SVM-Based Channel Estimation and Data Detection for One-Bit Massive MIMO Systems

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

Abstract—The use of low-resolution Analog-to-Digital Converters (ADCs) is a practical solution for reducing cost and power consumption for massive Multiple-Input-Multiple-Output (MIMO) systems. However, the severe nonlinearity of low-resolution ADCs causes significant distortions in the received signals and makes the channel estimation and data detection tasks much more challenging. In this paper, we show how *Support Vector Machine* (*SVM*), a well-known supervised-learning technique in machine learning, can be exploited to provide efficient and robust channel estimation and data detection in massive MIMO systems with one-bit ADCs. First, the problem of channel estimation for uncorrelated channels is formulated as a conventional SVM problem. The objective function of this SVM problem is then modified for estimating spatially correlated channels. Next, a two-stage detection algorithm is proposed where SVM is further exploited in the first stage. The performance of the proposed data detection method is very close to that of Maximum-Likelihood (ML) data detection when the channel is perfectly known. We also propose an SVM-based joint Channel Estimation and Data Detection (CE-DD) method, which makes use of both the to-be-decoded data vectors and the pilot data vectors to improve the estimation and detection performance. Finally, an extension of the proposed methods to OFDM systems with frequency-selective fading channels is presented. Simulation results show that the proposed methods are efficient and robust, and also outperform existing ones.

Index Terms—Channel estimation, data detection, machine learning, massive MIMO, one -bit ADCs, support vector machine.

I. INTRODUCTION

THE development of wireless communications systems has been moving toward the use of more and more antennas at the transceivers. Massive Multiple-Input-Multiple-Output (MIMO) technology is a result of this development and is now considered to be one of the disruptive technologies of 5G networks [1], [2]. The first and foremost benefit of massive MIMO is the significant increase in the spatial degrees of freedom obtained by combining tens to hundreds of antennas at the base station. This benefit of spatial degrees of freedom helps improve the throughput and energy efficiency by several orders of magnitude over conventional MIMO systems [3], [4]. However, the use of many antennas at the base station also poses a number of problems. More specifically, a massive MIMO system requires many Radio-Frequency (RF) chains and Analog-to-Digital Converters (ADCs) to support a massive number of antennas. This causes significant increases in hardware complexity, system cost, and power consumption.

Recently, low-resolution ADCs have attracted significant research interest and are considered to be a promising solution for the aforementioned problems. This is due to the simple structure and low power consumption of low-resolution ADCs. As reported in [5], the power consumption of an ADC is exponentially proportional to its resolution. Hence, using low-resolution ADCs can significantly reduce the power consumption of the system. The simplest architecture involving 1-bit ADCs requires only one comparator and does not require an Automatic Gain Control (AGC). Thus, 1-bit ADCs are an attractive potential solution for the problems of hardware complexity, system cost, and power consumption.

One major drawback of 1-bit ADCs is the induced distortion, since only the *sign* of the real and imaginary parts of the received signals is retained. This severe nonlinearity makes the channel estimation and data detection tasks much more challenging. MIMO channel estimation with 1-bit ADCs has been studied intensively in a number of papers with different scenarios, e.g., [6]–[23]. Maximum-Likelihood (ML) and Least-Squares (LS) channel estimators were proposed in [6] and [7], respectively. The Bussgang decomposition is exploited in [8] to form a Bussgang-based Minimum Mean-Squared Error (BMMSE) channel estimator. The work in [9] proposes a BMMSE channel estimator for massive MIMO systems with 1-bit spatial sigma-delta ADCs in a spatially oversampled array or for sectorized users. Channel estimation with temporally oversampled 1-bit ADCs is studied in [10] and [11]. The use of spatial and temporal oversampling 1-bit ADCs was shown to help improve the channel estimation accuracy but requires more resources and computations due to the oversampling process. A channel estimation method based on Support Vector Machine (SVM) with 1-bit ADCs, referred to as soft-SVM, was presented

in [12]. Angular-domain estimation for MIMO channels with 1-bit ADCs was studied in [13]–[15]. Other scenarios involving spatially/temporally correlated channels or multi-cell processing with pilot contamination were investigated in [16] and [17], respectively. For sparse millimeter-wave MIMO channels, the ML and maximum a posteriori (MAP) channel estimation problems were studied in [18] and [19], respectively. Taking into account the sparsity of such channels, the 1-bit ADC channel estimation problem has been formulated as a compressed sensing problem in [20]–[22]. Several performance bounds on the channel estimation of mmWave 1-bit massive MIMO channels were reported in [23]. In this paper, we focus on a more general channel model which is not assumed to be sparse without any oversampling.

There have also been several channel estimators proposed for 1-bit massive MIMO systems based on deep neural networks (DNN), e.g., [24]–[27]. However, these DNN-based estimators require a highly complicated *offline* training process using a large data set. In addition, the DNNs do not provide physical meaning or insights into the structure of the estimators. In this paper, our approach is based on the use of an *SVM*, which not only has low computational complexity but also provides an insightful connection with likelihood-based channel estimators. Moreover, the proposed SVM-based approach does not require training data beyond what would typically be used in a standard MIMO channel estimation method. It is also worth noting that the work in [26] is restricted to systems with only one single-antenna user and the work in [24] requires multiple OFDM symbols in the training sequence. Our proposed methods are applicable for multi-user systems and only one OFDM symbol is required in the training process.

Data detection in massive MIMO systems with 1-bit ADCs has also been studied intensively in the literature, e.g., [6], [28]–[37]. The one-bit ML detection problem is formulated in [6]. For large-scale systems where ML detection is impractical, the authors in [6] proposed a so-called near-ML (nML) data detection method. The ML and nML methods are however non-robust at high Signal-to-Noise Ratios (SNRs) when Channel State Information (CSI) is imperfectly known. A One-bit Sphere Decoding (OSD) technique was proposed in [28]. However, the OSD technique requires a preprocessing stage whose computational complexity for each channel realization is exponentially proportional to both the number of receive and transmit antennas. The exponential computational complexity of OSD makes it difficult to implement in large scale MIMO systems. Generalized Approximate Message Passing (GAMP) and Bayes inference are exploited in [29] but the proposed method is sophisticated and expensive to implement. A number of linear receivers for massive MIMO systems with 1-bit ADCs are presented in [30] and several learning-based methods are also proposed in [31]–[34]. The linear receivers in [30] are easy to implement but their performance is often limited by an error floor. The learning-based methods in [31]–[33] are blind detection methods for which CSI is not required, but they are restricted to MIMO systems with a small number of transmit antennas and only low-dimensional constellations. Several other data detection approaches were proposed in [34]–[37], but they are only applicable in systems where either a Cyclic Redundancy Check (CRC) [34]–[36] or

an error correcting code such as Low-Density Parity-Check (LDPC) code [37] is available.

In this paper, we propose channel estimation and data detection methods which are efficient, robust, and applicable to large-scale systems without the need for CRC or error correcting codes. Our work is based on SVM, a well-known supervised-learning technique in machine learning [38]. Since SVM problems can be solved by very efficient algorithms [39]–[43], the proposed methods can be implemented in an efficient manner. There are several prior works on the application of SVM to channel estimation and data detection problems, e.g., [44], [45]. However, these works consider either SISO or SIMO channels with full-resolution ADCs. In this paper, we focus on massive MIMO with one-bit ADCs. Our earlier work reported in [46] examined SVM-based channel estimation and data detection methods for 1-bit MIMO systems with independent and identically distributed (i.i.d.) and flat-fading channels. This paper extends the study in [46] with further considerations of spatially correlated and frequency-selective fading channels and presents the following contributions:

- An SVM-based channel estimation method for uncorrelated channels is first proposed by formulating the 1-bit ADC channel estimation problem as an SVM problem. Unlike the soft-SVM method in [12], the proposed method exploits the original idea of SVM by maximizing the margin achieved by the linear discriminator. For spatially correlated channels, we develop a new channel estimation problem by revising the conventional SVM objective function. Numerical results show that the high-SNR Normalized Mean-Squared Error (NMSE) floor of the proposed channel estimation methods is lower than that of the BMMSE method proposed in [8], which outperforms other existing methods.

- We then propose a two-stage SVM-based data detection method, where the first stage is also formulated as an SVM problem. A second stage is then employed to refine the solution from the first stage. Simulation results show that the performance of the proposed method is very close to that of the ML detection method if perfect CSI is available. With imperfect CSI, the proposed data detection method is shown to be robust and to also outperform existing methods. We then consider an SVM-based joint Channel Estimation and Data Detection (CE-DD) method where the to-be-decoded data vectors and pilot data vectors are both exploited to refine the estimated channel and thus improve the data detection performance.

- Finally, an extension of the proposed methods to OFDM systems with frequency-selective fading channels is derived. Numerical results show that the proposed SVM-based methods significantly outperform existing ones. For example, the high-SNR NMSE floor of the proposed SVM-based channel estimation method is about 3-dB lower that of the BMMSE method.

The rest of this paper is organized as follows: Section II introduces the assumed system model. In Section III, we present linear SVM for binary classification and the proposed methods for flat-fading channels. Extension of the proposed methods to
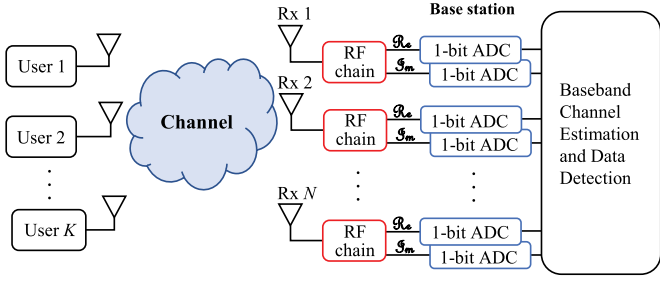
Fig. 1. Block diagram of a massive MIMO system with K single-antenna users and an N-antenna base station equipped with $2N$ 1-bit ADCs.

OFDM systems with frequency-selective fading is presented in Section IV. Numerical results are given in Section V, and Section VI concludes the paper.

Notation: Upper-case and lower-case boldface letters denote matrices and column vectors, respectively. $\mathbb{E}[\cdot]$ represents expectation. Depending on the context, the operator $|\cdot|$ is used to denote the absolute value of a number, or the cardinality of a set. $\|\cdot\|$ denotes the ℓ_2-norm of a vector. The transpose and conjugate transpose are denoted by $[\cdot]^T$ and $[\cdot]^H$, respectively. The notation $\Re\{\cdot\}$ and $\Im\{\cdot\}$ respectively denotes 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. $\mathbb{R}$ and $\mathbb{C}$ denote the set of real and complex numbers, respectively, and j is the unit imaginary number satisfying $j^2 = -1$. $\mathcal{N}(\cdot, \cdot)$ and $\mathcal{CN}(\cdot, \cdot)$ represent the real and the complex normal distributions respectively, where the first argument is the mean and the second argument is the variance or the covariance matrix. The operator $\mathrm{blockdiag}(\mathbf{A}_1, \ldots, \mathbf{A}_n)$ represents a block diagonal matrix, whose main-diagonal blocks are $\mathbf{A}_1, \ldots, \mathbf{A}_n$.

II. SYSTEM MODEL

We consider a massive MIMO system as illustrated in Fig. 1 with K single-antenna users and an N-antenna base station, 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 kth user under the power constraint $\mathbb{E}[|\bar{x}_k|^2] = 1$, $k \in \mathcal{K} = \{1, 2, \ldots, K\}$. Let $\bar{\mathbf{H}} \in \mathbb{C}^{N \times K}$ denote the channel, which for the moment 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 i.i.d. as $\bar{z}_i \sim \mathcal{CN}(0, N_0)$, and N_0 is the noise power. Each analog received signal $\bar{r}_i$ is then quantized by a pair of 1-bit ADCs. Hence, we have the received signal

$$\bar{\mathbf{y}} = \mathrm{sign}(\bar{\mathbf{r}}) = \mathrm{sign}\left(\Re\{\bar{\mathbf{r}}\}\right) + j\,\mathrm{sign}\left(\Im\{\bar{\mathbf{r}}\}\right) \tag{2}$$

where $\mathrm{sign}(\cdot)$ represents the 1-bit ADC with $\mathrm{sign}(a) = +1$ if $a \geq 0$ and $\mathrm{sign}(a) = -1$ if $a < 0$. The operator $\mathrm{sign}(\cdot)$ of a matrix or vector is applied separately to every element of that matrix or vector. The SNR is defined as $\rho = 1/N_0$. Recall that

each user is transmitting at unit average power. In addition, we assume that the norm of the channel from any given user to any given antenna is also normalized to 1. Thus, the SNR definition here is per-user per-antenna SNR.

III. PROPOSED SVM-BASED CHANNEL ESTIMATION AND DATA DETECTION WITH 1-BIT ADCs

A. Linear SVM for Binary Classification

Consider a binary classification problem with a training data set of P data pairs $\mathcal{D} = \{(\mathbf{x}_q, y_q)\}_{q=1,\ldots,P}$ where $\mathbf{x}_q$ is a training data point and $y_q \in \{\pm 1\}$ is an associated class label. Note that $\{\mathbf{x}_q\}$ here are vectors of real elements. The data set $\mathcal{D}$ is said to be linearly separable if and only if there exists a linear function $f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b$ such that $\forall q \in \{1, 2, \ldots, P\}$, $f(\mathbf{x}_q) > 0$ if $y_q = +1$ and $f(\mathbf{x}_q) < 0$ if $y_q = -1$. Here, $\mathbf{w}$ and b are referred to as the weight vector and the bias, respectively. In other words, the hyperplane $f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b = 0$ divides the space into two regions where $f(\mathbf{x}) = 0$ acts as the *decision boundary*. The margin of the hyperplane $f(\mathbf{x}) = 0$ with respect to $\mathcal{D}$ is defined as

$$m_{\mathcal{D}}(f) = \frac{2}{\|\mathbf{w}\|}. \tag{3}$$

The SVM technique seeks to find $\mathbf{w}$ and b such that the margin $m_{\mathcal{D}}(f)$ is maximized. The optimization problem can be expressed as [38]

$$\begin{aligned} \underset{\{\mathbf{w},b\}}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{w}\|^2 \\ \text{subject to} \quad & y_q(\mathbf{w}^T\mathbf{x}_q + b) \geq 1, \quad q = 1, 2, \ldots, P. \end{aligned} \tag{4}$$

In case the training data set $\mathcal{D}$ is not linearly separable, a generalized optimization problem is considered as follows:

$$\begin{aligned} \underset{\{\mathbf{w},b,\xi_q\}}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{q=1}^{P}\ell(\xi_q) \\ \text{subject to} \quad & y_q(\mathbf{w}^T\mathbf{x}_q + b) \geq 1 - \xi_q, \\ & \xi_q \geq 0, \quad q = 1, 2, \ldots, P. \end{aligned} \tag{5}$$

Here, $\{\xi_q\}$ are slack variables and $C > 0$ is a parameter that "controls the trade-off between the slack variable penalty and the margin" [38], and $\ell(\xi_q)$ is a function of ξ_q. In the SVM literature, two common forms of $\ell(\xi_q)$ are $\ell(\xi_q) = \xi_q$ and $\ell(\xi_q) = \xi_q^2$, which are often referred to as ℓ_1-norm SVM and ℓ_2-norm SVM, respectively.

An illustrative example for the SVM problem is given in Fig. 2. The larger the margin is, the farther the data points are from the hyperplane and so the better the classification is. This is the key point for the SVM approach, to find a hyperplane that maximizes the margin, which is equivalent to minimizing the norm of the weight vector.

The optimization problems (4) and (5) can be solved by very efficient algorithms [39]–[42]. For example, if the weight vector is sparse, the complexity of the algorithm in [39] scales linearly in both the number of features (size of the weight vector $\mathbf{w}$) and the number of training samples $|\mathcal{D}|$. For arbitrary weight
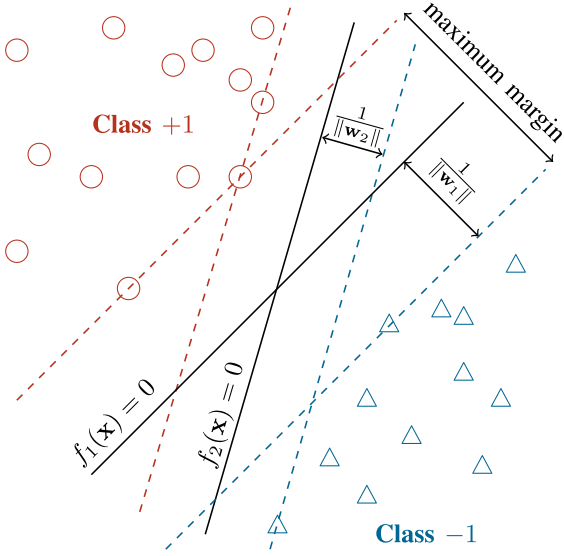
Fig. 2. An illustrative example of SVM. The hyperplane $f_2(\mathbf{x}) = \mathbf{w}_2^T \mathbf{x} + b_2 = 0$ correctly classifies the data points but its margin is not the largest possible. The hyperplane $f_1(\mathbf{x}) = \mathbf{w}_1^T \mathbf{x} + b_1 = 0$ not only correctly classifies the data points and its margin is also the maximum, thus f_1 is the SVM solution.

vectors, the complexity of the algorithms in [40]–[42] scales linearly in the number of features and super-linearly in the number of training samples. A good review of efficient methods for solving (4) and (5) can also be found in [43].

B. Proposed SVM-Based Channel Estimation

1) Uncorrelated Channels: First, we consider uncorrelated channels where the channel elements are assumed to be i.i.d. as $\mathcal{CN}(0, 1)$. In order to estimate the channel, a pilot sequence $\bar{\mathbf{X}}_t \in \mathbb{C}^{K \times T_t}$ of length T_t is used to generate the training data

$$\bar{\mathbf{Y}}_t = \text{sign}\left(\bar{\mathbf{H}}\bar{\mathbf{X}}_t + \bar{\mathbf{Z}}_t\right). \quad (6)$$

For convenience in later derivations, we convert the notation in (6) to the real domain as

$$\mathbf{Y}_t = \text{sign}\left(\mathbf{H}_t \mathbf{X}_t + \mathbf{Z}_t\right), \quad (7)$$

where

$$\mathbf{Y}_t = \left[\Re\{\bar{\mathbf{Y}}_t\}, \Im\{\bar{\mathbf{Y}}_t\}\right] = [\mathbf{y}_{t,1}, \mathbf{y}_{t,2}, \dots, \mathbf{y}_{t,N}]^T, \quad (8)$$

$$\mathbf{H}_t = \left[\Re\{\bar{\mathbf{H}}\}, \Im\{\bar{\mathbf{H}}\}\right] = [\mathbf{h}_{t,1}, \mathbf{h}_{t,2}, \dots, \mathbf{h}_{t,N}]^T, \quad (9)$$

$$\mathbf{Z}_t = \left[\Re\{\bar{\mathbf{Z}}_t\}, \Im\{\bar{\mathbf{Z}}_t\}\right] = [\mathbf{z}_{t,1}, \mathbf{z}_{t,2}, \dots, \mathbf{z}_{t,N}]^T, \text{ and} \quad (10)$$

$$\mathbf{X}_t = \begin{bmatrix} \Re\{\bar{\mathbf{X}}_t\} & \Im\{\bar{\mathbf{X}}_t\} \\ -\Im\{\bar{\mathbf{X}}_t\} & \Re\{\bar{\mathbf{X}}_t\} \end{bmatrix} = [\mathbf{x}_{t,1}, \mathbf{x}_{t,2}, \dots, \mathbf{x}_{t,2T_t}]. \quad (11)$$

Note that $\mathbf{y}_{t,i}^T \in \{\pm 1\}^{1 \times 2T_t}$, $\mathbf{h}_{t,i}^T \in \mathbb{R}^{1 \times 2K}$, and $\mathbf{z}_{t,i}^T \in \mathbb{R}^{1 \times 2T_t}$ with $i \in \{1, 2, \dots, N\}$ represent the ith rows of $\mathbf{Y}_t$, $\mathbf{H}_t$, and $\mathbf{Z}_t$, respectively. However, $\mathbf{x}_{t,n} \in \mathbb{R}^{2K \times 1}$ with $n \in \{1, 2, \dots, 2T_t\}$ is the nth column of $\mathbf{X}_t$.

It can be seen from (9) that estimating $\{\mathbf{h}_{t,i}\}_{i=1,2,\dots,N}$ is equivalent to estimating $\bar{\mathbf{H}}$. Here, we formulate the channel estimation problem in terms of $\mathbf{h}_{t,i}$. Let

$$\mathbf{y}_{t,i} = [y_{t,i,1}, \dots, y_{t,i,2T_t}]^T \text{ and } \mathbf{z}_{t,i} = [z_{t,i,1}, \dots, z_{t,i,2T_t}]^T,$$

then we have

$$y_{t,i,n} = \text{sign}\left(\mathbf{h}_{t,i}^T \mathbf{x}_{t,n} + z_{t,i,n}\right). \quad (12)$$

We stress that the estimation of $\mathbf{h}_{t,i}$ in (12) can be interpreted as an SVM binary classification problem. More specifically, $\{\mathbf{x}_{t,n}, y_{t,i,n}\}_{n=1,\dots,2T_t}$ plays the role of the training data set $\mathcal{D}$. The channel $\mathbf{h}_{t,i}$ acts as the weight vector and $z_{t,i,n}$ can be viewed as the bias. Hence, we can follow the SVM classification formulation to estimate $\mathbf{h}_{t,i}$ by solving the following optimization problem:

$$
\begin{aligned}
\underset{\{\mathbf{h}_{t,i}, \xi_n\}}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{h}_{t,i}\|^2 + C \sum_{n=1}^{2T_t} \ell(\xi_n) \\
\text{subject to} \quad & y_{t,i,n} \mathbf{h}_{t,i}^T \mathbf{x}_{t,n} \geq 1 - \xi_n, \\
& \xi_n \geq 0, \quad n = 1, 2, \dots, 2T_t.
\end{aligned}
\quad (13)
$$

Here, the bias is discarded because the $\{z_{t,i,n}\}$ are random noise with zero mean. In addition, at infinite SNR, (12) becomes $y_{t,i,n} = \text{sign}(\mathbf{h}_{t,i}^T \mathbf{x}_{t,n})$, which has no bias.

Physical interpretation of (13): Recall from Section III-A that minimizing the norm of the weight vector is equivalent to maximizing the margin, and so (13) can be read as maximizing the margin subject to the sign constraints. The larger the margin is, the farther the data points (pilot vectors $\mathbf{x}_{t,n}$) are from the hyperplane $\mathbf{h}_{t,i}^T \mathbf{x} = 0$, and so the larger the terms $y_{t,i,n} \mathbf{h}_{t,i}^T \mathbf{x}_{t,n}$ are. In other words, increasing the margin makes the sign constraints satisfied more strongly.

Let $\tilde{\mathbf{h}}_{t,i}$ denote the solution of (13). This solution provides an estimate of the channel "direction," but the magnitude of $\tilde{\mathbf{h}}_{t,i}$ is determined by the definition of the SVM margin, which in turn defines the sign constraints in (13). In fact, the instantaneous magnitude of $\mathbf{h}_{t,i}$ is not identifiable [23] since $\alpha \mathbf{h}_{t,i}$ for any $\alpha > 0$ will produce the same data set $\{y_{t,i,n}\}$:

$$y_{t,i,n} = \text{sign}\left(\mathbf{h}_{t,i}^T \mathbf{x}_{t,n}\right) = \text{sign}\left(\alpha \mathbf{h}_{t,i}^T \mathbf{x}_{t,n}\right), \text{ with } \alpha > 0.$$

Since in our model we assume that the $2K$ elements of $\mathbf{h}_{t,i}$ are each independent with variance $1/2$, we will scale the SVM solution so that the corresponding channel estimate has a squared norm of K:

$$\hat{\mathbf{h}}_{t,i} = \frac{\sqrt{K}\tilde{\mathbf{h}}_{t,i}}{\|\tilde{\mathbf{h}}_{t,i}\|}. \quad (14)$$

Here, we have assumed that the channel variance of $1/2$ is known *a priori*, so that the scaling step above is possible. We have found that this choice for the scaling provides the best estimation accuracy.

It should be noted that (13) only depends on a single index i, and so its solution is the estimate for the ith row of the channel matrix $\bar{\mathbf{H}}$, i.e., the channel vector from the K users to the ith receive antenna. This means we have N separate optimization problems of the same form (13), which is an advantage of the proposed SVM-based method since these N optimization problems can be solved in parallel.

Remark 1: The soft-SVM method in [12] does not maximize the margin, but instead calculates $\mathbf{h}_{t,i}$ such that the condition $y_{t,i,n}\mathbf{h}_{t,i}^T\mathbf{x}_{t,n} > 0$ is satisfied for as many n as possible. However, due to the noise component $z_{t,i,n}$, the condition $y_{t,i,n}\mathbf{h}_{t,i}^T\mathbf{x}_{t,n} > 0$ may not be satisfied even with the true channel vector $\mathbf{h}_{t,i}$. Our proposed method exploits the original idea of SVM by maximizing the margin achieved by the linear discriminator. The introduction of the slack variables in the problem circumvents the strict constraint $y_{t,i,n}\mathbf{h}_{t,i}^T\mathbf{x}_{t,n} > 0$.

Remark 2: Without slack variables, the problem in (13)

$$\underset{\{\mathbf{h}_{t,i}\}}{\text{minimize}} \quad \frac{1}{2}\|\mathbf{h}_{t,i}\|^2$$
$$\text{subject to} \quad y_{t,i,n}\mathbf{h}_{t,i}^T\mathbf{x}_{t,n} \geq 1, \quad n = 1, 2, \ldots, 2T_t, \tag{15}$$

is similar to the form in (4). For $\mathbf{h}_{t,i} \sim \mathcal{N}(\mathbf{0}, \frac{1}{2}\mathbf{I})$ we have

$$p(\mathbf{h}_{t,i}) = \frac{1}{\sqrt{\pi^{2K}}} \exp\left\{-\|\mathbf{h}_{t,i}\|^2\right\}, \tag{16}$$

and hence the optimization problem in (15) can be read as maximizing the pdf of $\mathbf{h}_{t,i}$ subject to the constraints $y_{t,i,n}\mathbf{h}_{t,i}^T\mathbf{x}_{t,n} \geq 1$ for $n = 1, 2, \ldots, 2T_t$. Thus, the SVM approach can be interpreted as finding the channel $\mathbf{h}_{t,i}$ that attains the highest likelihood under the constraints realized by the measured data. We will use this observation next to modify the SVM-based channel estimator when the channel is spatially correlated. Note that the work in [12] only considers uncorrelated channels.

2) Spatially Correlated Channels: We let $\bar{\mathbf{H}} = [\bar{\mathbf{h}}_1, \ldots, \bar{\mathbf{h}}_K]$, and so $\bar{\mathbf{h}}_k \in \mathbb{C}^{N \times 1}$ is the kth column of $\bar{\mathbf{H}}$. Here, we assume that the elements of $\bar{\mathbf{h}}_k$ are correlated, or in other words that the channels associated with different antennas are correlated. Let $\bar{\mathbf{h}}_k \sim \mathcal{CN}(\mathbf{0}, \bar{\mathbf{C}}_k)$ and $\bar{\mathbf{h}} = \text{vec}(\bar{\mathbf{H}})$, then $\bar{\mathbf{h}} \sim \mathcal{CN}(\mathbf{0}, \bar{\mathbf{C}})$ where $\bar{\mathbf{C}} = \text{blockdiag}(\bar{\mathbf{C}}_1, \bar{\mathbf{C}}_2, \ldots, \bar{\mathbf{C}}_K)$. We assume that the covariance matrices $\bar{\mathbf{C}}_k$ are known *a priori*. In practical systems, the covariance changes relatively slowly and thus can be obtained by a sample covariance estimator [16].

The pdf of $\bar{\mathbf{h}}$ is

$$p(\bar{\mathbf{h}}) = \frac{1}{\pi^{KN}\sqrt{\det(\bar{\mathbf{C}})}} \exp\left\{-\bar{\mathbf{h}}^H\bar{\mathbf{C}}^{-1}\bar{\mathbf{h}}\right\} \tag{17}$$

$$= \frac{1}{\pi^{KN}\sqrt{\det(\bar{\mathbf{C}})}} \exp\left\{-\sum_{k=1}^{K}\bar{\mathbf{h}}_k^H\bar{\mathbf{C}}_k^{-1}\bar{\mathbf{h}}_k\right\}. \tag{18}$$

The exponent term in (17) becomes a sum in (18) because $\bar{\mathbf{C}}$ is a block diagonal matrix, whose main-diagonal blocks are $\bar{\mathbf{C}}_1, \bar{\mathbf{C}}_2, \ldots, \bar{\mathbf{C}}_K$. Letting

$$\mathbf{h}_k = \begin{bmatrix} \Re\{\bar{\mathbf{h}}_k\} \\ \Im\{\bar{\mathbf{h}}_k\} \end{bmatrix} \text{ and } \mathbf{C}_k = \begin{bmatrix} \Re\{\bar{\mathbf{C}}_k\} & -\Im\{\bar{\mathbf{C}}_k\} \\ \Im\{\bar{\mathbf{C}}_k\} & \Re\{\bar{\mathbf{C}}_k\} \end{bmatrix},$$

the exponent term in (18) can be written as $\sum_{k=1}^{K}\mathbf{h}_k^T\mathbf{C}_k^{-1}\mathbf{h}_k$.

To maximize the likelihood of $\bar{\mathbf{h}}$ subject to the constraints $y_{t,i,n}\mathbf{h}_{t,i}^T\mathbf{x}_{t,n} \geq 1$ with $i = 1, 2, \ldots, N$ and $n = 1, 2, \ldots, 2T_t$,

we follow the intuition in (15) to formulate the following optimization problem:

$$\underset{\{\bar{\mathbf{H}}\}}{\text{minimize}} \quad \frac{1}{2}\sum_{k=1}^{K}\mathbf{h}_k^T\mathbf{C}_k^{-1}\mathbf{h}_k$$
$$\text{subject to} \quad y_{t,i,n}\mathbf{h}_{t,i}^T\mathbf{x}_{t,n} \geq 1, \tag{19}$$
$$i = 1, 2, \ldots, N \text{ and } n = 1, 2, \ldots, 2T_t.$$

In the above optimization problem, it is important to note that $\mathbf{h}_k \in \mathbb{R}^{2N \times 1}$ represents the kth column of $\bar{\mathbf{H}}$, but $\mathbf{h}_{t,i}^T$ represents the ith row of $\bar{\mathbf{H}}$. This means the objective function of (19) depends on the columns of $\bar{\mathbf{H}}$, but the constraints depend on the rows of $\bar{\mathbf{H}}$. Therefore, we cannot decompose (19) into smaller independent problems. In other words, the whole channel matrix $\bar{\mathbf{H}}$ has to be jointly estimated.

We note that the margin $\mathbf{h}_k^T\mathbf{C}_k^{-1}\mathbf{h}_k$ in (19) is measured using the Mahalanobis distance [47] rather than the Euclidean metric used in the standard SVM approach. The Mahalanobis distance in this context is the distance from $\mathbf{h}_k$ to the distribution $\mathcal{N}(\mathbf{0}, \mathbf{C}_k)$. The optimization problem in (19) can also be generalized by including slack variables as

$$\underset{\{\bar{\mathbf{H}}, \xi_{i,n}\}}{\text{minimize}} \quad \frac{1}{2}\sum_{k=1}^{K}\mathbf{h}_k^T\mathbf{C}_k^{-1}\mathbf{h}_k + C\sum_{i=1}^{N}\sum_{n=1}^{2T_t}\ell(\xi_{i,n})$$
$$\text{subject to} \quad y_{t,i,n}\mathbf{h}_{t,i}^T\mathbf{x}_{t,n} \geq 1 - \xi_{i,n} \text{ with } \xi_{i,n} \geq 0, \tag{20}$$
$$i = 1, 2, \ldots, N \text{ and } n = 1, 2, \ldots, 2T_t.$$

Since (20) is based on Remark 2, it can be interpreted as finding the channel $\bar{\mathbf{H}}$ that attains the highest likelihood under the sign constraints realized by the measured data.

Although the form of the objective function in (20) is different from that in conventional SVM problems, (20) can still be solved efficiently since it is a convex optimization problem. Let $\tilde{\mathbf{H}}$ be the solution of (20), then the channel estimate $\hat{\mathbf{H}}$ is defined as

$$\hat{\mathbf{H}} = \sqrt{\text{trace}\{\bar{\mathbf{C}}\}}\frac{\tilde{\mathbf{H}}}{\|\tilde{\mathbf{H}}\|_{\text{F}}},$$

where $\|\cdot\|_{\text{F}}$ denotes the Frobenius norm. This normalization step is similar to that for the case of uncorrelated channels, except a different coefficient $\sqrt{KN}$ is used since we jointly estimate the whole channel matrix and $\mathbb{E}[\|\bar{\mathbf{H}}\|_{\text{F}}] = \sqrt{KN}$.

C. Proposed Two-Stage SVM-Based Data Detection

In this section, we propose a two-stage SVM-based method for data detection with 1-bit ADCs. We first formulate the data detection as an SVM problem. A second stage is then employed to refine the solution from the first stage. Let $\bar{\mathbf{X}}_d = [\bar{\mathbf{x}}_{d,1}, \bar{\mathbf{x}}_{d,2}, \ldots, \bar{\mathbf{x}}_{d,T_d}] \in \mathbb{C}^{K \times T_d}$ be the transmitted data sequence of length T_d. The received data signal is given as

$$\bar{\mathbf{Y}}_d = \text{sign}\left(\bar{\mathbf{H}}\bar{\mathbf{X}}_d + \bar{\mathbf{Z}}_d\right). \tag{21}$$

The above equation is also converted to the real domain as

$$\mathbf{Y}_d = \text{sign}\left(\mathbf{H}_d\mathbf{X}_d + \mathbf{Z}_d\right) \tag{22}$$

where

$$\mathbf{Y}_{\mathrm{d}} = \begin{bmatrix} \Re\{\bar{\mathbf{Y}}_{\mathrm{d}}\} \\ \Im\{\bar{\mathbf{Y}}_{\mathrm{d}}\} \end{bmatrix} = [\mathbf{y}_{\mathrm{d},1}, \mathbf{y}_{\mathrm{d},2}, \dots, \mathbf{y}_{\mathrm{d},T_{\mathrm{d}}}], \qquad (23)$$

$$\mathbf{X}_{\mathrm{d}} = \begin{bmatrix} \Re\{\bar{\mathbf{X}}_{\mathrm{d}}\} \\ \Im\{\bar{\mathbf{X}}_{\mathrm{d}}\} \end{bmatrix} = [\mathbf{x}_{\mathrm{d},1}, \mathbf{x}_{\mathrm{d},2}, \dots, \mathbf{x}_{\mathrm{d},T_{\mathrm{d}}}], \qquad (24)$$

$$\mathbf{Z}_{\mathrm{d}} = \begin{bmatrix} \Re\{\bar{\mathbf{Z}}_{\mathrm{d}}\} \\ \Im\{\bar{\mathbf{Z}}_{\mathrm{d}}\} \end{bmatrix} = [\mathbf{z}_{\mathrm{d},1}, \mathbf{z}_{\mathrm{d},2}, \dots, \mathbf{z}_{\mathrm{d},T_{\mathrm{d}}}], \text{ and} \qquad (25)$$

$$\mathbf{H}_{\mathrm{d}} = \begin{bmatrix} \Re\{\bar{\mathbf{H}}\} & -\Im\{\bar{\mathbf{H}}\} \\ \Im\{\bar{\mathbf{H}}\} & \Re\{\bar{\mathbf{H}}\} \end{bmatrix} = [\mathbf{h}_{\mathrm{d},1}, \mathbf{h}_{\mathrm{d},2}, \dots, \mathbf{h}_{\mathrm{d},2N}]^{T}. \quad (26)$$

Here, $\mathbf{y}_{\mathrm{d},m} \in \{\pm 1\}^{2N \times 1}$, $\mathbf{x}_{\mathrm{d},m} \in \mathbb{R}^{2K \times 1}$, and $\mathbf{z}_{\mathrm{d},m} \in \mathbb{R}^{2N \times 1}$ with $m \in \{1, 2, \dots, T_{\mathrm{d}}\}$ are the mth columns of $\mathbf{Y}_{\mathrm{d}}$, $\mathbf{X}_{\mathrm{d}}$, and $\mathbf{Z}_{\mathrm{d}}$, respectively. However, $\mathbf{h}_{\mathrm{d},i'}^{T} \in \mathbb{R}^{1 \times 2K}$ with $i' \in \{1, 2, \dots, 2N\}$ represents the i'^{th} row of $\mathbf{H}_{\mathrm{d}}$.

It can be noted that the real and imaginary parts in (8)–(11) are stacked side-by-side, but they are stacked on top of each other in (23)–(26). This is due to the exchange in the role of the channel and the data matrices. In the formulation for channel estimation in (8)–(11), each row of the channel matrix is treated as the weight vector and the columns of the pilot data matrix are used as the training data points. On the other hand, the data detection formulation in (23)–(26) treats each column of the to-be-decoded data matrix as the weight vector and the rows of the channel matrix as the training data points.

It should also be noted that the pilot sequence and the data sequence are assumed to experience the same block-fading channel. Although the two channel matrices $\mathbf{H}_{\mathrm{t}}$ in (9) and $\mathbf{H}_{\mathrm{d}}$ in (26) are constructed differently, they still depend on the same channel $\bar{\mathbf{H}}$. Let

$$\mathbf{y}_{\mathrm{d},m} = [y_{\mathrm{d},m,1}, y_{\mathrm{d},m,2}, \dots, y_{\mathrm{d},m,2N}]^{T} \text{ and}$$

$$\mathbf{z}_{\mathrm{d},m} = [z_{\mathrm{d},m,1}, z_{\mathrm{d},m,2}, \dots, z_{\mathrm{d},m,2N}]^{T},$$

then we have

$$y_{\mathrm{d},m,i'} = \mathrm{sign}\left(\mathbf{h}_{\mathrm{d},i'}^{T}\mathbf{x}_{\mathrm{d},m} + z_{\mathrm{d},m,i'}\right). \qquad (27)$$

It is observed that the estimation of $\mathbf{x}_{\mathrm{d},m}$ can also be interpreted as an SVM binary classification problem. More specifically, we can treat $\mathbf{x}_{\mathrm{d},m}$ as the weight vector and the set $\{\hat{\mathbf{h}}_{\mathrm{d},i'}, y_{\mathrm{d},m,i'}\}_{i'=1,\dots,2N}$ as the training set, where $\hat{\mathbf{h}}_{\mathrm{d},i'}$ is the channel estimate of $\mathbf{h}_{\mathrm{d},i'}$ obtained as explained above. The following optimization problem provides the first-stage in finding $\mathbf{x}_{\mathrm{d},m}$:

$$\begin{aligned} \underset{\{\mathbf{x}_{\mathrm{d},m}, \xi_{i'}\}}{\mathrm{minimize}} \quad & \frac{1}{2}\|\mathbf{x}_{\mathrm{d},m}\|^{2} + C \sum_{i'=1}^{2N} \ell(\xi_{i'}) \\ \mathrm{subject\ to} \quad & y_{\mathrm{d},m,i'}\mathbf{x}_{\mathrm{d},m}^{T}\hat{\mathbf{h}}_{\mathrm{d},i'} \geq 1 - \xi_{i'}, \\ & \xi_{i'} \geq 0, \quad i' = 1, 2, \dots, 2N, \end{aligned} \qquad (28)$$

where the bias is discarded as in the channel estimation problem. Similar to the physical interpretation of (13), the detection problem (28) can be read as finding a weight vector $\mathbf{x}_{\mathrm{d},m}$ that

maximizes the margin subject to the sign constraints. The larger the margin is, the farther the data points (channel estimates $\hat{\mathbf{h}}_{\mathrm{d},i'}$) are from the hyperplane $\mathbf{x}_{\mathrm{d},m}^{T}\mathbf{h} = 0$, and so the larger the terms $y_{\mathrm{d},m,i'}\mathbf{x}_{\mathrm{d},m}^{T}\hat{\mathbf{h}}_{\mathrm{d},i'}$ are. In other words, the sign constraints are satisfied more strongly as the margin increases.

Let $\tilde{\mathbf{x}}_{\mathrm{d},m}$ denote the solution of (28) and let $\dot{\mathbf{x}}_{\mathrm{d},m}$ be the normalized version of $\tilde{\mathbf{x}}_{\mathrm{d},m}$ as

$$\dot{\mathbf{x}}_{\mathrm{d},m} = \frac{\sqrt{K}\tilde{\mathbf{x}}_{\mathrm{d},m}}{\|\tilde{\mathbf{x}}_{\mathrm{d},m}\|}. \qquad (29)$$

This normalization step is also used in [6] in order to make the power of the estimated signal equal the power of the transmitted signal.

Let $\dot{\mathbf{x}}_{\mathrm{d},m} = [\dot{x}_{\mathrm{d},m,1}, \dots, \dot{x}_{\mathrm{d},m,2K}]^{T}$, and define the first-stage detected data vector $\check{\mathbf{x}}_{\mathrm{d},m} = [\check{x}_{\mathrm{d},m,1}, \dots, \check{x}_{\mathrm{d},m,K}]^{T}$ obtained using symbol-by-symbol detection as

$$\check{x}_{\mathrm{d},m,k} = \arg\min_{x \in \mathcal{M}} |(\dot{x}_{\mathrm{d},m,k} + j\dot{x}_{\mathrm{d},m,k+K}) - x|, \qquad (30)$$

where $k \in \mathcal{K}$ and $\mathcal{M}$ represents the signal constellation (e.g., QPSK or 16-QAM). The solution to (30) is referred to as the stage 1 solution. To further improve the detection performance, a simple but efficient second detection stage is proposed as follows.

First, a candidate set $\mathcal{X}_{k}$ for each $\bar{x}_{\mathrm{d},m,k}$ is created using $\check{x}_{\mathrm{d},m,k}$ and $\dot{x}_{\mathrm{d},m,k} + j\dot{x}_{\mathrm{d},m,k+K}$ as

$$\mathcal{X}_{k} = \left\{ \acute{x} \in \mathcal{M} \,\middle|\, \frac{|(\dot{x}_{\mathrm{d},m,k} + j\dot{x}_{\mathrm{d},m,k+K}) - \acute{x}|}{|(\dot{x}_{\mathrm{d},m,k} + j\dot{x}_{\mathrm{d},m,k+K}) - \check{x}_{\mathrm{d},m,k}|} < \gamma \right\} \tag{31}$$

where $\gamma \geq 1$ is a parameter that controls the size of $\mathcal{X}_{k}$. Then the candidate set $\mathcal{X}$ for $\mathbf{x}_{\mathrm{d},m}$ is obtained as

$$\mathcal{X} = \left\{ [\acute{x}_{1}, \acute{x}_{2}, \dots, \acute{x}_{K}]^{T} \mid \acute{x}_{k} \in \mathcal{X}_{k}, \forall k \in \mathcal{K} \right\}. \qquad (32)$$

The above candidate set formation was introduced in [6]. However, the detected data vector in [6] is obtained by searching over $\mathcal{X}$ using the ML criterion, and the resulting performance is susceptible to imperfect CSI at high SNRs. This susceptibility has been reported via numerical results in [32] and [31], but no justification was given. We provide an explanation for this issue in Appendix A. To deal with the issue, we adopt here a different criterion referred to as *minimum weighted Hamming distance* [28]. Suppose that $\mathcal{X} = \{\acute{\mathbf{x}}_{1}, \acute{\mathbf{x}}_{2}, \dots, \acute{\mathbf{x}}_{|\mathcal{X}|}\}$ and let $\dot{\mathbf{x}}_{l} = [\Re\{\acute{\mathbf{x}}_{l}\}^{T}, \Im\{\acute{\mathbf{x}}_{l}\}^{T}]^{T}$ with $l \in \{1, 2, \dots, |\mathcal{X}|\}$. The second-stage detected data vector $\hat{\mathbf{x}}_{\mathrm{d},m}$ is defined as $\hat{\mathbf{x}}_{\mathrm{d},m} = \acute{\mathbf{x}}_{\hat{l}}$ where

$$\hat{l} = \arg\min_{l \in \{1, \dots, |\mathcal{X}|\}} d_{\mathrm{w}}\left(\mathbf{y}_{\mathrm{d},m}, \mathrm{sign}(\hat{\mathbf{H}}_{\mathrm{d}}\dot{\mathbf{x}}_{l})\right). \qquad (33)$$

Here, $\hat{\mathbf{H}}_{\mathrm{d}}$ is the channel estimate of $\mathbf{H}_{\mathrm{d}}$ and $d_{\mathrm{w}}(\cdot, \cdot)$ is the weighted Hamming distance defined in [28]. Note that the complexity of stage 2 depends on $|\mathcal{X}|$, which is controlled by γ. In particular, the complexity increases as γ increases. Therefore, a proper value of γ should be chosen in order to keep $|\mathcal{X}|$ from growing too large, but still remain large enough for $\mathcal{X}$ to have a high probability of containing the true transmitted signal vector so that the detection performance can be significantly improved.

The minimum weighted Hamming distance criterion above was shown to be statistically efficient [28]. However, the OSD method proposed in [28] requires a preprocessing stage whose computational complexity is proportional to $2^{N_s}|\mathcal{M}|^K$ for each channel realization. Here $N_s = 2N/G$ where $G \geq 1$ is an integer. The exponential computational complexity of OSD is a significant drawback in large-scale system implementation. The proposed SVM-based data detection method in this paper can address this complexity issue since the optimization problem (28) can be solved by very efficient algorithms [39], [43], [48].

D. An SVM-Based Joint CE-DD

In 1-bit ADC systems, the channel estimation accuracy can be improved by increasing the length of the pilot training sequence, but not necessarily by increasing the SNR [8]. For this reason, we consider an SVM-based joint CE-DD method to effectively improve the channel estimate without lengthening the pilot training sequence. The idea is to use the detected data vectors from the two-stage SVM-based method together with the pilot data vectors to obtain a refined channel estimate and then use this refined channel estimate to improve the data detection performance.

Let $\hat{\mathbf{X}}_d$ be the detected version of $\bar{\mathbf{X}}_d$ using the proposed two-stage data detection method and let

$$\hat{\mathbf{X}}_{d2} = \begin{bmatrix} \Re\{\hat{\mathbf{X}}_d\} & \Im\{\hat{\mathbf{X}}_d\} \\ -\Im\{\hat{\mathbf{X}}_d\} & \Re\{\hat{\mathbf{X}}_d\} \end{bmatrix} = [\hat{\mathbf{x}}_{d2,1}, \ldots, \hat{\mathbf{x}}_{d2,2T_d}], \quad (34)$$

$$\mathbf{Y}_{d2} = \begin{bmatrix} \Re\{\bar{\mathbf{Y}}_d\}, \Im\{\bar{\mathbf{Y}}_d\} \end{bmatrix} = [\mathbf{y}_{d2,1}, \ldots, \mathbf{y}_{d2,N}]^T, \quad (35)$$

where $\mathbf{y}_{d2,i} = [y_{d2,i,1}, y_{d2,i,2}, \ldots, y_{d2,i,2T_d}]^T$, $i = 1, \ldots, N$. The channel estimate can be refined by solving the following optimization problem:

$$\begin{aligned} \underset{\{\mathbf{h}_{t,i}, \xi_{t,n}, \xi_{d,m}\}}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{h}_{t,i}\|^2 + C\left(\sum_{n=1}^{2T_t}\ell(\xi_{t,n}) + \sum_{m=1}^{2T_d}\ell(\xi_{d,m})\right) \\ \text{subject to} \quad & y_{t,i,n}\mathbf{h}_{t,i}^T\mathbf{x}_{t,n} \geq 1 - \xi_{t,n}, \\ & y_{d2,i,m}\mathbf{h}_{t,i}^T\hat{\mathbf{x}}_{d2,m} \geq 1 - \xi_{d,m}, \\ & \xi_{t,n} \geq 0, \quad n = 1, 2, \ldots, 2T_t, \\ & \xi_{d,m} \geq 0, \quad m = 1, 2, \ldots, 2T_d. \end{aligned}$$
$$(36)$$

In the optimization problem above, we use two sets of slack variables $\{\xi_{t,n}\}$ and $\{\xi_{d,m}\}$, which correspond to the pilot sequence and the data sequence, respectively. This is just for notational convenience, as the two sets of slack variables play the same role. The physical interpretation of (36) is exactly the same as that of (13), the only difference here is that (36) involves more training data points. The refined channel estimate obtained by solving (36) can now be used for data detection again in (28) and (33). Note that the channel estimate obtained by (13) can be used as the initial solution to (36) so that the algorithm will more quickly converge to the optimal solution. Similarly, $\hat{\mathbf{X}}_d$ can also be used as the initial solution when solving (28) with the refined channel estimate. Numerical results in Section V show that this strategy will hit a certain performance bound as T_d increases.

IV. EXTENSION TO OFDM SYSTEMS WITH FREQUENCY-SELECTIVE FADING CHANNELS

In conventional OFDM systems with high-resolution ADCs, the orthogonality of the received signals on different subcarriers is preserved, so a full-resolution wide-band OFDM system can be converted into multiple narrow-band representations thanks to the use of the IFFT operation at the receiver side. Such a conversion significantly simplifies the channel estimation and data detection tasks. However, in one-bit OFDM systems, the orthogonality in the received signals is not preserved due to the severe non-linearity of one-bit ADCs [49]. Therefore, using the IFFT to convert the one-bit OFDM system into multiple narrow-band representations results in significant performance degradation. This is the fundamental motivation for the use of time-domain signal processing in one-bit OFDM systems, for example as in [8], where the Bussgang decomposition was utilized. In this section, we elaborate on how our SVM approach can be used for channel estimation and data detection for one-bit OFDM systems with frequency-selective fading channels.

Consider an uplink multiuser OFDM system with N_c subcarriers. Denote $\bar{\mathbf{x}}_k^{\mathrm{FD}} \in \mathbb{C}^{N_c \times 1}$ as the OFDM symbol from the kth user in the frequency domain. Throughout the paper, we use the superscripts "TD" and "FD" to refer to Time Domain and Frequency Domain, respectively. A cyclic prefix (CP) of length N_{cp} is added and the number of channel taps L_{tap} is assumed to satisfy $L_{\mathrm{tap}} - 1 \leq N_{\mathrm{cp}} \leq N_c$. It is also assumed that L_{tap} is known. After removing the CP, the quantized received signal at the ith antenna in the time domain is given by

$$\bar{\mathbf{y}}_i^{\mathrm{TD}} = \mathrm{sign}\left(\sum_{k=1}^K \bar{\mathbf{G}}_{i,k}^{\mathrm{TD}}\mathbf{F}^H\bar{\mathbf{x}}_k^{\mathrm{FD}} + \bar{\mathbf{z}}_i^{\mathrm{TD}}\right) \quad (37)$$

where $\mathbf{F}$ is the DFT matrix of size $N_c \times N_c$; $\bar{\mathbf{G}}_{i,k}^{\mathrm{TD}}$ is a circulant matrix whose first column is $\bar{\mathbf{g}}_{i,k}^{\mathrm{TD}} = [(\bar{\mathbf{h}}_{i,k}^{\mathrm{TD}})^T, 0, \ldots, 0]^T$; and $\bar{\mathbf{h}}_{i,k}^{\mathrm{TD}}$ is the channel vector of the kth user containing the L_{tap} channel taps, which are assumed to be i.i.d. and distributed as $\mathcal{CN}(0, \frac{1}{L_{\mathrm{tap}}})$. We also assume block-fading channels where the first OFDM symbol is used for channel estimation and the other OFDM symbols in the block-fading interval are for data transmission. Thus, the problem of channel estimation and data detection are studied separately.

If silent subcarriers are considered, the representation in equation (37) is still valid. This is because the sign operation is performed in the time domain. The only difference this would produce in our current model is that several elements in the frequency domain signal vector $\bar{\mathbf{x}}_k^{\mathrm{FD}}$ can be nullified (set to zero) for the guard subcarriers. Nevertheless, the SVM-based channel estimation and data detection using the active subcarriers in the time domain in the following derivations remain valid.

A. Proposed SVM-Based Channel Estimation in OFDM Systems With Frequency-Selective Fading Channels

Denote $\bar{\phi}_k^{\mathrm{TD}} = \mathbf{F}^H\bar{\mathbf{x}}_k^{\mathrm{FD}}$ and the training matrix $\bar{\mathbf{\Phi}}_k^{\mathrm{TD}}$ as a circulant matrix with first column equal to $\bar{\phi}_k^{\mathrm{TD}}$. We can reorganize

the system model in (37) as follows:

$$\bar{\mathbf{y}}_i^{\text{TD}} = \text{sign}\left(\sum_{k=1}^{K} \bar{\mathbf{\Phi}}_k^{\text{TD}} \bar{\mathbf{g}}_{i,k}^{\text{TD}} + \bar{\mathbf{z}}_i^{\text{TD}}\right)$$

$$= \text{sign}\left(\sum_{k=1}^{K} \bar{\mathbf{\Phi}}_{k,L_{\text{tap}}}^{\text{TD}} \bar{\mathbf{h}}_{i,k}^{\text{TD}} + \bar{\mathbf{z}}_i^{\text{TD}}\right)$$

$$= \text{sign}\left(\bar{\mathbf{\Phi}}_L^{\text{TD}} \bar{\mathbf{h}}_i^{\text{TD}} + \bar{\mathbf{z}}_i^{\text{TD}}\right) \quad (38)$$

where $\bar{\mathbf{\Phi}}_{k,L_{\text{tap}}}^{\text{TD}}$ is the matrix corresponding to the first L_{tap} columns of $\bar{\mathbf{\Phi}}_k^{\text{TD}}$, $\bar{\mathbf{\Phi}}_{L_{\text{tap}}}^{\text{TD}} = [\bar{\mathbf{\Phi}}_{1,L_{\text{tap}}}^{\text{TD}}, \bar{\mathbf{\Phi}}_{2,L_{\text{tap}}}^{\text{TD}}, \ldots, \bar{\mathbf{\Phi}}_{K,L_{\text{tap}}}^{\text{TD}}]$, and $\bar{\mathbf{h}}_i^{\text{TD}} = [(\bar{\mathbf{h}}_{i,1}^{\text{TD}})^T, (\bar{\mathbf{h}}_{i,2}^{\text{TD}})^T, \ldots, (\bar{\mathbf{h}}_{i,K}^{\text{TD}})^T]^T$.

We also convert (38) into the real domain as

$$\mathbf{y}_i^{\text{TD}} = \text{sign}\left(\mathbf{\Phi}_{L_{\text{tap}}}^{\text{TD}} \mathbf{h}_i^{\text{TD}} + \mathbf{z}_i^{\text{TD}}\right) \quad (39)$$

where

$$\mathbf{y}_i^{\text{TD}} = \left[\Re\{\bar{\mathbf{y}}_i^{\text{TD}}\}^T, \Im\{\bar{\mathbf{y}}_i^{\text{TD}}\}^T\right]^T,$$

$$\mathbf{h}_i^{\text{TD}} = \left[\Re\{\bar{\mathbf{h}}_i^{\text{TD}}\}^T, \Im\{\bar{\mathbf{h}}_i^{\text{TD}}\}^T\right]^T,$$

$$\mathbf{z}_i^{\text{TD}} = \left[\Re\{\bar{\mathbf{z}}_i^{\text{TD}}\}^T, \Im\{\bar{\mathbf{z}}_i^{\text{TD}}\}^T\right]^T, \text{ and}$$

$$\mathbf{\Phi}_{L_{\text{tap}}}^{\text{TD}} = \begin{bmatrix} \Re\{\bar{\mathbf{\Phi}}_{L_{\text{tap}}}^{\text{TD}}\} & -\Im\{\bar{\mathbf{\Phi}}_{L_{\text{tap}}}^{\text{TD}}\} \\ \Im\{\bar{\mathbf{\Phi}}_{L_{\text{tap}}}^{\text{TD}}\} & \Re\{\bar{\mathbf{\Phi}}_{L_{\text{tap}}}^{\text{TD}}\} \end{bmatrix}.$$

Denote $\mathbf{y}_i^{\text{TD}} = [y_{i,1}^{\text{TD}}, y_{i,2}^{\text{TD}}, \ldots, y_{i,2N_c}^{\text{TD}}]^T$ and $\mathbf{\Phi}_{L_{\text{tap}}}^{\text{TD}} = [(\phi_1^{\text{TD}})^T, (\phi_2^{\text{TD}})^T, \ldots, (\phi_{2N_c}^{\text{TD}})^T]^T$, leading to the following SVM problem for estimating the OFDM channel using one-bit ADCs:

$$\begin{aligned}
\underset{\{\mathbf{h}_i^{\text{TD}}, \xi_n\}}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{h}_i^{\text{TD}}\|^2 + C \sum_{n=1}^{2N_c} \ell(\xi_n) \\
\text{subject to} \quad & y_{i,n}^{\text{TD}}\left(\mathbf{h}_i^{\text{TD}}\right)^T \phi_n^{\text{TD}} \geq 1 - \xi_n, \\
& \xi_n \geq 0, \quad n = 1, 2, \ldots, 2N_c.
\end{aligned} \quad (40)$$

Denoting $\tilde{\mathbf{h}}_i^{\text{TD}}$ as the solution of (40), then we estimate $\mathbf{h}_i^{\text{TD}}$ as

$$\hat{\mathbf{h}}_i^{\text{TD}} = \frac{\sqrt{K}\tilde{\mathbf{h}}_i^{\text{TD}}}{\|\tilde{\mathbf{h}}_i^{\text{TD}}\|}. \quad (41)$$

Frequency-selective channel estimation methods using one-bit ADCs have been previously proposed in [8], [49], and [24] based on the Bussgang decomposition, Additive Quantization Noise Model (AQNM), and deep learning, respectively. The deep learning method in [24] was shown to outperform the methods of [8], [49] at low SNRs, but its performance tends to degrade as the SNR increases. In addition, the method in [24] requires a training sequence that contains many OFDM symbols, which are required to be orthogonal between different users. In our proposed method, only one OFDM symbol is used in the training phase and all users send their training symbols concurrently.

B. Proposed SVM-Based Data Detection in OFDM Systems With Frequency-Selective Fading Channels

In this section, we describe how SVM can also be used for data detection in OFDM systems with frequency-selective fading channels. We can rewrite the received quantized vector in (37) as

$$\bar{\mathbf{y}}_i^{\text{TD}} = \text{sign}\left(\bar{\mathbf{G}}_i^{\text{FD}} \bar{\mathbf{x}}^{\text{FD}} + \bar{\mathbf{z}}_i^{\text{TD}}\right) \quad (42)$$

where $\bar{\mathbf{G}}_i^{\text{FD}} = [\bar{\mathbf{G}}_{i,1}^{\text{TD}} \mathbf{F}^H, \ldots, \bar{\mathbf{G}}_{i,K}^{\text{TD}} \mathbf{F}^H] \in \mathbb{C}^{N_c \times N_c K}$ and $\bar{\mathbf{x}}^{\text{FD}} = [(\bar{\mathbf{x}}_1^{\text{FD}})^T, \ldots, (\bar{\mathbf{x}}_K^{\text{FD}})^T]^T$ is the transmitted symbol vector from the K users over N_c subcarriers. By stacking all the received signal vectors $\{\bar{\mathbf{y}}_i^{\text{TD}}\}_{i=1,\ldots,N}$ in a column vector, we have the following equation:

$$\bar{\mathbf{y}}^{\text{TD}} = \text{sign}\left(\bar{\mathbf{G}}^{\text{FD}} \bar{\mathbf{x}}^{\text{FD}} + \bar{\mathbf{z}}^{\text{TD}}\right) \quad (43)$$

where $\bar{\mathbf{y}}^{\text{TD}} = [(\bar{\mathbf{y}}_1^{\text{TD}})^T, (\bar{\mathbf{y}}_2^{\text{TD}})^T, \ldots, (\bar{\mathbf{y}}_N^{\text{TD}})^T]^T$ and $\bar{\mathbf{G}}^{\text{FD}} = [(\bar{\mathbf{G}}_1^{\text{FD}})^T, (\bar{\mathbf{G}}_2^{\text{FD}})^T, \ldots, (\bar{\mathbf{G}}_N^{\text{FD}})^T]^T$. Let $\mathbf{y}^{\text{TD}}$, $\mathbf{G}^{\text{FD}}$, and $\mathbf{x}^{\text{FD}}$ be the real-valued versions of $\bar{\mathbf{y}}^{\text{TD}}$, $\bar{\mathbf{G}}^{\text{FD}}$, and $\bar{\mathbf{x}}^{\text{FD}}$, respectively. Converting (43) to the real domain as in (23)–(26), we can formulate an SVM problem by treating the rows of $\mathbf{G}^{\text{FD}}$ as the feature vectors, the elements of $\mathbf{y}^{\text{TD}}$ as the binary indicators and $\mathbf{x}^{\text{FD}}$ as the weight vector. The solution of the SVM problem then provides the detected data.

V. NUMERICAL RESULTS

This section presents numerical results to show the superiority of the proposed methods against existing ones. For the simulations we set $C = 1$ and parameter γ for the second stage of the SVM-based detection method as $\gamma = \min\{\frac{\rho}{10} + 1.5, 3\}$ for QPSK and $\gamma = \min\{\frac{\rho}{10} + 1.3, 1.5\}$ for 16-QAM where ρ is the SNR. These values of γ are chosen empirically to make sure that $|\mathcal{X}|$ is not too large, but still large enough for $\mathcal{X}$ to have a high chance of containing the true transmitted signal vector. The length of the block-fading interval is assumed to be 500 (i.e., $T_t + T_d = 500$) unless otherwise stated. Such an assumption is not stringent for the frequency ranges (e.g., FR1 and FR2) used in 5G systems even with high user mobility, since the high Doppler will be offset by increases in bandwidth and sampling rate.

It should also be noted that the channels considered in all figures of this section are i.i.d uncorrelated, except Fig. 6. For flat-fading channel estimation, the kth row of the training matrix $\mathbf{X}_t$ is the $(k+1)^{\text{th}}$ column of the discrete Fourier transform (DFT) matrix of size $T_t \times T_t$. For frequency-selective fading channel estimation, we use orthogonal pilot sequences similar to those in [49, Eq. (23)]. Results in this section are obtained using the ℓ_2-norm SVM formulation as we have found that it provides better performance compared to the ℓ_1-norm formulation. For solving the proposed SVM-based channel estimation and data detection problems, we use the Scikit-learn library [50].

Fig. 3 presents a performance comparison of different channel estimation methods in terms of NMSE, defined here as $\text{NMSE} = \mathbb{E}[\|\hat{\mathbf{H}} - \bar{\mathbf{H}}\|_F^2]/(KN)$, where $\hat{\mathbf{H}}$ is an estimate of the channel $\bar{\mathbf{H}}$. It can first be seen that the soft-SVM method performs worse than the other methods. The error floor of the proposed SVM-based channel estimator is lower than that of the BMMSE estimator in [8] and the error floor of the proposed
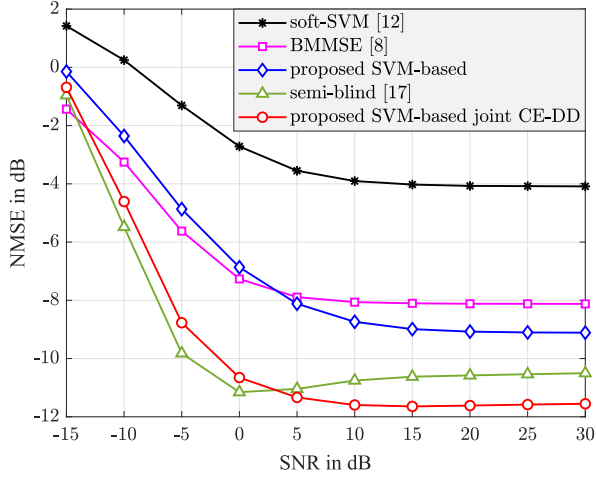
Fig. 3.   NMSE comparison between different channel estimators with $K = 4$, $N = 32$, $T_t = 20$, and $T_d = 100$.

Fig. 4.   NMSE comparison between BMMSE and the proposed SVM-based channel estimator with $K = 4$, $N = 32$, and $T_t \in \{20, 40\ 100\}$.

TABLE I
COMPUTATIONAL COMPLEXITY COMPARISON OF VARIOUS CHANNEL
ESTIMATORS. HERE, N_{iter} IS THE NUMBER OF ITERATIONS AND $\kappa(\cdot)$ IS A
SUPER-LINEAR FUNCTION

Method	Complexity
Soft-SVM [12]	$\mathcal{O}(KNT_t N_{\text{iter}})$
BMMSE [8]	$\mathcal{O}(KN^2 T_t)$
SVM-based	$\mathcal{O}(KNT_t \kappa(T_t))$
Semi-blind [17]	$\mathcal{O}(KN^2 T_b N_{\text{iter}})$
SVM-based joint CE-DD	$\mathcal{O}(KNT_b \kappa(T_b))$

Fig. 5.   Effect of T_d on the NMSE of the proposed SVM-based joint CE-DD with $K = 4$, $N = 32$, and $T_t = 20$ at $\rho = 30$ dB.

SVM-based joint CE-DD method is also lower than that of the semi-blind channel estimator in [17]. It should be noted that the semi-blind channel estimator is an extension of the BMMSE estimator when the training data set is augmented with some initially detected data vectors. The channel estimators in [8] and [17] perform well at low SNRs. However, they are outperformed by the proposed SVM-based channel estimators at higher SNRs because they use the Bussgang decomposition to obtain a linearized system model that assumes Gaussian inputs to the one-bit quantizers, an assumption that is accurate at low SNRs but less likely to be accurate as the SNR increases. The computational complexity order of the channel estimators studied in these examples is given in Table I.

In Fig. 4, we compare the NMSE of BMMSE with the NMSE of the proposed SVM-based method for different values of T_t. It is observed that the high-SNR error floor of the BMMSE method quickly reaches a bound as T_t increases. However, the performance of the proposed SVM-based method improves as T_t increases. The error floor of BMMSE even with $T_t = 100$ is still higher than that of the proposed SVM-based method with a much shorter training sequence ($T_t = 20$). The results in Fig. 4 show that increasing T_t can help improve the channel estimation accuracy. However, the spectral efficiency of the system is adversely affected as a result. Thus, the proposed
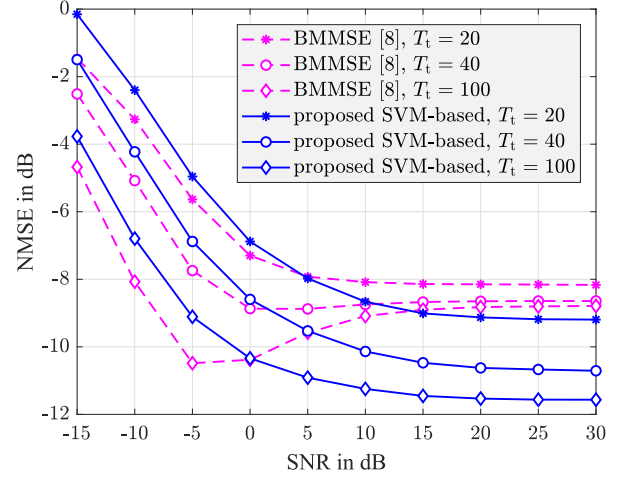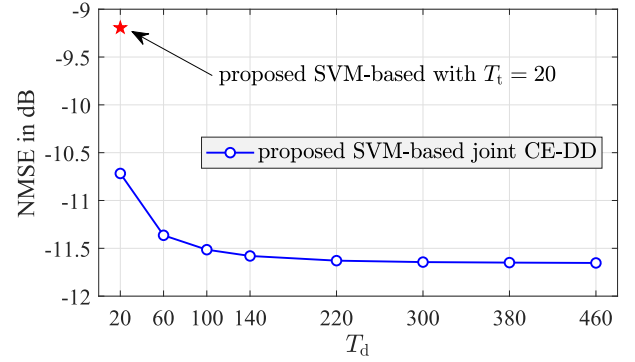
SVM-based joint CE-DD method can help improve both the channel estimation performance and the spectral efficiency.

We study the effect of T_d on the NMSE of the proposed SVM-based joint CE-DD method in Fig. 5. It can be seen that as T_d increases, the channel estimation performance of the SVM-based joint CE-DD method reaches a bound. It is also seen that with a data segment of only about 150 time slots, the channel estimation accuracy can asymptotically reach the bound, which is much better than the performance of using only the training sequence (the red star symbol). This observation suggests that when T_d is large, the complexity of the SVM-based joint CE-DD method can be reduced by using only a portion of the data block for refining the channel estimate without reducing the estimation accuracy.

Fig. 6 presents channel estimation results for spatially correlated channels. We use the same typical urban channel model as in [8]. The power angle spectrum of the channel model follows a Laplacian distribution with an angle spread of $10°$. The simulation results indicate the performance advantage of the proposed SVM-based solution over the BMMSE method at high SNR, and thus justify the SVM-based problem formulation in (20).
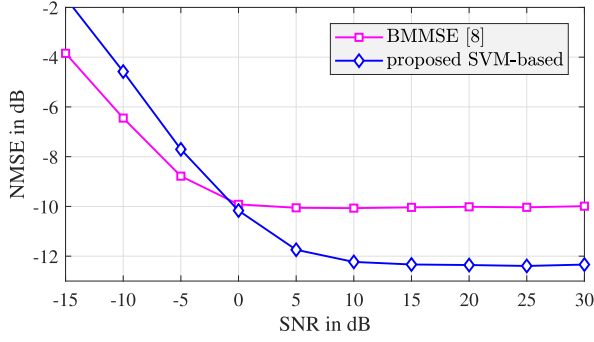
Fig. 6. NMSE comparison between the BMMSE channel estimator and the proposed SVM-based channel estimator for spatially correlated channels with $K = 4$, $N = 32$, and $T_t = 20$.
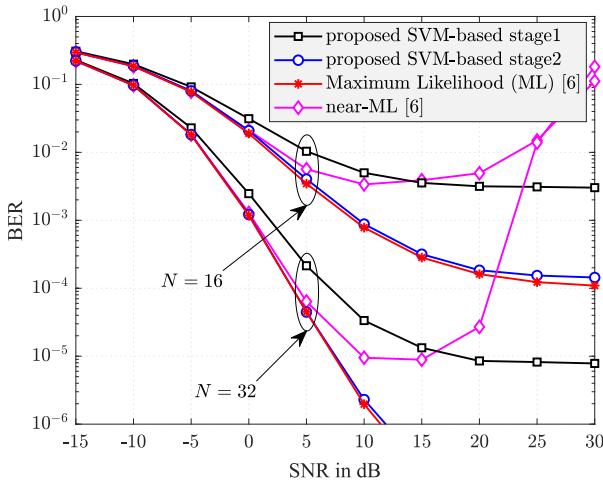

Fig. 7. Performance comparison between the proposed two-stage SVM-based data detection method and ML detection [6] with perfect CSI, QPSK modulation, and $K = 4$. The average cardinalities of $\mathcal{X}$ for $N = 16$ and $N = 32$ are 2.9352 and 1.6140, respectively.
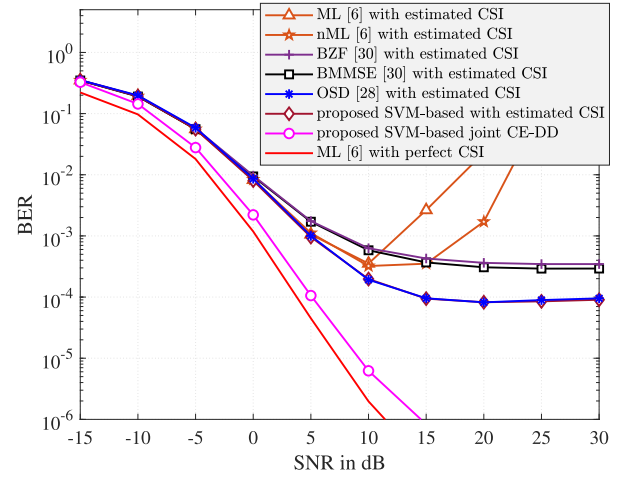

Fig. 8. Performance comparison between two proposed data detection methods and other existing methods with estimated CSI, QPSK modulation, $N = 32$, $K = 4$, $T_t = 20$, and $T_d = 480$.

TABLE II
COMPUTATIONAL COMPLEXITY COMPARISON OF DATA DETECTION METHODS: $T_b = T_t + T_d$, N_{iter} IS THE NUMBER OF ITERATIONS, $\kappa(N)$ IS A SUPER-LINEAR FUNCTION, AND $GN_s = 2N$

Method	Preprocessing	Detection Stage
BZF [30]	$\mathcal{O}(K^2 N)$	$\mathcal{O}(KNT_d)$
BMMSE [30]	$\mathcal{O}(\max\{KN^2, N^{2.373}\})$	
OSD [28]	$\mathcal{O}(2^{N_s} KN\vert\mathcal{M}\vert^K)$	$\mathcal{O}(KNGLT_d)$
ML [6]	$\mathcal{O}(KN\vert\mathcal{M}\vert^K)$	$\mathcal{O}(N\vert\mathcal{M}\vert^K T_d)$
nML [6]	–	$\mathcal{O}(KNN_{iter}T_d)$
SVM-based	–	$\mathcal{O}(KN\kappa(N)T_d)$
SVM-based joint CE-DD	–	$\mathcal{O}(KN\kappa(T_b)T_b)$

In Fig. 7, the proposed two-stage SVM-based data detection method is compared with the ML and nML detection methods for the case of perfect CSI. It is observed that the performance of the proposed method is very close to that of the ML method after two stages. It should be noted that the ML method performs well but it is an exhaustive-search method and so its computational complexity is prohibitively high for large-scale systems. While the nML method is applicable for large-scale systems, it is not robust at high SNRs. This non-robustness occurs regardless of the quality of the CSI, since nML depends on the gradient of a fractional form whose numerator and denominator both rapidly approach zero. It should also be noted that the average cardinalities of $\mathcal{X}$ for $N = 16$ and $N = 32$ are 2.9352 and 1.6140, respectively. This means the second stage of the proposed method is relatively simple to implement since it only has to search over a few candidates.

For the case of imperfect CSI, a bit-error-rate (BER) comparison is provided in Fig. 8, where the estimated CSI is obtained by the SVM-based channel estimator. Here, the SVM-based CE-DD method can be compared with other methods because it also starts with CSI estimated by the SVM-based channel estimator. Note that BZF and BMMSE are Bussgang-based zero forcing (BZF) and Bussgang-based minimum mean square error (BMMSE) linear receivers, respectively [30]. It is seen that both the ML and nML detection methods are non-robust at high SNRs with imperfect CSI. The susceptibility of ML was also reported in [32] and [31]. An explanation for the susceptibility of ML detection can be found in Appendix A. It is also observed that the proposed SVM-based and OSD detection methods give the same performance. However, the complexity order of the proposed SVM-based method is much lower than that of the OSD method as can be seen in Table II. Note that the OSD method requires the choice of two parameters N_s and L. Here, we set $N_s = 8$ and $L = 8$ since this choice provides the best performance. The proposed SVM-based joint CE-DD algorithm significantly outperforms other methods and its performance is quite close to the performance of the ML method with perfect CSI. This performance enhancement is due to the refined channel estimate obtained by solving (36).

Although the SVM-based and OSD methods give the same performance, the computational complexity of the SVM-based approach is much lower than that of OSD. This is illustrated in
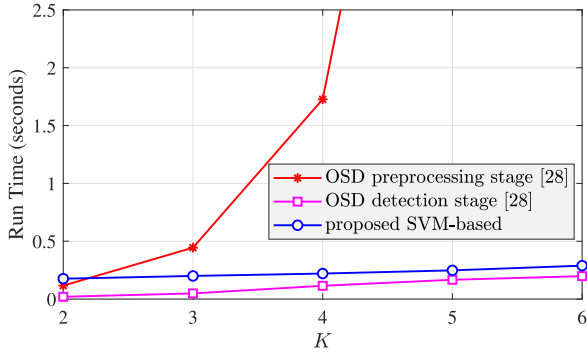
Fig. 9. Run time comparison between OSD and the proposed SVM-based detection method with QPSK modulation, $T_d = 500$, $N = 32$, and K varies.
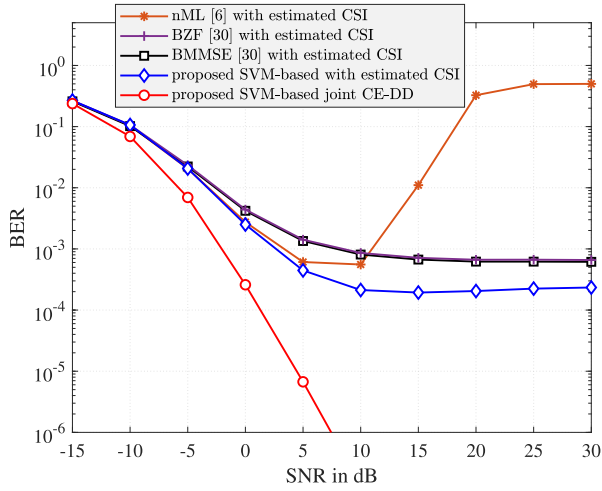

Fig. 10. Performance comparison between two proposed data detection methods and other existing methods with estimated CSI, QPSK modulation, $N = 64$, $K = 8$, $T_t = 40$, and $T_d = 460$.
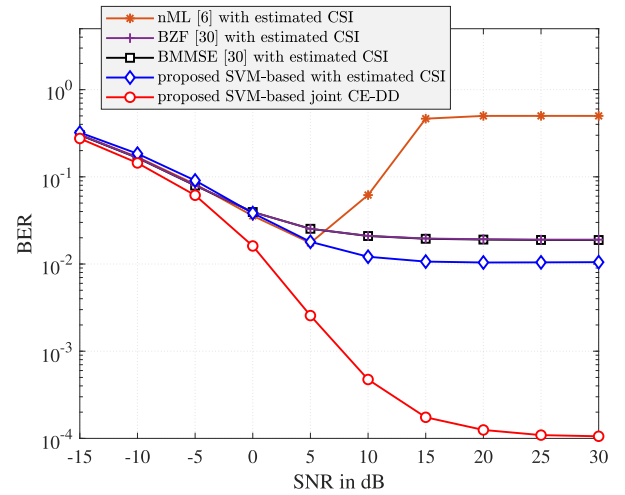

Fig. 11. Performance comparison between two proposed data detection methods and other existing methods with estimated CSI, 16-QAM modulation, $N = 128$, $K = 8$, $T_t = 40$, and $T_d = 460$.
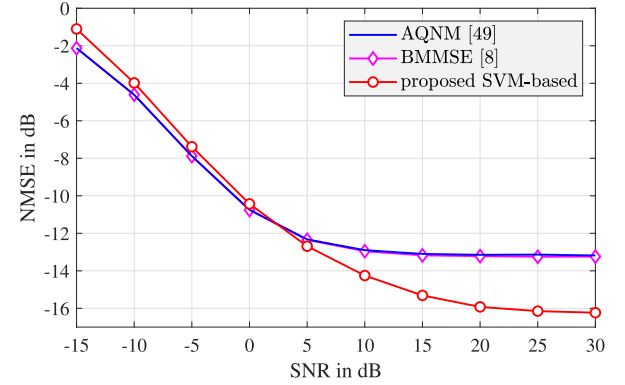

Fig. 12. NMSE comparison between different channel estimators for an OFDM system in a frequency-selective channel with $N_c = 256$, $K = 2$, $N = 16$, and $L_{tap} = 8$.
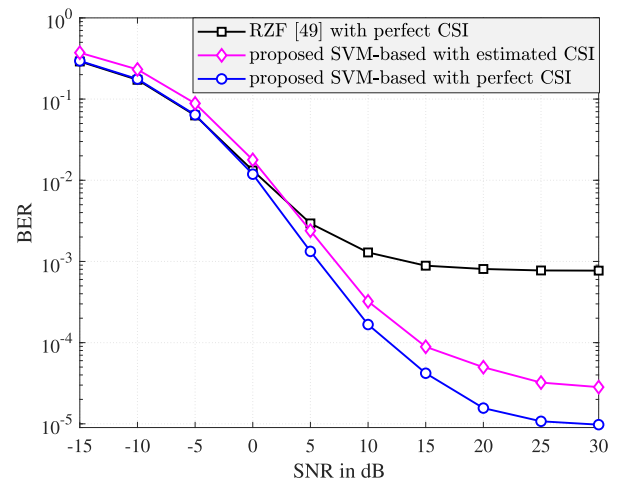

Fig. 13. BER comparison between different data detection methods for an OFDM system in a frequency-selective channel with $N_c = 256$, QPSK modulation, $K = 2$, $N = 16$, and $L_{tap} = 8$.

Fig. 9. We calculate the average run time required to perform data detection over a block-fading interval of 500 slots. Note that the OSD method contains two stages: a preprocessing stage and a detection stage. It is observed that the OSD method has a low-complexity detection stage. Interestingly, Fig. 9 indicates that the run time of proposed SVM-based method is comparable to that of the OSD detection stage. However, the OSD method requires a high-complexity preprocessing stage, which scales exponentially with the number of users. This makes the total complexity of the OSD method much higher than that of the SVM-based method, as observed in the figure. For the simulations of the OSD method in Fig. 9, we also set $N_s = 8$. The value of parameter L that achieves the best performance is 4, 4, 8, 16, and 32 for $K = 2, 3, 4, 5,$ and 6, respectively.

Fig. 10 and Fig. 11 provide BER comparisons between the proposed SVM-based data detection methods and other existing methods with QPSK and 16-QAM modulations using the CSI estimated by the SVM-based channel estimator. Due to their high computational complexity, we are not able to provide the BER of the ML and OSD detection methods. Instead, the performance

of the nML method and other linear receivers are provided as alternatives. The proposed methods not only outperform the existing methods but are also robust at high SNRs.

Finally, channel estimation and data detection results for OFDM systems with frequency-selective fading channels are given in Fig. 12 and Fig. 13, respectively. It is observed that the BMMSE channel estimator [8] slightly outperforms the AQNM-based channel estimator [49], but both of these methods have higher NMSE than the proposed SVM-based channel estimator at high SNRs. More specifically, the high-SNR error floor of the SVM-based method is about 3-dB lower that that of the BMMSE and the AQNM-based methods. In Fig. 13, data detection results show that the SVM-based method considerably outperforms the Regularized Zero-Forcing (RZF) of [49]. At high SNRs, the BER of the RZF method even with perfect CSI is much higher than the BER of the SVM-based method with estimated CSI.

VI. CONCLUSION

In this paper, we have shown how linear SVM, a well-known machine learning technique, can be exploited to provide efficient and robust channel estimation and data detection. We proposed SVM-based channel estimation methods for both uncorrelated and spatially correlated channels, a two-stage SVM-based data detection method, and an SVM-based joint CE-DD method. Extension of the proposed methods to OFDM systems with frequency-selective fading channels was also derived. The key idea is to formulate the channel estimation and data detection problems as SVM problems so that they can be efficiently solved. Simulation results revealed the superiority of the proposed methods against existing ones and the gain is greatest for moderate to high SNR regimes.

APPENDIX A
EXPLANATION FOR THE SUSCEPTIBILITY OF ML DETECTION AT HIGH SNRs WITH IMPERFECT CSI

The ML detection method of [6] is defined as

$$\hat{\mathbf{x}}_{\mathrm{d},m}^{\mathtt{ML}} = \arg \max_{\bar{\mathbf{x}} \in \mathcal{M}^{N_{\mathrm{t}}}} \underbrace{\prod_{i=1}^{2N} \Phi\left(\sqrt{2\rho} y_{\mathrm{d},m,i} \hat{\mathbf{h}}_{\mathrm{d},i}^T \mathbf{x}\right)}_{\mathcal{L}(\mathbf{x})}, \qquad (44)$$

where $\mathbf{x} = [\Re\{\bar{\mathbf{x}}\}^T, \Im\{\bar{\mathbf{x}}\}^T]^T$, $\mathcal{L}(\mathbf{x})$ is the likelihood function, and $\Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\tau^2/2} d\tau$ is the cumulative distribution function of the standard Gaussian random variable. It is clear that as $\rho \to \infty$, we have

$$\begin{cases} \Phi\left(\sqrt{2\rho} y_{\mathrm{d},m,i} \hat{\mathbf{h}}_{\mathrm{d},i}^T \mathbf{x}\right) \to 0 \text{ if } y_{\mathrm{d},m,i} \hat{\mathbf{h}}_{\mathrm{d},i}^T \mathbf{x} < 0, \\ \Phi\left(\sqrt{2\rho} y_{\mathrm{d},m,i} \hat{\mathbf{h}}_{\mathrm{d},i}^T \mathbf{x}\right) \to 1 \text{ if } y_{\mathrm{d},m,i} \hat{\mathbf{h}}_{\mathrm{d},i}^T \mathbf{x} > 0. \end{cases}$$

This means, as $\rho \to \infty$, $\mathcal{L}(\mathbf{x}) = 0$ if there exists at least one index i such that $y_{\mathrm{d},m,i} \hat{\mathbf{h}}_{\mathrm{d},i}^T \mathbf{x} < 0$ and $\mathcal{L}(\mathbf{x}) = 1$ if $y_{\mathrm{d},m,i} \hat{\mathbf{h}}_{\mathrm{d},i}^T \mathbf{x} > 0$ for all i.

Now, suppose that a vector $\bar{\mathbf{x}}^\star$ was transmitted and let $\mathbf{x}^\star = [\Re\{\bar{\mathbf{x}}^\star\}^T, \Im\{\bar{\mathbf{x}}^\star\}^T]^T$. If the CSI is perfectly known, i.e., $\hat{\mathbf{h}}_{\mathrm{d},i} = \mathbf{h}_{\mathrm{d},i}$, we have $y_{\mathrm{d},m,i} \hat{\mathbf{h}}_{\mathrm{d},i}^T \mathbf{x}^\star > 0$ for all i because

$y_{\mathrm{d},m,i} = \mathrm{sign}(\mathbf{h}_{\mathrm{d},i}^T \mathbf{x}^\star) = \mathrm{sign}(\hat{\mathbf{h}}_{\mathrm{d},i}^T \mathbf{x}^\star)$ as $\rho \to \infty$. In other words, $\mathcal{L}(\mathbf{x}^\star) = 1$ if the CSI is perfectly known at infinite SNR. However, if the CSI is not known perfectly, i.e., $\hat{\mathbf{h}}_{\mathrm{d},i} \neq \mathbf{h}_{\mathrm{d},i}$, there is a non-zero probability that $y_{\mathrm{d},m,i} = \mathrm{sign}(\mathbf{h}_{\mathrm{d},i}^T \mathbf{x}^\star) \neq \mathrm{sign}(\hat{\mathbf{h}}_{\mathrm{d},i}^T \mathbf{x}^\star)$, which means $y_{\mathrm{d},m,i} \mathrm{sign}(\hat{\mathbf{h}}_{\mathrm{d},i}^T \mathbf{x}^\star) < 0$. This causes $\mathcal{L}(\mathbf{x}^\star) = 0$. For any $\mathbf{x} \neq \mathbf{x}^\star$, it is possible that $y_{\mathrm{d},m,i} = \mathrm{sign}(\mathbf{h}_{\mathrm{d},i}^T \mathbf{x}^\star) \neq \mathrm{sign}(\hat{\mathbf{h}}_{\mathrm{d},i}^T \mathbf{x})$, which also leads to $\mathcal{L}(\mathbf{x}) = 0$. Hence, detection errors occur. The above explanation is argued at infinite SNR, but it is also valid for high SNRs because the function $\Phi(t)$ approaches 0 very fast.

To remove the product in (44), one may argue to transform the function $\mathcal{L}(\mathbf{x})$ into a sum of $\log$ functions as follows:

$$\hat{\mathbf{x}}_{\mathrm{d},m}^{\mathtt{ML}} = \arg \max_{\bar{\mathbf{x}} \in \mathcal{M}^{N_{\mathrm{t}}}} \underbrace{\sum_{i=1}^{2N} \log \Phi\left(\sqrt{2\rho} y_{\mathrm{d},m,i} \hat{\mathbf{h}}_{\mathrm{d},i}^T \mathbf{x}\right)}_{\mathcal{L}(\mathbf{x})}. \qquad (45)$$

However, the function $\mathcal{L}(\mathbf{x})$ in (45) still depends on $\Phi(\cdot)$ and can involve $\log(0)$. The proposed SVM-based data detection method is robust against imperfect CSI since it does not depend on the $\Phi(\cdot)$ function and information about the SNR is not required either.

We note that the OSD method in [28] is also robust against imperfect CSI thanks to the use of the approximation $1 - \Phi(t) \approx \frac{1}{2} e^{-0.374 t^2 - 0.777 t}$ for non-negative t. This approximation helps remove the effect of $\log \Phi(\cdot)$ in (45) since $\log e^a = a$. However, the OSD method has higher computational complexity than the proposed SVM-based methods.

REFERENCES

[1] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[2] J. G. Andrews *et al.*, "What will 5G be?," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[3] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[4] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

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

[6] 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, May 2016.

[7] C. Risi, D. Persson, and E. G. Larsson, "Massive MIMO with 1-bit ADC," *CoRR*, 2014. [Online]. Available: https://arxiv.org/abs/1404.7736

[8] 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, Aug. 2017.

[9] 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.*, Brighton, United Kingdom, May 2019, pp. 4484–4488.

[10] 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. Int. ITG Workshop Smart Antennas*, Vienna, Austria, Apr. 2019, pp. 1–5.

[11] 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.*, Brighton, United Kingdom, May 2019, pp. 4669–4673.

[12] K. Gao, N. J. Estes, B. Hochwald, J. Chisum, and J. N. Laneman, "Power-performance analysis of a simple one-bit transceiver," in *Proc. Inf. Theory Appl. Workshop*, San Diego, CA, USA, Feb. 2017, pp. 1–10.

[13] F. Liu, H. Zhu, C. Li, J. Li, P. Wang, and P. Orlik, "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.

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

[15] 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.

[16] 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, 2019.

[17] 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 Tran. Veh. Technol.*, vol. 68, no. 8, pp. 7854–7868, Aug. 2019.

[18] 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.

[19] 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, 2019.

[20] J. Mo, P. Schniter, and R. W. Heath, "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.

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

[22] 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 Glob. Commun. Conf.*, San Diego, CA, USA, Dec. 2015, pp. 1–6.

[23] 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.

[24] 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.

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

[26] 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.

[27] 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, Jul. 2020.

[28] Y. Jeon, N. Lee, S. 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.

[29] 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.

[30] L. V. Nguyen and D. H. N. Nguyen, "Linear receivers for massive MIMO systems with one-bit ADCs," *CoRR*, 2019. [Online]. Available: https://arXiv.org/abs/1907.06664

[31] 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. 2019.

[32] Y. Jeon, S. 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.

[33] S. Kim, M. So, N. Lee, and S. Hong, "Semi-supervised learning detector for MU-MIMO systems with one-bit ADCs," in *Proc. IEEE Int. Conf. Commun. Workshops*, Shanghai, China, May 2019, pp. 1–6.

[34] 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.

[35] S. H. Song, S. 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.*, Marrakesh, Morocco, Apr. 2019, pp. 1–5.

[36] Y. Cho and S. 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, Feb. 2019.

[37] 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.

[38] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2006.

[39] T. Joachims, "Training linear SVMs in linear time," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*. Philadelphia, PA, USA: ACM, Aug. 2006, pp. 217–226.

[40] J. Platt, "Sequential minimal optimization: A fast algorithm for training support vector machines," Microsoft Research, Tech. Rep. MSR-TR-98-14, 1999.

[41] T. Joachims, "Making large-scale SVM learning practical," in *Adv. Kernel Methods - Support Vector Learn.*, B. Scholkopf A. Smola, Eds. MIT Press, 1998, pp. 44–56.

[42] C. W. Hsu and C. J. Lin, "A simple decomposition method for support vector machines," *Mach. Learn.*, vol. 46, pp. 291–314, 2002.

[43] L. Bottou and C.-J. Lin, "Support vector machine solvers," *Large scale Kernel Mach.*, vol. 3, no. 1, pp. 301–320, 2007.

[44] M. J. F. Garcia, J. L. Rojo-Alvarez, F. Alonso-Atienza, and M. Martinez-Ramon, "Support vector machines for robust channel estimation in OFDM," *IEEE Signal Process. Lett.*, vol. 13, no. 7, pp. 397–400, Jul. 2006.

[45] O. M. Abdul-Latif and J. Dubois, "LS-SVM detector for RMSGC diversity in SIMO channels," in *Proc. IEEE Int. Symp. Signal Process. Appl.*, Sharjah, United Arab Emirates, Feb. 2007, pp. 1–4.

[46] L. V. Nguyen, D. H. N. Nguyen, and A. L. Swindlehurst, "SVM-based channel estimation and data detection for massive MIMO systems with one-bit ADCs," in *Proc. IEEE Int. Conf. Commun.*, Dublin, Ireland, Ireland, Jun. 2020, pp. 1–6.

[47] P. C. Mahalanobis, "On the generalized distance in statistics," in *Proc. Nat. Inst. Sci. India*, vol. 12, pp. 49–55, 1936.

[48] S. S. Keerthi and D. DeCoste, "A modified finite newton method for fast solution of large scale linear SVMs," *J. Mach. Learn. Res.*, vol. 6, pp. 341–361, Mar. 2005.

[49] C. Mollén, J. Choi, E. G. Larsson, and R. W. Heath, "Uplink performance of wideband massive MIMO with one-bit ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 87–100, Jan. 2017.

[50] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.

Ly V. Nguyen (Student Member, IEEE) received the B.Eng. degree in electronics and telecommunications from the University of Engineering and Technology, Vietnam National University, Hanoi, Vietnam, in 2014 and the M.Sc. degree in advanced wireless communications systems from CentraleSupélec, Paris-Saclay University, Gif-sur-Yvette, France, in 2016. He is currently working toward the Ph.D. degree in a Joint Doctoral Program in computational science with the San Diego State University, San Diego, CA, USA, and University of California, Irvine, CA, USA. His research interests include wireless communications, signal processing, and machine learning. He was the recipient of the Best Paper Award from the 2020 IEEE International Conference on Communications.

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. From 1990 to 2007, he was with the Department of Electrical and Computer Engineering, BYU, where from 2003 to 2006, he was the Department Chair. From 1996 to 1997, he held a joint appointment as a Visiting Scholar with Uppsala University, Uppsala, Sweden and with the KTH Royal Institute of Technology, Stockholm, Sweden. From 2006 to 2007, he was 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 Irvine, Irvine, CA, USA. From 2013 to 2016, he was the Associate Dean for Research and Graduate Studies with the Samueli School of Engineering, University of California Irvine. From 2014 to 2017, he was also a Hans Fischer Senior Fellow with the Institute for Advanced Studies, Technical University of Munich, Munich, Germany. In 2016, he was elected as a Foreign Member of the Royal Swedish Academy of Engineering Sciences, Stockholm, Sweden. He has authored or coauthored more than 300 publications in his areas of research, which include array signal processing for radar, wireless communications, and biomedical applications. He was the inaugural Editor-in-Chief of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING. 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 and 2010 IEEE Signal Processing Society's Best Paper Awards, and the 2017 IEEE Signal Processing Society Donald G. Fink Overview Paper Award.

Duy H. N. Nguyen (Senior Member, IEEE) received the B.Eng. degree (Hons.) in electrical engineering from the Swinburne University of Technology, Hawthorn, VIC, Australia, in 2005, the M.Sc. degree in electrical engineering from the University of Saskatchewan, Saskatoon, SK, Canada, in 2009, and the Ph.D. degree in electrical engineering from McGill University, Montréal, QC, Canada, in 2013. From 2013 to 2015, he held a joint appointment as a Research Associate with McGill University and a Postdoctoral Research Fellow with the Institut National de la Recherche Scientifique, Université du Québec, Montréal, QC, Canada. In 2015, he was a Research Assistant with the University of Houston, Houston, TX, USA. In 2016, he was a Postdoctoral Research Fellow with the University of Texas at Austin, Austin, TX, USA. Since 2016, he has been an Assistant 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 is currently an Associate Editor for the *EURASIP Journal on Wireless Communications and Networking*. He is a TPC member for a number of flagship IEEE conferences, including the ICC, GLOBECOM, and INFOCOM. He was the recipient of the Australian Development Scholarship, the FRQNT Doctoral Fellowship and Postdoctoral Fellowship, and the NSERC Post-doctoral Fellowship. He was also the recipient of the Best Paper Award at the 2020 IEEE International Conference on Communications.