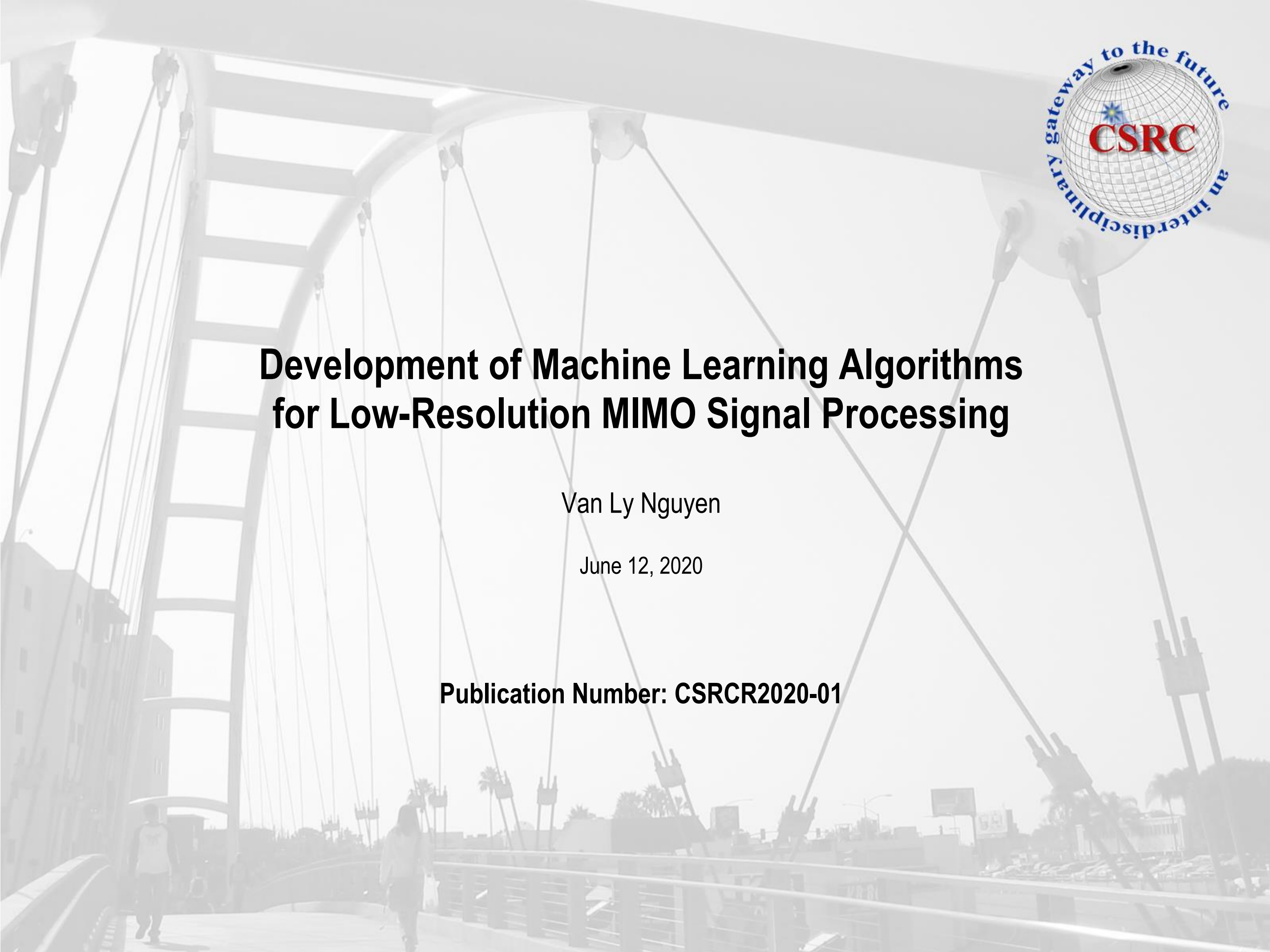


Development of Machine Learning Algorithms for Low-Resolution MIMO Signal Processing

Van Ly Nguyen

June 12, 2020

Publication Number: CSRCR2020-01

SAN DIEGO STATE
UNIVERSITY

Computational Science Research Center
College of Sciences
5500 Campanile Drive
San Diego, CA 92182-1245
(619) 594-3430

SAN DIEGO STATE UNIVERSITY
&
UNIVERSITY OF CALIFORNIA, IRVINE

Development of Machine Learning Algorithms for Low-Resolution MIMO Signal Processing

RESEARCH REPORT

submitted in partial satisfaction of the requirements
for the Research Report Examination

DOCTOR OF PHILOSOPHY

in Computational Science

by

Van Ly Nguyen

Research Report Committee:
Professor Duy H. N. Nguyen (SDSU), Advisor
Professor A. Lee Swindlehurst (UCI), Co-Advisor
Professor Ashkan Ashrafi (SDSU)
Professor Filippo Capolino (UCI)
Professor Ender Ayanoglu (UCI)

2020

TABLE OF CONTENTS

Page

LIST OF FIGURES **iv**

REFEREED PUBLICATIONS **vi**

ABSTRACT **vii**

1 Introduction **1**

2 Literature Survey and Research Statement **4**
2.1 Literature Survey 4
2.2 Research Statement 6

3 Supervised and semi-supervised learning for MIMO blind detection with low-resolution ADCs **8**
3.1 System Model 8
3.2 Blind Detection Problem 10
3.2.1 Full-space Training Method 11
3.2.2 Subspace Training Method 11
3.3 Proposed Learning Methods 13
3.3.1 Proposed Supervised Learning Method 14
3.3.2 Proposed Semi-supervised Learning Method 17
3.4 Performance Analysis with One-bit ADCs 20
3.4.1 VER Analysis at Low SNRs 21
3.4.2 VER Analysis as $\rho \to \infty$ 23
3.4.3 Transmit Signal Design 25
3.5 Simulations and Results 27
3.5.1 Numerical Evaluation of the Proposed Methods 27
3.5.2 Validation of Performance Analysis 32

4 SVM-based channel estimation and data detection for massive MIMO systems with one-bit ADCs **36**
4.1 System Model 36
4.2 Proposed SVM-based Channel Estimation and Data Detection with 1-bit ADCs 38
4.2.1 Linear SVM for Binary Classification 38
4.2.2 Proposed SVM-based Channel Estimation 39

4.2.3 Proposed Two-Stage SVM-based Data Detection 44
4.2.4 Proposed SVM-based Joint CE-DD 47
4.3 Extension to OFDM systems with Frequency-Selective Fading Channels . . . 49
4.3.1 Proposed SVM-based Channel Estimation in OFDM Systems with Frequency-Selective Fading Channels 50
4.3.2 Proposed SVM-based Data Detection in OFDM Systems with Frequency-Selective Fading Channels 51
4.4 Numerical Results 52

5 Conclusion 60

Bibliography 62

Appendix A Proof of Proposition 3.2 68

Appendix B Proof of Theorem 3.1 70

Appendix C Proof of Proposition 3.4 71

Appendix D Explanation for the susceptibility of ML detection at high SNRs with imperfect CSI 73

LIST OF FIGURES

Page

3.1 Block diagram of a MIMO communication system with low-resolution ADC at the receiver. 9

3.2 Usage of CRC for multiple data segments in each block-fading interval. . . . 14

3.3 Effect of L_{t} on MCD and the proposed methods with 1-bit ADCs, $N_{\mathrm{r}} = 16$ and BPSK modulation. 28

3.4 Performance improvement for different iterations with 1-bit ADCs, BPSK modulation, $N_{\mathrm{r}} = 16$ and $L_{\mathrm{t}} = 3$. 29

3.5 Performance comparison between blind and coherent detection with 1-bit ADCs, QPSK modulation, $N_{\mathrm{r}} = 16$ and $L_{\mathrm{t}} = 3$. 30

3.6 Performance of the proposed methods for different numbers of receive antennas N_{r} and ADC resolutions b with $L_{\mathrm{t}} = 3$. 31

3.7 Validation of the analytical pairwise VER in (3.31) and the analytical VER in (3.32) at low SNRs with $N_{\mathrm{t}} = 2$, $N_{\mathrm{r}} = 16$, and BPSK modulation. 32

3.8 Validation of the analytical VER at infinite SNR in Propositions 3.4, 3.5, and 3.6. 33

3.9 Validation of the transmit signal design with $N_{\mathrm{t}} = 6$, $N_{\mathrm{r}} = 16$, $\tilde{K} = 4$, and BPSK modulation. 33

3.10 Spectral efficiency versus $\tilde{K}$ with $N_{\mathrm{t}} = 8$, $N_{\mathrm{r}} = 16$, QPSK modulation, $L_{\mathrm{t}} = 3$, and $T_{\mathrm{b}} = 500$. 35

4.1 Block diagram of a massive MIMO system with U single-antenna users and an N-antenna base station equipped with $2N$ 1-bit ADCs. 37

4.2 NMSE comparison between different channel estimators with $U = 4$, $N = 32$, and $T_{\mathrm{t}} = 20$. 53

4.3 NMSE comparison between BMMSE and the proposed SVM-based channel estimator with $U = 4$, $N = 32$, and $T_{\mathrm{t}} \in \{20, 40, 100\}$. 53

4.4 Effect of T_{d} on the NMSE of the proposed SVM-based joint CE-DD with $U = 4$, $N = 32$, and $T_{\mathrm{t}} = 20$ at $\varrho = 30$ dB. 54

4.5 NMSE comparison between the BMMSE channel estimator and the proposed SVM-based channel estimator for spatially correlated channels with $U = 4$, $N = 32$, and $T_{\mathrm{t}} = 20$. 54

4.6 Performance comparison between the proposed two-stage SVM-based data detection method and ML detection [1] with perfect CSI, QPSK modulation, and $U = 4$. The average cardinalities of $\mathcal{X}$ for $N = 16$ and $N = 32$ are 2.9352 and 1.6140, respectively. 55
4.7 Performance comparison between two proposed data detection methods and other existing methods with estimated CSI, QPSK modulation, $N = 32$, $U = 4$, and $T_{\mathrm{t}} = 20$. 55
4.8 Run time comparison between OSD and the proposed SVM-based detection method with QPSK modulation, $N = 32$, and U varies. 57
4.9 Performance comparison between two proposed data detection methods and other existing methods with estimated CSI, QPSK modulation, $N = 64$, $U = 8$, and $T_{\mathrm{t}} = 40$. 57
4.10 Performance comparison between two proposed data detection methods and other existing methods with estimated CSI, 16-QAM modulation, $N = 128$, $U = 8$, and $T_{\mathrm{t}} = 40$. 58
4.11 NMSE comparison between different channel estimators for an OFDM system in a frequency-selective channel with $U = 2$, $N = 16$, and $L = 8$. 58
4.12 BER comparison between different data detection methods for an OFDM system in a frequency-selective channel with $N_{\mathrm{c}} = 256$, QPSK modulation, $U = 2$, $N = 16$, and $L = 8$. 59

REFEREED JOURNAL PUBLICATIONS

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.

L. V. Nguyen, A. L. Swindlehurst, and D. H. N. Nguyen, "SVM-based channel estimation and data detection for one-bit massive MIMO systems," *submitted to IEEE Trans. Signal Process.*, arXiv:2003.10678, Mar. 2020.

REFEREED CONFERENCE PUBLICATIONS

L. V. Nguyen, D. T. Ngo, N. H. Tran, and D. H. N. Nguyen, "Learning methods for MIMO blind detection with low-resolution ADCs," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kansas City, MO, USA, May 2018.

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," *accepted to IEEE Int. Conf. Commun. (ICC)*, Dublin, Ireland, June 2020.

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. This report shows that machine learning can be very useful for addressing the channel estimation and data detection problems in MIMO systems with low-resolution ADCs.

First, the blind detection problem in MIMO systems with low-resolution ADCs is studied. Two learning methods, which employ a sequence of pilot symbol vectors as the initial training data, are proposed. The first method exploits the use of a cyclic redundancy check (CRC) to obtain more training data, which helps improve the detection accuracy. The second method is based on the perspective that the to-be-decoded data can itself assist the learning process, so no further training information is required except the pilot sequence. For the case of 1-bit ADCs, a performance analysis of the vector error rate for the proposed methods is derived. Based on the analytical results, a criterion for designing transmitted signals is also presented. Simulation results show that the proposed learning methods outperform existing techniques and are also more robust.

Next, Support Vector Machine (SVM) – a well-known supervised-learning technique in machine learning – is exploited to provide efficient and robust channel estimation and data detection in massive MIMO systems with 1-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. An SVM-based joint Channel Estimation and Data Detection (CE-DD) method is also proposed. The proposed SVM-based joint CE-DD method 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 SVM-based methods are efficient and robust, and also outperform existing ones.

Chapter 1

Introduction

Wireless communications have a long history of more than 100 years dating back to the invention of the first photophone by Alexander Graham Bell and Charles Sumner Tainter in 1880. The viability of the photophone was significantly reduced due to its operational requirement of sunlight and a clear line of sight between the transmitter and the receiver. More than a decade later, in 1894, the first wireless telegraph system using radio waves was developed by Guglielmo Marconi. However, the revolution of wireless communications did not really begin until the 1990s when the semiconductor technology achieved advanced developments. With millions of electronic components packed in a single chip, advanced digital signal processing techniques and algorithms were implementable, and thus paved the way for a booming period of different wireless systems/networks such as radio and television broadcasting, radar communications, satellite communications, cellular networks, WiFi, and Bluetooth.

Today, most of the wireless communications systems use electromagnetic waves as the means of communication where the waves are transmitted and received by antenna elements. A system equipped with multiple antennas is referred to as a "Multiple Input Multiple Output" (MIMO) system. The first commercial MIMO technology was introduced by Iospan Wireless Inc. in 2001. MIMO is now included in many wireless standards [2] thanks to significant benefits obtained by the two main techniques of MIMO including: (i) spatial diversity which combats fading effects to reduce communication errors, and (ii) spatial multiplexing which exploits multipath to achieve higher data rates. The development of MIMO systems has been moving toward the use of more and more antennas at the transceivers. Massive MIMO technology is a result of this development and is now considered to be one of the disruptive technologies of 5G networks [3, 4]. 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 [5, 6]. 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 [7], 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). In addition, a massive number of active antennas and a high sampling rate demand prohibitively high bandwidth on the fronthaul link between the baseband processing unit and the RF chains. For example, a receiver that is equipped with 100 antennas, where each antenna employs two separate ADCs for the in-phase and quadrature components, and where each ADC samples at a rate of 5 GS/s with 10-bit precision would produce 10 Terabit/s of data, which is much higher than the rates of the common public radio interface in todays fiber-optical fronthaul links [8]. Thus, low-resolution ADCs are an attractive potential solution for the problems of hardware complexity, system cost, and power consumption.

This report presents a study on the use of low-resolution ADCs in massive MIMO systems, particularly on the channel estimation and data detection problems. Chapter 2 reviews the literature on MIMO systems with low-resolution ADCs and states the research problem. Two learning methods for MIMO blind detection with low-resolution ADCs are proposed in Chapter 3. A performance analysis for the proposed learning methods and a criterion for transmit signal design are also provided in Chapter 3. Next, in Chapter 4, channel estimation and data detection methods based on Support Vector Machine (SVM) for massive MIMO systems with 1-bit ADCs are proposed. Finally, Chapter 5 presents the conclusion of the report.

Chapter 2

Literature Survey and Research Statement

2.1 Literature Survey

One of the first studies on MIMO systems with low-resolution ADCs is in [9], which shows that the mutual information of 1-bit ADC MIMO systems degraded by only a factor of $2/\pi$ compared to systems with infinite-resolution ADCs. Since then, a lot more attention and efforts have been spent on this research topic. The capacity in case of correlated noise and spatially correlated channels are studied in [10] and [11], respectively. Bounds on the high SNR capacity are derived in [12]. Capacity analysis with Channel State Information at Transmitter (CSIT) is carried on in [13]. An approximate uplink achievable rate for massive MIMO systems is calculated in [14] by using the Additive Quantization Noise Model (AQNM). The achievable rate of hybrid analog-digital MIMO architectures is investigated in [15,16]. A study of achievable rate for mixed-ADC massive MIMO systems is in [17], which is extended for frequency-selective channels in [18]. A capacity lower bound for wideband massive MIMO systems with a large number of channel taps is derived in [19]. Bussgang decomposition is used for throughput analysis in [20, 21].

One major drawback of low-resolution ADCs is the significant distortions in the received signals. The severe distortions make the channel estimation and data detection tasks much more challenging compared to conventional systems with high-resolution ADCs. MIMO channel estimation with low-resolution ADCs has been studied intensively in a number of papers with different scenarios, e.g., [1, 21–40]. Maximum-Likelihood (ML) and Least-Squares (LS) 1-bit channel estimators were proposed in [1] and [22], respectively. The Bussgang decomposition is exploited in [21] to form a Bussgang-based Minimum Mean-Squared Error (BMMSE) 1-bit channel estimator. The work in [23] 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 [24] and [25]. 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 [26]. Deep learning is applied to estimate the uplink massive MIMO channels with mixed-resolution ADCs [27]. Angular-domain estimation for MIMO channels with 1-bit ADCs was studied in [28–30]. Other scenarios involving spatially/temporally correlated channels or multi-cell processing with pilot contamination were investigated in [31] and [32], respectively. For sparse millimeter-wave MIMO channels, the ML and maximum a posteriori (MAP) 1-bit channel estimation problems were studied in [33] and [34], 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 [35–37]. Several performance bounds on the channel estimation of mmWave massive MIMO channels with 1-bit ADCs were reported in [38]. The works in [39, 40] address the sparse channel estimation problem in massive MIMO systems where both hybrid analog-digital processing and low-resolution ADCs are utilized.

Data detection in MIMO systems with low-resolution ADCs has also been studied intensively in the literature, e.g., [1,41–51]. The one-bit ML detection problem is formulated in [1]. For large-scale systems where ML detection is impractical, the authors in [1] 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. ML detection with low-resolution ADCs is studied in [41,42], where the ML detection problem in [41] was relaxed to a convex optimization program for it to be solvable by low-complexity algorithms. A One-bit Sphere Decoding (OSD) technique was proposed in [43]. 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 [44] 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 [45] and several learning-based methods are also proposed in [46–48]. The linear receivers in [45] are easy to implement but their performance is often limited by an error floor. The learning-based methods in [46,47] 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 [48–51], but they are only applicable in systems where either a Cyclic Redundancy Check (CRC) [48–50] or an error correcting code such as Low-Density Parity-Check (LDPC) code [51] is available.

2.2 Research Statement

This report is concerned with the channel estimation and data detection problems in MIMO systems with low-resolution ADCs. The primary research motivation is to show that machine learning can be used for efficiently addressing the severe nonlinearity caused by low-resolution

ADCs. Blind detection in MIMO systems with low-resolution ADCs is first studied. Blind detection here means information about the CSI is unavailable. When the CSI is unknown, the channel is treated as a black box and learning methods can be exploited for addressing the blind detection problem. Then, the application of SVM to the channel estimation and data detection problems in massive MIMO systems with 1-bit ADCs is studied.

Throughout the report, we use the following notation: Upper-case and lower-case boldface letters denote matrices and column vectors, respectively. The notation $\mathbf{1}$ is a vector where every element is equal to one. $\mathbb{E}[\cdot]$ represents expectation and $\mathbb{P}[\cdot]$ is the probability of some event. $\mathbb{I}[\cdot]$ represents the indicator function, which equals 1 if the argument event is true and equals 0 otherwise. Depending on the context, the operator $|\cdot|$ is used to denote the absolute value of a real 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 operator $\mathrm{mod}(a, b)$ calculates a modulo b. The notations $\mathrm{Var}[\cdot]$ and $\mathrm{Cov}[\cdot, \cdot]$ denote the variance and covariance, respectively. The integral $\Phi(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-t^2/2} dt$ is the cumulative distribution function of the standard normal random variable. The notation $\Re\{\cdot\}$ and $\Im\{\cdot\}$ respectively denotes the real and imaginary parts of the complex argument. If $\Re\{\cdot\}$, $\Im\{\cdot\}$ or $\Phi(\cdot)$ 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 $\mathbf{A}_1, \ldots, \mathbf{A}_n$.

Chapter 3

Supervised and semi-supervised learning for MIMO blind detection with low-resolution ADCs

This chapter focuses on the blind detection problem in MIMO systems with low-resolution ADCs. Blind detection in this context means detection without information about the CSI. The system model is first presented in Section 3.1 and the blind detection problem is stated in Section 3.2. Then, a supervised learning method and a semi-supervised learning method are proposed in Section 3.3. A performance analysis for the case of 1-bit ADCs and a criterion for transmit signal design are presented in Section 3.4. Finally, simulations and results can be found in Section 3.5.

3.1 System Model

The considered MIMO system, as illustrated in Figure 3.1, has N_{t} transmit antennas and N_{r} receive antennas, where it is assumed that $N_{\mathrm{r}} \geq N_{\mathrm{t}}$. Let $\mathbf{x}[n] = [x_1[n], \ldots, x_{N_{\mathrm{t}}}[n]]^T \in \mathbb{C}^{N_{\mathrm{t}}}$ be the transmitted signal vector at time slot n, where $x_i[n]$ is the symbol transmitted at the i^{th} transmit antenna. Each symbol $x_i[n]$ is drawn from a constellation $\mathcal{M}$ with a constellation

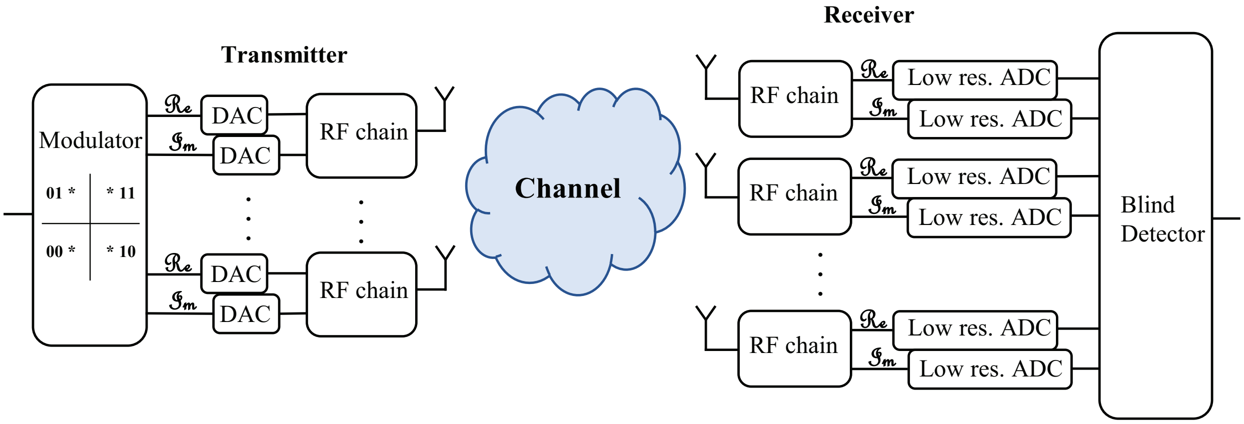


Figure 3.1: Block diagram of a MIMO communication system with low-resolution ADC at the receiver.

size of $M = |\mathcal{M}|$ under the power constraint $\mathbb{E}[|x_i[n]|^2] = 1$. The channel is assumed to be block-fading, and each block-fading interval lasts for T_b time slots. Hence, the channel $\mathbf{H} = [h_{n_\mathrm{r}n_\mathrm{t}}] \in \mathbb{C}^{N_\mathrm{r}\times N_\mathrm{t}}$ remains constant over T_b time slots. For the analysis and simulations, we assume a Rayleigh fading channel with independent and identically distributed (i.i.d.) elements and $h_{n_\mathrm{r}n_\mathrm{t}} \sim \mathcal{CN}(0,1)$, but the proposed algorithms are applicable to any channel model. The system model in each block-fading interval is

$$\mathbf{r}[n] = \mathbf{H}\mathbf{x}[n] + \mathbf{z}[n], \tag{3.1}$$

where $\mathbf{r}[n] = [r_1[n],\ldots,r_{N_\mathrm{r}}[n]]^T \in \mathbb{C}^{N_\mathrm{r}}$ is the analog received signal vector, and $\mathbf{z}[n] = [z_1[n],\ldots,z_{N_\mathrm{r}}[n]]^T \in \mathbb{C}^{N_\mathrm{r}}$ is the noise vector. The noise elements are assumed to be i.i.d. with $z_i[n] \sim \mathcal{CN}(0, N_0)$. CSI is unavailable at both the transmitter and receiver sides, i.e., $\mathbf{H}$ is unknown. The signal-to-noise ratio (SNR) is defined as $\rho = N_\mathrm{t}/N_0$.

The considered system employs an ADC that performs b-bit uniform scalar quantization, $b \in \{1,2,3,\ldots\}$. The b-bit ADC model is characterized by a set of 2^b-1 thresholds denoted as $\{\tau_1,\tau_2,\ldots,\tau_{2^b-1}\}$. Without loss of generality, we can assume $-\infty = \tau_0 < \tau_1 < \ldots < \tau_{2^b-1} < \tau_{2^b} = \infty$. Let Δ be the step size, so the threshold of a uniform quantizer is given as

$$\tau_l = (-2^{b-1}+l)\Delta, \text{ for } l \in \mathcal{L} = \{1,2,\ldots,2^b-1\}. \tag{3.2}$$

Let $Q_b(.)$ denote the element-wise quantizer, so that the quantization output is defined as

$$Q_b(r) = \begin{cases} \tau_l - \frac{\Delta}{2} & \text{if } r \in (\tau_{l-1}, \tau_l] \text{ with } l \in \mathcal{L}, \\ (2^b - 1)\frac{\Delta}{2} & \text{if } r \in (\tau_{2^b-1}, \tau_{2^b}]. \end{cases} \tag{3.3}$$

It should be noted that this mid-rise uniform quantizer satisfies $Q_b(-r) = -Q_b(r), \forall r$. The step size Δ is chosen to minimize the distortion between the quantized and non-quantized signals. The optimal value of Δ depends on the distribution of the input signals [52]. For standard Gaussian signals, the optimal step size $\Delta_{\text{opt}}^{\text{standard}}$ can be found numerically as in [53]. For non-standard complex Gaussian signals with variance $\sigma^2 \neq 1$, the optimal step size for each real/imaginary signal component can be computed as $\Delta_{\text{opt}} = \sqrt{\sigma^2/2}\Delta_{\text{opt}}^{\text{standard}}$. Hence, the optimal step size in the considered system is $\Delta_{\text{opt}} = \sqrt{(N_{\text{t}} + N_0)/2}\Delta_{\text{opt}}^{\text{standard}}$. The variance of the analog received signals $N_{\text{t}} + N_0$ is assumed to be known at the receiver.

The real and imaginary parts of each received symbol are applied to two separate ADCs. Hence, if $\mathbf{y}[n] = \big[y_1[n], \dots, y_{N_{\text{r}}}[n]\big]^T \in \mathbb{C}^{N_{\text{r}}}$ is the quantized version of the received signal vector $\mathbf{r}[n]$, then $\mathbf{y}[n] = Q_b(\mathbf{r}[n])$ in which $\Re\{y_i[n]\} = Q_b(\Re\{r_i[n]\})$ and $\Im\{y_i[n]\} = Q_b(\Im\{r_i[n]\})$ for all $i \in \mathcal{N}_{\text{r}} = \{1, 2, \dots, N_{\text{r}}\}$.

3.2 Blind Detection Problem

This section describes the blind detection problem for the block-fading channel. The first T_{t} time slots of each block fading interval contain the training symbol sequence while the remaining $T_{\text{d}} = T_{\text{b}} - T_{\text{t}}$ time slots comprise the data symbol sequence. Let $\check{\mathcal{X}} = \{\check{\mathbf{x}}_1, \check{\mathbf{x}}_2, \dots, \check{\mathbf{x}}_K\}$ denote the set of all possible transmitted symbol vectors with $K = M^{N_{\text{t}}}$ and let $\mathcal{K} = \{1, 2, \dots, K\}$. Hereafter, a possible transmitted symbol vector is called a *label*. We first revisit the MCD method presented in [54], which serves as a baseline for the study of this chapter. The input-output relations to be learned in the MCD method are $\big\{\mathbb{E}\big[\mathbf{y}|\mathbf{x} = \check{\mathbf{x}}_k\big], k \in \mathcal{K}\big\}$, in which $\mathbb{E}\big[\mathbf{y}|\mathbf{x} = \check{\mathbf{x}}_k\big]$ represents the centroid of the received quantized signal given that the label $\check{\mathbf{x}}_k$ is transmitted. The MCD data detection is given by

$$f(\mathbf{y}[n]) = \operatorname*{argmin}_{k\in\mathcal{K}} \left\| \mathbf{y}[n] - \mathbb{E}\big[\mathbf{y}|\mathbf{x} = \check{\mathbf{x}}_k\big] \right\|_2, \tag{3.4}$$

where $\mathbf{y}[n]$ is the received data symbol vector at time slot n with $n \in \{T_\mathrm{t}+1, \dots, T_\mathrm{b}\}$. Thus, the MCD approach identifies the index of the transmitted label as the one whose centroid is closest to the received vector. Denote $\check{\mathbf{y}}_k = \mathbb{E}\big[\mathbf{y}|\mathbf{x} = \check{\mathbf{x}}_k\big]$; each $\check{\mathbf{y}}_k$ is called a *representative vector* for the label $\check{\mathbf{x}}_k$. There are K representative vectors $\check{\mathcal{Y}} = \{\check{\mathbf{y}}_1, \check{\mathbf{y}}_2, \dots, \check{\mathbf{y}}_K\}$. Thus, the MCD method has to learn $\check{\mathcal{Y}}$ in order to perform the detection task. We now present two MCD training methods from [46, 54] that help the receiver empirically learn $\check{\mathcal{Y}}$.

3.2.1 Full-space Training Method

Since the transmitted signal space $\check{\mathcal{X}}$ contains K labels, a straightforward method to help the receiver learn $\check{\mathcal{Y}}$ is using a training sequence that contains all the labels, where each label is repeated a number of times. Hence, the training symbol matrix can be represented as $\mathbf{X}_\mathrm{t} = [\check{\mathbf{X}}_1, \check{\mathbf{X}}_2, \dots, \check{\mathbf{X}}_K]$, where $\check{\mathbf{X}}_k = [\check{\mathbf{x}}_k, \dots, \check{\mathbf{x}}_k] \in \mathbb{C}^{N_\mathrm{t}\times L_\mathrm{t}}$ consists of L_t labels $\check{\mathbf{x}}_k$, $k \in \mathcal{K}$. Using this training method, the representative vector $\check{\mathbf{y}}_k$ can be learned empirically as

$$\check{\mathbf{y}}_k = \frac{1}{L_\mathrm{t}} \sum_{t=1}^{L_\mathrm{t}} \mathbf{y}[(k-1)L_\mathrm{t} + t], \tag{3.5}$$

where $\mathbf{Y}_\mathrm{t} = \big[\mathbf{y}[1], \dots, \mathbf{y}[T_\mathrm{t}]\big] = Q_b(\mathbf{H}\mathbf{X}_\mathrm{t} + \mathbf{Z}_\mathrm{t})$. The length of the training sequence is $T_\mathrm{t} = KL_\mathrm{t}$. This training method has been employed in [54].

3.2.2 Subspace Training Method

It is worth noting that the training sequence does not need to cover all the labels for the receiver to learn $\check{\mathcal{Y}}$ when $\mathcal{M}$ satisfies either of the following two conditions:

- *Condition* 1: $-x \in \mathcal{M}$, $\forall x \in \mathcal{M}$.

- *Condition* 2: $\alpha x \in \mathcal{M}$, $\forall x \in \mathcal{M}$ and $\forall \alpha \in \{-1, j, -j\}$.

Although Condition 2 implies Condition 1 when $\alpha = -1$, i.e., any $\mathcal{M}$ satisfying Condition 2 will also satisfy Condition 1, we maintain these as two separate conditions for convenience in our later derivations. Examples of $\mathcal{M}$ for Condition 1 are BPSK, 8-QAM and for Condition 2 are QPSK, 16-QAM.

If Condition 1 is satisfied, $-\check{\mathbf{x}}_k \in \check{\mathcal{X}}$ for all $\check{\mathbf{x}}_k \in \check{\mathcal{X}}$. The set of all labels can be written as

$$\check{\mathcal{X}} = \{\check{\mathcal{X}}_{\text{ha}}, -\check{\mathcal{X}}_{\text{ha}}\}, \tag{3.6}$$

where $\check{\mathcal{X}}_{\text{ha}} = \{\check{\mathbf{x}}_1, \ldots, \check{\mathbf{x}}_{K/2}\}$. Without loss of generality, it is assumed that $\check{\mathbf{x}}_{k+K/2} = -\check{\mathbf{x}}_k$ with $k \in \{1, \ldots, K/2\}$. If Condition 2 is satisfied, then $\alpha\check{\mathbf{x}}_k \in \check{\mathcal{X}}$ for all $\check{\mathbf{x}}_k \in \check{\mathcal{X}}$ and $\alpha \in \{-1, j, -j\}$. The set of all labels can be written as

$$\check{\mathcal{X}} = \{\check{\mathcal{X}}_{\text{fo}}, -\check{\mathcal{X}}_{\text{fo}}, j\check{\mathcal{X}}_{\text{fo}}, -j\check{\mathcal{X}}_{\text{fo}}\}, \tag{3.7}$$

where $\check{\mathcal{X}}_{\text{fo}} = \{\check{\mathbf{x}}_1, \ldots, \check{\mathbf{x}}_{K/4}\}$. It is then assumed that $\check{\mathbf{x}}_{k+K/4} = -\check{\mathbf{x}}_k$, $\check{\mathbf{x}}_{k+K/2} = j\check{\mathbf{x}}_k$, and $\check{\mathbf{x}}_{k+3K/4} = -j\check{\mathbf{x}}_k$ for $k \in \{1, \ldots, K/4\}$. The subscripts 'ha' and 'fo' here stand for 'half' and 'fourth', indicating the first one-half and the first one-fourth of the set $\check{\mathcal{X}}$, respectively.

The work in [46] showed that if the transmitter employs QAM modulation and the quantization function satisfies $Q_b(-r) = -Q_b(r)$ for any $r \in \mathbb{R}$, then the length of the training sequence can be reduced to $T_{\text{t}} = KL_{\text{t}}/4$. In Proposition 3.1 below, we generalize this result for any modulation scheme.

Proposition 3.1. *Given any constellation $\mathcal{M}$, if the quantizer $Q_b(.)$ is symmetric, i.e., $Q_b(-r) = -Q_b(r)\ \forall r \in \mathbb{R}$, the length of the training sequence T_{t} can be reduced to*

$$T_{\text{t}} = \begin{cases} \frac{1}{2}KL_{\text{t}} & \text{if Condition 1 holds,} \\ \frac{1}{4}KL_{\text{t}} & \text{if Condition 2 holds.} \end{cases} \tag{3.8}$$

Proof. For any two labels $\check{\mathbf{x}}_{k_1}$ and $\check{\mathbf{x}}_{k_2} = -\check{\mathbf{x}}_{k_1}$, we have

$$
\begin{aligned}
p(\mathbf{y}|\mathbf{x} = \check{\mathbf{x}}_{k_2}) = \mathbb{P}\big[\mathbf{y} = Q_b(\mathbf{H}\mathbf{x}_{k_2} + \mathbf{z})\big] = \mathbb{P}\big[\mathbf{y} = Q_b(-\mathbf{H}\mathbf{x}_{k_1} - \mathbf{z})\big] = \mathbb{P}\big[-\mathbf{y} = Q_b(\mathbf{H}\mathbf{x}_{k_1} + \mathbf{z})\big]& \\
= p(-\mathbf{y}|\mathbf{x} = \check{\mathbf{x}}_{k_1}).& \qquad (3.9)
\end{aligned}
$$

Therefore, $\check{\mathbf{y}}_{k_2} = -\check{\mathbf{y}}_{k_1}$ since

$$
\begin{aligned}
\check{\mathbf{y}}_{k_2} = \mathbb{E}\big[\mathbf{y}|\mathbf{x} = \check{\mathbf{x}}_{k_2}\big] &= \sum \mathbf{y}p(\mathbf{y}|\mathbf{x} = \check{\mathbf{x}}_{k_2}) = \sum \mathbf{y}p(-\mathbf{y}|\mathbf{x} = \check{\mathbf{x}}_{k_1}) \\
&= -\sum \dot{\mathbf{y}}p(\dot{\mathbf{y}}|\mathbf{x} = \check{\mathbf{x}}_{k_1}) \qquad (3.10) \\
&= -\mathbb{E}\big[\mathbf{y}|\mathbf{x} = \check{\mathbf{x}}_{k_1}\big] = -\check{\mathbf{y}}_{k_1}, \qquad (3.11)
\end{aligned}
$$

where (3.10) is obtained by setting $\dot{\mathbf{y}} = -\mathbf{y}$ and (3.11) holds because the sample spaces of $\dot{\mathbf{y}}$ and $\mathbf{y}$ are the same. Hence, the representative vectors satisfy $\check{\mathbf{y}}_{k+K/2} = -\check{\mathbf{y}}_k$ with $k \in \{1, \ldots, K/2\}$ if Condition 1 holds. This means the training sequence only needs to cover $\check{\mathcal{X}}_{\text{ha}}$ to help the receiver learn all K representative vectors in $\check{\mathcal{Y}}$. Similarly, when Condition 2 holds, we can also show that $\check{\mathbf{y}}_{k+K/4} = -\check{\mathbf{y}}_k$, $\check{\mathbf{y}}_{k+K/2} = j\check{\mathbf{y}}_k$, and $\check{\mathbf{y}}_{k+3K/4} = -j\check{\mathbf{y}}_k$ with $k \in \{1, \ldots, K/4\}$, and so the training sequence only needs to contain $\check{\mathcal{X}}_{\text{fo}}$. It should be noted that the proof for Condition 2 requires that $Q_b(jc) = jQ_b(c), \forall c \in \mathbb{C}$, which is satisfied by the quantizer being used. □

3.3 Proposed Learning Methods

The MCD detection method is simple but it has a primary drawback – its detection accuracy heavily depends on the length of the training sequence. If the training sequence cannot provide accurate representative vectors in (3.5), then detection errors will appear in (3.4). In fact, a short training sequence often results in poor estimation of the representative vectors. In order to improve the detection accuracy *without* lengthening the training sequence, the

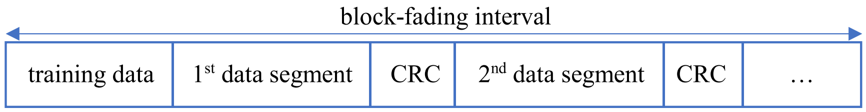


Figure 3.2: Usage of CRC for multiple data segments in each block-fading interval.

idea is to use the training sequence as an initial guide for the learning process, and then find more precise representative vectors by exploiting other information.

3.3.1 Proposed Supervised Learning Method

In practical communications systems, error control mechanisms such as the CRC can be used to determine whether a segment of data is correctly decoded or not. This approach has been exploited to mitigate the effect of imperfect CSI on the ML detection for low-resolution ADCs [55, 56]. An error correcting code was also used to update the weights in a neural network as the channel changes, assuming perfect ADCs [57].

In the proposed method, should the CRC be available, it can be exploited for blind detection as follows: Data detection is first performed by the MCD using the training sequence, then the correctly decoded data confirmed by the CRC is used to augment the training set. As a result, the representative vectors obtained from the training sequence in (3.5) can be refined and the incorrectly decoded data can be re-evaluated by the MCD data detection. The process of CRC checking, updating the representative vectors, and data detection is repeated until no further correctly decoded segment is found.

In the system considered, we assume the use of the CRC for multiple data segments as illustrated in Figure 3.2. Suppose there are S segments in one block-fading interval, and each segment contains a data segment and a CRC block. Let L_{CRC} and L_{data} denote the length of the CRC and the length of each data segment in bits, respectively. Thus, we have

$$S \times (L_{\mathrm{data}} + L_{\mathrm{CRC}}) = T_{\mathrm{d}} \times N_{\mathrm{t}} \times \log_2(M). \tag{3.12}$$

We also assume that $L_{\mathrm{data}} + L_{\mathrm{CRC}}$ is a multiple of $N_{\mathrm{t}} \log_2 M$. This means the number of

Algorithm 1: Supervised Learning Decoding.

```
1  Set $u_n = \lfloor (n-1)/L_{\mathrm{t}} \rfloor + 1$ and $c_n = 1$ for $1 \leq n \leq T_{\mathrm{t}}$;
2  Initialize $u_n = 0$ and $c_n = 0$ for $T_{\mathrm{t}} < n \leq T_{\mathrm{b}}$;
3  Set $\mathcal{C} = \varnothing$, $\mathcal{S} = \{1, 2, \ldots, S\}$, $iter = 0$, and $done = false$;
4  Find $\check{\mathcal{Y}}$ using (3.13) with the above inital setting;
5  while $done = false$ do
6      foreach $s \in \mathcal{S}$ do
7          foreach $\mathbf{y}[n] \in \mathbf{Y}_s$ do
8              Set $u_n = f(\mathbf{y}[n])$;
9          end
10         if CRC confirms the correct detection of $\mathbf{Y}_s$ then
11             Set $\mathcal{C} = \mathcal{C} \cup \{s\}$;
12             foreach $\mathbf{y}[n] \in \mathbf{Y}_s$ do
13                 Set $c_n = 1$;
14             end
15         end
16         Update $\check{\mathcal{Y}}$ using (3.13);
17     end
18     Set $iter = iter + 1$;
19     Set $\mathcal{S} = \mathcal{S} \backslash \mathcal{C}$, then set $\mathcal{C} = \varnothing$;
20     if $\mathcal{S} = \varnothing$ or $iter = iter_{\max}$ or no change in $\mathbf{u}$ then
21         $done = true$;
22     end
23 end
```

bits in a segment is a multiple of the number bits in a transmitted vector. The decoding algorithm of this proposed method is presented in Algorithm 1. The detailed explanation of Algorithm 1 is as follows.

Let $\mathbf{u} = [u_1, \ldots, u_{T_{\mathrm{b}}}]$ denote the vector of decoded indices where $u_n \in \mathcal{K}$ with $n \in \{1, \ldots, T_{\mathrm{b}}\}$ is the decoded index of received signal $\mathbf{y}[n]$. Here, we can set $u_n = \lfloor (n-1)/L_{\mathrm{t}} \rfloor + 1$ for $1 \leq n \leq T_{\mathrm{t}}$ (line 1) due to the training sequence and we can initialize $u_n = 0$ for $T_{\mathrm{t}} < n \leq T_{\mathrm{b}}$ (line 2). Let $\mathbf{c} = [c_1, \ldots, c_{T_{\mathrm{b}}}]$ denote the vector of binary values where $c_n = 1$ if the CRC confirms a correct detection of $\mathbf{y}[n]$, otherwise $c_n = 0$. Note that $c_n = 0$ does not imply an incorrect detection of $\mathbf{y}[n]$. Instead, it implies that the CRC cannot confirm a correct detection of $\mathbf{y}[n]$. Since the first T_{t} time slots are for the training sequence, we can set $c_n = 1$ for $1 \leq n \leq T_{\mathrm{t}}$ (line 1) and initialize $c_n = 0$ for $T_{\mathrm{t}} < n \leq T_{\mathrm{b}}$ (line 2). Let s denote the index of the segments, $s \in \{1, \ldots, S\}$, and let $\mathbf{Y}_s$ denote the s^{th} received data segment.

After the detection of each segment, $\check{\mathbf{y}}_k$ can be refined as (line 16):

$$\check{\mathbf{y}}_k = \frac{\sum_{n=1}^{T_\text{b}} \big(\mathbb{I}[u_n = k] + c_n \gamma(n,k)\big)\mathbf{y}[n]}{\sum_{n=1}^{T_\text{b}} \big(\mathbb{I}[u_n = k] + c_n \mathbb{I}[\gamma(n,k) \neq 0]\big)} \tag{3.13}$$

where $\mathbb{I}$ is the indicator function, and $\gamma(n,k)$ is a function of n and k defined as follows:

- *Condition* 1: $\gamma(n,k) = -\mathbb{I}[u_n = \bar{k}]$ with

$$\bar{k} = \begin{cases} k + \frac{K}{2} & \text{if } k \leq \frac{K}{2}, \\ k - \frac{K}{2} & \text{if } k > \frac{K}{2}. \end{cases} \tag{3.14}$$

- *Condition* 2:

  Let $\mathcal{K}_1 = \big\{1, \dots, \frac{K}{4}\big\}$, $\mathcal{K}_2 = \big\{\frac{K}{4} + 1, \dots, \frac{K}{2}\big\}$, $\mathcal{K}_3 = \big\{\frac{K}{2} + 1, \dots, \frac{3K}{4}\big\}$, and $\mathcal{K}_4 = \big\{\frac{3K}{4} + 1, \dots, K\big\}$;

$$\begin{aligned}
&\text{if } k \in \mathcal{K}_1, \text{ let } \bar{k}_1 = k + \frac{K}{4}, \bar{k}_2 = k + \frac{K}{2}, \bar{k}_3 = k + \frac{3K}{4}, \\
&\text{if } k \in \mathcal{K}_2, \text{ let } \bar{k}_1 = k - \frac{K}{4}, \bar{k}_2 = k + \frac{K}{2}, \bar{k}_3 = k + \frac{K}{4}, \\
&\text{if } k \in \mathcal{K}_3, \text{ let } \bar{k}_1 = k + \frac{K}{4}, \bar{k}_2 = k - \frac{K}{4}, \bar{k}_3 = k - \frac{K}{2}, \\
&\text{if } k \in \mathcal{K}_4, \text{ let } \bar{k}_1 = k - \frac{K}{4}, \bar{k}_2 = k - \frac{3K}{4}, \bar{k}_3 = k - \frac{K}{2},
\end{aligned}$$

$$\gamma(n,k) = -\mathbb{I}[u_n = \bar{k}_1] - j\mathbb{I}[u_n = \bar{k}_2] + j\mathbb{I}[u_n = \bar{k}_3]. \tag{3.15}$$

Intuitively, the representative vector $\check{\mathbf{y}}_k$ in (3.13) is updated by using received vectors whose decoded indices are k and ones that are decoded correctly (confirmed by the CRC) with decoded indices $\bar{k}$ for Condition 1 or $\bar{k}_1$, $\bar{k}_2$, $\bar{k}_3$ for Condition 2.

The refined representative vectors are then used to perform data detection on the next segment (back to lines 7–9). In the first iteration, the next segment is $\mathbf{Y}_{s+1}$, which has not been decoded before. In the subsequent iterations, the next segment is one that has not been successfully decoded. Iterations here are accounted for by the **while** loop. The process of CRC checking, updating the representative vectors and data detection is repeated until all segments are decoded correctly or no change in $\mathbf{u}$ is found or a maximum number of iterations is reached (line 20).

3.3.2 Proposed Semi-supervised Learning Method

In this part we propose a semi-supervised learning method. This proposed method is based on the K-means clustering technique [58]. The idea is to use the training sequence as an initial guidance to find coarse estimates of the representative vectors. Based on these coarse estimates, the received data vectors are then self-classified iteratively.

The K-means clustering technique aims to partition data into a number of clusters. However, in this communication context, the decoding task is not just to partition the received data into clusters but also to assign labels to the clusters, which can be done by using the training sequence. In addition, we take into account the constraints $\check{\mathbf{y}}_{k+K/2} = -\check{\mathbf{y}}_k$, $k = 1, \ldots, K/2$, if Condition 1 holds; and the constraints $\check{\mathbf{y}}_{k+K/4} = -\check{\mathbf{y}}_k$, $\check{\mathbf{y}}_{k+K/2} = j\check{\mathbf{y}}_k$, $\check{\mathbf{y}}_{k+3K/4} = -j\check{\mathbf{y}}_k$, $k = 1, \ldots, K/4$, if Condition 2 holds. These constraints can be adopted because clusters are formed based on their centroids, which are also referred to as the representative vectors $\{\check{\mathbf{y}}_k\}$ in this paper.

First, we introduce a set of binary variables $\beta_{n,k} \in \{0, 1\}$ to indicate which of the K labels that the received vector $\mathbf{y}[n]$ belongs to. Specifically, if a received vector $\mathbf{y}[n]$ belongs to label k, then $\beta_{n,k} = 1$ and $\beta_{n,l} = 0$ $\forall l \neq k$. We have the following optimization problems:

- *Condition 1*:

$$\begin{aligned} \underset{\{\beta_{n,k}\},\{\check{\mathbf{y}}_k\}}{\text{minimize}} \quad & J = \sum_{n=1}^{T_\mathrm{b}} \sum_{k=1}^{K} \beta_{n,k} \|\mathbf{y}[n] - \check{\mathbf{y}}_k\|^2 \\ \text{subject to} \quad & \check{\mathbf{y}}_{k+\frac{K}{2}} = -\check{\mathbf{y}}_k, \quad k = 1, \ldots, K/2. \end{aligned} \tag{3.16}$$

  The objective function in (3.16) is called the *distortion measure* [58]. This problem can be rewritten as

$$\underset{\{\beta_{n,k}\},\{\check{\mathbf{y}}_k\}}{\text{minimize}} \quad J_1 \tag{3.17}$$

where

$$J_1 = \sum_{n=1}^{T_\mathrm{b}} \sum_{k=1}^{\frac{K}{2}} \big(\beta_{n,k}\|\mathbf{y}[n] - \check{\mathbf{y}}_k\|^2 + \beta_{n,k+\frac{K}{2}}\|\mathbf{y}[n] + \check{\mathbf{y}}_k\|^2\big). \tag{3.18}$$

Problem (3.17) can be solved iteratively in which each iteration finds $\{\beta_{n,k}\}$ based on fixed $\{\check{\mathbf{y}}_k\}$ and vice versa. If $\{\check{\mathbf{y}}_k\}$ are fixed, J_1 is a linear function of $\{\beta_{n,k}\}$. It can be seen that the solutions $\{\beta_{n,k}\}$ are independent of n, so they can be found separately. With any $n \in \{T_\mathrm{t}+1, \dots, T_\mathrm{b}\}$, the optimization problem for $\{\beta_{n,k}\}$ is

$$\underset{\{\beta_{n,k}\}}{\text{minimize}} \quad \sum_{k=1}^{K} \beta_{n,k}\|\mathbf{y}[n] - \check{\mathbf{y}}_k\|^2, \tag{3.19}$$

whose solution is found by setting $\beta_{n,k} = 1$ for the k associated with the minimum value of $\|\mathbf{y}[n] - \check{\mathbf{y}}_k\|^2$. The solutions $\{\beta_{n,k}\}$ can be written as

$$\beta_{n,k} = \begin{cases} 1 & \text{if } k = \operatorname{argmin}_{k'} \|\mathbf{y}[n] - \check{\mathbf{y}}_{k'}\|^2, \\ 0 & \text{otherwise.} \end{cases} \tag{3.20}$$

It should be noted that $\beta_{n,k} = 1$ whenever $n \le T_\mathrm{t}$ and $k = \lfloor (n-1)/L_\mathrm{t} \rfloor + 1$ because the labels of the received training vectors are known at the receiver. When the $\{\beta_{n,k}\}$ are fixed, J_1 becomes a quadratic function of $\{\check{\mathbf{y}}_k\}$. Hence the solutions $\{\check{\mathbf{y}}_k\}$ can be found by finding the derivative of J_1 with respect to $\check{\mathbf{y}}_k$:

$$\frac{\partial J_1}{\partial \check{\mathbf{y}}_k} = \sum_{n=1}^{T_\mathrm{b}} \beta_{n,k}\big(-\mathbf{y}[n]^H + \check{\mathbf{y}}_k^H\big) + \beta_{n,k+\frac{K}{2}}\big(\mathbf{y}[n]^H + \check{\mathbf{y}}_k^H\big), \tag{3.21}$$

when being set to 0 yields

$$\check{\mathbf{y}}_k = \frac{\sum_n \big(\beta_{n,k} - \beta_{n,k+\frac{K}{2}}\big)\mathbf{y}[n]}{\sum_n \big(\beta_{n,k} + \beta_{n,k+\frac{K}{2}}\big)}, \quad k = 1, \dots, \frac{K}{2}. \tag{3.22}$$

Equation (3.22) says that the representative vector $\check{\mathbf{y}}_k$, with $k \le K/2$, is calculated by using the received vectors that not only belong to cluster k but also to cluster $k + K/2$.

- *Condition 2*:

$$\begin{aligned}
&\underset{\{\beta_{n,k}\},\{\check{\mathbf{y}}_k\}}{\text{minimize}} \quad J = \sum_{n=1}^{T_\mathrm{b}} \sum_{k=1}^{K} \beta_{n,k} \|\mathbf{y}[n] - \check{\mathbf{y}}_k\|^2 \\
&\text{subject to} \quad \check{\mathbf{y}}_{k+\frac{K}{4}} = -\check{\mathbf{y}}_k, \quad \check{\mathbf{y}}_{k+\frac{K}{2}} = j\check{\mathbf{y}}_k, \quad \check{\mathbf{y}}_{k+\frac{3K}{4}} = -j\check{\mathbf{y}}_k \\
&\qquad\qquad\quad k = 1, \dots, K/4.
\end{aligned} \tag{3.23}$$

The optimization problem (3.23) can also be rewritten as

$$\underset{\{\beta_{n,k}\},\{\check{\mathbf{y}}_k\}}{\text{minimize}} \quad J_2 \tag{3.24}$$

where

$$\begin{aligned}
J_2 = \sum_{n=1}^{T_\mathrm{b}} \sum_{k=1}^{\frac{K}{4}} \big(&\beta_{n,k} \|\mathbf{y}[n] - \check{\mathbf{y}}_k\|^2 + \beta_{n,k+\frac{K}{4}} \|\mathbf{y}[n] + \check{\mathbf{y}}_k\|^2 \\
&+ \beta_{n,k+\frac{K}{2}} \|\mathbf{y}[n] - j\check{\mathbf{y}}_k\|^2 + \beta_{n,k+\frac{3K}{4}} \|\mathbf{y}[n] + j\check{\mathbf{y}}_k\|^2\big)
\end{aligned} \tag{3.25}$$

Applying the same technique as in Condition 1 to this problem, we can find $\beta_{n,k}$ from (3.20) and

$$\check{\mathbf{y}}_k = \frac{\sum_n \big(\beta_{n,k} - \beta_{n,k+\frac{K}{4}} - j\beta_{n,k+\frac{K}{2}} + j\beta_{n,k+\frac{3K}{4}}\big)\mathbf{y}[n]}{\sum_n \big(\beta_{n,k} + \beta_{n,k+\frac{K}{4}} + \beta_{n,k+\frac{K}{2}} + \beta_{n,k+\frac{3K}{4}}\big)}, \quad k = 1, \dots, \frac{K}{4}. \tag{3.26}$$

Equation (3.26) also points out that the representative vector $\check{\mathbf{y}}_k$, with $k \leq K/4$, is found by using the received vectors that not only belong to cluster k but also to clusters $k + K/4$, $k + K/2$ and $k + 3K/4$.

The decoding algorithm for this semi-supervised learning method is presented in Algorithm 2. Coarse estimation of the representative vectors is first obtained by using the training sequence (line 2). Then clustering is applied on all of the received data vectors (line 5). Depending on whether Condition 1 or Condition 2 is satisfied, the representative vectors are updated (lines 7-8 or lines 11-12). The process of clustering the received data

Algorithm 2: Semi-supervised Learning Decoding.

```
1  Initialize $done = false$, $iter = 0$;
2  Find $\mathcal{Y}$ using the training sequence;
3  while $done = false$ do
4      $iter = iter + 1$;
5      Perform (3.20);
6      if Condition 1 holds then
7          Perform (3.22);
8          Set $\check{\mathbf{y}}_{k+\frac{K}{2}} = -\check{\mathbf{y}}_k$, with $k = 1, \ldots, K/2$;
9      end
10     if Condition 2 holds then
11         Perform (3.26);
12         Set $\check{\mathbf{y}}_{k+\frac{K}{4}} = -\check{\mathbf{y}}_k$, $\check{\mathbf{y}}_{k+\frac{K}{2}} = j\check{\mathbf{y}}_k$, $\check{\mathbf{y}}_{k+\frac{3K}{4}} = -j\check{\mathbf{y}}_k$, with $k = 1, \ldots, K/4$;
13     end
14     if convergent or $iter = iter_{\max}$ then
15         $done = true$;
16     end
17 end
```

vectors and updating the representative vectors is repeated until convergence or the number of iterations exceeds a maximum value (line 15). Convergence is achieved if the solutions $\{\beta_{n,k}\}$ are the same for two successive iterations. Convergence of Algorithm 2 is assured because after each iteration, the value of the objective function does not increase. However, the point of convergence is not guaranteed to be a global optimum.

3.4 Performance Analysis with One-bit ADCs

This section presents a performance analysis of the proposed methods for the case of 1-bit ADCs. The analysis is applicable for any blind detection scheme for MIMO receivers with low-resolution ADCs and for Rayleigh fading channels, independent of the channel realization. We assume that all symbol vectors in $\check{\mathcal{X}}$ are a priori equally likely to be transmitted. The objective is to characterize the VER. Since the performance of the proposed methods for 1-bit ADCs is independent of the step size Δ, we choose $\Delta = 2$ so that the quantization function becomes the sign$(\cdot)$ function, where sign$(a) = +1$ if $a \geq 0$ and sign$(a) = -1$ if

$a < 0$. If a is a complex number, then $\text{sign}(a) = \text{sign}(\Re\{a\}) + j\,\text{sign}(\Im\{a\})$. The operator $\text{sign}(\cdot)$ of a matrix or vector is applied separately to every element of that matrix or vector.

3.4.1 VER Analysis at Low SNRs

Here, an approximate pairwise VER at low SNRs for the Rayleigh fading channel is presented. First, using the Bussgang decomposition, the system model $\mathbf{y} = Q_b(\mathbf{r})$ can be rewritten as $\mathbf{y} = \mathbf{F}\mathbf{r} + \mathbf{e}$ [10] where $\mathbf{e}$ is the quantization distortion and

$$\mathbf{F} = \sqrt{\frac{2}{\pi}}\,\text{diag}(\boldsymbol{\Sigma}_r)^{-\frac{1}{2}}. \tag{3.27}$$

The term $\boldsymbol{\Sigma}_r = \mathbf{H}\mathbf{H}^H + N_0\mathbf{I}$ is the covariance matrix of $\mathbf{r}$. Let $\mathbf{A} = \mathbf{F}\mathbf{H}$ and $\mathbf{w} = \mathbf{F}\mathbf{z} + \mathbf{e}$, then the system model becomes

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}, \tag{3.28}$$

where $\mathbf{A} = \sqrt{2/\pi}\,\text{diag}(\boldsymbol{\Sigma}_r)^{-\frac{1}{2}}\mathbf{H}$ and the effective noise $\mathbf{w} = [w_1, w_2, \ldots, w_{N_\text{r}}]^T$ is modeled as Gaussian [10] with zero mean and covariance matrix

$$\boldsymbol{\Sigma}_w = \frac{2}{\pi}\Big[\arcsin\Big(\text{diag}(\boldsymbol{\Sigma}_r)^{-\frac{1}{2}}\boldsymbol{\Sigma}_r\,\text{diag}(\boldsymbol{\Sigma}_r)^{-\frac{1}{2}}\Big) - \text{diag}(\boldsymbol{\Sigma}_r)^{-\frac{1}{2}}\boldsymbol{\Sigma}_r\,\text{diag}(\boldsymbol{\Sigma}_r)^{-\frac{1}{2}} + N_0\,\text{diag}(\boldsymbol{\Sigma}_r)^{-1}\Big]. \tag{3.29}$$

Note that the operation $\arcsin(\cdot)$ of a matrix is applied element-wise on that matrix. The representative vector $\check{\mathbf{y}}_k$ now becomes $\check{\mathbf{y}}_k = \mathbf{A}\check{\mathbf{x}}_k$.

In the low SNR regime, the approximation $\boldsymbol{\Sigma}_r \approx \boldsymbol{\Sigma}_z$ holds [10], where $\boldsymbol{\Sigma}_z = N_0\mathbf{I}$ is the covariance matrix of $\mathbf{z}$. This approximation leads to $\mathbf{A} \approx \sqrt{2/(N_0\pi)}\mathbf{H}$ and $\boldsymbol{\Sigma}_w \approx \mathbf{I}$. Let $\boldsymbol{\upsilon} = [\upsilon_1, \ldots, \upsilon_{N_\text{r}}]^T = \check{\mathbf{y}}_k - \check{\mathbf{y}}_{k'}$, where $\upsilon_i = \sqrt{2/(N_0\pi)}\mathbf{h}_i^T(\check{\mathbf{x}}_k - \check{\mathbf{x}}_{k'})$ with $\mathbf{h}_i$ being the i^{th} column of $\mathbf{H}$. Since $\mathbf{H}$ is comprised of i.i.d. Gaussian random variables $\mathcal{CN}(0,1)$, υ_i is also Gaussian of zero mean with variance

$$\sigma_{kk'}^2 = \frac{2}{N_0\pi}\|\check{\mathbf{x}}_k - \check{\mathbf{x}}_{k'}\|_2^2. \tag{3.30}$$

Denote $P_{\check{\mathbf{x}}_k \to \check{\mathbf{x}}_{k'}}$ as the pairwise vector error probability of confusing $\check{\mathbf{x}}_k$ with $\check{\mathbf{x}}_{k'}$ when $\check{\mathbf{x}}_k$ is transmitted and when $\check{\mathbf{x}}_k$ and $\check{\mathbf{x}}_{k'}$ are the only two hypotheses [59]. The following proposition establishes the relationship between $P_{\check{\mathbf{x}}_k \to \check{\mathbf{x}}_{k'}}$ and $\sigma^2_{kk'}$.

Proposition 3.2. *$P_{\check{\mathbf{x}}_k \to \check{\mathbf{x}}_{k'}}$ at low SNR can be approximated as*

$$P_{\check{\mathbf{x}}_k \to \check{\mathbf{x}}_{k'}} \approx 1 - \Phi\Big(\sqrt{N_\mathrm{r}/(1 + 2/\sigma^2_{kk'})}\Big). \tag{3.31}$$

Proof. Please refer to Appendix A. □

The result in Proposition 3.2 clearly shows the dependency of the pairwise VER on the Euclidean distance between the two symbol vectors $\check{\mathbf{x}}_k$ and $\check{\mathbf{x}}_{k'}$. We now proceed to obtain an upper bound on the VER, denoted as P^{ver}_ρ, at low SNR assuming a priori equally likely $\check{\mathbf{x}}_1, \ldots, \check{\mathbf{x}}_K$. The VER is defined as

$$P^{\mathrm{ver}}_\rho = \sum_{k=1}^{K} \mathbb{P}[\hat{\mathbf{x}} \neq \check{\mathbf{x}}_k, \mathbf{x} = \check{\mathbf{x}}_k]$$

where $\hat{\mathbf{x}}$ is the detected symbol vector and $\mathbb{P}[\hat{\mathbf{x}} \neq \check{\mathbf{x}}_k, \mathbf{x} = \check{\mathbf{x}}_k]$ is the probability that $\check{\mathbf{x}}_k$ was transmitted but the detected symbol vector is not $\check{\mathbf{x}}_k$.

Proposition 3.3. *P^{ver}_ρ at low SNR is upper-bounded as*

$$P^{\mathrm{ver}}_\rho \leq \frac{1}{K} \sum_{k=1}^{K} \sum_{k' \neq k}^{K} \Big[1 - \Phi\Big(\sqrt{N_r/(1 + 2/\sigma^2_{kk'})}\Big)\Big]. \tag{3.32}$$

Proof. The bound on P^{ver}_ρ is obtained via the union bound

$$P^{\mathrm{ver}}_\rho = \sum_{k=1}^{K} \mathbb{P}[\hat{\mathbf{x}} \neq \check{\mathbf{x}}_k, \mathbf{x} = \check{\mathbf{x}}_k] = \frac{1}{K} \sum_{k=1}^{K} \mathbb{P}[\hat{\mathbf{x}} \neq \check{\mathbf{x}}_k \mid \mathbf{x} = \check{\mathbf{x}}_k] \leq \frac{1}{K} \sum_{k=1}^{K} \sum_{k' \neq k}^{K} P_{\check{\mathbf{x}}_k \to \check{\mathbf{x}}_{k'}}$$

and the application of Proposition 3.2. □

The probability $\mathbb{P}[\hat{\mathbf{x}} \neq \check{\mathbf{x}}_k \,|\, \mathbf{x} = \check{\mathbf{x}}_k]$ is invariant to $\check{\mathbf{x}}_k$ for the case of PSK modulation.

Without loss of generality, we assume that $\check{\mathbf{x}}_1$ was transmitted, so that the VER simplifies to

$$P_\rho^{\text{ver}} \leq \sum_{k\neq 1}^{K} \left[1 - \Phi\Big(\sqrt{N_{\text{r}}/(1+2/\sigma_{1k}^2)}\Big)\right]. \tag{3.33}$$

We note that this result is valid for low SNRs. In the following analysis, we characterize the VER at a very high SNR, i.e., $\rho \to \infty$.

3.4.2 VER Analysis as $\rho \to \infty$

Here, the VER as $\rho \to \infty$ is evaluated. Let $\mathbf{g}_k = [g_{k,1}, \ldots, g_{k,N_{\text{r}}}]^T = \mathbf{H}\check{\mathbf{x}}_k$, then

$$\mathbb{P}[\Re\{y_i\} = +1 \mid \mathbf{x} = \check{\mathbf{x}}_k] = \Phi(\sqrt{2\rho/N_{\text{t}}}\,\Re\{g_{k,i}\}), \tag{3.34}$$

$$\mathbb{P}[\Im\{y_i\} = +1 \mid \mathbf{x} = \check{\mathbf{x}}_k] = \Phi(\sqrt{2\rho/N_{\text{t}}}\,\Im\{g_{k,i}\}). \tag{3.35}$$

The true representative vectors are

$$\check{\mathbf{y}}_k = \mathbb{E}\big[\mathbf{y} \mid \mathbf{x} = \check{\mathbf{x}}_k\big] = 2\Phi(\sqrt{2\rho/N_{\text{t}}}\mathbf{g}_k) - (\mathbf{1} + j\mathbf{1}) \tag{3.36}$$

which becomes $\text{sign}(\mathbf{g}_k)$ as $\rho \to \infty$. It is possible for a given realization of $\mathbf{H}$ that more than one symbol vector will lead to the same representative vector: $\text{sign}(\mathbf{g}_k) = \text{sign}(\mathbf{g}_{k'})$ with $k \neq k'$, and in such cases a detection error will occur regardless of the detection scheme. In the following, we analyze the probability that $\text{sign}(\mathbf{g}_k) = \text{sign}(\mathbf{g}_{k'})$. The analysis is applicable for the cases of BPSK and QPSK modulation.

To facilitate the analysis, we convert the notation into the real domain as follows:

$$\check{\mathbf{x}}_k^{\Re} = [\check{x}_{k,1}^{\Re}, \check{x}_{k,2}^{\Re}, \ldots, \check{x}_{k,2N_{\text{t}}}^{\Re}]^T = [\Re\{\check{\mathbf{x}}_k\}^T, \Im\{\check{\mathbf{x}}_k\}^T]^T,$$
$$\mathbf{g}_k^{\Re} = [g_{k,1}^{\Re}, g_{k,2}^{\Re}, \ldots, g_{k,2N_{\text{r}}}^{\Re}]^T = [\Re\{\mathbf{g}_k\}^T, \Im\{\mathbf{g}_k\}^T]^T.$$

We first consider BPSK modulation, i.e., $\mathcal{M} = \{\pm 1\}$. In this case, $\Im\{\check{\mathbf{x}}_k\} = \mathbf{0}$.

Theorem 3.1. *Given $d = \|\check{\mathbf{x}}_k^{\Re} - \check{\mathbf{x}}_{k'}^{\Re}\|_0$ as the Hamming distance between the two labels, then*

$$\mathbb{P}\big[\mathrm{sign}(\mathbf{g}_k) = \mathrm{sign}(\mathbf{g}_{k'})\big] = \left[\frac{2}{\pi}\arctan\sqrt{\frac{N_{\mathrm{t}} - d}{d}}\right]^{2N_{\mathrm{r}}}. \tag{3.37}$$

Proof. Please refer to Appendix B. □

As $\rho \to \infty$, the effect of the AWGN can be ignored. Thus, $\mathbb{P}\big[\check{\mathbf{y}}_k = \check{\mathbf{y}}_{k'}\big] = \mathbb{P}\big[\mathrm{sign}(\mathbf{g}_k) = \mathrm{sign}(\mathbf{g}_{k'})\big]$. An upper bound on the VER is established in the following proposition.

Proposition 3.4. *With BPSK modulation, the asymptotic VER at high SNR is upper-bounded as*

$$P_{\rho\to\infty}^{\mathrm{ver}} \leq \frac{1}{2}\sum_{d=1}^{N_{\mathrm{t}}}\binom{N_{\mathrm{t}}}{d}\left[\frac{2}{\pi}\arctan\sqrt{\frac{N_{\mathrm{t}} - d}{d}}\right]^{2N_{\mathrm{r}}}. \tag{3.38}$$

Proof. Please refer to Appendix C. □

Proposition 3.5. *With BPSK modulation and $N_{\mathrm{t}} = 2$, the upper bound in* (3.38) *is tight.*

Proof. For BPSK modulation and $N_{\mathrm{t}} = 2$, let $\check{\mathbf{x}}_1^{\Re} = [1, 1, 0, 0]$, $\check{\mathbf{x}}_2^{\Re} = [1, -1, 0, 0]$, $\check{\mathbf{x}}_3^{\Re} = [-1, 1, 0, 0]$, $\check{\mathbf{x}}_4^{\Re} = [-1, -1, 0, 0]$. Herein, $\check{\mathbf{x}}_1^{\Re} = -\check{\mathbf{x}}_4^{\Re}$ and $\check{\mathbf{x}}_2^{\Re} = -\check{\mathbf{x}}_3^{\Re}$, resulting in $\check{\mathbf{y}}_1 = -\check{\mathbf{y}}_4$ and $\check{\mathbf{y}}_2 = -\check{\mathbf{y}}_3$ as $\rho \to \infty$. Hence, events $\check{\mathbf{y}}_1 = \check{\mathbf{y}}_2$ and $\check{\mathbf{y}}_1 = \check{\mathbf{y}}_3$ are mutually exclusive while event $\check{\mathbf{y}}_1 = \check{\mathbf{y}}_4$ does not exist. This proposition thus follows as a direct consequence of the proof for Proposition 3.4 given in Appendix C. □

For the case of QPSK modulation, the Hamming distance $d = \|\check{\mathbf{x}}_k^{\Re} - \check{\mathbf{x}}_{k'}^{\Re}\|_0$ between any two labels can be as large as $2N_{\mathrm{t}}$. Following the same derivation as in Theorem 3.1 and Proposition 3.4, an upper-bound for the asymptotic VER at high SNR can be established by the following proposition.

Proposition 3.6. *With QPSK modulation, the asymptotic VER at high SNR is upper-*

bounded as

$$P^{\mathrm{ver}}_{\rho\to\infty} \le \frac{1}{2}\sum_{d=1}^{2N_{\mathrm{t}}}\binom{2N_{\mathrm{t}}}{d}\left[\frac{2}{\pi}\arctan\sqrt{\frac{2N_{\mathrm{t}}-d}{d}}\right]^{2N_{\mathrm{r}}}. \tag{3.39}$$

3.4.3 Transmit Signal Design

Thus far it has been assumed that the transmitter uses all K possible labels for transmission. However, as K grows large, the training task for all the K labels becomes impractical, since the block fading interval T_{b} is finite. In this section, we consider a system where the transmitter employs only a subset of $\tilde{K}$ labels among the K possible labels for both the training and data transmission phases. The rest of the $K-\tilde{K}$ labels are unused. While using only $\tilde{K}$ labels reduces the transmission rate as compared to using all the K possible labels, the VER can be improved. In many 5G networks, e.g., Machine-to-Machine (M2M) communication systems, the priority is on the reliability, not the rate [3]. In addition, the reduction in training time with small $\tilde{K}$ may help improve the system throughput.

The design problem is how to choose $\tilde{K}$ labels among the K labels. To address this problem, we rely on Proposition 3.4 and Proposition 3.6. These propositions reveal that the VER at infinite SNR is inversely proportional to the Hamming distances between the labels. Thus, the following criterion for choosing the transmit signals is proposed:

$$\mathcal{X}^{\star} = \mathop{\arg\max}_{\mathcal{X}\subset\check{\mathcal{X}}^{\Re}} \min_{1\le k_1<k_2\le\tilde{K}} \|\mathbf{x}_{k_1}-\mathbf{x}_{k_2}\|_0, \tag{3.40}$$

where $\mathcal{X}=\{\mathbf{x}_1,\ldots,\mathbf{x}_{\tilde{K}}\}$ denote the set of $\tilde{K}$ different labels for transmission, and $\check{\mathcal{X}}^{\Re}=\{\check{\mathbf{x}}^{\Re}_1,\ldots,\check{\mathbf{x}}^{\Re}_K\}$. This design criterion aims to maximize the minimum pairwise Hamming distance among the $\tilde{K}$ labels. Note that the proposed criterion is also applicable for low SNRs because as shown in Proposition 3.3, the VER is inversely proportional to the Euclidean distance, which is analogous to the Hamming distance for BPSK and QPSK, albeit with some scaling factor. It should be noted that the proposed criterion does not rely on a

Algorithm 3: Transmit Signal Design.

```
1  Randomly generate $N_{\text{set}}$ initial sets $\{\mathcal{X}_i, i = 1, \ldots, N_{\text{set}}\}$;
2  for $i = 1 : N_{\text{set}}$ do
3      $done = false$;
4      while $done = false$ do
5          Let $flag = 1$;
6          Set $\mathcal{X}' = \check{\mathcal{X}} \backslash \mathcal{X}_i = \{\mathbf{x}'_1, \ldots, \mathbf{x}'_{K-\tilde{K}}\}$;
7          for $k_1 = 1 : \tilde{K}$ do
8              for $k_2 = 1 : K - \tilde{K}$ do
9                  Let $\hat{\mathcal{X}}_i = \big(\mathcal{X}_i \backslash \{\mathbf{x}_{k_1}\}\big) \cup \{\mathbf{x}'_{k_2}\}$;
10                 if $d_{\min}(\hat{\mathcal{X}}_i) > d_{\min}(\mathcal{X}_i)$ then
11                     Set $\mathcal{X}_i = \hat{\mathcal{X}}$ and $flag = 0$;
12                     Exit both for loops;
13                 end
14             end
15         end
16         if $flag = 1$ then
17             Set $done = true$ and $\mathcal{X}_i^\star = \mathcal{X}_i$;
18         end
19     end
20 end
21 $\mathcal{X}^\star = \arg\max_{\mathcal{X}_i^\star} d_{\min}(\mathcal{X}_i^\star)$;
```

specific channel realization, so the design task can be carried out off-line.

Problem (3.40) can be solved by exhaustive search when $\binom{K}{\tilde{K}}$ is not too large. When the exhaustive search is not possible, we propose a simple greedy algorithm, whose pseudocode can be found in Algorithm 3. Here, $d_{\min}(\mathcal{X})$ denotes the minimum pairwise Hamming distance among the labels in $\mathcal{X}$ and $\mathcal{X}'$ in line 6 denotes the set of labels, which is not used for transmission. The principle of Algorithm 3 is as follows:

- Generate N_{set} initial sets $\{\mathcal{X}_i\}_{i=1,\ldots,N_{\text{set}}}$, where each set $\mathcal{X}_i$ contains $\tilde{K}$ different labels randomly chosen from $\check{\mathcal{X}}^{\Re}$.

- For each initial set $\mathcal{X}_i$, find $\mathbf{x}' \in \mathcal{X}'$ such that when an element of $\mathcal{X}_i$ is replaced by $\mathbf{x}'$, the value of the objective function in (3.40), i.e., the minimum Hamming distance, is increased. This is repeated until no further increase in the objective function is possible after evaluating all replacements.

- Each initial set $\mathcal{X}_i$ produces a corresponding solution $\mathcal{X}_i^\star$ as in line 17. The solution $\mathcal{X}^\star$ of (3.40) is obtained by selecting the $\mathcal{X}_i^\star$ whose objective function value is largest (line 21).

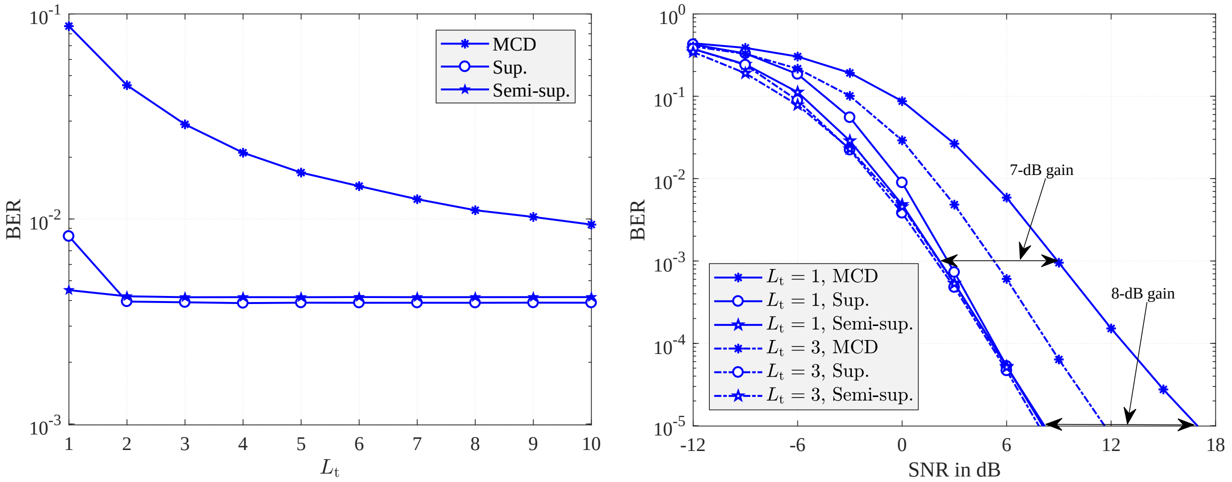
Note that the larger N_{set} is, the more likely Algorithm 3 will find the optimal solution.

3.5 Simulations and Results

3.5.1 Numerical Evaluation of the Proposed Methods

Monte Carlo simulations are used to numerically evaluate the performance of the proposed methods. The simulation settings are as follows. The number of transmit antennas N_{t} is set to be 2 unless otherwise stated. The data phase contains $T_{\text{d}} = 500$ time slots. In the supervised learning method, a 24-bit CRC as in the 3GPP Long Term Evolution (LTE) standard [60] is adopted. The generator of the CRC in the simulation is $z^{24}+z^{23}+z^{14}+z^{12}+z^{8}+1$, and the length of each data segment is 16 bits. Thus, the length of each coded segment is 40 bits. This is the minimum length in the 3GPP LTE standard. In all figures, 'Sup.' and 'Semi-sup.' stand for the supervised learning and semi-supervised learning methods, respectively.

The effect of the training sequence length L_{t} on MCD and the two proposed methods is first studied (Figure 3.3). BPSK modulation with $N_{\text{r}} = 16$ and 1-bit ADCs are used. Figure 3.3a shows the change of the BER as L_{t} varies. An interesting observation is that the performance of the proposed methods is much less affected by L_{t} than the MCD method. Hence, the length of the training sequence can be reduced without causing much degradation on the performances of the proposed methods. This is illustrated more clearly in Figure 3.3b, where we carry out the simulation for $L_{\text{t}} = 1$ and $L_{\text{t}} = 3$, still with BPSK modulation, 1-bit ADCs and $N_{\text{r}} = 16$. It can be seen from Figure 3.3b that, as L_{t} is reduced from 3 to 1, the BER of MCD is significantly degraded while the BERs of the proposed methods experience only a small degradation at low SNRs and do not change at higher SNRs. This leads to a



(a) L_t varies and $\rho = 0$ dB.

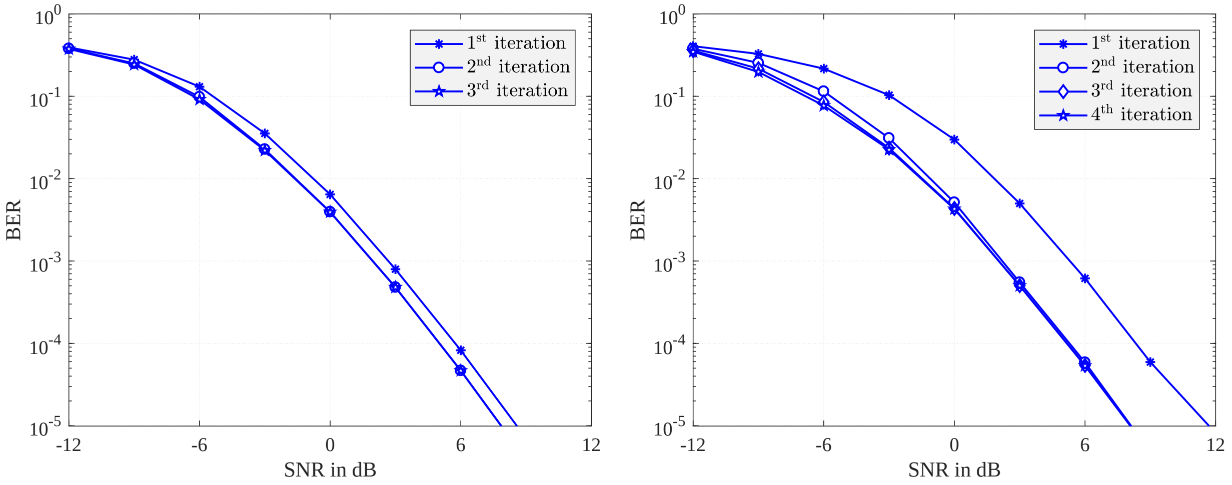
(b) $L_t = 1$ and $L_t = 3$, ρ varies.

Figure 3.3: Effect of L_t on MCD and the proposed methods with 1-bit ADCs, $N_r = 16$ and BPSK modulation.

significant improvement for the proposed methods as compared to MCD, for example, about a 7-dB gain at a BER of 10^{-3} and 8-dB at a BER of 10^{-5} when $L_t = 1$. Even for moderately long training sequences, e.g., $L_t = 3$, the gain of the proposed methods is still considerable, from 3-dB to 4-dB.

The results in Figure 3.3 can be explained as follows. The performance of MCD is susceptible to L_t because its detection accuracy relies on the representative vectors estimated only from the training sequence. Therefore, if L_t is small, the representative vectors are not estimated correctly and so the performance can be degraded significantly. On the other hand, the proposed methods are much less dependent on L_t because they use the training sequence only as the initial guide for the detection task. Compared to the semi-supervised learning method, the supervised learning method is slightly more dependent on L_t because it depends on detection results from the training sequence.

Since the proposed methods work iteratively, numerous simulations are performed to evaluate the improvement in BER over the iterations. Simulation results are shown in Figure 3.4. For the supervised learning method, Figure 3.4a, it can be seen that the BER converges after only 2 iterations. For the semi-supervised learning method, Figure 3.4b, there is considerable improvement between the first and the second iterations, but then the



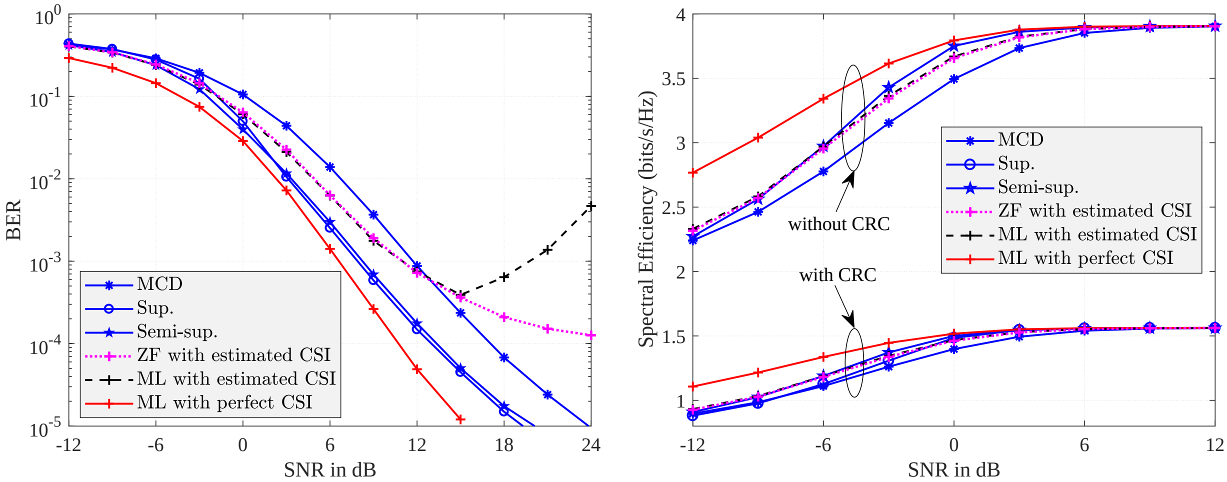
(a) Supervised learning method.

(b) Semi-supervised learning method.

Figure 3.4: Performance improvement for different iterations with 1-bit ADCs, BPSK modulation, $N_\mathrm{r} = 16$ and $L_\mathrm{t} = 3$.

third and the fourth iterations give approximately the same performance. It is therefore preferred to limit the maximum number of iterations to 3 in the semi-supervised learning method. It should be noted that the BER on the first iteration of the semi-supervised learning method is actually the BER of the MCD method because the first iteration only exploits the training sequence.

Figure 3.5 compares the aforementioned blind detection methods with several coherent detection methods. The simulation uses 1-bit ADCs, QPSK modulation, $N_\mathrm{r} = 16$ and $L_\mathrm{t} = 3$. For coherent detection, CSI is first estimated by the Bussgang Linear Minimum Mean Squared Error (BLMMSE) method proposed in [21]. The length of the training sequence in the blind detection methods is 12, so we also set the length of the pilot sequence for the channel estimation to 12. The ZF detection method is presented in [21]. The ML method for 1-bit ADCs is provided in [1]. A performance comparison in terms of BER is given in Figure 3.5a, which shows that the proposed methods outperform the ZF and ML methods with estimated CSI. It is also seen that the BER of the proposed methods is quite close the BER of ML detection with perfect CSI. Here, it is observed that a significant increase in the BER at high SNRs for the ML method with estimated CSI. This observation was also



(a) BER

(b) Spectral efficiency

Figure 3.5: Performance comparison between blind and coherent detection with 1-bit ADCs, QPSK modulation, $N_\mathrm{r} = 16$ and $L_\mathrm{t} = 3$.

reported in [54]. In comparing the two proposed methods in Figure 3.5a and Figure 3.3, should the CRC be available, it is more beneficial to use the supervised learning method for better BER performance.

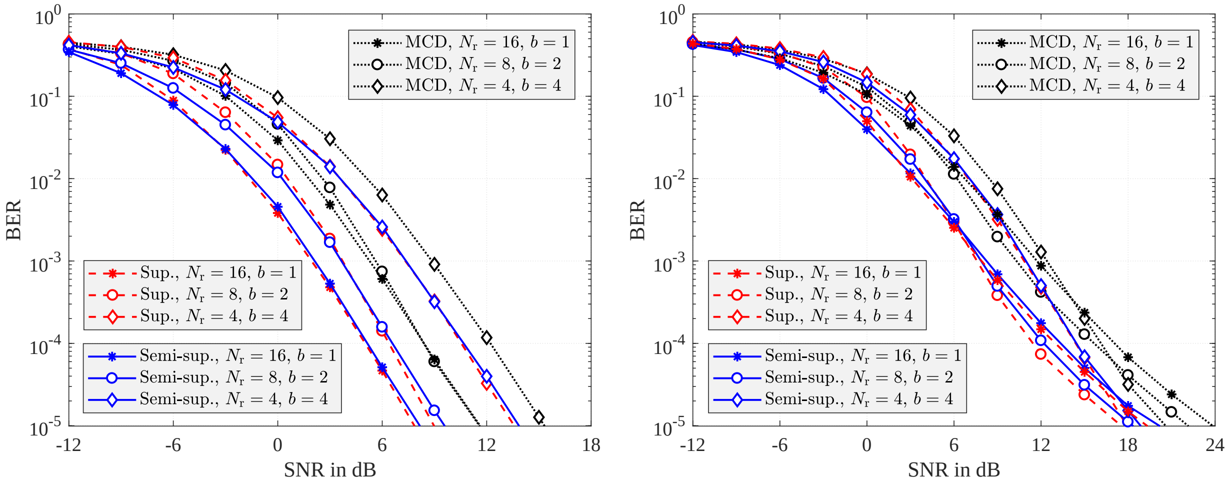
Figure 3.5b provides a comparison in terms of spectral efficiency η, defined as the average number of information bits received correctly per block-fading interval T_b. We determine η for the case without CRC as

$$\eta = \frac{T_\mathrm{d}}{T_\mathrm{b}} \times (1 - \mathrm{BER}) \times N_\mathrm{t} \times \log_2 M$$

and for the case with CRC as

$$\eta = \frac{L_\mathrm{data}}{L_\mathrm{data} + L_\mathrm{CRC}} \times \frac{T_\mathrm{d}}{T_\mathrm{b}} \times (1 - \mathrm{BER}) \times N_\mathrm{t} \times \log_2 M.$$

Figure 3.5b indicates a proportional drop in the spectral efficiency due to the use of CRC. Note that the supervised learning method can only be applied in systems where the CRC is available but the other methods can be used in any system regardless of the CRC. Thus, should the CRC be eliminated for improved spectral efficiency, the semi-supervised method



(a) BPSK modulation.

(b) QPSK modulation.

Figure 3.6: Performance of the proposed methods for different numbers of receive antennas N_r and ADC resolutions b with $L_\mathrm{t} = 3$.

provides better performance than MCD. It also performs slightly better than conventional coherent detection with estimated CSI. The small performance gap observed in Figure 3.5b is due to the small difference in BER performance in the SNR region between -12 to 12 dB, as shown in Figure 3.5a. At high SNR, while the proposed method performs much better than other methods in terms of BER, its effect on the throughput η is negligible since $1 - \mathrm{BER} \approx 1$.

To study the trade-off between N_r and b, the proposed methods are evaluated in three different scenarios: (i) $N_\mathrm{r} = 4, b = 4$; (ii) $N_\mathrm{r} = 8, b = 2$; and (iii) $N_\mathrm{r} = 16, b = 1$. This is to ensure the same number of bits after the ADCs for baseband processing. The number of label repetitions L_t is set to be 3. The simulation results are shown in Figure 3.6, with BPSK in Figure 3.6a and QPSK in Figure 3.6b. For BPSK modulation, the best performance is achieved by scenario (iii) for all methods. Hence, this suggests the use of more receive antennas and fewer bits in the ADCs when BPSK modulation is employed. However, for QPSK modulation, there is a trade-off between scenarios (ii) and (iii). For low SNRs, the setting $N_\mathrm{r} = 16$ and $b = 1$ gives better performance, but for high SNRs, the best results are with $N_\mathrm{r} = 8$ and $b = 2$. The results in Figure 3.6 also show that the proposed methods

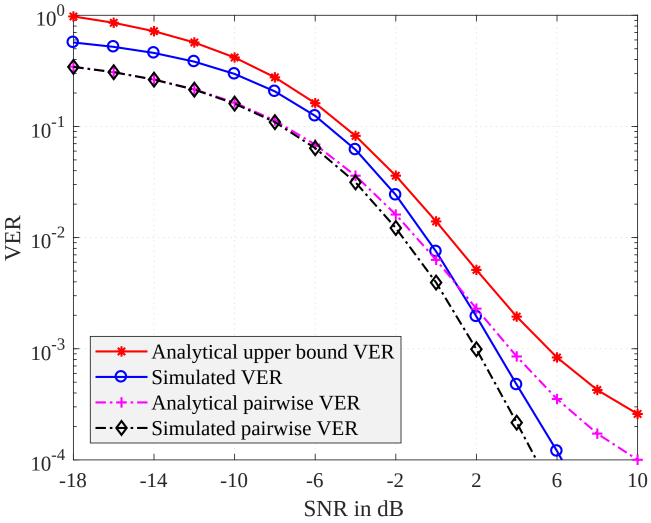


Figure 3.7: Validation of the analytical pairwise VER in (3.31) and the analytical VER in (3.32) at low SNRs with $N_{\mathrm{t}} = 2$, $N_{\mathrm{r}} = 16$, and BPSK modulation.

outperform the MCD method in all three scenarios.

3.5.2 Validation of Performance Analysis

This section presents a validation on the performance analyses in Section 3.4. Figure 3.7 provides the analytical approximate pairwise VER in (3.31) and the VER in (3.32). the setting of $N_{\mathrm{t}} = 2$, $N_{\mathrm{r}} = 16$, and BPSK modulation is used. The two labels used to examine the pairwise VER are $\check{\mathbf{x}}_k = [+1, +1]^T$ and $\check{\mathbf{x}}_{k'} = [+1, -1]^T$. It can be seen that our approximate pairwise VER is very close to the simulated pairwise VER at low SNRs, typically with SNRs less than 0-dB. However, as the SNR increases, the approximate pairwise VER tends to diverge from the true pairwise VER because the approximation $\boldsymbol{\Sigma}_r \approx \boldsymbol{\Sigma}_z$ is inapplicable for high SNRs. The simulation results also show that the analytical VER is quite close to the true VER at low SNRs.

Validation of the high SNR expressions for the analytical VER is given in Figure 3.8 with $N_{\mathrm{r}} = 8$. The horizontal lines represent the analytical upper bounds on the VER at infinite SNR. For the case of BPSK and $N_{\mathrm{t}} = 2$, it can be seen that the simulated VER approaches the horizontal solid line as the SNR increases and then they match at very high SNRs. This

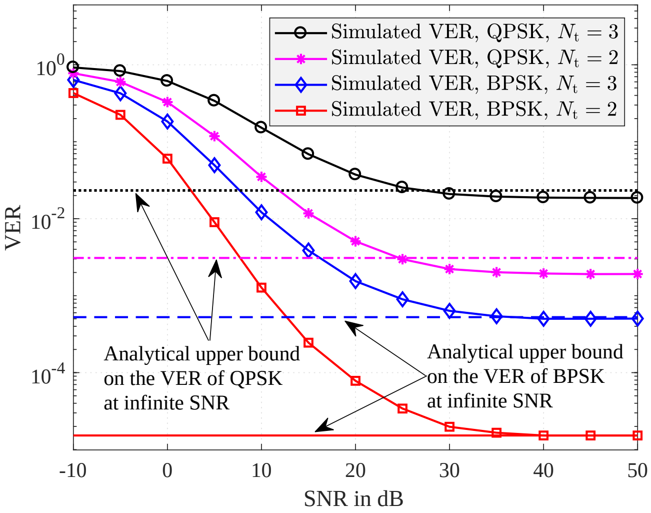


Figure 3.8: Validation of the analytical VER at infinite SNR in Propositions 3.4, 3.5, and 3.6.

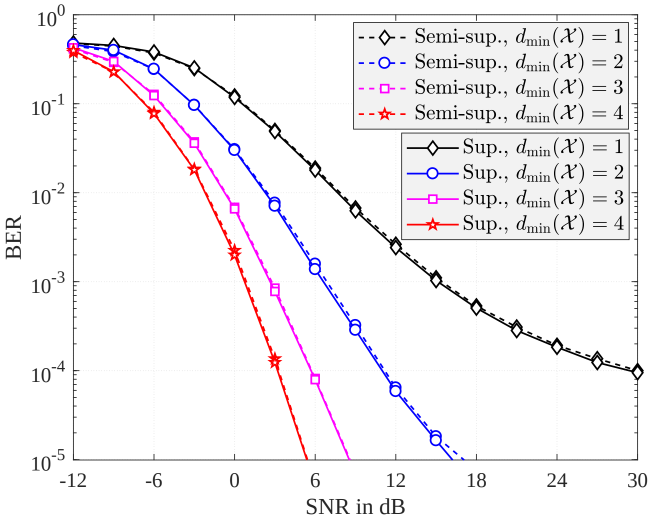


Figure 3.9: Validation of the transmit signal design with $N_\mathrm{t}=6$, $N_\mathrm{r}=16$, $\tilde{K}=4$, and BPSK modulation.

validates the result of Proposition 3.5 indicating that the bound is tight in the case of BPSK and $N_{\mathrm{t}} = 2$. With BPSK and $N_{\mathrm{t}} = 3$, the horizontal dashed line is just slightly higher than the floor of the simulated VER. For QPSK modulation, there is a small gap between the horizontal lines and the floors of the simulated VER. These observations validate the analytical upper-bound results in Proposition 3.4 and Proposition 3.6.

Figure 3.9 provides a validation for the proposed transmit signal design based on the minimum Hamming distance in Section 3.4.3. With different selections of the label sets $\mathcal{X}$, the BER performance in Figure 3.9 improves as $d_{\min}(\mathcal{X})$ increases, which validates the analysis. In this particular simulation scenario ($N_{\mathrm{t}} = 6$, $N_{\mathrm{r}} = 16$, $\tilde{K} = 4$, and BPSK modulation), the minimum Hamming distance of an optimal set can be found to be 4. The proposed Algorithm 3 then helps select an optimal set $\mathcal{X}^{\star}$ with $d_{\min}(\mathcal{X}^{\star}) = 4$. Hence, the curves with star markers in Figure 3.9 also represent the BER obtained by $\mathcal{X}^{\star}$.

As $\tilde{K}$ is increased, the data rate also increases, but the BER will degrade. Thus, there is a specific value for $\tilde{K}$ that provides the best compromise for the spectral efficiency. Figure 3.10 illustrates the change of spectral efficiency with respect to $\tilde{K}$ at different SNR values. The simulations are carried out with $N_{\mathrm{t}} = 8$, $N_{\mathrm{r}} = 16$, QPSK modulation, $L_{\mathrm{t}} = 3$, and $\tilde{K} \in \{4, 8, 16, 32, 64, 128\}$. The maximum number of time slots for the block-fading interval is $T_{\mathrm{b}} = 500$. The availability of the CRC is assumed so that the supervised learning method can be compared with other methods. The lengths of the data segment for $\tilde{K} \in \{4, 8, 64, 128\}$ and $\tilde{K} \in \{16, 32\}$ are 18 bits and 16 bits, respectively. This is to ensure that the number of bits in a segment is a multiple of the number bits in a transmitted vector. The length of the data block T_{d} is also set to be a multiple of $(L_{\mathrm{CRC}} + L_{\mathrm{data}})/\log_2 \tilde{K}$. The spectral efficiency is then computed as

$$\eta = \frac{L_{\mathrm{data}}}{L_{\mathrm{CRC}} + L_{\mathrm{data}}} \times \frac{T_{\mathrm{d}}}{T_{\mathrm{d}} + T_{\mathrm{t}}} \times (1 - \mathrm{BER}) \times \log_2 \tilde{K}.$$

For each value of $\tilde{K}$, Algorithm 3 is applied to find the solution $\mathcal{X}^{*}$ of (3.40). It is found

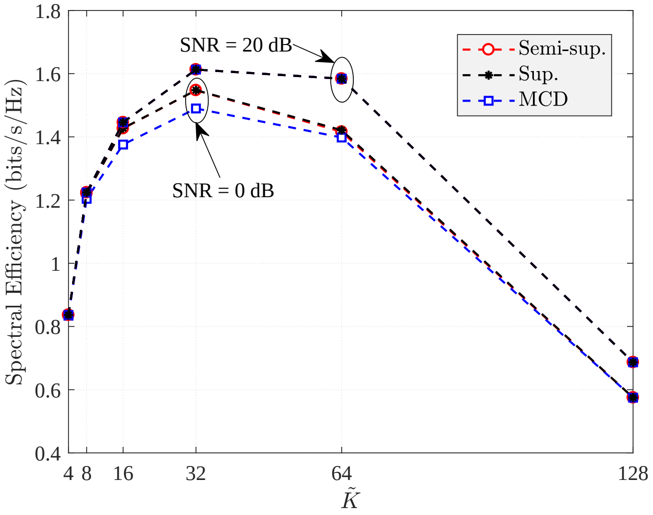


Figure 3.10: Spectral efficiency versus $\tilde{K}$ with $N_{\mathrm{t}} = 8$, $N_{\mathrm{r}} = 16$, QPSK modulation, $L_{\mathrm{t}} = 3$, and $T_{\mathrm{b}} = 500$.

that the symbol vectors of $\mathcal{X}^*$ do not satisfy Condition 2, and so the full-space training method is used. The simulation results in Figure 3.10 show that increasing $\tilde{K}$ does not necessarily improve the spectral efficiency, due to the increased training overhead. There is thus an optimal value of $\tilde{K} = 32$ in this scenario. It is also seen that at low SNR the spectral efficiencies of the proposed methods are higher than that of MCD.

Chapter 4

SVM-based channel estimation and data detection for massive MIMO systems with one-bit ADCs

This chapter presents an SVM-based approach for channel estimation and data detection in massive MIMO systems with 1-bit ADCs. The system model is first presented in Section 4.1. SVM-based methods for flat-fading channels are then proposed in Section 4.2. Section 4.3 includes an extension of the proposed methods to OFDM sysems with frequency-selective fading channels. Finally, numerical results are provided in Section 4.4.

4.1 System Model

The considered massive MIMO system is illustrated in Figure 4.1 with U single-antenna users and an N-antenna base station, where it is assumed that $N \geq U$. Let $\bar{\mathbf{x}} = [\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_U]^T \in \mathbb{C}^U$ denote the transmitted signal vector, where $\bar{x}_u$ is the signal transmitted from the u^{th} user under the power constraint $\mathbb{E}[|\bar{x}_u|^2] = 1$, $u \in \mathcal{U} = \{1, 2, \ldots, U\}$. Let $\bar{\mathbf{H}} \in \mathbb{C}^{N \times U}$ denote the channel, which for the moment is assumed to be block flat fading. Let $\bar{\mathbf{r}} = [\bar{r}_1, \bar{r}_2, \ldots, \bar{r}_N]^T \in$

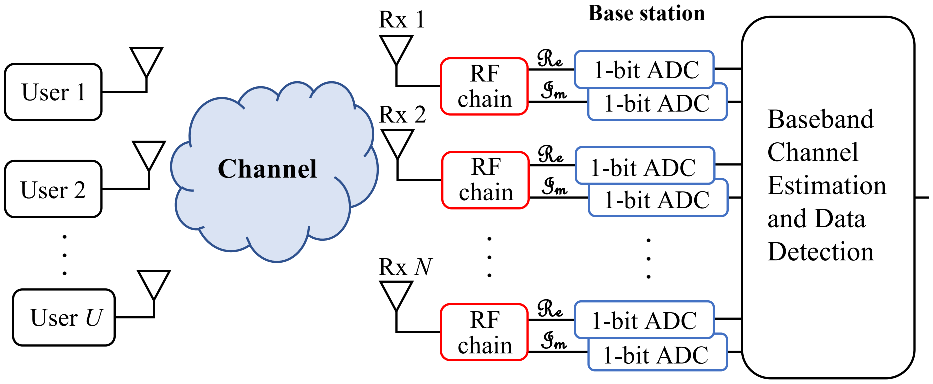


Figure 4.1: Block diagram of a massive MIMO system with U single-antenna users and an N-antenna base station equipped with $2N$ 1-bit ADCs.

$\mathbb{C}^N$ be the unquantized received signal vector at the base station, which is given as

$$\bar{\mathbf{r}} = \bar{\mathbf{H}}\bar{\mathbf{x}} + \bar{\mathbf{z}}, \tag{4.1}$$

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

$$\bar{\mathbf{y}} = \text{sign}(\bar{\mathbf{r}}) = \text{sign}\left(\Re\{\bar{\mathbf{r}}\}\right) + j\,\text{sign}\left(\Im\{\bar{\mathbf{r}}\}\right) \tag{4.2}$$

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

4.2 Proposed SVM-based Channel Estimation and Data Detection with 1-bit ADCs

4.2.1 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 $f(\mathbf{x}) = \boldsymbol{\omega}^T\mathbf{x} + \delta$ 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, $\boldsymbol{\omega}$ and δ are referred to as the weight vector and the bias, respectively. In other words, the hyperplane $f(\mathbf{x}) = \boldsymbol{\omega}^T\mathbf{x} + \delta = 0$ 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

$$m_{\mathcal{D}}(f) = \frac{2}{\|\boldsymbol{\omega}\|}. \tag{4.3}$$

The SVM technique seeks to find $\boldsymbol{\omega}$ and δ such that the margin $m_{\mathcal{D}}(f)$ is maximized. The optimization problem can be expressed as [58]

$$\begin{aligned} \underset{\{\boldsymbol{\omega},\delta\}}{\text{minimize}} \quad & \frac{1}{2}\|\boldsymbol{\omega}\|^2 \\ \text{subject to} \quad & y_q(\boldsymbol{\omega}^T\mathbf{x}_q + \delta) \geq 1, \quad q = 1, 2, \ldots, P. \end{aligned} \tag{4.4}$$

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

$$\begin{aligned} \underset{\{\boldsymbol{\omega},\delta,\xi_q\}}{\text{minimize}} \quad & \frac{1}{2}\|\boldsymbol{\omega}\|^2 + C\sum_{q=1}^{P}\xi_q \\ \text{subject to} \quad & y_q(\boldsymbol{\omega}^T\mathbf{x}_q + \delta) \geq 1 - \xi_q, \\ & \xi_q \geq 0, \quad q = 1, 2, \ldots, P. \end{aligned} \tag{4.5}$$

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" [58]. The optimization problems (4.4) and (4.5) can be solved by very efficient algorithms [61,62]. For example, if the weight vector is sparse, the complexity of the algorithm in [61] scales linearly in both the number of features (size of the weight vector $\boldsymbol{\omega}$) and the number of training samples $|\mathcal{D}|$. For arbitrary weight vectors, the complexity of the algorithm in [62] scales linearly in the number of training samples and quadratically in the number of features for the worst case. A good review of efficient methods for solving (4.4) and (4.5) can also be found in [63].

4.2.2 Proposed SVM-based Channel Estimation

Uncorrelated Channels

First, uncorrelated channels are considered. The channel elements are assumed to be i.i.d. as $\mathcal{CN}(0,1)$. In order to estimate the channel, a pilot sequence $\bar{\mathbf{X}}_{\mathrm{t}} \in \mathbb{C}^{U \times T_{\mathrm{t}}}$ of length T_{t} is used to generate the training data

$$\bar{\mathbf{Y}}_{\mathrm{t}} = \operatorname{sign}\left(\bar{\mathbf{H}}\bar{\mathbf{X}}_{\mathrm{t}} + \bar{\mathbf{Z}}_{\mathrm{t}}\right). \tag{4.6}$$

For convenience in later derivations, we convert the notation in (4.6) to the real domain as

$$\mathbf{Y}_{\mathrm{t}} = \operatorname{sign}\left(\mathbf{H}_{\mathrm{t}}\mathbf{X}_{\mathrm{t}} + \mathbf{Z}_{\mathrm{t}}\right), \tag{4.7}$$

where

$$\mathbf{Y}_{\mathrm{t}} = \left[\Re\{\bar{\mathbf{Y}}_{\mathrm{t}}\}, \Im\{\bar{\mathbf{Y}}_{\mathrm{t}}\}\right] = \left[\mathbf{y}_{\mathrm{t},1}, \mathbf{y}_{\mathrm{t},2}, \ldots, \mathbf{y}_{\mathrm{t},N}\right]^T, \tag{4.8}$$

$$\mathbf{H}_{\mathrm{t}} = \left[\Re\{\bar{\mathbf{H}}\}, \Im\{\bar{\mathbf{H}}\}\right] = \left[\mathbf{h}_{\mathrm{t},1}, \mathbf{h}_{\mathrm{t},2}, \ldots, \mathbf{h}_{\mathrm{t},N}\right]^T, \tag{4.9}$$

$$\mathbf{Z}_{\mathrm{t}} = \left[\Re\{\bar{\mathbf{Z}}_{\mathrm{t}}\}, \Im\{\bar{\mathbf{Z}}_{\mathrm{t}}\}\right] = \left[\mathbf{z}_{\mathrm{t},1}, \mathbf{z}_{\mathrm{t},2}, \ldots, \mathbf{z}_{\mathrm{t},N}\right]^T, \tag{4.10}$$

and

$$\mathbf{X}_{\mathrm{t}} = \begin{bmatrix} \Re\{\bar{\mathbf{X}}_{\mathrm{t}}\} & \Im\{\bar{\mathbf{X}}_{\mathrm{t}}\} \\ -\Im\{\bar{\mathbf{X}}_{\mathrm{t}}\} & \Re\{\bar{\mathbf{X}}_{\mathrm{t}}\} \end{bmatrix} = [\mathbf{x}_{\mathrm{t},1}, \mathbf{x}_{\mathrm{t},2}, \ldots, \mathbf{x}_{\mathrm{t},2T_{\mathrm{t}}}]. \tag{4.11}$$

Note that $\mathbf{y}_{\mathrm{t},i}^T \in \{\pm 1\}^{1\times 2T_{\mathrm{t}}}$, $\mathbf{h}_{\mathrm{t},i}^T \in \mathbb{R}^{1\times 2U}$, and $\mathbf{z}_{\mathrm{t},i}^T \in \mathbb{R}^{1\times 2T_{\mathrm{t}}}$ with $i \in \{1, 2, \ldots, N\}$ represent the i^{th} rows of $\mathbf{Y}_{\mathrm{t}}$, $\mathbf{H}_{\mathrm{t}}$, and $\mathbf{Z}_{\mathrm{t}}$, respectively. However, $\mathbf{x}_{\mathrm{t},n} \in \mathbb{R}^{2U\times 1}$ with $n \in \{1, 2, \ldots, 2T_{\mathrm{t}}\}$ is the n^{th} column of $\mathbf{X}_{\mathrm{t}}$.

It can be seen from (4.9) that estimating $\{\mathbf{h}_{\mathrm{t},i}\}_{i=1,2,\ldots,N}$ is equivalent to estimating $\bar{\mathbf{H}}$. Here, the channel estimation problem is formulated in terms of $\mathbf{h}_{\mathrm{t},i}$. Let

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

then we have

$$y_{\mathrm{t},i,n} = \operatorname{sign}\left(\mathbf{h}_{\mathrm{t},i}^T \mathbf{x}_{\mathrm{t},n} + z_{\mathrm{t},i,n}\right). \tag{4.12}$$

It is stressed that the estimation of $\mathbf{h}_{\mathrm{t},i}$ in (4.12) can be interpreted as an SVM binary classification problem. More specifically, $\{\mathbf{x}_{\mathrm{t},n}, y_{\mathrm{t},i,n}\}_{n=1,\ldots,2T_{\mathrm{t}}}$ plays the role of the training data set $\mathcal{D}$. The channel $\mathbf{h}_{\mathrm{t},i}$ acts as the weight vector and $z_{\mathrm{t},i,n}$ can be viewed as the bias. Hence, the SVM classification formulation can be exploited to estimate $\mathbf{h}_{\mathrm{t},i}$ by solving the following optimization problem:

$$\begin{aligned} \underset{\{\mathbf{h}_{\mathrm{t},i},\xi_n\}}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{h}_{\mathrm{t},i}\|^2 + C\sum_{n=1}^{2T_{\mathrm{t}}} \xi_n \\ \text{subject to} \quad & y_{\mathrm{t},i,n}\mathbf{h}_{\mathrm{t},i}^T \mathbf{x}_{\mathrm{t},n} \geq 1 - \xi_n, \\ & \xi_n \geq 0, \quad n = 1, 2, \ldots, 2T_{\mathrm{t}}. \end{aligned} \tag{4.13}$$

Here, the bias is discarded because the $\{z_{\mathrm{t},i,n}\}$ are random noise with zero mean. In addition, at infinite SNR, (4.12) becomes $y_{\mathrm{t},i,n} = \operatorname{sign}\left(\mathbf{h}_{\mathrm{t},i}^T \mathbf{x}_{\mathrm{t},n}\right)$, which has no bias. It should be noted that (4.13) only depends on a single index i, and so its solution is the estimate for the i^{th} row of the channel matrix $\bar{\mathbf{H}}$, i.e., the channel vector from the U users to the ith receive antenna. This means we have N separate optimization problems of the same form (4.13), which is an advantage of the proposed SVM-based method since these N optimization problems can be solved in parallel.

Let $\tilde{\mathbf{h}}_{\mathrm{t},i}$ denote the solution of (4.13). This solution provides an estimate of the channel "direction", but the magnitude of $\tilde{\mathbf{h}}_{\mathrm{t},i}$ is determined by the definition of the SVM margin, which in turn defines the inequality constraints in (4.13). In fact, the instantaneous magnitude of $\mathbf{h}_{\mathrm{t},i}$ is not identifiable [38] since $a\mathbf{h}_{\mathrm{t},i}$ for any $a > 0$ will produce the same data set $\{y_{t,i,n}\}$:

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

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

$$\hat{\mathbf{h}}_{\mathrm{t},i} = \frac{\sqrt{U}\tilde{\mathbf{h}}_{\mathrm{t},i}}{\|\tilde{\mathbf{h}}_{\mathrm{t},i}\|}. \tag{4.14}$$

This rescaling choice is found to provide the best estimation accuracy.

Remark 1: The soft-SVM method in [26] does not maximize the margin, but instead calculates $\mathbf{h}_{\mathrm{t},i}$ such that the condition $y_{\mathrm{t},i,n}\mathbf{h}_{\mathrm{t},i}^T \mathbf{x}_{\mathrm{t},n} > 0$ is satisfied for as many n as possible. However, due to the noise component $z_{\mathrm{t},i,n}$, the condition $y_{\mathrm{t},i,n}\mathbf{h}_{\mathrm{t},i}^T \mathbf{x}_{\mathrm{t},n} > 0$ may not be satisfied even with the true channel vector $\mathbf{h}_{\mathrm{t},i}$. The 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 $y_{\mathrm{t},i,n}\mathbf{h}_{\mathrm{t},i}^T \mathbf{x}_{\mathrm{t},n} > 0$.

Remark 2: Without slack variables, the problem in (4.13)

$$\begin{aligned} \underset{\{\mathbf{h}_{\mathrm{t},i}\}}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{h}_{\mathrm{t},i}\|^2 \\ \text{subject to} \quad & y_{\mathrm{t},i,n}\mathbf{h}_{\mathrm{t},i}^T \mathbf{x}_{\mathrm{t},n} \geq 1, \quad n = 1, 2, \ldots, 2T_{\mathrm{t}}, \end{aligned} \tag{4.15}$$

is similar to the form in (4.4). For $\mathbf{h}_{\mathrm{t},i} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ we have

$$p(\mathbf{h}_{\mathrm{t},i}) = \frac{1}{\sqrt{(2\pi)^{2U}}} \exp\left\{-\frac{1}{2}\|\mathbf{h}_{\mathrm{t},i}\|^2\right\}, \tag{4.16}$$

and hence the optimization problem in (4.15) can be read as maximizing the pdf of $\mathbf{h}_{\mathrm{t},i}$ subject to the constraints $y_{\mathrm{t},i,n}\mathbf{h}_{\mathrm{t},i}^T\mathbf{x}_{\mathrm{t},n} \geq 1$ for $n = 1, 2, \ldots, 2T_{\mathrm{t}}$. Thus, the SVM approach can be interpreted as finding the channel $\mathbf{h}_{\mathrm{t},i}$ that attains the highest likelihood under the constraints realized by the measured data. This observation will be used next to modify the SVM-based channel estimator when the channel is spatially correlated. Note that the work in [26] only considers uncorrelated channels.

Spatially Correlated Channels

Let $\bar{\mathbf{H}} = [\bar{\mathbf{h}}_1, \ldots, \bar{\mathbf{h}}_U]$, and so $\bar{\mathbf{h}}_u \in \mathbb{C}^{N\times 1}$ is the u^{th} column of $\bar{\mathbf{H}}$. Here, it is assumed that the elements of $\bar{\mathbf{h}}_u$ are correlated, or in other words that the channels associated with different antennas are correlated. Let $\bar{\mathbf{h}}_u \sim \mathcal{CN}(\mathbf{0}, \bar{\mathbf{C}}_u)$ and $\bar{\mathbf{h}} = \mathrm{vec}(\bar{\mathbf{H}})$, then $\bar{\mathbf{h}} \sim \mathcal{CN}(\mathbf{0}, \bar{\mathbf{C}})$ where $\bar{\mathbf{C}} = \mathrm{blockdiag}(\bar{\mathbf{C}}_1, \bar{\mathbf{C}}_2, \ldots, \bar{\mathbf{C}}_U)$. The pdf of $\bar{\mathbf{h}}$ is

$$\begin{aligned} p(\bar{\mathbf{h}}) &= \frac{1}{\pi^{UN}\sqrt{\det(\bar{\mathbf{C}})}} \exp\left\{-\bar{\mathbf{h}}^H\bar{\mathbf{C}}^{-1}\bar{\mathbf{h}}\right\} && (4.17)\\ &= \frac{1}{\pi^{UN}\sqrt{\det(\bar{\mathbf{C}})}} \exp\left\{-\sum_{u=1}^{U}\bar{\mathbf{h}}_u^H\bar{\mathbf{C}}_u^{-1}\bar{\mathbf{h}}_u\right\}. && (4.18)\end{aligned}$$

The exponent term in (4.17) becomes a sum in (4.18) because $\bar{\mathbf{C}}$ is a block diagonal matrix, whose main-diagonal blocks are $\bar{\mathbf{C}}_1, \bar{\mathbf{C}}_2, \ldots, \bar{\mathbf{C}}_U$. Letting

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

the exponent term in (4.18) can be rewritten as $\sum_{u=1}^{U}\mathbf{h}_u^T\mathbf{C}_u^{-1}\mathbf{h}_u$.

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

$1, 2, \ldots, N$ and $n = 1, 2, \ldots, 2T_\mathrm{t}$, we can follow the intuition in (4.15) to formulate the following optimization problem:

$$\begin{aligned} \underset{\{\bar{\mathbf{H}}\}}{\text{minimize}} \quad & \frac{1}{2}\sum_{u=1}^{U} \|\mathbf{h}_u^T \mathbf{C}_u^{-1} \mathbf{h}_u\|^2 \\ \text{subject to} \quad & y_{\mathrm{t},i,n} \mathbf{h}_{\mathrm{t},i}^T \mathbf{x}_{\mathrm{t},n} \geq 1, \\ & i = 1, 2, \ldots, N \text{ and } n = 1, 2, \ldots, 2T_\mathrm{t}. \end{aligned} \tag{4.19}$$

In the above optimization problem, it is important to note that $\mathbf{h}_u \in \mathbb{R}^{2N \times 1}$ represents the u^{th} column of $\bar{\mathbf{H}}$, but $\mathbf{h}_{\mathrm{t},i}^T$ represents the i^{th} row of $\bar{\mathbf{H}}$. This means the objective function of (4.19) depends on the columns of $\bar{\mathbf{H}}$, but the constraints depend on the rows of $\bar{\mathbf{H}}$. Therefore, we cannot decompose (4.19) into smaller independent problems. In other words, the whole channel matrix $\bar{\mathbf{H}}$ has to be jointly estimated.

Note that the margin $\mathbf{h}_u^T \mathbf{C}_u^{-1} \mathbf{h}_u$ in (4.19) is measured using the Mahalanobis distance [64] rather than the Euclidean metric used in the standard SVM approach. The optimization problem in (4.19) can also be generalized by including slack variables as

$$\begin{aligned} \underset{\{\bar{\mathbf{H}}, \xi_{i,n}\}}{\text{minimize}} \quad & \frac{1}{2}\sum_{u=1}^{U} \|\mathbf{h}_u^T \mathbf{C}_u^{-1} \mathbf{h}_u\|^2 + C \sum_{i=1}^{N} \sum_{n=1}^{2T_\mathrm{t}} \xi_{i,n} \\ \text{subject to} \quad & y_{\mathrm{t},i,n} \mathbf{h}_{\mathrm{t},i}^T \mathbf{x}_{\mathrm{t},n} \geq 1 - \xi_{i,n} \text{ with } \xi_{i,n} \geq 0, \\ & i = 1, 2, \ldots, N \text{ and } n = 1, 2, \ldots, 2T_\mathrm{t}. \end{aligned} \tag{4.20}$$

Although the form of the objective function in (4.20) is different from that in conventional SVM problems, (4.20) can still be solved efficiently since it is a convex optimization problem. Let $\tilde{\mathbf{H}}$ be the solution of (4.20), then the channel estimate $\hat{\mathbf{H}}$ is defined as

$$\hat{\mathbf{H}} = \frac{\sqrt{UN}\tilde{\mathbf{H}}}{\|\tilde{\mathbf{H}}\|_\mathrm{F}},$$

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

4.2.3 Proposed Two-Stage SVM-based Data Detection

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

$$\bar{\mathbf{Y}}_{\mathrm{d}} = \operatorname{sign}\left(\bar{\mathbf{H}}\bar{\mathbf{X}}_{\mathrm{d}} + \bar{\mathbf{Z}}_{\mathrm{d}}\right). \tag{4.21}$$

The above equation is also converted to the real domain as

$$\mathbf{Y}_{\mathrm{d}} = \operatorname{sign}\left(\mathbf{H}_{\mathrm{d}}\mathbf{X}_{\mathrm{d}} + \mathbf{Z}_{\mathrm{d}}\right) \tag{4.22}$$

where

$$\mathbf{Y}_{\mathrm{d}} = \begin{bmatrix} \Re\{\bar{\mathbf{Y}}_{\mathrm{d}}\} \\ \Im\{\bar{\mathbf{Y}}_{\mathrm{d}}\} \end{bmatrix} = [\mathbf{y}_{\mathrm{d},1}, \mathbf{y}_{\mathrm{d},2}, \ldots, \mathbf{y}_{\mathrm{d},T_{\mathrm{d}}}], \tag{4.23}$$

$$\mathbf{X}_{\mathrm{d}} = \begin{bmatrix} \Re\{\bar{\mathbf{X}}_{\mathrm{d}}\} \\ \Im\{\bar{\mathbf{X}}_{\mathrm{d}}\} \end{bmatrix} = [\mathbf{x}_{\mathrm{d},1}, \mathbf{x}_{\mathrm{d},2}, \ldots, \mathbf{x}_{\mathrm{d},T_{\mathrm{d}}}], \tag{4.24}$$

$$\mathbf{Z}_{\mathrm{d}} = \begin{bmatrix} \Re\{\bar{\mathbf{Z}}_{\mathrm{d}}\} \\ \Im\{\bar{\mathbf{Z}}_{\mathrm{d}}\} \end{bmatrix} = [\mathbf{z}_{\mathrm{d},1}, \mathbf{z}_{\mathrm{d},2}, \ldots, \mathbf{z}_{\mathrm{d},T_{\mathrm{d}}}], \text{ and} \tag{4.25}$$

$$\mathbf{H}_{\mathrm{d}} = \begin{bmatrix} \Re\{\bar{\mathbf{H}}\} & -\Im\{\bar{\mathbf{H}}\} \\ \Im\{\bar{\mathbf{H}}\} & \Re\{\bar{\mathbf{H}}\} \end{bmatrix} = [\mathbf{h}_{\mathrm{d},1}, \mathbf{h}_{\mathrm{d},2}, \ldots, \mathbf{h}_{\mathrm{d},2N}]^T. \tag{4.26}$$

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

represents the i'^{th} row of $\mathbf{H}_{\text{d}}$.

It can be noted that the real and imaginary parts in (4.8)–(4.11) are stacked side-by-side, but they are stacked on top of each other in (4.23)–(4.26). This is due to the exchange in the role of the channel and the data matrices. In the formulation for channel estimation in (4.8)–(4.11), 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 (4.23)–(4.26) 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 $\mathbf{H}_{\text{t}}$ in (4.9) and $\mathbf{H}_{\text{d}}$ in (4.26) are constructed differently, they still depend on the same channel $\bar{\mathbf{H}}$. Let

$$\begin{aligned}\mathbf{y}_{\text{d},m} &= [y_{\text{d},m,1}, y_{\text{d},m,2}, \ldots, y_{\text{d},m,2N}]^T \text{ and}\\ \mathbf{z}_{\text{d},m} &= [z_{\text{d},m,1}, z_{\text{d},m,2}, \ldots, z_{\text{d},m,2N}]^T,\end{aligned}$$

then we have

$$y_{\text{d},m,i'} = \operatorname{sign}\left(\mathbf{h}_{\text{d},i'}^T \mathbf{x}_{\text{d},m} + z_{\text{d},m,i'}\right). \tag{4.27}$$

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

$$\begin{aligned}\underset{\{\mathbf{x}_{\text{d},m},\xi_{i'}\}}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{x}_{\text{d},m}\|^2 + C\sum_{i=1}^{2N}\xi_{i'}\\ \text{subject to} \quad & y_{\text{d},m,i'}\mathbf{x}_{\text{d},m}^T\hat{\mathbf{h}}_{\text{d},i'} \geq 1 - \xi_{i'},\\ & \xi_{i'} \geq 0, \quad i' = 1, 2, \ldots, 2N,\end{aligned} \tag{4.28}$$

where the bias is discarded as in the channel estimation problem. Let $\tilde{\mathbf{x}}_{\text{d},m}$ denote the solution of (4.28) and let $\grave{\mathbf{x}}_{\mathrm{d},m}$ be the normalized version of $\tilde{\mathbf{x}}_{\mathrm{d},m}$ as

$$\grave{\mathbf{x}}_{\mathrm{d},m} = \frac{\sqrt{U}\tilde{\mathbf{x}}_{\mathrm{d},m}}{\|\tilde{\mathbf{x}}_{\mathrm{d},m}\|}. \tag{4.29}$$

This normalization step is also used in [1] in order to make the power of the estimated signal equal the power of the transmitted signal.

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

$$\check{x}_{\mathrm{d},m,u} = \underset{x \in \mathcal{M}}{\arg\min} \left| (\grave{x}_{\mathrm{d},m,u} + j\grave{x}_{\mathrm{d},m,u+U}) - x \right|, \tag{4.30}$$

where $u \in \mathcal{U}$ and $\mathcal{M}$ represents the signal constellation (e.g., QPSK or 16-QAM). The solution to (4.30) 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 $\mathcal{X}_u$ for each $\bar{x}_{\mathrm{d},m,u}$ is created using $\check{x}_{\mathrm{d},m,u}$ and $\grave{x}_{\mathrm{d},m,u} + j\grave{x}_{\mathrm{d},m,u+U}$ as

$$\mathcal{X}_u = \left\{ \acute{x} \in \mathcal{M} \middle| \frac{|(\grave{x}_{\mathrm{d},m,u} + j\grave{x}_{\mathrm{d},m,u+U}) - \acute{x}|}{|(\grave{x}_{\mathrm{d},m,u} + j\grave{x}_{\mathrm{d},m,u+U}) - \check{x}_{\mathrm{d},m,u}|} < \nu \right\} \tag{4.31}$$

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

$$\mathcal{X}_{\mathrm{d},m} = \left\{ [\acute{x}_1, \acute{x}_2, \ldots, \acute{x}_U]^T \mid \acute{x}_u \in \mathcal{X}_u, \forall u \in \mathcal{U} \right\}. \tag{4.32}$$

The above candidate set formation was introduced in [1]. However, the detected data vector in [1] is obtained by searching over $\mathcal{X}_{\mathrm{d},m}$ 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 [46], but no justification was given. An explanation for this issue is provided in Appendix D. To deal with the issue, here a different criterion referred to as *minimum weighted Hamming distance* [43] is adopted. Suppose that $\mathcal{X}_{\mathrm{d},m} = \{\acute{\mathbf{x}}_1, \acute{\mathbf{x}}_2, \ldots, \acute{\mathbf{x}}_{|\mathcal{X}_{\mathrm{d},m}|}\}$ and let $\dot{\mathbf{x}}_l = [\Re\{\acute{\mathbf{x}}_l\}^T, \Im\{\acute{\mathbf{x}}_l\}^T]^T$ with $l \in \{1, 2, \ldots, |\mathcal{X}_{\mathrm{d},m}|\}$. 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

$$\hat{l} = \underset{l \in \{1,\ldots,|\mathcal{X}_{\mathrm{d},m}|\}}{\arg\min}\ d_{\mathrm{w}}\left(\mathbf{y}_{\mathrm{d},m}, \mathrm{sign}(\hat{\mathbf{H}}_{\mathrm{d}}\dot{\mathbf{x}}_l)\right). \tag{4.33}$$

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 [43].

The minimum weighted Hamming distance criterion above was shown to be statistically efficient [43]. However, the OSD method proposed in [43] requires a preprocessing stage whose computational complexity is proportional to $2^{N_{\mathrm{s}}}|\mathcal{M}|^{N_{\mathrm{t}}}$ for each channel realization. Here $N_{\mathrm{s}} = 2N/G$ 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 (4.28) can be solved by very efficient algorithms [61–63].

4.2.4 Proposed 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 [21]. For this reason, an SVM-based joint CE-DD method is here proposed 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 $\hat{\mathbf{X}}_{\mathrm{d}}$ be the detected version of $\bar{\mathbf{X}}_{\mathrm{d}}$ using the proposed two-stage data detection method and let

$$\hat{\mathbf{X}}_{\text{d2}} = \begin{bmatrix} \Re\{\hat{\mathbf{X}}_{\text{d}}\} & \Im\{\hat{\mathbf{X}}_{\text{d}}\} \\ -\Im\{\hat{\mathbf{X}}_{\text{d}}\} & \Re\{\hat{\mathbf{X}}_{\text{d}}\} \end{bmatrix} = [\hat{\mathbf{x}}_{\text{d2},1}, \ldots, \hat{\mathbf{x}}_{\text{d2},2T_{\text{d}}}], \tag{4.34}$$

$$\mathbf{Y}_{\text{d2}} = \left[\Re\{\bar{\mathbf{Y}}_{\text{d}}\}, \Im\{\bar{\mathbf{Y}}_{\text{d}}\}\right] = [\mathbf{y}_{\text{d2},1}, \ldots, \mathbf{y}_{\text{d2},N}]^T, \tag{4.35}$$

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

$$\begin{aligned} \underset{\{\mathbf{h}_{\text{t},i}, \xi_{\text{t},n}, \xi_{\text{d},m}\}}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{h}_{\text{t},i}\|^2 + C\left(\sum_{n=1}^{2T_{\text{t}}} \xi_{\text{t},n} + \sum_{m=1}^{2T_{\text{d}}} \xi_{\text{d},m}\right) \\ \text{subject to} \quad & y_{\text{t},i,n}\mathbf{h}_{\text{t},i}^T\mathbf{x}_{\text{t},n} \geq 1 - \xi_{\text{t},n}, \\ & y_{\text{d2},i,m}\mathbf{h}_{\text{t},i}^T\hat{\mathbf{x}}_{\text{d2},m} \geq 1 - \xi_{\text{d},m}, \\ & \xi_{\text{t},n} \geq 0, \quad n = 1, 2, \ldots, 2T_{\text{t}}, \\ & \xi_{\text{d},m} \geq 0, \quad m = 1, 2, \ldots, 2T_{\text{d}}. \end{aligned} \tag{4.36}$$

In the optimization problem above, we use two sets of slack variables $\{\xi_{\text{t},n}\}$ and $\{\xi_{\text{d},m}\}$, 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 refined channel estimate obtained by solving (4.36) can now be used for data detection again in (4.28) and (4.33). Note that the channel estimate obtained by (4.13) can be used as the initial solution to (4.36) so that the algorithm will more quickly converge to the optimal solution. Similarly, $\hat{\mathbf{X}}_{\text{d}}$ can also be used as the initial solution when solving (4.28) with the refined channel estimate. Numerical results in Section 4.4 show that this strategy will hit a certain performance bound as T_{d} increases.

4.3 Extension to OFDM systems with Frequency-Selective Fading Channels

This section develops SVM-based channel estimation and SVM-based data detection for OFDM systems with frequency-selective fading channels. Consider an uplink multiuser OFDM system with N_{c} subcarriers. Denote $\bar{\mathbf{x}}_u^{\text{FD}} \in \mathbb{C}^{N_{\text{c}} \times 1}$ as the OFDM symbol from the u^{th} 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 is assumed to satisfy $L-1 \leq N_{\text{cp}} \leq N_{\text{c}}$. It is assumed that L is known. After removing the CP, the quantized received signal at the i^{th} antenna in the time domain is given by

$$\bar{\mathbf{y}}_i^{\text{TD}} = \text{sign}\left(\sum_{u=1}^{U} \bar{\mathbf{G}}_{i,u}^{\text{TD}} \mathbf{\Gamma}^H \bar{\mathbf{x}}_u^{\text{FD}} + \bar{\mathbf{z}}_i^{\text{TD}}\right) \tag{4.37}$$

where $\mathbf{\Gamma}$ is the DFT matrix of size $N_{\text{c}} \times N_{\text{c}}$; $\bar{\mathbf{G}}_{i,u}^{\text{TD}}$ is a circulant matrix whose first column is $\bar{\mathbf{g}}_{i,u}^{\text{TD}} = [(\bar{\mathbf{h}}_{i,u}^{\text{TD}})^T, 0, \ldots, 0]^T$; and $\bar{\mathbf{h}}_{i,u}^{\text{TD}}$ is the channel vector of the u^{th} user containing the L channel taps, which are assumed to be i.i.d. and distributed as $\mathcal{CN}(0, \frac{1}{L})$. 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.

4.3.1 Proposed SVM-based Channel Estimation in OFDM Systems with Frequency-Selective Fading Channels

Denote $\bar{\boldsymbol{\phi}}_u^{\mathrm{TD}} = \boldsymbol{\Gamma}^H \bar{\mathbf{x}}_u^{\mathrm{FD}}$ and the training matrix $\bar{\boldsymbol{\Phi}}_u^{\mathrm{TD}}$ as a circulant matrix with first column equal to $\bar{\boldsymbol{\phi}}_u^{\mathrm{TD}}$. The system model in (4.37) can be reorganized as follows:

$$\begin{aligned}
\bar{\mathbf{y}}_i^{\mathrm{TD}} &= \mathrm{sign}\left(\sum_{u=1}^{U} \bar{\boldsymbol{\Phi}}_u^{\mathrm{TD}} \bar{\mathbf{g}}_{i,u}^{\mathrm{TD}} + \bar{\mathbf{z}}_i^{\mathrm{TD}}\right) \\
&= \mathrm{sign}\left(\sum_{u=1}^{U} \bar{\boldsymbol{\Phi}}_{u,L}^{\mathrm{TD}} \bar{\mathbf{h}}_{i,u}^{\mathrm{TD}} + \bar{\mathbf{z}}_i^{\mathrm{TD}}\right) \\
&= \mathrm{sign}\left(\bar{\boldsymbol{\Phi}}_L^{\mathrm{TD}} \bar{\mathbf{h}}_i^{\mathrm{TD}} + \bar{\mathbf{z}}_i^{\mathrm{TD}}\right) \qquad (4.38)
\end{aligned}$$

where $\bar{\boldsymbol{\Phi}}_{u,L}^{\mathrm{TD}}$ is the matrix corresponding to the first L columns of $\bar{\boldsymbol{\Phi}}_u^{\mathrm{TD}}$, $\bar{\boldsymbol{\Phi}}_L^{\mathrm{TD}} = [\bar{\boldsymbol{\Phi}}_{1,L}^{\mathrm{TD}}, \ldots, \bar{\boldsymbol{\Phi}}_{U,L}^{\mathrm{TD}}]$, and $\bar{\mathbf{h}}_i^{\mathrm{TD}} = [(\bar{\mathbf{h}}_{i,1}^{\mathrm{TD}})^T, \ldots, (\bar{\mathbf{h}}_{i,U}^{\mathrm{TD}})^T]^T$.

We also convert (4.38) into the real domain as

$$\mathbf{y}_i^{\mathrm{TD}} = \mathrm{sign}\left(\boldsymbol{\Phi}_L^{\mathrm{TD}} \mathbf{h}_i^{\mathrm{TD}} + \mathbf{z}_i^{\mathrm{TD}}\right) \qquad (4.39)$$

where

$$\begin{aligned}
\mathbf{y}_i^{\mathrm{TD}} &= \left[\Re\{\bar{\mathbf{y}}_i^{\mathrm{TD}}\}^T, \Im\{\bar{\mathbf{y}}_i^{\mathrm{TD}}\}^T\right]^T, \\
\mathbf{h}_i^{\mathrm{TD}} &= \left[\Re\{\bar{\mathbf{h}}_i^{\mathrm{TD}}\}^T, \Im\{\bar{\mathbf{h}}_i^{\mathrm{TD}}\}^T\right]^T, \\
\mathbf{z}_i^{\mathrm{TD}} &= \left[\Re\{\bar{\mathbf{z}}_i^{\mathrm{TD}}\}^T, \Im\{\bar{\mathbf{z}}_i^{\mathrm{TD}}\}^T\right]^T, \text{ and} \\
\boldsymbol{\Phi}_L^{\mathrm{TD}} &= \begin{bmatrix} \Re\{\bar{\boldsymbol{\Phi}}_L^{\mathrm{TD}}\} & -\Im\{\bar{\boldsymbol{\Phi}}_L^{\mathrm{TD}}\} \\ \Im\{\bar{\boldsymbol{\Phi}}_L^{\mathrm{TD}}\} & \Re\{\bar{\boldsymbol{\Phi}}_L^{\mathrm{TD}}\} \end{bmatrix}.
\end{aligned}$$

Denote $\mathbf{y}_i^{\mathrm{TD}} = [y_{i,1}^{\mathrm{TD}}, y_{i,2}^{\mathrm{TD}}, \ldots, y_{i,2N_c}^{\mathrm{TD}}]^T$ and $\boldsymbol{\Phi}_L^{\mathrm{TD}} = \left[(\boldsymbol{\phi}_1^{\mathrm{TD}})^T, (\boldsymbol{\phi}_2^{\mathrm{TD}})^T, \ldots, (\boldsymbol{\phi}_{2N_c}^{\mathrm{TD}})^T\right]^T$, leading to the following SVM problem for estimating the OFDM channel using one-bit ADCs:

$$\begin{aligned} \underset{\{\mathbf{h}_i^{\mathrm{TD}},\xi_n\}}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{h}_i^{\mathrm{TD}}\|^2 + C\sum_{n=1}^{2N_{\mathrm{c}}}\xi_n \\ \text{subject to} \quad & y_{i,n}^{\mathrm{TD}}\left(\mathbf{h}_i^{\mathrm{TD}}\right)^T \boldsymbol{\phi}_n^{\mathrm{TD}} \geq 1-\xi_n, \\ & \xi_n \geq 0, \quad n = 1,2,\ldots,2N_{\mathrm{c}}. \end{aligned} \tag{4.40}$$

Denoting $\tilde{\mathbf{h}}_i^{\mathrm{TD}}$ as the solution of (4.40), then $\mathbf{h}_i^{\mathrm{TD}}$ can be estimated as

$$\hat{\mathbf{h}}_i^{\mathrm{TD}} = \frac{\sqrt{U}\tilde{\mathbf{h}}_i^{\mathrm{TD}}}{\|\tilde{\mathbf{h}}_i^{\mathrm{TD}}\|}. \tag{4.41}$$

Frequency-selective channel estimation methods using one-bit ADCs have been previously proposed in [19, 21], and [65] based on the Bussgang decomposition, Additive Quantization Noise Model (AQNM), and deep learning, respectively. The deep learning method in [65] was shown to outperform the methods of [19, 65] at low SNRs, but its performance tends to degrade as the SNR increases. In addition, the method in [65] requires a training sequence that contains many OFDM symbols, which are required to be orthogonal between different users. In the proposed method, only one OFDM symbol is used in the training phase and all users send their training symbols concurrently.

4.3.2 Proposed SVM-based Data Detection in OFDM Systems with Frequency-Selective Fading Channels

This section describes how SVM can also be used for data detection in OFDM systems with frequency-selective fading channels. The received quantized vector in (4.37) can be rewritten as

$$\bar{\mathbf{y}}_i^{\mathrm{TD}} = \operatorname{sign}\left(\bar{\mathbf{G}}_i^{\mathrm{FD}}\bar{\mathbf{x}}^{\mathrm{FD}} + \bar{\mathbf{z}}_i^{\mathrm{TD}}\right) \tag{4.42}$$

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

$$\bar{\mathbf{y}}^{\mathrm{TD}} = \mathrm{sign}\left(\bar{\mathbf{G}}^{\mathrm{FD}}\bar{\mathbf{x}}^{\mathrm{FD}} + \bar{\mathbf{z}}^{\mathrm{TD}}\right) \tag{4.43}$$

where $\bar{\mathbf{y}}^{\mathrm{TD}} = \left[(\bar{\mathbf{y}}_1^{\mathrm{TD}})^T, (\bar{\mathbf{y}}_2^{\mathrm{TD}})^T, \ldots, (\bar{\mathbf{y}}_N^{\mathrm{TD}})^T\right]^T$ and $\bar{\mathbf{G}}^{\mathrm{FD}} = \left[(\bar{\mathbf{G}}_1^{\mathrm{FD}})^T, (\bar{\mathbf{G}}_2^{\mathrm{FD}})^T, \ldots, (\bar{\mathbf{G}}_N^{\mathrm{FD}})^T\right]^T$. Let $\mathbf{y}^{\mathrm{TD}}$, $\mathbf{G}^{\mathrm{FD}}$, and $\mathbf{x}^{\mathrm{FD}}$ be the real-valued versions of $\bar{\mathbf{y}}^{\mathrm{TD}}$, $\bar{\mathbf{G}}^{\mathrm{FD}}$, and $\bar{\mathbf{x}}^{\mathrm{FD}}$, respectively. Converting (4.43) to the real domain as in (4.23)–(4.26), we can formulate an SVM problem by treating the rows of $\mathbf{G}^{\mathrm{FD}}$ as the feature vectors, the elements of $\mathbf{y}^{\mathrm{TD}}$ as the binary indicators and $\mathbf{x}^{\mathrm{FD}}$ as the weight vector. The solution of the SVM problem then provides the detected data.

4.4 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 ν for the second stage of the SVM-based detection method as $\nu = \min\left\{\frac{\varrho}{10} + 1.5, 3\right\}$ for QPSK and $\nu = \min\left\{\frac{\varrho}{10} + 1.3, 1.5\right\}$ for 16-QAM where ϱ is the SNR. The length of the block-fading interval is set to 500 (i.e., $T_\mathrm{t} + T_\mathrm{d} = 500$) unless otherwise stated. For solving the proposed SVM-based channel estimation and data detection problems, the Scikit-learn machine learning library [66] is used.

Figure 4.2 presents a performance comparison of different channel estimation methods in terms of NMSE, defined here as $\mathrm{NMSE} = \mathbb{E}\left[\|\hat{\mathbf{H}} - \bar{\mathbf{H}}\|_\mathrm{F}^2\right]/(UN)$, 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, and the proposed SVM-based joint CE-DD method significantly improves the channel estimation accuracy. This is due to the help of the to-be-decoded data

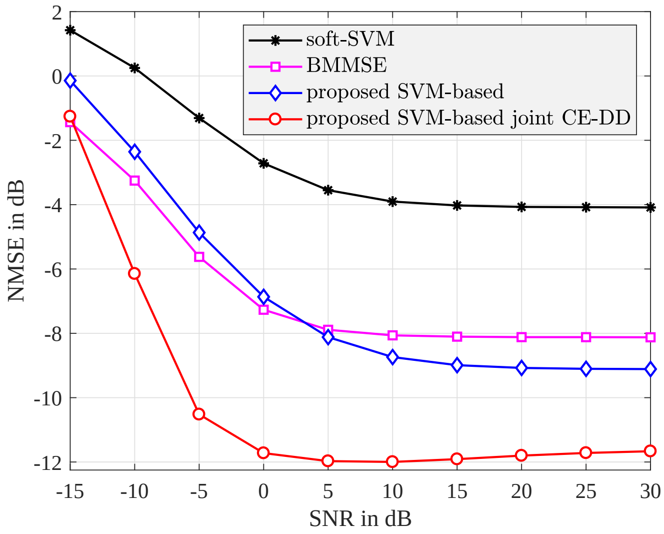


Figure 4.2: NMSE comparison between different channel estimators with $U = 4$, $N = 32$, and $T_{\mathrm{t}} = 20$.

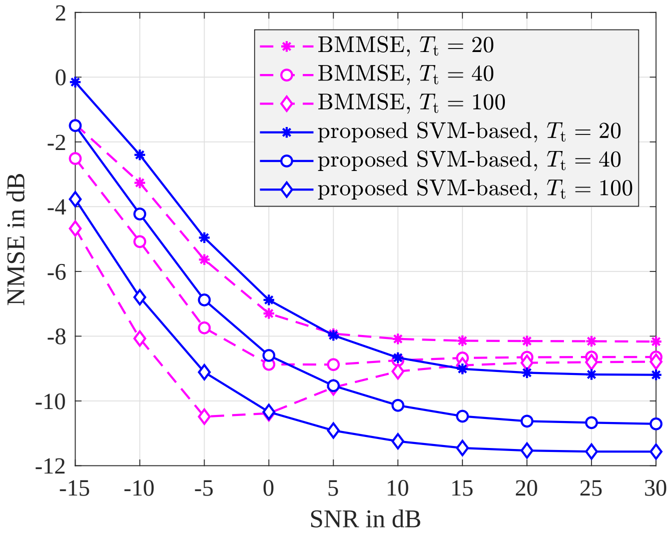


Figure 4.3: NMSE comparison between BMMSE and the proposed SVM-based channel estimator with $U = 4$, $N = 32$, and $T_{\mathrm{t}} \in \{20, 40, 100\}$.

in refining the channel estimate.

Figure 4.3 compares 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 $T_{\mathrm{t}} = 100$ is still higher than that of the proposed SVM-based method with a much shorter training

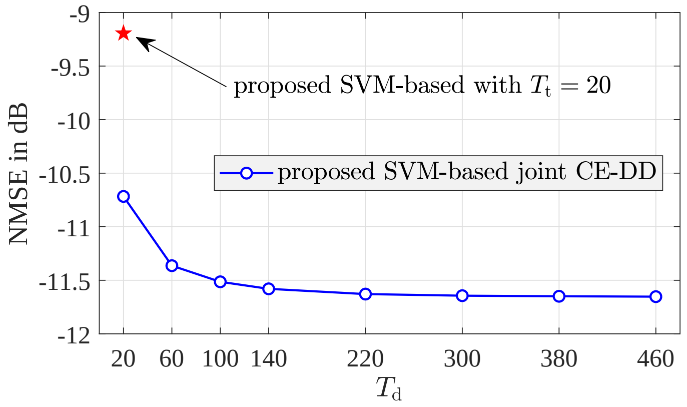


Figure 4.4: Effect of T_{d} on the NMSE of the proposed SVM-based joint CE-DD with $U = 4$, $N = 32$, and $T_{\mathrm{t}} = 20$ at $\varrho = 30$ dB.

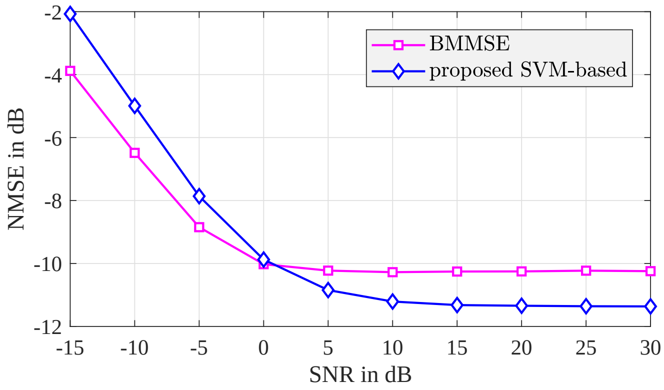


Figure 4.5: NMSE comparison between the BMMSE channel estimator and the proposed SVM-based channel estimator for spatially correlated channels with $U = 4$, $N = 32$, and $T_{\mathrm{t}} = 20$.

sequence ($T_{\mathrm{t}} = 20$). The results in Figure 4.3 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.

The effect of T_{d} on the NMSE of the proposed SVM-based joint CE-DD method is studied in Figure 4.4. 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).

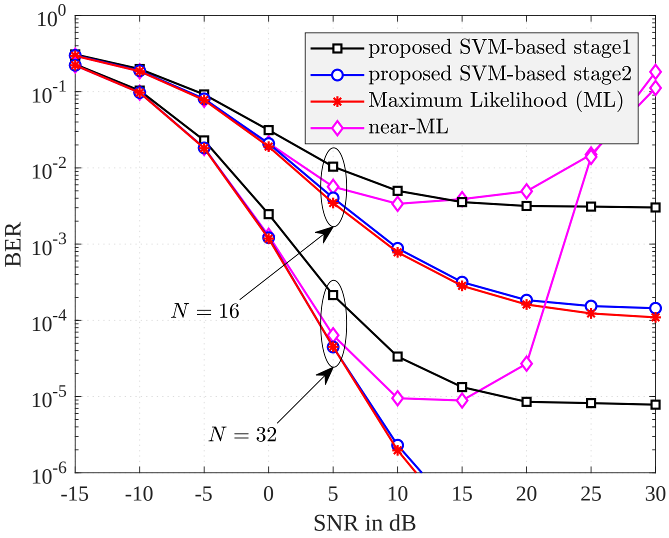


Figure 4.6: Performance comparison between the proposed two-stage SVM-based data detection method and ML detection [1] with perfect CSI, QPSK modulation, and $U = 4$. The average cardinalities of $\mathcal{X}$ for $N = 16$ and $N = 32$ are 2.9352 and 1.6140, respectively.

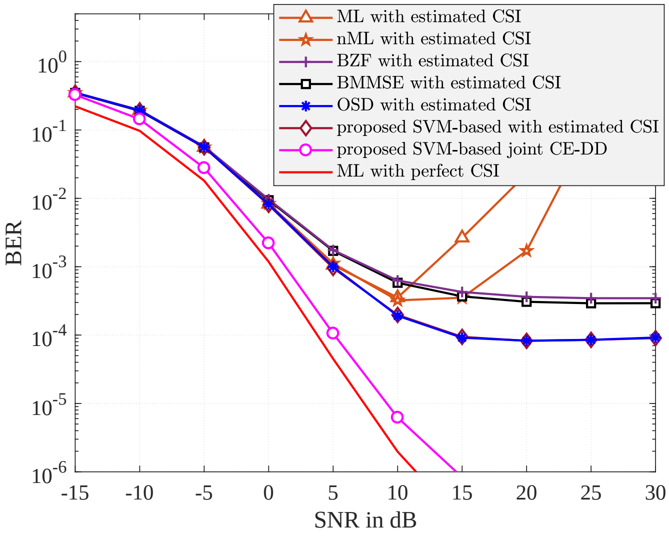


Figure 4.7: Performance comparison between two proposed data detection methods and other existing methods with estimated CSI, QPSK modulation, $N = 32$, $U = 4$, and $T_{\mathrm{t}} = 20$.

Figure 4.5 presents channel estimation results for spatially correlated channels. We use the same typical urban channel model as in [21]. 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 (4.20).

In Figure 4.6, 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 Figure 4.7, where the estimated CSI is obtained by the SVM-based channel estimator. Here, the SVM-based joint CE-DD method can be compared with other methods because it also starts with CSI estimated by the SVM-based channel estimator. 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 [46]. An explanation for the susceptibility of ML detection can be found in Appendix D. It is also observed that the proposed SVM-based and OSD detection methods give the same performance. However, 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 (4.36).

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 Figure 4.8. The average run time required to perform data detection over a block-fading interval of 500 slots is calculated. Note that the OSD method contains two stages: a preprocessing stage and a detection stage. It is observed that the OSD method

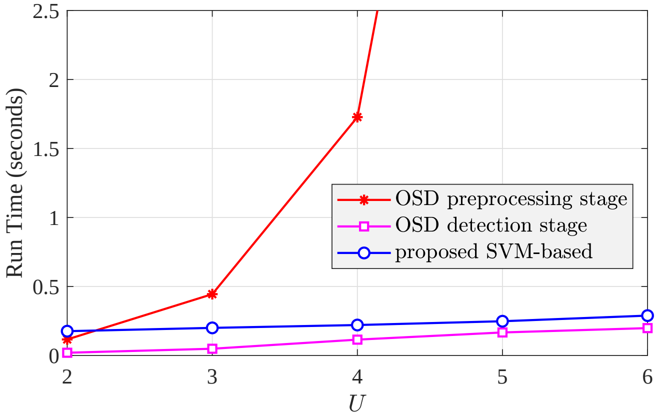


Figure 4.8: Run time comparison between OSD and the proposed SVM-based detection method with QPSK modulation, $N = 32$, and U varies.

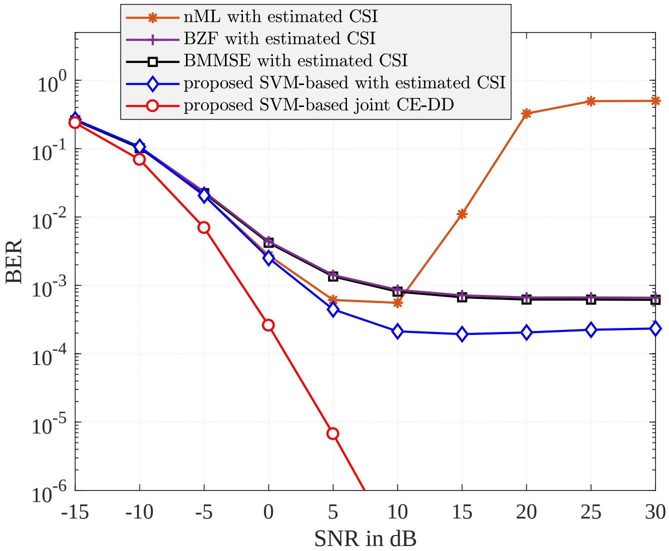


Figure 4.9: Performance comparison between two proposed data detection methods and other existing methods with estimated CSI, QPSK modulation, $N = 64$, $U = 8$, and $T_{\mathrm{t}} = 40$.

has a low-complexity detection stage. Interestingly, Figure 4.8 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.

Figure 4.9 and Figure 4.10 provide BER comparisons between the proposed SVM-based data detection methods and other existing methods with QPSK and 16-QAM modulations

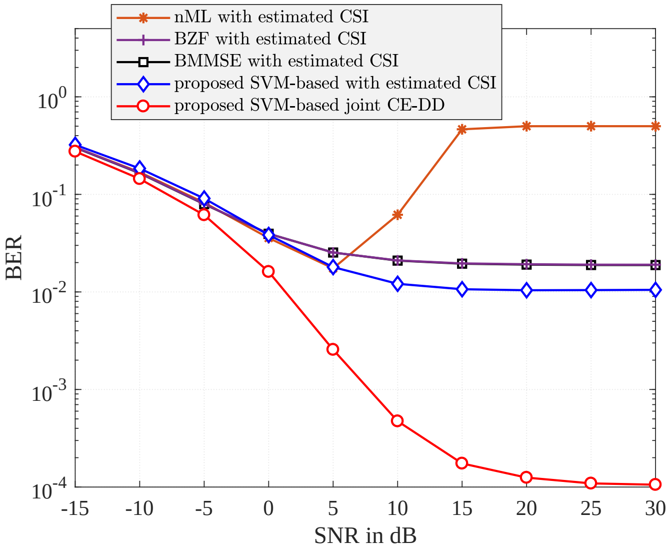


Figure 4.10: Performance comparison between two proposed data detection methods and other existing methods with estimated CSI, 16-QAM modulation, $N = 128$, $U = 8$, and $T_{\mathrm{t}} = 40$.

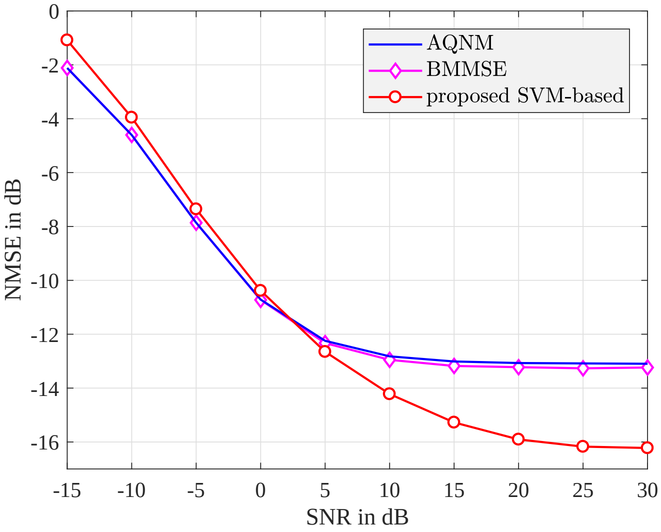


Figure 4.11: NMSE comparison between different channel estimators for an OFDM system in a frequency-selective channel with $U = 2$, $N = 16$, and $L = 8$.

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.

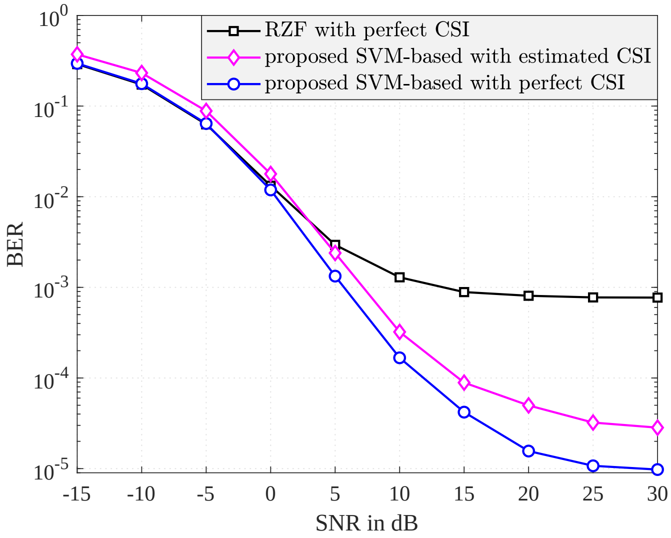


Figure 4.12: BER comparison between different data detection methods for an OFDM system in a frequency-selective channel with $N_\mathrm{c} = 256$, QPSK modulation, $U = 2$, $N = 16$, and $L = 8$.

Finally, channel estimation and data detection results for OFDM systems with frequency-selective fading channels are given in Figure 4.11 and Figure 4.12, respectively. It is observed that the BMMSE channel estimator [21] slightly outperforms the AQNM-based channel estimator [19], 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 Figure 4.12, data detection results show that the SVM-based method considerably outperforms the Regularized Zero-Forcing (RZF) of [19]. 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.

Chapter 5

Conclusion

This report has shown that the channel estimation and data detection problems in MIMO systems with low-resolution ADCs can be addressed effectively by machine learning-based methods. Chapter 3 proposed two new learning methods for enhancing the performance of blind detection in MIMO systems with low-resolution ADCs. The supervised learning method exploits the use of CRC in practical systems to gain more training data. The semi-supervised learning method is based on the perspective that the to-be-decoded data can itself help the detection task thanks to grouping of received symbol vectors for the same transmitted signal. Numerical results demonstrate the performance improvement and robustness of our proposed methods over existing techniques. Numerical results also show that the two proposed learning methods require only a few iterations to converge. A performance analysis for the proposed methods has also been carried out by evaluating the VER in different SNR regimes. In addition, a new criterion for the transmit signal design problem has also been proposed.

Chapter 4 showed how linear SVM, a well-known machine learning technique, can be exploited to provide efficient and robust channel estimation and data detection. 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 were proposed. 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 SVM-based methods against existing ones and the gain is greatest for moderate to high SNR regimes.

Bibliography

[1] 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.

[2] J. R. Hampton, *Introduction to MIMO communications.* Cambridge University Press, 2013.

[3] 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.

[4] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Select. Areas in Commun.*, vol. 32, no. 6, pp. 1065–1082, June 2014.

[5] 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. Select. Areas in Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[6] 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.

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

[8] *Common Public Radio Interface (CPRI); Interface Specification, CPRI Specification v6.0*, Ericsson AB, Huawei Technol., NEC Corp., Alcatel Lucent, and Nokia Siemens Netw., Aug. 2013.

[9] A. Mezghani and J. A. Nossek, "On ultra-wideband MIMO systems with 1-bit quantized outputs: Performance analysis and input optimization," in *Proc. IEEE Int. Symp. Inf. Theory*, 2007, pp. 1286–1289.

[10] 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*, 2012.

[11] P. Dong, H. Zhang, W. Xu, G. Y. Li, and X. You, "Performance analysis of multiuser massive MIMO with spatially correlated channels using low-precision ADC," *IEEE Commun. Letters*, vol. 22, no. 1, pp. 205–208, Jan. 2018.

[12] J. Mo and R. W. Heath, "High SNR capacity of millimeter wave MIMO systems with one-bit quantization," in *Proc. Inf. Theory and Applications Workshop*, 2014.

[13] ——, "Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5498–5512, Oct. 2015.

[14] L. Fan, S. Jin, C. Wen, and H. Zhang, "Uplink achievable rate for massive MIMO systems with low-resolution ADC," *IEEE Commun. Letters*, vol. 19, no. 12, pp. 2186–2189, Dec. 2015.

[15] J. Mo, A. Alkhateeb, S. Abu-Surra, and R. W. Heath, "Hybrid architectures with few-bit ADC receivers: Achievable rates and energy-rate tradeoffs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2274–2287, Apr. 2017.

[16] K. Roth and J. A. Nossek, "Achievable rate and energy efficiency of hybrid and digital beamforming receivers with low resolution ADC," *IEEE J. Select. Areas in Commun.*, vol. 35, no. 9, pp. 2056–2068, Sept. 2017.

[17] N. Liang and W. Zhang, "Mixed-ADC massive MIMO," *IEEE J. Select. Areas in Commun.*, vol. 34, no. 4, pp. 983–997, Apr. 2016.

[18] ——, "Mixed-ADC massive MIMO uplink in frequency-selective channels," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4652–4666, Nov. 2016.

[19] 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.

[20] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "Throughput analysis of massive MIMO uplink with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 4038–4051, June 2017.

[21] 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.

[22] C. Risi, D. Persson, and E. G. Larsson, "Massive MIMO with 1-bit ADC," *arXiv:1404.7736 [cs.IT]*, 2014.

[23] S. Rao, A. L. Swindlehurst, and H. Pirzadeh, "Massive MIMO channel estimation with 1-bit spatial sigma-delta ADCs," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Process.*, Brighton, United Kingdom, May 2019, pp. 4484–4488.

[24] 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 on Smart Antennas*, Vienna, Austria, April 2019.

[25] 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. on Acoustics, Speech and Signal Process.*, Brighton, United Kingdom, May 2019, pp. 4669–4673.

[26] K. Gao, N. J. Estes, B. Hochwald, J. Chisum, and J. N. Laneman, "Power-performance analysis of a simple one-bit transceiver," in *Proc. Information Theory and Applications Workshop*, San Diego, CA, USA, Feb. 2017.

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

[28] 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. (Early Access)*, 2020.

[29] I. 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.*, Abu Dhabi, United Arab Emirates, Dec. 2018.

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

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

[32] 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.

[33] 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, June 2018.

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

[35] 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.

[36] 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, Systems and Computers*, Pacific Grove, CA, USA, Nov. 2016, pp. 768–773.

[37] 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.*, San Diego, CA, USA, Dec. 2015.

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

[39] Y. Ding, S. Chiu, and B. D. Rao, “Bayesian channel estimation algorithms for massive MIMO systems with hybrid analog-digital processing and low-resolution ADCs,” *IEEE J. Select. Topics in Signal Process.*, vol. 12, no. 3, pp. 499–513, June 2018.

[40] A. Kaushik, E. Vlachos, J. Thompson, and A. Perelli, “Efficient channel estimation in millimeter wave hybrid MIMO systems with low resolution ADCs,” in *Proc. European Signal Processing Conference*, Rome, Italy, Sept. 2018, pp. 1825–1829.

[41] S. Wang, Y. Li, and J. Wang, “Convex optimization based multiuser detection for uplink large-scale MIMO under low-resolution quantization,” in *Proc. IEEE Int. Conf. Commun.*, Sydney, NSW, Australia, June 2014, pp. 4789–4794.

[42] A. Mezghani, M. Khoufi, and J. A. Nossek, “Maximum likelihood detection for quantized MIMO systems,” in *Proc. Int. ITG Workshop on Smart Antennas*, Vienna, Austria, Feb. 2008, pp. 278–284.

[43] 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, July 2018.

[44] 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.

[45] L. V. Nguyen and D. H. N. Nguyen, “Linear receivers for massive MIMO systems with one-bit ADCs,” *arXiv preprint arXiv:1907.06664*, 2019.

[46] 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.

[47] 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.

[48] Y. 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.

[49] 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. and Networking Conf.*, Marrakesh, Morocco, Apr. 2019.

[50] 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. 27 172–27 182, Feb. 2019.

[51] 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. Letters*, vol. 7, no. 3, pp. 476–479, June 2018.

[52] D. Hui and D. L. Neuhoff, “Asymptotic analysis of optimal fixed-rate uniform scalar quantization,” *IEEE Trans. Inform. Theory*, vol. 47, no. 3, pp. 957–977, Mar. 2001.

[53] N. Al-Dhahir and J. M. Cioffi, “On the uniform ADC bit precision and clip level computation for a Gaussian signal,” *IEEE Trans. Signal Process.*, vol. 44, no. 2, pp. 434–438, Feb. 1996.

[54] Y. S. Jeon, S. N. Hong, and N. Lee, “Blind detection for MIMO systems with low-resolution ADCs using supervised learning,” in *Proc. IEEE Int. Conf. Commun.*, Paris, France, May 2017.

[55] Y. Jeon, M. So, and N. Lee, “Reinforcement-learning-aided ML detector for uplink massive MIMO systems with low-precision ADCs,” in *Proc. IEEE Wireless Commun. and Networking Conf.*, Barcelona, Spain, Apr. 2018.

[56] Y. Jeon, H. Lee, and N. Lee, “Robust MLSD for wideband SIMO systems with one-bit ADCs: Reinforcement-Learning Approach,” in *Proc. IEEE Int. Conf. Commun. Workshops*, Kansas City, MO, USA, May 2018.

[57] S. Schibisch, S. Cammerer, S. Dorner, J. Hoydis, and S. ten Brink, “Online label recovery for deep learning-based communication through error correcting codes,” in *Proc. Int. Symp. Wireless Commun. Systems*, Lisbon, Portugal, Aug. 2018.

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

[59] D. Tse and P. Viswanath, *Fundamentals of wireless communication.* Cambridge University Press, 2005.

[60] *Multiplexing and Channel Coding*, 3GPP Std. TS36.212, 2012.

[61] T. Joachims, “Training linear SVMs in linear time,” in *Proc. the ACM SIGKDD international conference on Knowledge discovery and Data mining.* Philadelphia, PA, USA: ACM, Aug. 2006, pp. 217–226.

[62] S. S. Keerthi and D. DeCoste, “A modified finite Newton method for fast solution of large scale linear SVMs,” *Journal of Machine Learning Research*, vol. 6, pp. 341–361, Mar. 2005.

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

[64] P. C. Mahalanobis, "On the generalized distance in statistics," in *Proc. National Institute of Science of India*, 1936.

[65] E. Balevi and J. G. Andrews, "Two-stage learning for uplink channel estimation in one-bit massive MIMO," *arXiv preprint arXiv:1911.12461*, 2019.

[66] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, Oct. 2011.

Appendix A

Proof of Proposition 3.2

We first express $P_{\check{\mathbf{x}}_k \to \check{\mathbf{x}}_{k'}}$ as follows:

$$\begin{aligned} P_{\check{\mathbf{x}}_k \to \check{\mathbf{x}}_{k'}} &= \mathbb{P}\Big[\|\mathbf{y} - \check{\mathbf{y}}_k\|_2^2 \geq \|\mathbf{y} - \check{\mathbf{y}}_{k'}\|_2^2 \mid \mathbf{x} = \check{\mathbf{x}}_k\Big] \\ &= \mathbb{P}\Big[\|\boldsymbol{\upsilon}\|_2^2 + 2\Re\{\boldsymbol{\upsilon}^H \mathbf{w}\} \leq 0\Big] \\ &= \mathbb{P}\Big[\sum_{i=1}^{N_\mathrm{r}} \big(|\upsilon_i|^2 + 2\Re\{\upsilon_i^* w_i\}\big) \leq 0\Big]. \end{aligned} \tag{A.1}$$

By letting $\varepsilon_i = |\upsilon_i|^2 + 2\Re\{\upsilon_i^* w_i\}$, (A.1) becomes

$$P_{\check{\mathbf{x}}_k \to \check{\mathbf{x}}_{k'}} = \mathbb{P}\Big[\sum_{i=1}^{N_\mathrm{r}} \varepsilon_i \leq 0\Big]. \tag{A.2}$$

In order to approximate the probability in (A.2), we need to compute the mean and variance of ε_i. The mean of ε_i is

$$\mathbb{E}[\varepsilon_i] = \mathbb{E}\big[|\upsilon_i|^2 + 2\Re\{\upsilon_i^* w_i\}\big] = \mathbb{E}\big[|\upsilon_i|^2\big] = \sigma_{kk'}^2. \tag{A.3}$$

The variance of ε_i is given as

$$\sigma_{\varepsilon_i}^2 = \mathrm{Var}\left[|\upsilon_i|^2\right] + \mathrm{Var}\left[2\Re\{\upsilon_i^* w_i\}\right] + 2\,\mathrm{Cov}\left(|\upsilon_i|^2, 2\Re\{\upsilon_i^* w_i\}\right). \tag{A.4}$$

The first term in the right-hand side of (A.4) is

$$\mathrm{Var}\left[|\upsilon_i|^2\right] = \mathbb{E}\big[|\upsilon_i|^4\big] - \mathbb{E}\big[|\upsilon_i|^2\big]^2 = \sigma_{kk'}^4. \tag{A.5}$$

The second term in the right-hand side of (A.4) is

$$\mathrm{Var}\left[2\Re\{\upsilon_i^* w_i\}\right] = \mathrm{Var}\left[\upsilon_i^* w_i\right] + \mathrm{Var}\left[\upsilon_i w_i^*\right] + 2\,\mathrm{Cov}\left(\upsilon_i^* w_i, \upsilon_i w_i^*\right). \tag{A.6}$$

Since $\mathrm{Var}\left[\upsilon_i^* w_i\right] = \mathrm{Var}\left[\upsilon_i w_i^*\right] = \mathbb{E}\big[|\upsilon_i|^2\big] = \sigma_{kk'}^2$, and $\mathrm{Cov}\left(\upsilon_i^* w_i, \upsilon_i w_i^*\right) = 0$, we have

$$\mathrm{Var}\left[2\Re\{\upsilon_i^* w_i\}\right] = 2\sigma_{kk'}^2. \tag{A.7}$$

The last term in the right-hand side of (A.4) is

$$\mathrm{Cov}\left(|\upsilon_i|^2, 2\Re\{\upsilon_i^* w_i\}\right) = \mathbb{E}\big[|\upsilon_i|^2 2\Re\{\upsilon_i^* w_i\}\big] + \mathbb{E}\big[|\upsilon_i|^2\big]\mathbb{E}\big[2\Re\{\upsilon_i^* w_i\}\big] = 0, \tag{A.8}$$

since $\mathbb{E}\big[|\upsilon_i|^2 2\Re\{\upsilon_i^* w_i\}\big] = \mathbb{E}\big[|\upsilon_i|^2(\upsilon_i^* w_i + \upsilon_i w_i^*)\big] = 0$ and $\mathbb{E}\big[2\Re\{\upsilon_i^* w_i\}\big] = \mathbb{E}\big[\upsilon_i^* w_i\big] + \mathbb{E}\big[\upsilon_i w_i^*\big] = 0$.

Substituting the results in (A.5), (A.7), and (A.8) into (A.4) yields the variance of ε_i as

$$\sigma_{\varepsilon_i}^2 = \sigma_{kk'}^4 + 2\sigma_{kk'}^2. \tag{A.9}$$

The variables $\{\varepsilon_i\}_{i=1,\dots,N_\mathrm{r}}$ are i.i.d. because of the i.i.d. elements in $\mathbf{H}$. Hence, by the central limit theorem, the variable $\sum_{i=1}^{N_\mathrm{r}} \varepsilon_i$ in (A.2) can be approximated by a Gaussian random variable with mean $N_\mathrm{r}\sigma_{kk'}^2$ and variance $N_\mathrm{r}(\sigma_{kk'}^4 + 2\sigma_{kk'}^2)$. Finally, the probability in (A.2) can be approximated as

$$P_{\check{\mathbf{x}}_k \to \check{\mathbf{x}}_{k'}} \approx \Phi\left(\frac{-N_\mathrm{r}\sigma_{kk'}^2}{\sqrt{N_\mathrm{r}(\sigma_{kk'}^4 + 2\sigma_{kk'}^2)}}\right) = 1 - \Phi\left(\sqrt{N_\mathrm{r}/(1 + 2/\sigma_{kk'}^2)}\right). \tag{A.10}$$

Appendix B

Proof of Theorem 3.1

For two labels $\check{\mathbf{x}}_k^{\Re}$ and $\check{\mathbf{x}}_{k'}^{\Re}$, we can always find two disjoint index sets $\mathcal{I}_\mathrm{c}$ and $\mathcal{I}_\mathrm{d}$ such that $\check{x}_{k,i}^{\Re} = \check{x}_{k',i}^{\Re} \neq 0$, $\forall i \in \mathcal{I}_\mathrm{c}$, and $\check{x}_{k,i}^{\Re} = -\check{x}_{k',i}^{\Re}$ $\forall i \in \mathcal{I}_\mathrm{d}$. We denote $d = |\mathcal{I}_\mathrm{d}|$ as the Hamming distance between the two labels $\check{\mathbf{x}}_1^{\Re}$ and $\check{\mathbf{x}}_k^{\Re}$. Note that $d \leq N_\mathrm{t}$ and $|\mathcal{I}_\mathrm{c}| = N_\mathrm{t} - d$ for BPSK signaling. The two vectors $\mathbf{g}_1^{\Re}$ and $\mathbf{g}_k^{\Re}$ can now be expressed as $\mathbf{g}_k^{\Re} = \mathbf{g}_\mathrm{c} + \mathbf{g}_\mathrm{d}$ and $\mathbf{g}_{k'}^{\Re} = \mathbf{g}_\mathrm{c} - \mathbf{g}_\mathrm{d}$, where $\mathbf{g}_\mathrm{c}$ and $\mathbf{g}_\mathrm{d}$ are the summations of the $N_\mathrm{t} - d$ and d columns of $\mathbf{H}$ corresponding to the indices given in $\mathcal{I}_\mathrm{c}$ and $\mathcal{I}_\mathrm{d}$, respectively. For Rayleigh fading with unit variance, $\mathbf{g}_\mathrm{c}$ is $\mathcal{N}(\mathbf{0}, \frac{N_\mathrm{t}-d}{2}\mathbf{I}_{2N_\mathrm{r}})$ and $\mathbf{g}_\mathrm{d}$ is $\mathcal{N}(\mathbf{0}, \frac{d}{2}\mathbf{I}_{2N_\mathrm{r}})$.

The probability that $\mathrm{sign}(g_{1,i}^{\Re}) = \mathrm{sign}(g_{k,i}^{\Re})$ is given as

$$\mathbb{P}\big[\,\mathrm{sign}(g_{k,i}^{\Re}) = \mathrm{sign}(g_{k',i}^{\Re})\big] = \frac{2}{\pi}\arctan\sqrt{\frac{N_\mathrm{t}-d}{d}}. \tag{B.1}$$

This is obtained by applying a result in [1], which states that if $a \sim \mathcal{N}(0, \sigma_a^2)$ and $b \sim \mathcal{N}(0, \sigma_b^2)$ then

$$\mathbb{P}\big[\,\mathrm{sign}(a+b) = \mathrm{sign}(a-b)\big] = \frac{2}{\pi}\arctan\frac{\sigma_a}{\sigma_b}. \tag{B.2}$$

Due to the independence between the events $\mathrm{sign}(g_{k,i}^{\Re}) = \mathrm{sign}(g_{k',i}^{\Re})$, for $i = 1, 2, \ldots, 2N_\mathrm{r}$, the result in (3.37) thus follows.

Appendix C

Proof of Proposition 3.4

Without loss of generality, we assume that $\check{\mathbf{x}}_1^{\Re} = [\mathbf{1}_{N_\mathrm{t}}^T, \mathbf{0}_{N_\mathrm{t}}^T]^T$ was transmitted. Denote E_k, $1 < k \le K$, as the event $\check{\mathbf{y}}_1 = \check{\mathbf{y}}_k$. The detection error event E is then defined as $E = \bigcup_{k>1} E_k$. We want to find the VER given event E and subsequently prove that $P_{\rho \to \infty}^{\mathrm{ver}} \le \frac{1}{2} \sum_{k>1}^{K} \mathbb{P}(E_k)$. We note that $E_2, \ldots, E_K$ are not necessarily mutually exclusive nor independent. However, we can combine $E_2, \ldots, E_K$ into larger events $G_1, \ldots, G_L$ that are mutually exclusive. Herein, the rule for forming G_ℓ is as follows:

1. If E_k is mutually exclusive with all other events, then $E_k \subset G_1$.

2. If a pair of events E_k and E_m intersect, i.e., $E_k \cap E_m \neq \varnothing$, but $E_k \cup E_m$ is mutually exclusive with all other events, then $(E_k \cup E_m) \subset G_2$.

3. $G_3, \ldots, G_L$ are then formed in a similar fashion.

Certainly, if $E_k \subset G_\ell$, then $E_k \cap G_{\ell'} = \varnothing$, for $\ell' \neq \ell$. This combining strategy effectively partitions E into mutually exclusive events $G_1, \ldots, G_L$. The VER is calculated as:

1. If event $E_k \subset G_1$ has occurred, the receiver would erroneously pick the detected vector $\hat{\mathbf{x}}_k^{\Re} \neq \check{\mathbf{x}}_1^{\Re}$ with a probability of 1/2, i.e., VER = 1/2.

2. For any two events $E_k, E_m \subset G_2$ and $E_k \cap E_m \neq \varnothing$, we consider the following three partitions of $E_k \cup E_m$:

- If $E_k \cap E_m^{\mathrm{c}}$ has occurred, VER $= 1/2$.

- If $E_k^{\mathrm{c}} \cap E_m$ has occurred, VER $= 1/2$.

- If $E_k \cap E_m$ has occurred, the receiver would erroneously pick the detected vector as either $\hat{\mathbf{x}}_k^{\Re}$ or $\hat{\mathbf{x}}_m^{\Re}$ with a probability of $2/3$, i.e., VER $= 2/3$.

We then have

$$\begin{aligned} & \frac{1}{2}\mathbb{P}[E_k \cap E_m^{\mathrm{c}}] + \frac{1}{2}\mathbb{P}[E_k^{\mathrm{c}} \cap E_m] + \frac{2}{3}\mathbb{P}[E_k \cap E_m] \\ \leq \; & \frac{1}{2}\mathbb{P}[E_k \cap E_m^{\mathrm{c}}] + \frac{1}{2}\mathbb{P}[E_k^{\mathrm{c}} \cap E_m] + \mathbb{P}[E_k \cap E_m] = \frac{1}{2}\mathbb{P}[E_k] + \frac{1}{2}\mathbb{P}[E_m]. \end{aligned} \tag{C.1}$$

3. The same principle of partitioning can be applied for events in $G_3, \ldots, G_L$ to calculate the VER.

Therefore, $P_{\rho\to\infty}^{\mathrm{ver}}$ is upper-bounded as

$$\begin{aligned} P_{\rho\to\infty}^{\mathrm{ver}} & \leq \sum_{E_k \subset G_1} \frac{1}{2}\mathbb{P}[E_k] + \sum_{E_k \subset G_2} \frac{1}{2}\mathbb{P}[E_k] + \ldots \\ & = \frac{1}{2}\sum_{k>1}^{K} \mathbb{P}[E_k]. \end{aligned} \tag{C.2}$$

The inequality presented in the proposition follows by combining the result in Theorem 3.1 and noting that there are $\binom{N_{\mathrm{t}}}{d}$ labels with Hamming distance d from $\check{\mathbf{x}}_1^{\Re}$. If the error event E is comprised of only mutual events $E_2, \ldots, E_K$, the inequality (C.2) becomes $P_{\rho\to\infty}^{\mathrm{ver}} = \sum_{k=2}^{K} \frac{1}{2}\mathbb{P}[E_k]$. Thus, the VER upper-bound becomes tight in this case.

Appendix D

Explanation for the susceptibility of ML detection at high SNRs with imperfect CSI

The ML detection method of [1] is defined as

$$\hat{\mathbf{x}}_{\mathrm{d},m}^{\mathtt{ML}} = \underset{\bar{\mathbf{x}} \in \mathcal{M}^U}{\arg\max} \underbrace{\prod_{i=1}^{2N} \Phi\left(\sqrt{2\varrho} y_{\mathrm{d},m,i} \hat{\mathbf{h}}_{\mathrm{d},i}^T \mathbf{x}\right)}_{\mathcal{P}(\mathbf{x})}, \tag{D.1}$$

where $\mathbf{x} = [\Re\{\bar{\mathbf{x}}\}^T, \Im\{\bar{\mathbf{x}}\}^T]^T$, $\mathcal{P}(\mathbf{x})$ 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 $\varrho \to \infty$, we have

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

This means, as $\varrho \to \infty$, $\mathcal{P}(\mathbf{x}) = 0$ 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 $\mathcal{P}(\mathbf{x}) = 1$ 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 $\varrho \to \infty$. In other words, $\mathcal{P}(\mathbf{x}^\star) = 1$ 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 $\mathcal{P}(\mathbf{x}^\star) = 0$. 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 $\mathcal{P}(\mathbf{x}) = 0$. Hence, detection errors occur. The above explanation is argued at infinite SNR, but it is also valid for high SNRs because $\Phi(t)$ approaches 0 very fast.

To remove the product in (D.1), one may argue to transform the function $\mathcal{L}(\mathbf{x})$ into a sum of log functions as follows:

$$\hat{\mathbf{x}}_{\mathrm{d},m}^{\mathtt{ML}} = \arg\max_{\bar{\mathbf{x}}\in\mathcal{M}^U} \underbrace{\sum_{i=1}^{2N} \log\Phi\left(\sqrt{2\varrho}\,y_{\mathrm{d},m,i}\hat{\mathbf{h}}_{\mathrm{d},i}^T\mathbf{x}\right)}_{\mathcal{P}(\mathbf{x})}. \tag{D.2}$$

However, the function $\mathcal{P}(\mathbf{x})$ in (D.2) 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 [43] is also robust against imperfect CSI thanks to the use of the approximation $\Phi(t) \approx \frac{1}{2}e^{-0.374t^2-0.777t}$ for non-negative t. This approximation helps remove the effect of $\log\Phi(\cdot)$ in (D.2) since $\log e^a = a$. However, the OSD method has higher computational complexity than the proposed SVM-based methods.