink MU-MIMO systems with one-bit ADCS," *IEEE Access*, vol. 7, pp. 27172–27182, 2019.
[36] Z. Shao, R. C. de Lamare, and L. T. N. Landau, "Iterative detection and decoding for large-scale multiple-antenna systems with 1-bit ADCs," *IEEE Wireless Commun. Lett.*, vol. 7, no. 3, pp. 476–479, Jun. 2018.
[37] C. K. Wen, C. J. Wang, S. Jin, K. K. Wong, and P. Ting, "Bayes-optimal joint channel-and-data estimation for massive MIMO with low-precision ADCs," *IEEE Trans. Signal Process.*, vol. 64, no. 10, pp. 2541–2556, May 2016.
[38] S. S. Thoota and C. R. Murthy, "Variational Bayes' joint channel estimation and soft symbol decoding for uplink massive MIMO systems with low resolution ADCs," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3467–3481, May 2021.
[39] Y.-S. Jeon, S.-N. Hong, and N. Lee, "Supervised-learning-aided communication framework for MIMO systems with low-resolution ADCs," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7299–7313, Aug. 2018.
[40] L. V. Nguyen, D. T. Ngo, N. H. Tran, A. L. Swindlehurst, and D. H. N. Nguyen, "Supervised and semi-supervised learning for MIMO blind detection with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2427–2442, Apr. 2020.
[41] L. V. Nguyen, D. H. N. Nguyen, and A. L. Swindlehurst, "DNN-based detectors for massive MIMO systems with low-resolution ADCs," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Montreal, QC, Canada, Jun. 2021, pp. 1–6.
[42] J. Max, "Quantizing for minimum distortion," *IRE Trans. Inf. Theory*, vol. 6, no. 1, pp. 7–12, Mar. 1960.
[43] J. W. Pratt, "Concavity of the log likelihood," *J. Amer. Stat. Assoc.*, vol. 76, no. 373, pp. 103–106, Mar. 1981.
[44] S. R. Bowling, M. T. Khasawneh, S. Kaewkuekool, and B. R. Cho, "A logistic approximation to the cumulative normal distribution," *J. Ind. Eng. Manage.*, vol. 2, no. 1, pp. 114–127, Jul. 2009.
[45] J. R. Hershey, J. Le Roux, and F. Weninger, "Deep unfolding: Model-based inspiration of novel deep architectures," 2014, *arXiv:1409.2574*.
[46] A. Mezghani and J. A. Nossek, "Capacity lower bound of MIMO channels with output quantization and correlated noise," in *Proc. IEEE Int. Symp. Inf. Theory*, Cambridge, MA, USA, Jul. 2012, pp. 1–5.
[47] N. T. Nguyen and K. Lee, "Deep learning-aided Tabu search detection for large MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 6, pp. 4262–4275, Jun. 2020.
[48] 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, Aug. 2020.
[49] A. Mezghani, M.-S. Khoufi, and J. A. Nossek, "Maximum likelihood detection for quantized MIMO systems," in *Proc. Int. ITG Workshop Smart Antennas*, Vienna, Austria, Feb. 2008, pp. 278–284.
[50] M. Abadi *et al.* (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. [Online]. Available: https://www.tensorflow.org/
[51] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[52] M. Abdelghany, A. A. Farid, M. E. Rasekh, U. Madhow, and M. J. W. Rodwell, "A design framework for all-digital mmwave massive MIMO with per-antenna nonlinearities," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 5689–5701, Sep. 2021.
[53] O. Ozdogan, E. Bjornson, and E. G. Larsson, "Massive MIMO with spatially correlated rician fading channels," *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3234–3250, May 2019.
[54] N. Samuel, T. Diskin, and A. Wiesel, "Learning to detect," *IEEE Trans. Signal Process.*, vol. 67, no. 10, pp. 2554–2564, May 2019.
[55] 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, Oct. 2021.

Ly V. Nguyen (Member, IEEE) received the B.Eng. degree in electronics and telecommunications from the Vietnam National University (VNU) University of Engineering and Technology, Hanoi, Vietnam, in 2014, the M.Sc. degree in advanced wireless communications systems from CentraleSupélec, Paris-Saclay University, France, in 2016, and the joint Ph.D. degree in computational science from San Diego State University and the University of California at Irvine, Irvine, CA, USA, in 2022. His research interests include wireless communications, signal processing, and machine learning. He received the Best Paper Award at the 2020 IEEE International Conference on Communications (ICC).

Duy H. N. Nguyen (Senior Member, IEEE) received the B.Eng. degree (Hons.) from the Swinburne University of Technology, Hawthorn, VIC, Australia, in 2005, the M.Sc. degree from the University of Saskatchewan, Saskatoon, SK, Canada, in 2009, and the Ph.D. degree from McGill University, Montreal, QC, Canada, in 2013, all in electrical engineering. From 2013 to 2015, he held a joint appointment as a Research Associate with McGill University and a Post-Doctoral Research Fellow with the Institut National de la Recherche Scientifique, Université du Québec, Montreal. He was a Research Assistant with the University of Houston, Houston, TX, USA, in 2015; and a Post-Doctoral Research Fellow with The University of Texas at Austin, Austin, TX, USA, in 2016. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, San Diego State University, San Diego, CA, USA. His current research interests include resource allocation in wireless networks, signal processing for communications, convex optimization, game theory, and machine learning. He has been serving as a TPC Member for a number of flagship IEEE conferences, including ICC, GLOBECOM, and INFOCOM. He was a recipient of the Australian Development Scholarship, the FRQNT Doctoral Fellowship and Post-Doctoral Fellowship, and the NSERC Post-Doctoral Fellowship. He received the Best Paper Award at the 2020 IEEE International Conference on Communications (ICC). He is currently an Associate Editor of the *EURASIP Journal on Wireless Communications and Networking*.

A. Lee Swindlehurst (Fellow, IEEE) received the B.S. and M.S. degrees in electrical engineering from Brigham Young University (BYU), Provo, UT, USA, in 1985 and 1986, respectively, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 1991. He was with the Department of Electrical and Computer Engineering, BYU, from 1990 to 2007. During 1996 to 1997, he held a joint appointment as a Visiting Scholar with Uppsala University, Uppsala, Sweden; and the Royal Institute of Technology, Stockholm, Sweden. From 2006 to 2007, he was on leave working as the Vice President of Research for ArrayComm LLC, San Jose, CA, USA. Since 2007, he has been a Professor with the Electrical Engineering and Computer Science Department, University of California at Irvine, Irvine, CA, USA. During 2014 to 2017, he was also a Hans Fischer Senior Fellow with the Institute for Advanced Studies, Technical University of Munich, Munich, Germany. His research interests include array signal processing for radar, wireless communications, and biomedical applications, and has more than 350 publications in these areas. In 2016, he was elected as a Foreign Member of the Royal Swedish Academy of Engineering Sciences (IVA). He was the recipient of the 2000 IEEE W. R. G. Baker Prize Paper Award; the 2006 IEEE Communications Society Stephen O. Rice Prize in the Field of Communication Theory; the 2006, 2010, and 2021 IEEE Signal Processing Society Best Paper Awards; the 2017 IEEE Signal Processing Society Donald G. Fink Overview Paper Award; and the Best Paper award at the 2020 IEEE International Conference on Communications. He was the Inaugural Editor-in-Chief of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING.