CSI-Free Geometric Symbol Detection via Semi-supervised Learning and Ensemble Learning

Journal:	*IEEE Transactions on Communications*
Manuscript ID	TCOM-TPS-21-1382.R2
Manuscript Type:	Transactions Paper Submissions
Date Submitted by the Author:	27-Aug-2022
Complete List of Authors:	Zhang, Jianjun Masouros, Christos; UCL, Electronics Huang, Yongming; Southeast University, School of Information Science and Engineering
Keyword:	MIMO systems, Detectors, Unsupervised learning, Learning systems

CSI-Free Geometric Symbol Detection via Semi-supervised Learning and Ensemble Learning

Jianjun Zhang, *Member, IEEE*, Christos Masouros, *Senior Member, IEEE* and Yongming Huang, *Senior Member, IEEE*

Abstract

Symbol detection (SD) plays an important role in a digital communication system. However, most SD algorithms require channel state information (CSI), which is often difficult to estimate accurately. As a consequence, it is challenging for these SD algorithms to approach the performance of the maximum likelihood detection (MLD) algorithm. To address this issue, we employ both semi-supervised learning and ensemble learning to design a flexible parallelizable approach in this paper. First, we prove theoretically that the proposed algorithms can arbitrarily approach the performance of the MLD algorithm with perfect CSI. Second, to enable parallel implementation and also enhance design flexibility, we further propose a parallelizable approach for multi-output systems. Finally, comprehensive simulation results are provided to demonstrate the effectiveness and superiority of the designed algorithms. In particular, the proposed algorithms approach the performance of the MLD algorithm with perfect CSI, and outperform it when the CSI is imperfect. Interestingly, a detector constructed with received signals from only two receiving antennas (less than the size of the whole receiving antenna array) can also provide good detection performance.

Index Terms

MIMO detection, semi-supervised learning, ensemble learning, parallel detection, MIMO communications.

The work was supported by the Engineering and Physical Sciences Research Council, UK under project EP/S028455/1.

J. Zhang and C. Masouros are with the Department of Electronic & Electrical Engineering, University College London, London WC1E7JE, U.K. (E-mail: {jianjun.zhang,c.masouros} @ucl.ac.uk). Y. Huang is with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China. (Email: huangym@seu.edu.cn).

I. Introduction

Multiple-input multiple-output (MIMO) techniques that deploy multiple antennas at both transmitter (Tx) and receiver (Rx) have been receiving considerable attentions in wireless communications, thanks to the high spectral efficiency and link reliability [1]. Symbol detection (SD) is a fundamental function or task of the MIMO receiver, which is responsible for reliably recovering transmitted symbols from observed channel outputs [2]. Due to the importance of SD, a variety of SD algorithms have been proposed, e.g., with the goal of balancing the performance and complexity [3]–[6]. Among the existing detection algorithms, the maximum likelihood detection (MLD) algorithm can achieve optimal performance, which, however, has a prohibitive computational complexity as the number of antennas gets large. To tackle this issue, several sub-optimal but low-complexity algorithms (e.g., zero-forcing or minimum mean-squared error SD algorithms), are proposed, which, however, often incur large performance degradation compared with MLD. Although theoretically more receiving antennas provide larger diversity/array gains [7], SD is very challenging in the large-scale MIMO setting.

Based on approximate message passing (AMP) and/or expectation propagation [8], iterative MIMO detection algorithms have been proposed [5], [6], which can achieve good performance with moderate computational complexity. The key of the two algorithms approximates the posterior probability by utilizing the central limit theorem or factorized Gaussian distributions, which can achieve Bayes-optimal performance under certain conditions. However, the practical performance of the iterative detectors is far from their theoretical counterparts for more typical small-scale MIMO systems. Moreover, the performance deteriorates seriously for imperfect channel state information (CSI) or correlated channels [9]–[12]. Although good performance can be achieved for large-scale systems, it is challenging to estimate CSI.

Thanks to the powerful learning abilities, machine learning (ML) based algorithms have been developed to enhance physical layer wireless communications, varying from channel estimation, precoding, SD and CSI feedback [13]–[16]. Within various applications of ML, SD is almost the most typical, natural and mature one, as SD is essentially a classification problem. The existing learning-based SD algorithms can fall into different categories, which depends on the taxonomy or standard considered. The availability of CSI has an important impact on SD design. A profound taxonomy is defined via the standard - to what degree the CSI is available. Then, the

learning-based SD algorithms fall into two categories, i.e., CSI-aided SD designs [17]–[20] and CSI-agnostic SD designs [21]–[26]. In particular, learning-aided SD designs have been proposed to handle the case of few pilots (e.g., in the scenario of Internet-of-Things [21]) based on meta-learning [21], active learning scheme [22], self-supervision [23], and so on. For the second category, it can be further classified into two sub-categories, i.e., SD designs with CSI parameter unknown (but model available) [25], [26] and SD designs with CSI model even unavailable [21]–[24] (e.g., generative-adversarial-network enabled SD design in [22] applicable to non-Gaussian or time-varying channel models).

Another widely used taxonomy is concerned with underlying ML paradigm. Based on this taxonomy, the learning-based SD algorithms fall into two categories, i.e., data-driven SD designs [17]–[20], [27]–[31] and model-driven SD designs [23], [32]–[38]. The key idea of the data-driven SD algorithms is to construct a deep neural network (DNN) to predict key parameters of existing SD algorithms or transmitted symbols themselves. As an example, a DNN is constructed to predict the decoding radius of the sphere decoding algorithm [19], while a network, referred to as DetNet, is trained to predict transmitted symbols directly [17]. In particular, a weight scaling framework for DNN-based MIMO detection has been recently proposed [31], which can achieve accuracy-complexity scalability during inference. The foundation of the data-driven design paradigm is the universal approximation ability of DNNs. More precisely, the desired prediction function is treated as a black-box and driven by "big data".

In general, the data-driven black-box based algorithms require a huge number of training samples and suffer from poor interpretability. To overcome these drawbacks, model-driven algorithms, implemented via algorithm unfolding (AU), have been proposed recently. AU unfolds iterations of an existing iterative algorithm into a DNN-analogous layer-wise structure and optimizes relevant hyper-parameters via gradient descent and back-propagation methods [23], [32]–[37], [39]. By unfolding the orthogonal AMP algorithm, a model-driven deep learning network, referred to as OAMP-Net2, was proposed in [34] for MIMO detection. Recently, ViterbiNet has been proposed in [24], by integrating deep learning into the classical Viterbi algorithm. Similar to the Viterbi algorithm, an appealing feature of ViterbiNet is that CSI is not required, by exploiting the underlying Markovian structure of finite-memory causal channels.

In the previous literatures, most SD algorithms depend on CSI. The use of CSI causes at least two issues. First, estimating CSI occupies precious communication resources, which, as a result,

reduces communication efficiency. Second, it is almost impossible to estimate CSI accurately in practice, which inevitably limits detection performance. In fact, imperfect CSI makes it difficult to approach the performance of the MLD algorithm. To address these issues, SD algorithms without requiring CSI have been recently proposed in [24]–[26]. In particular, label-assisted SD approach is proposed in [25], where the phenomenon that received signals form multiple clusters is exploited to design SD algorithms. Although it is promising to achieve good performance, theoretical analysis is still absent. The Viterbi algorithm, which is the foundation of the algorithms in [24], limits the application scope and extension, although CSI is not required in [24].

In this paper, we incorporate semi-supervised learning (SSL) and ensemble learning (EL) to design a parallelizable CSI-free geometric detection approach. First, we theoretically prove that the proposed algorithms can arbitrarily approach the performance of the optimal MLD algorithm with perfect CSI. Simulation results further confirm that they outperform the MLD algorithm with imperfect CSI. Second, to enhance design flexibility and further reduce computational complexity, we propose a flexible parallelizable multi-antenna detection approach by incorporating SSL and EL. Finally, simulation results are provided to confirm the effectiveness and superiority of the designed SD algorithms. The main contributions are summarized as follows:

- By exploiting the clustering feature of received signals, we propose an efficient CSI-free SD approach. In particular, we theoretically show that the proposed SD approach can arbitrarily approach the performance of the optimal MLD algorithm with perfect CSI. In contrast to most SD algorithms whose detection performance depends on the quality of estimated CSI and/or the performance of CSI estimation algorithm, our approach is not affected by these factors.
- To enhance design flexibility and further reduce computational complexity, we incorporate SSL and EL to propose a parallelizable detection approach. Compared with the original approach, the novel approach is highly flexible, e.g., few modifications are involved when the number of receiving antennas changes, which facilitates the incorporation or integration of receiving antenna selection and MIMO detection.
- Comprehensive simulation results are provided to demonstrate the effectiveness and superiority of the proposed algorithms. Interestingly, a detector constructed with received signals from only two receiving antennas (less than the size of the whole receiving antenna array) provides a good detection performance as well.

The remainder of this paper is organized as follows. The system model of MIMO detection is described in Section II. Efficient SD algorithms and performance analysis are elaborated in Section III. An EL based SD approach is proposed in Section IV to enhance the original design. Simulation results and conclusions are given in Section V and Section VII, respectively. To improve readability, the proofs are deferred to appendices.

Notations: Bold uppercase $\mathbf{A}$ and bold lowercase $\mathbf{a}$ denote matrices and column vectors, respectively. Non-bold letters A, a denote scalars. $\mathbf{A}(i,:)$ and $\mathbf{A}(:,j)$ represent the i-th row and j-th column of matrix $\mathbf{A}$, respectively. Caligraphic letters $\mathcal{A}$ stand for sets. $\mathbb{E}(\cdot)$ and $(\cdot)^{\mathrm{H}}$ denote the mathematical expectation and Hermitian operators, respectively. $\mathbb{I}\{\cdot\}$ and $\mathrm{card}(\mathcal{A})$ represent the indicator function and cardinality of $\mathcal{A}$, respectively. $(\cdot)^\star$ represents an optimal quantity, e.g., an optimal solution. $\mathcal{CN}(\mathbf{m},\mathbf{R})$ stands for a complex Gaussian random vector with mean vector $\mathbf{m}$ and covariance matrix $\mathbf{R}$.

II. System Model

Consider a MIMO communication system, where Tx and Rx are equipped with N_{T} and N_{R} antennas, respectively. The task of Rx is to recover symbol vectors transmitted over a block-fading channel. Let $\mathbf{H}_k \in \mathbb{C}^{N_{\mathrm{R}}\times N_{\mathrm{T}}}$ represent channel matrix in time-block k. As shown in Fig. 1, the channel matrix keeps fixed within each time-block while varying across different time-blocks. Within each time-block, T symbol vectors are transmitted. Let $\mathbf{y}_{k,i}$ denote channel output at time-slot $i \in \{1,\cdots,T\}$ in time-block k. Similarly, the transmitted symbol vector and received noise vector at time-slot i in time-block k is denoted by $\mathbf{s}_{k,i}$ and $\mathbf{w}_{k,i}$, respectively.

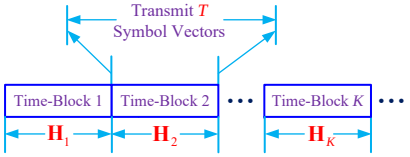

Fig. 1. An illustration of system model and time-block structure.

Without loss of generality, we concentrate on an arbitrary but fixed time-block, and omit the first subscript k throughout the paper. The number of data streams is equal to the number of transmit antennas. The input-output relationship is given by

$$\mathbf{y}_i = \sqrt{p}\mathbf{H}\mathbf{s}_i + \mathbf{w}_i, \tag{1}$$

where p denotes transmit power and $\mathbf{w}_i \sim \mathcal{CN}(0, \mathbf{I})$ denotes received random noise vector. The j-th component of $\mathbf{s}_i$ is assumed to be uniformly distributed over a constellation $\mathcal{C}_j$ of size $\mathrm{card}(\mathcal{C}_j) = M_j$. Hence, $\mathbf{s}_i$ is uniformly distributed over set $\mathcal{C} = \mathcal{C}_1 \times \cdots \times \mathcal{C}_{N_\mathrm{T}}$ of size $M_1 \times \cdots M_{N_\mathrm{T}}$. The symbol vector set $\mathcal{C}$ is denoted by $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_M\}$. Given CSI $\mathbf{H}$ and symbol vector $\mathbf{s}_i = \mathbf{c}_j$, the received signals $\{\mathbf{y}_i\}$ are independent (i.e., conditionally independent), no matter whether the CSI $\mathbf{H}$ across different time-blocks is independent or correlated.

The goal of this paper is to design an efficient SD algorithm to recover transmitted symbol vectors $\{\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_T\}$ from channel outputs $\{\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_T\}$ without knowing the channel matrix $\mathbf{H}$, i.e., CSI is unavailable for Rx. To address this issue, we will employ SSL and EL to design an efficient SD approach, in particular, based on geometrical intuition.

III. CSI-Free Geometric Detection Approach via Semi-Supervised Learning

In this section, we propose efficient CSI-free geometric SD algorithms via SSL. In particular, we show theoretically that the proposed algorithms can approach the performance of the MLD algorithm with perfect CSI.

Before proceeding, we first explain the CSI-free geometric SD approach from the perspective of ML. In contrast to supervised learning, where each sample has a label, and unsupervised learning, where no sample has a label, in SSL only a very small part of data samples have labels while most samples have no label. Two key operations are involved in SSL to obtain an effective learning model (e.g., a decision criterion). First, SSL extracts and excavates latent structure or feature information from unlabeled samples. In general, the unlabeled samples are easy to obtain, and thus the amount is often very large. Second, SSL incorporates the structure or feature information extracted from the unlabeled samples and the label information provided by the labeled samples. Specific to the CSI-free SD approach, the underlying principle is that received signals present the clustering feature, which relaxes the requirement of CSI.

A. CSI-Free Detection via SSL - MISO Case

To obtain an intuitive understanding, the MISO case (i.e., $N_\mathrm{R} = 1$ in (1)) is considered in this section. Then, the input-output relationship in (1) is simplified as

$$y_i = \sqrt{p}\mathbf{h}^\mathrm{H}\mathbf{s}_i + w_i. \tag{2}$$

As mentioned earlier, an important feature of the received signals $\mathcal{Y} = \{y_1, y_2, \cdots, y_T\}$ is that $\mathcal{Y}$ [1] often forms $\mathrm{card}(\mathcal{C}) = M$ clusters (typically, for moderate or high SNR settings), based on which an incomplete decision criterion can be obtained. Assisted by necessary supervised information, the incomplete decision criterion can be enhanced into a complete decision criterion. As shown in Fig. 2, the CSI-free SD method consists of three steps: (1) transmit supervised identification signal (SIS); (2) construct complete decision criterion; and (3) recover all transmitted symbol vectors.

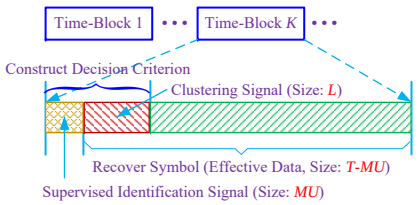

Fig. 2. Structure of time-block and key steps of CSI-free detection approach. Each time-block corresponds to a learning operation block. In particular, the CSI is assumed to be static within each time-block.

1) Transmit Supervised Identification Signal: At the beginning of each time-block, vector $\mathbf{c}_1$ is transmitted first and repeated U times, which yields received signals $y_1, \cdots, y_U$. The received signals are collected into $\mathcal{Y}_{\mathbf{c}_1} = \{y_1, \cdots, y_U\}$. Then, vector $\mathbf{c}_2$ is transmitted and repeated U times, which yields received signals $\mathcal{Y}_{\mathbf{c}_2} = \{y_{U+1}, y_{U+2}, \cdots, y_{2U}\}$. The other symbol vectors in $\mathcal{C}$ are operated similarly, which yields received signal sets $\mathcal{Y}_{\mathbf{c}_3}, \cdots, \mathcal{Y}_{\mathbf{c}_M}$. Note that because of the predefined transmission mode, we can easily recognize the transmitted symbol vectors. An example is provided in Fig. 3-(a).

The signals $\mathcal{Y}_{\mathbf{c}_1}, \cdots, \mathcal{Y}_{\mathbf{c}_M}$ are referred to as SISs. The important role of SIS will be clear later. With $\mathcal{Y}_{\mathbf{c}_1}, \cdots, \mathcal{Y}_{\mathbf{c}_M}$ available, one may try to construct a decision criterion. However, because the number of available samples is very small and the samples are corrupted by noise, the constructed decision criterion is highly inaccurate and unreliable. Although increasing U helps to construct

[1]Here, we ignore the mathematical rigor of the definition of the set, i.e., a set should not contain repetitive elements. In fact, the probability that two (or more) received signals (e.g., y_i and y_j) are equal is zero due to the continuous distribution of $\{w_i\}$. Even if they are equal they should not be merged, because each received signal corresponds to a transmitted symbol vector. The elements within $\mathcal{Y}$ should also be recorded chronologically, so as to recover the transmitted symbols correctly, which naturally introduces an order.

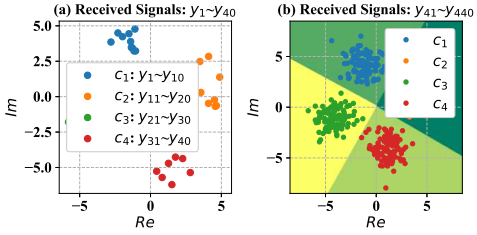

Fig. 3. A visualization of the CSI-free SD approach ($N_{\mathrm{T}} = 1$ and QPSK modulation with $\mathcal{C} = \{e^{j\pi/4}, e^{j3\pi/4}, e^{j5\pi/4}, e^{j7\pi/4}\}$): (a) the constellation of received signals $y_1 \sim y_{40}$; (b) the constellation of received signals $y_{41} \sim y_{440}$. Note that the symbols corresponding to $y_1 \sim y_{40}$ are transmitted as per a predefined mode, while the symbols corresponding to $y_{41} \sim y_{440}$ are transmitted randomly (depending on specific effective data.)

a better decision criterion, the transmission efficiency is also reduced, since $\mathcal{Y}_{\mathbf{c}_1}, \cdots, \mathcal{Y}_{\mathbf{c}_M}$ do not contain effective data.

2) Construct Complete Decision Criterion: Note that because received signals corresponding to effective data also present the clustering feature, we can construct an "incomplete" decision criterion based on the received signals corresponding to effective data. Let $\{y_{MU+1}, y_{MU+2}, \cdots, y_{MU+L}\}$ represent the received signals used for clustering, where L denotes the size of the set. Since enough data samples are available and more data samples usually yield better performance, L can and should be as large as possible, if the computing resources are allowed and the resultant time-delay is tolerant. To better describe the proposed approach, the concepts of Voronoi cell (VC) and Voronoi tessellation (VT) are used here, whose definitions are given below [40].

Definition 1. *The Voronoi cell $\mathcal{V}(\boldsymbol{x})$ of a point $\boldsymbol{x}$ of a point set $\Phi \subset \mathbb{R}^d$ consists of locations of $\mathbb{R}^d$ whose distance to $\boldsymbol{x}$ is not greater than the distance to any other point in Φ, i.e.,*

$$\mathcal{V}(\boldsymbol{x}) = \left\{\boldsymbol{y} \in \mathbb{R}^d \,|\, \|\boldsymbol{y} - \boldsymbol{x}\| < \|\boldsymbol{y} - \boldsymbol{z}\|, \forall\, \boldsymbol{z} \in \Phi \backslash \{\boldsymbol{x}\}\right\}. \tag{3}$$

A Voronoi tessellation is a decomposition of space $\mathbb{R}^d$ into the Voronoi cells of a point set, but ignoring the boundaries.

Essentially, invoking a clustering algorithm to $\{y_{MU+1}, y_{MU+2}, \cdots, y_{MU+L}\}$ divides the complex plane $\mathbb{C}$ into M disjoint regions (i.e., VCs) $\mathcal{R}_1, \mathcal{R}_2, \cdots, \mathcal{R}_M$, i.e.,

$$\mathcal{R}_1 \cup \mathcal{R}_2 \cup \cdots \cup \mathcal{R}_M = \mathbb{C} \quad \text{and} \quad \mathcal{R}_i \cap \mathcal{R}_j = \emptyset\, (\forall i \neq j), \tag{4}$$

which form a classification criterion. An example is provided in Fig. 3-(b). However, the obtained criterion is incomplete from the perspective of SD, since symbols cannot be recovered based on the criterion. In fact, for a received signal y_n, we cannot identify which symbol vector in $\mathcal{C}$ it corresponds to, even though we know that it lies in a VC, e.g., $\mathcal{R}_i$.

To obtain a complete decision criterion, we need to associate each symbol vector in $\mathcal{C}$ to a VC. We propose two simple but efficient association methods. Note that it is possible that two or more $\{\mathbf{c}_i\}$ are assigned to one cluster, e.g., when the SNR is too low. But, the probability decreases dramatically as the SNR increases. The first one is voting method. Specifically, the voting method associates symbol vector $\mathbf{c}_j \in \mathcal{C}$ to VC $\mathcal{R}_i$ if and only if

$$\sum_{y_n \in \mathcal{Y}_{\mathbf{c}_j}} \mathbb{I}\{y_n \in \mathcal{R}_i\} > \sum_{y_n \in \mathcal{Y}_{\mathbf{c}_j}} \mathbb{I}\{y_n \in \mathcal{R}_k\}, \quad (\forall\, k \neq i). \tag{5}$$

The second one is centroid method. For each SIS set $\mathcal{Y}_{\mathbf{c}_j}$, we first compute its centroid $w_{\mathbf{c}_j}$, i.e.,

$$w_{\mathbf{c}_j} = \frac{1}{U} \sum_{y_n \in \mathcal{Y}_{\mathbf{c}_j}} y_n. \tag{6}$$

The centroid method associates symbol vector $\mathbf{c}_j \in \mathcal{C}$ to VC $\mathcal{R}_i$ if and only if $w_{\mathbf{c}_j} \in \mathcal{R}_i$. For either of the two association methods, a complete decision criterion is formally denoted by $\{(\mathbf{c}_j, \mathcal{R}_i) \,|\, \mathbf{c}_j \in \mathcal{C}\}$, i.e., $\mathbf{c}_j$ is associated to $\mathcal{R}_i$.

3) Recover Transmitted Symbol Vectors: With a complete decision criterion $\{(\mathbf{c}_j, \mathcal{R}_i) \,|\, \mathbf{c}_j \in \mathcal{C}\}$ available, the transmitted symbol vectors can be recovered from the received signals easily: the symbol vector corresponding to an arbitrary received signal y_n is $\mathbf{c}_j$ if and only if $y_n \in \mathcal{R}_i$. Note that when constructing the complete decision criterion, the symbol vectors corresponding to received signals $\{y_{MU+1}, \cdots, y_{MU+L}\}$ have already been recovered.

Algorithm 1: CSI-Free Geometric SD Algorithm (MISO Case)
1: **input:** $\mathcal{C}$ with $\mathrm{card}(\mathcal{C}) = M$ - set of transmitted symbol vectors; U - number of SISs; L - number of signals used for clustering
2: **transmit** SISs for each symbol vector in $\mathcal{C}$
3: **construct** decision criterion
(a) **choose** L received signals (of effective data vectors)
(b) **invoke** clustering algorithm to generate an incomplete decision criterion $\Longrightarrow \{\mathcal{R}_1, \cdots, \mathcal{R}_M\}$
(c) **associate** each symbol vector in $\mathcal{C}$ to a VC to obtain a complete decision criterion
4: **recover** all symbol vectors from received signals based on the complete decision criterion
5: **output:** recovered symbol vectors $\{\hat{\mathbf{s}}_{MU+1}, \cdots, \hat{\mathbf{s}}_T\}$

For clarity, the CSI-free geometric SD algorithm is summarized in Algorithm 1. The input includes the set of transmitted symbol vectors (i.e., the Cartesian product space of constellations), number of SISs U and number of received signals used for clustering L. In step 2, SISs are transmitted for identification. In step 3-(a), L received signals corresponding to effective data $\{y_{MU+1}, y_{MU+2}, \cdots, y_{MU+L}\}$ are chosen for clustering. A clustering algorithm (e.g., the k-means algorithm) is invoked to divide the complex plane $\mathbb{C}$ into M VCs (or a VT) in step 3-(b). In step 3-(c), SISs are utilized to associate each symbol vector in $\mathcal{C}$ to a VC. In step 4, all symbol vectors are recovered based on the obtained complete decision criterion.

From the view of ML, SISs can be regarded as the training samples, which do not convey effective information. Hence, as shown in Fig. 2, the transmission efficiency can be reasonably defined as MU/T, with T denoting the size of total transmitted symbol vectors within each time-block. As shown in Theorem 2, since U is often very small, the number of training samples is also very small. Although there exist many clustering algorithms, in this paper we focus on the k-means algorithm [41], which is one of the most widely used clustering algorithms in practice. [2] For completeness, the algorithm is provided in Appendix A, which clearly shows that the clustering algorithm (and thus the corresponding SD algorithm) generates VCs or VTs. Note that similar to the vector quantization [42], which is model agnostic, the developed SD approach is also applicable to other channel models, e.g., any memoryless MIMO channel model.

Note that Algorithm 1 can be directly extended to handle MIMO case. In fact, many algorithms (e.g., the k-means or Lloyd-Max algorithm) mainly use the metric/distance structure of vector space $\mathbb{C}^{N_{\mathrm{T}}}$ and there exists a well-defined metric/distance function on $\mathbb{C}^{N_{\mathrm{T}}}$. Moreover, the key operation of the derived SD algorithm is to cluster complex vectors of high dimension.

B. Performance Analysis and Optimization

In this subsection, we will show that the designed algorithm can approach the MLD algorithm with perfect CSI under some mild conditions. It is assumed that all symbol vectors in $\mathcal{C}$ are

[2]The k-means algorithm is, in fact, closely related to vector quantization design [42], especially the well-known Lloyd-Max algorithm. The reason why it is chosen here is that its terseness facilitates theoretical analysis. But, it should be noted that since the proposed SD approach is essentially a SD framework, it accommodates various clustering algorithms (e.g., KNN, Gaussian mixture model (GMM) and spectral clustering [43]) and can be easily extended by choosing other clustering algorithms, which may yield better SD algorithms. As an example, when the GMM-based clustering algorithm [43] is incorporated into the SD framework, the SD algorithm yielded can provide soft outputs instead of hard decisions.

transmitted equiprobably. Since a decision criterion is essentially equivalent to a VT, the key of the proof is to identify and prove the following two key hypotheses (also facts):

- The "difference" between the VTs generated by the MLD algorithm with perfect CSI and the k-means clustering algorithm can be arbitrarily small in some sense.
- Each symbol vector in $\mathcal{C}$ can be correctly associated to the corresponding VC with a high probability.

To begin with, we investigate the first hypothesis and prove its validity (as a fact) rigorously. Note that since the k-means algorithm generates a VT, we need to show that the MLD algorithm also generates a VT. In view that a VT is uniquely determined by a point set, we can equivalently consider the point set. Let $\{w_{\mathbf{c}_j} \,|\, \mathbf{c}_j \in \mathcal{C}, j = 1, \cdots, M\}$ be the point set generated by the k-means algorithm. The VCs generated by the two algorithms (i.e., the MLD algorithm and the k-means algorithm) are characterized in the following lemma.

Lemma 1. *(1) Let $\hat{\mathbf{c}}(y)$ represent the estimated symbol vector of received signal y. Then, the MLD algorithm with perfect CSI generates a VT with point set $\{w_{\mathbf{c}_j}^{MLD} = \sqrt{p}\mathbf{h}^H\mathbf{c}_j \,|\, j = 1, \cdots, M\}$, and $\hat{\mathbf{c}}(y) = \mathbf{c}_j$ if and only if $y \in \mathcal{V}(w_{\mathbf{c}_j}^{MLD})$. (2) Let $\varepsilon > 0$ be an arbitrary real number. If each symbol vector $\mathbf{c}_j \in \mathcal{C}$ is transmitted K_j times and each point $w_{\mathbf{c}_j}$ is generated by the centroid method, the following inequality holds*

$$\mathbb{P}(\max_j |w_{\mathbf{c}_j} - w_{\mathbf{c}_j}^{MLD}| \leq \varepsilon) \geq 1 - \sum_{j=1}^{M} K_j \exp\Big(-pd_{min}^2\Big) - \sum_{j=1}^{M} \exp\big(-K_j\varepsilon^2\big), \tag{7}$$

where the minimum distance d_{min} is given by

$$d_{min} = 0.5 \min\big\{|\mathbf{h}^H\mathbf{c}_j - \mathbf{h}^H\mathbf{c}_k| \,\big|\, \forall\, \mathbf{c}_j, \mathbf{c}_k \in \mathcal{C}, j \neq k\big\} > 0.$$

Proof: See Appendix B. ■

Based on Lemma 1, we can prove that the CSI-free geometric SD algorithm can approach the MLD algorithm with perfect CSI, as characterized in the following theorem.

Theorem 1. *The symbol error rate (SER) of the MLD algorithm with perfect CSI (or the CSI-free geometric SD algorithm) is denoted by Θ^{MLD} (or Θ). Then, for an arbitrary (but fixed) real number $\delta > 0$, there exist $\{K_j\}$ and $C > 0$ such that as long as $pd_{min}^2 > C$ (without requiring*

$p \to \infty$), $\mathbb{P}(|\Theta - \Theta^{MLD}| < \delta)$ can be lower bounded by

$$\mathbb{P}\big(\big|\Theta - \Theta^{MLD}\big| < \delta\big) \geq 1 - \sum_{j=1}^{M} \bigg(K_j \exp\big(-pd_{min}^2\big) + \exp\big(-K_j\varepsilon^2\big) \bigg). \tag{8}$$

Proof: Notice that (1) both the CSI-free geometric SD algorithm and the MLD algorithm are based on VCs (or VTs); and (2) the SER performance metric of the MLD algorithm is continuous with respect to $\{w_{\mathbf{c}_j}^{\text{MLD}}\}$. Therefore, as long as pd_{min}^2 and $\{K_j\}$ are sufficiently large, $\max_j |w_{\mathbf{c}_j} - w_{\mathbf{c}_j}^{\text{MLD}}|$ can be sufficiently small, with probability (at least) given in (8). As a result, the gap of SER performance between the two algorithms can also be arbitrarily small, with probability (at least) given in (8). ■

Theorem 1 indicates that the CSI-free geometric SD algorithm can arbitrarily approach the MLD algorithm with perfect CSI, as long as pd_{min}^2 and $\{K_j\}$ are sufficiently large. Mathematically, if the error measure $\delta > 0$ in Theorem 1 (or $\varepsilon > 0$ in Lemma 1) is very small (e.g., tends to zero), the required $\{K_j\}$ may be very large and, as a result, the entire dataset is even consumed for clustering. However, from the practical point of view, there is no need to choose a very small value for the error measure and the revealed insights also hold true for a reasonably large value. In general, for a reasonable value, the number of required data points is small or moderately large. In fact, since $\exp\big(-K_j\varepsilon^2\big)$ and $\exp\big(-pd_{\text{min}}^2\big)$ are both exponential functions and thus decay very quickly as $\{K_j\}$ and pd_{min}^2 become large, the performance gap between the two algorithms is small even for a moderate value. Moreover, even if $\{K_j\}$ may be very large, the transmission efficiency is not affected, because the data samples used for clustering correspond to effective data. In practice, the CSI-free geometric SD algorithm often outperforms the MLD algorithm with imperfect CSI, as demonstrated in Section V.

The number U has an important influence on system performance. On the one hand, if U is too small, a symbol vector may be incorrectly associated to a VC, which leads to bad performance. On the other hand, if U is too large, precious communication time resource will be wasted. Therefore, it is very desirable to provide an explicit expression to guide the choice of U.

To derive an analytical expression, we make two reasonable assumptions. The first one is that the two sets $\{w_{\mathbf{c}_j} \,|\, j = 1, \cdots, M\}$ and $\{w_{\mathbf{c}_j}^{\text{MLD}} = \sqrt{p}\mathbf{h}^{\text{H}}\mathbf{c}_j \,|\, j = 1, \cdots, M\}$ coincide. According to Lemma 1, $\max_j |w_{\mathbf{c}_j} - w_{\mathbf{c}_j}^{\text{MLD}}|$ can be arbitrarily small. Hence, it is reasonable to make this assumption, so as to simplify the derivation. The second one is that $d_{\text{min}} \geq V_0$, where $V_0 > 0$ is a threshold value. Then, a lower bound of U is analytically characterized in the following

theorem.

Theorem 2. *Based on the above two assumptions, the probability that all symbol vectors are correctly associated to the corresponding VCs, denoted by P_{AC}, is lower bounded by*

$$P_{AC} \geq 1 - M \exp\big(-pUd_{min}^2\big) \geq 1 - M \exp\big(-pUV_0^2\big). \tag{9}$$

For an arbitrary real number $0 < \varepsilon \ll 1$, $P_{AC} \geq 1 - \varepsilon$ if

$$U \geq \left\lceil \frac{\ln(M/\varepsilon)}{pV_0^2} \right\rceil. \tag{10}$$

Proof: See Appendix C. ■

The threshold value V_0 depends on the channel model, i.e., the distribution of channel vector $\mathbf{h}$. Thanks to the logarithmic term $\ln(M/\varepsilon)$ in (10), the number of required SISs (for each symbol vector) is still small, even if $\varepsilon > 0$ may be very small. However, the total number of SISs is $M\lceil p^{-1}V_0^{-2}\ln(M/\varepsilon)\rceil$, which may be large when M is too large (e.g., high-order modulation or large N_{T}). In the next subsection, we will propose an effective method to tackle this issue.

C. Extension

To reduce the overhead of transmitting SISs, a method has been proposed in [25]. The basic idea is as follows. Let matrix $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_M]$ collect all symbol vectors in $\mathcal{C}$. By decomposing $\mathbf{C}$ as $\mathbf{C}_{N_{\mathrm{T}}\times M} = \mathbf{U}_{N_{\mathrm{T}}\times N_{\mathrm{T}}}\mathbf{V}_{N_{\mathrm{T}}\times M}$ (e.g., via the QR-decomposition or full-rank decomposition), the columns of $\mathbf{U}$ are regarded as SISs. The received signals can be written as

$$\mathbf{r}^{\mathrm{H}} = \sqrt{p}\mathbf{h}^{\mathrm{H}}\mathbf{U} + \mathbf{w}, \tag{11}$$

where $\mathbf{w}$ is the received noise vector. With $\mathbf{r}$ available, $\mathbf{h}^{\mathrm{H}}\mathbf{c}_i$ can be estimated as $\mathbf{r}^{\mathrm{H}}\mathbf{v}_i/\sqrt{p}$, where $\mathbf{v}_i$ is the i-th column of $\mathbf{V}$.

It is not difficult to understand that the quality of $\mathbf{h}^{\mathrm{H}}\mathbf{U}$ estimated greatly affects the performance of the SD algorithm. To tackle this issue, we propose an efficient method by making full use of the received signals of effective data. Let $\mathbf{C}_{\mathrm{S}}$ (of size $N_{\mathrm{T}} \times N_{\mathrm{T}}$) represent an arbitrary sub-matrix of $\mathbf{C}$ such that: (1) each column of $\mathbf{C}_{\mathrm{S}}$ is chosen from $\mathcal{C}$; and (2) $\mathbf{C}_{\mathrm{S}}$ is invertible [3]. The columns of $\mathbf{C}_{\mathrm{S}}$ are collected into $\mathcal{C}_{\mathrm{S}}$. The method to estimate $\{\mathbf{h}^{\mathrm{H}}\mathbf{c}_j \,|\, \mathbf{c}_j \in \mathcal{C}\}$ is as follows.

[3]In practice, we also hope/require that the condition number of matrix $\mathbf{C}_{\mathrm{S}}$ can be as small as possible. This can be achieved via exhaustive search (if $\mathrm{card}(\mathcal{C})$ is small) or random search (if $\mathrm{card}(\mathcal{C})$ is large).

1) Transmit Supervised Identification Signal: Each element of $\mathcal{C}_{\rm S}$ is regarded as SIS and is transmitted U times with U given by

$$U \geq \lceil -p^{-1}V_0^{-2}\ln(\varepsilon)\rceil. \tag{12}$$

As per Theorem 2, the probability that an element of $\mathcal{C}_{\rm S}$ (e.g., $\mathbf{c}_i$) is correctly associated to its VC is greater than $1-\varepsilon$.

2) Reestimate $\{\mathbf{h}^H\mathbf{c}_i \,|\, \mathbf{c}_i \in \mathcal{C}_S\}$ More Accurately: Let $\hat{w}_{\mathbf{c}_i}$ represent the centroid of symbol vector $\mathbf{c}_i \in \mathcal{C}_{\rm S}$ estimated by averaging the received signals in terms of SIS. But, instead of using $\{\hat{w}_{\mathbf{c}_i} \,|\, \mathbf{c}_i \in \mathcal{C}_{\rm S}\}$ to estimate the centroids of symbol vectors in $\mathcal{C}\backslash\mathcal{C}_{\rm S}$, we estimate $\{\hat{w}_{\mathbf{c}_i} \,|\, \mathbf{c}_i \in \mathcal{C}_{\rm S}\}$ by incorporating the received signals of effective data, so as to achieve much higher accuracy.

The VC associated to $\mathbf{c}_i \in \mathcal{C}_{\rm S}$ is denoted by $\mathcal{R}_{j_i}$. Let $\mathcal{Y} = \{y_{UN_{\rm T}+1}, y_{UN_{\rm T}+2}, \cdots, y_T\}$ collect the received signals of effective data. To reestimate $w_{\mathbf{c}_j}$, we first choose a subset from $\mathcal{Y} \cap \mathcal{R}_{j_i}$ (or use the entire set $\mathcal{Y} \cap \mathcal{R}_{j_i}$) and denote the set by $\mathcal{Y}_{\mathbf{c}_i}$. By averaging all elements in $\mathcal{Y}_{\mathbf{c}_i}$, we can reestimate $w_{\mathbf{c}_i}$ more accurately. The newly estimated centroid of $w_{\mathbf{c}_i}$ is denoted by $\bar{w}_{\mathbf{c}_i}$.

3) Recover Other Centroids: With $\{\bar{w}_{\mathbf{c}_i} \,|\, \mathbf{c}_i \in \mathcal{C}_{\rm S}\}$ available, we can recover other centroids, i.e., $\{\bar{w}_{\mathbf{c}_j} \,|\, \mathbf{c}_j \in \mathcal{C}\backslash\mathcal{C}_{\rm S}\}$. Specifically, the centroid of each symbol vector $\mathbf{c}_j \in \mathcal{C}\backslash\mathcal{C}_{\rm S}$ is calculated or estimated as

$$\bar{w}_{\mathbf{c}_j} = \mathbf{g}\mathbf{C}_{\rm S}^{-1}\mathbf{c}_j, \tag{13}$$

where $\mathbf{g} = [\bar{w}_{\mathbf{c}_{i_1}}, \bar{w}_{\mathbf{c}_{i_2}}, \cdots, \bar{w}_{\mathbf{c}_{i_{N_{\rm T}}}}] \approx \mathbf{h}^{\rm H}\mathbf{C}_{\rm S}$ is a row vector collecting the centroids of symbol vectors in $\mathcal{C}_{\rm S}$, i.e., $\bar{w}_{\mathbf{c}_{i_t}}$ is the centroid of the t-th column vector of matrix $\mathbf{C}_{\rm S}$.

Remark 3.1 Compared with the use of $\{\hat{w}_{\mathbf{c}_i} \,|\, \mathbf{c}_i \in \mathcal{C}_{\rm S}\}$ to estimate $\{\hat{w}_{\mathbf{c}_j} \,|\, \mathbf{c}_j \in \mathcal{C}\backslash\mathcal{C}_{\rm S}\}$ directly, the proposed method greatly increases the estimation accuracy, without incurring any extra training overhead. Moreover, compared with the original method, the overhead of transmitting SIS has been greatly reduced, therefore improving the transmission efficiency from MU/T to $N_{\rm T}U/T$.

Based on the above discussion, we can obtain a more efficient SD algorithm, which can be applicable in more challenging cases. For clarity, the SD algorithm is summarized in Algorithm 2. The input also includes a real number $d > 0$, whose role will be clear later. In step 2, an appropriate sub-matrix $\mathbf{C}_{\rm S}$ is chosen from matrix $\mathbf{C}$. The criterion to choose $\mathbf{C}_{\rm S}$ in practice is that the condition number of $\mathbf{C}_{\rm S}$ should be as small as possible, although an arbitrary invertible matrix $\mathbf{C}_{\rm S}$ works as well. After transmitting SIS for $\mathcal{C}_{\rm S}$, we next reestimate the centroids, which

Algorithm 2: CSI-Free Geometric SD Algorithm (MISO Case)

1: **input:** $\mathcal{C}$ with $\mathrm{card}(\mathcal{C}) = M$ - set of transmitted symbol vectors; U - number of SISs; $d > 0$ - radius of disk in the complex plane

2: **choose** sub-matrix $\mathbf{C}_{\mathrm{S}}$ (of size $N_{\mathrm{T}} \times N_{\mathrm{T}}$) from matrix $\mathbf{C}$

3: **transmit** SISs for each symbol vector in $\mathcal{C}_{\mathrm{S}}$

4: **reestimate** centroids for symbol vectors in $\mathcal{C}_{\mathrm{S}}$ - for each $\mathbf{c}_i \in \mathcal{C}_{\mathrm{S}}$:

(a) **calculate** centroid of low accuracy $\hat{w}_{\mathbf{c}_i}$ via SIS received signals

(b) **choose** a subset from received signals for reestimating centroid $\mathcal{Y}_{\mathbf{c}_i} = \mathcal{Y} \cap \mathcal{B}(\hat{w}_{\mathbf{c}_i}, r)$

(c) **reestimate** centroid of high accuracy $\bar{w}_{\mathbf{c}_i}$ by averaging all elements in set $\mathcal{Y}_{\mathbf{c}_i}$

5: **estimate** centroids of other symbol vectors in $\mathcal{C} \backslash \mathcal{C}_{\mathrm{S}}$

6: **construct** a complete decision criterion based on all centroids

7: **recover** all symbol vectors based on the complete decision criterion

8: **output:** recovered symbol vectors $\{\hat{\mathbf{s}}_{N_{\mathrm{T}}U+1}, \cdots, \hat{\mathbf{s}}_T\}$

consists three steps. First, we estimate a coarse centroid $\hat{w}_{\mathbf{c}_i}$ for each $\mathbf{c}_i \in \mathcal{C}_{\mathrm{S}}$ by averaging the received signals in terms of SIS. Then, we choose an appropriate subset from the set of received signals $\mathcal{Y} = \{y_{UN_{\mathrm{T}}+1}, y_{UN_{\mathrm{T}}+2}, \cdots, y_T\}$. A simple but effective method is to choose the subset as $\mathcal{Y}_{\mathbf{c}_i} = \mathcal{Y} \cap \mathcal{B}(\hat{w}_{\mathbf{c}_i}, r)$ [4] for each $\mathbf{c}_i \in \mathcal{C}_{\mathrm{S}}$. Finally, we can obtain a more accurate centroid $\bar{w}_{\mathbf{c}_i}$ by averaging all elements in set $\mathcal{Y}_{\mathbf{c}_i}$. With $\{\bar{w}_{\mathbf{c}_i} \,|\, \mathbf{c}_i \in \mathcal{C}_{\mathrm{S}}\}$ available, we can further estimate the centroids for $\mathcal{C} \backslash \mathcal{C}_{\mathrm{S}}$ via (13). With all centroids available, we can construct a complete decision criterion and finally recover all transmitted symbol vectors.

Compared with Algorithm 1, another important advantage of the above SD algorithm is that there is no need to invoke the clustering algorithm, which significantly reduces the computational complexity.

In view that there is often no need to activate all receiving antennas (e.g., via receiving antenna selection), we will propose an EL-based SD algorithm in the next section. Compared with a SD algorithm extending directly from the previous algorithms, the EL-based algorithm has the following benefits. First, it is naturally and seamlessly compatible with receiving antenna selection. Second, it facilitates efficient parallel implementation and modular design, and provides better design flexibility. Third, clustering within the EL-based solution involves low dimensional

[4] $\mathcal{B}(z_0, d)$ is defined as $\mathcal{B}(z_0, d) = \{z \in \mathbb{C} \,|\, |z - z_0| \leq d\}$, i.e., the disk in the complex plane $\mathbb{C}$ with center z_0 and radius d. It is observed that d characterizes the number of received signals used for reestimating the centroid. An empirical formula to choose d is $d = c_0\sqrt{p}$, where c_0 is a proportionality constant and p denotes the transmit power.

data, leading to lower complexity.

IV. CSI-Free Geometric Detection Approach via Ensemble Learning

In this section, we propose a flexible CSI-free SD algorithm via EL, which enables both parallel implementation and flexible modular design. The key intuition underlying the EL-based algorithm is that although the number of clusters observed from a single antenna may be less than the product of sizes of the constellations (due to channel coupling effect), the overlapped clusters observed from one antenna may be separated when observed from another antenna. Based on this intuition, we can design an efficient SD algorithm by aggregating multiple MISO detectors. For convenience, we first briefly introduce EL and then design an efficient SD algorithm.

A. A Brief Introduction to Ensemble Learning

In contrast to ordinary ML approaches which try to construct only one learner from training dataset, to solve the same problem, EL tries to construct a number of learners and combines them together [44]. Therefore, EL is also referred to as committee-based learning [44]. As shown in Fig. 4, an EL model contains a set of learners, which are referred to as base learners (BLs). BLs can be decision tree, neural network or other kinds of learning algorithms. If an EL model uses a single base learning algorithm to produce homogeneous BLs, i.e., learners of the same type, it is referred to as homogeneous EL. An important advantage of homogeneous EL is that it has sufficient flexibility and facilitates modular design.

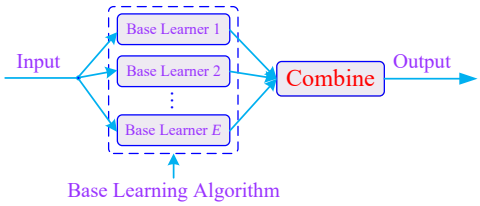

Fig. 4. An illustration of the typical architecture of ensemble learning.

In general, an EL model is constructed via two steps, i.e., (1) generate BLs and (2) combine them together. To obtain a good EL model, it is generally believed that the BLs should be as accurate as possible, and as diverse as possible [44]. Note that in a general ML setting, it is difficult to meet the two requirements simultaneously. To address this challenging issue, we

propose to regard a MIMO communication system as multiple MISO subsystems in this paper. Because the channels of different MISO subsystems experience independent fadings, the BLs (i.e., homogeneous BLs) generated by the same base learning algorithm keep sufficient diversity.

After generating a number of BLs, the other important step is to combine these BLs to generate a powerful learner, which has a strong generalization ability. Since the combination step plays a crucial role in EL, various combination methods have been proposed [44], among which voting is the most popular and fundamental combination method. A voting method usually consists of two key steps, i.e., (1) define a voting function (which can be either linear or nonlinear) and (2) specify a voting manner (e.g., the widely used weighted average method or the taking maximum method). Thanks to the simplicity and good performance, in the next subsection we will adopt the voting method to generate a complete learner.

B. CSI-Free Geometric SD via Ensemble Learning

To enable modular design and parallel implementation, the $N_{\mathrm{R}} \times N_{\mathrm{T}}$ MIMO system is regarded as N_{R} MISO subsystems and each MISO subsystem is associated with a BL. [5] By this means, the designed algorithm can inherit the advantages and features of homogeneous EL, typically, flexibility and modular design. Moreover, the designed algorithm is ideally compatible with receiving antenna selection techniques. For example, if the number of active receiving antennas is n, it is sufficient to design n BLs. In particular, since the BLs are homogeneous, there is no need to redesign an algorithm to produce the BLs. When increasing/decreasing receiving antennas, we only need to increase/decrease the homogeneous BL module units.

Without loss of generality, we assume that all receiving antennas are active. Let $\mathbf{h}_i$ denote the channel vector between the transmit antenna array and the i-th receiving antenna. The channel matrix can be compactly written as

$$\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_{N_{\mathrm{R}}}]^{\mathrm{H}}. \tag{14}$$

Let matrices $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_T]$, $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_T]$ and $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_T]$ collect (all) transmitted symbol vectors, received signal vectors and noise vectors, respectively. Then, the

[5] Although the independence is often assumed and/or required in an EL-based algorithm design, simulation results in Section V demonstrate that the developed EL-based SD algorithm works well even if these MISO subsystems are not independent (e.g., for correlated channels). Moreover, interesting conclusions and insights revealed in the previous sections also hold true.

input-output relationship can be written as

$$\mathbf{Y}_{N_{\mathrm{R}}\times T} = \sqrt{p}\mathbf{H}_{N_{\mathrm{R}}\times N_{\mathrm{T}}}\mathbf{S}_{N_{\mathrm{T}}\times T} + \mathbf{W}_{N_{\mathrm{R}}\times T}. \tag{15}$$

It is observed that designing a BL for a MISO subsystem is essentially equivalent to designing a detector that utilizes only one row (or a part of the row) of matrix $\mathbf{Y}$, e.g., $\mathbf{Y}(i,:)$. As the first step of employing EL, Algorithm 1 (or Algorithm 2) is chosen as the base learning algorithm to generate BLs. The BL corresponding to the i-th receiving antenna is denoted by RBL-i.

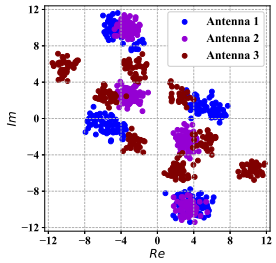

Fig. 5. Constellations or clusters of received signals: transmit power 10dB, $N_{\mathrm{R}} = N_{\mathrm{T}} = 3$ and BPSK. The number of clusters varies across different receiving antennas. Note that a detector constructed based on the received signals of antenna 1 or 2 fails to recover all symbols. But a detector designed based on the received signals of antenna 3 can recover all symbols.

Next, we propose an efficient combination method. If a high-order modulation is used and/or the SNR is relatively low, it may be difficult to identify all clusters of received signals of a MISO subsystem. The number of clusters may also vary across different receiving antennas for a MIMO system. In fact, because of the coupling or mixing effect of channel vectors, one cluster may correspond to multiple transmitted symbol vectors. As a result, multiple symbol vectors may associate to one VC. An example is provided in Fig. 5, where the number of receiving antennas is 3 and the numbers of clusters corresponding to the three antennas are 4, 4 and 8, respectively.

The received signals used by RBL-i for clustering are collected into $\mathcal{Y}^i = \{y^i_{MU+1}, y^i_{MU+2}, \cdots, y^i_{MU+L}\}$. By applying a clustering algorithm to $\mathcal{Y}^i$, the complex plane is divided into M_i ($1 \leq M_i \leq M$) disjoint VCs $\mathcal{R}^i_1, \cdots, \mathcal{R}^i_{M_i}$. By using an association method (e.g., the centroid method), an "ambiguous" decision criterion can be obtained, which is denoted by (for each RBL-i)

$$\{(\mathcal{S}^i_{k_j}, \mathcal{R}^i_j) \,|\, \mathcal{S}^i_{k_j} \subset \mathcal{C}, j = 1, \cdots, M_i\}. \tag{16}$$

Note that the decision criterion in (16) is ambiguous, because $\mathcal{S}_{k_j}^i$ may contain multiple transmitted symbol vectors [6] and these transmitted symbol vectors cannot be recovered in this case. Due to the fading effect of wireless channels, the ambiguity is uncontrollable and inevitable.

To address the above issue, we leverage the second important technique in EL, i.e., combination - to combine multiple weak BLs to generate a strong learner. To reduce complexities, the simple but efficient weighted voting combination method is considered here [44]. Each $\mathcal{R}_j^i$ is a VC, denoted by $\mathcal{V}(y^i_{\mathbf{c}_{k_j}})$, i.e., $\mathcal{R}_j^i = \mathcal{V}(y^i_{\mathbf{c}_{k_j}})$, where, as a representative element of $\mathcal{S}_{k_j}$, $\mathbf{c}_{k_j}$ can be an arbitrary element in $\mathcal{S}_{k_j}$. Let $\omega(\mathbf{c}, y)$ denote the weighted function, where y is a received signal and $\mathbf{c} \in \mathcal{C}$ is a symbol vector. The value of the weighted function exploits and characterizes implicit correlation among the received signals from different active antennas.

Let $\mathbf{y}$ denote an arbitrary received signal vector. The symbol vector can be estimated by

$$\hat{\mathbf{c}}(\mathbf{y}) = \arg\max_{\mathbf{c}\in\mathcal{C}} F(\{\omega(\mathbf{c}, y_l)\}), \tag{17}$$

where y_l is the l-th component of vector $\mathbf{y}$ and F denotes the combination function of EL. As an example, the combination function can be

$$F(\{\omega(\mathbf{c}, y_l)\}) = \sum_{l=1}^{N_{\mathrm{R}}} \omega(\mathbf{c}, y_l) \tag{18}$$

or

$$F(\{\omega(\mathbf{c}, y_l)\}) = \max_{1\le l\le N_{\mathrm{R}}} \omega(\mathbf{c}, y_l). \tag{19}$$

Note that the definition of $\omega(\mathbf{c}, y)$ is not unique. For example, $\omega(\mathbf{c}, y)$ can be a decreasing function with respect to $|y - y_{\mathbf{c}_{k_j}}|$, e.g.,

$$\omega(\mathbf{c}, y) = -|y - y_{\mathbf{c}_{k_j}}|, \ (\mathbf{c} \in \mathcal{S}_{k_j}^i) \tag{20}$$

or

$$\omega(\mathbf{c}, y) = -|y - y_{\mathbf{c}_{k_j}}|^2, \ (\mathbf{c} \in \mathcal{S}_{k_j}^i). \tag{21}$$

Remark 4.1 An important advantage of the EL-based SD approach is that it can be implemented modularly and in parallel, and thus it is sufficiently flexible. Typically, when the number of active receiving antennas changes (e.g., the receiving antenna selection technique is utilized), there is no need to redesign a SD algorithm. In general, if more receiving antennas are activated and involved in the ensemble process, better detection performance can be achieved.

[6] The size of each set $\mathcal{S}_{k_j}^i$ depends on multiple factors, e.g., the algorithm used for clustering and the clustering features of the used samples.

Algorithm 3: CSI-Free Geometric SD Algorithm via SSL+EL

1: **input:** $\mathcal{C}$ with $\mathrm{card}(\mathcal{C}) = M$ - set of transmitted symbol vectors; U - number of SISs; $\mathbf{Y}$ - matrix of received signals; $\mathcal{A}$ - set of active receiving antennas

2: **transmit** SISs for each symbol vector in $\mathcal{C}$

3: **generate** base learners: **for** each $i \in \mathcal{A}$, construct a decision criterion (maybe ambiguous) based on $\mathbf{Y}(i,:)$

4: **choose** weighted function to generate voting method

5: **recover** symbol vectors from received signals

6: **output:** recovered symbol vectors $\{\hat{\mathbf{s}}_{MU+1}, \cdots, \hat{\mathbf{s}}_T\}$

V. Simulation Results

In this section, we evaluate and verify the effectiveness and superiority of the proposed SD algorithms via simulation results. We consider the following two typical channel models:

- Uncorrelated Channel Model (UCM): All elements of channel matrix $\mathbf{H}$ (for the MISO case, $\mathbf{H}$ is degenerated into a vector) are mutually independent and each element h_{ij} is distributed as $h_{ij} \sim \mathcal{CN}(0,1)$.
- Correlated Channel Model (CCM): The classical exponential correlation model [34], [45] is chosen here. The (i,j)-th element of the channel covariance matrix $\mathbf{R}$, denoted by $r_{i,j}$, is given by $r_{i,j} = \rho^{|i-j|}$, where $\rho \in (0,1]$ indicates the strength of channel correlation.

The estimated channel matrix $\hat{\mathbf{H}}$, real channel matrix $\mathbf{H}$ and noisy channel matrix $\Delta\mathbf{H}$ satisfy the following relationship

$$\hat{\mathbf{H}} = \mathbf{H} + \Delta\mathbf{H}. \tag{22}$$

Each element of the noisy channel matrix $\Delta\mathbf{H}$ is distributed as $\Delta h_{ij} \sim \mathcal{CN}(0, \sigma^2_{\Delta h})$. The estimated CSI in (22) is formulated only for the benchmark techniques (e.g., the MLD algorithm), since the proposed SD algorithms do not need CSI. Throughout this section, each time-block is assumed to consist of 5×10^6 transmitted symbol vectors, i.e., $T = 5 \times 10^6$.

For comparison, the MLD algorithm (providing the limit SER performance) and a learning-based MIMO detection algorithm proposed in [17] (referred to as "DetNet" in [17], i.e., detection network) are chosen as benchmarks to evaluate our algorithms. For convenience, Algorithm 1 and Algorithm 2 are abbreviated as CSI-Free-GSD-MISO (i.e., CSI-free geometric SD algorithm for MISO) and CSI-Free-GSD-MISO-NC (i.e., CSI-Free-GSD-MISO but without clustering), respectively. For the MIMO case, the MIMO detection algorithm in [17], the CSI-free geometric SD algorithm extended from Algorithm 2 directly (via high dimensional data clustering) and the

CSI-free geometric SD algorithm based on both SSL and EL (i.e., Algorithm 3) are named as DetNet, CSI-Free-GSD-SSL-MIMO and CSI-Free-GSD-(SSL+EL)-MIMO, respectively.

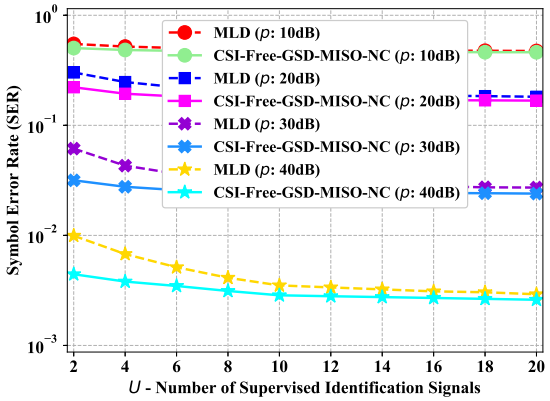

Fig. 6. The SER performance (of MLD and CSI-Free-GSD-MISO-NC) vs training overhead U: $N_{\mathrm{T}} = 4$, $N_{\mathrm{R}} = 1$, $d = 0.04\sqrt{p}$, $L = 2000$, UCM and QPSK modulation.

First, we consider the MISO case and evaluate MLD and CSI-Free-GSD-MISO-NC from the perspective of training overhead, as shown in Fig. 6. The two algorithms cost the same amount of training or pilot overhead, i.e., the number of total SISs and the size of total pilot signals are both UN_{T}. It is observed that CSI-Free-GSD-MISO-NC outperforms MLD. The reason for this is that CSI-Free-GSD-MISO-NC can fully exploit information from received signals of effective data to aid SD. Specifically, by exploiting the received signals of effective data to mitigate the impact of noise, the accuracy of reestimated centroids is further improved, which thus improves the detection performance. Besides the good SD performance, another important advantage of our approach is that there is no need to estimate CSI, which greatly simplifies system designs.

The SER performance of different SD algorithms for different channel conditions is provided in Fig. 7. It is seen that in the two cases the gap between the two performance curves (corresponding to CSI-Free-GSD-MISO and MLD with perfect CSI) is very small, which, in fact, coincides with our theoretical analysis. It is also observed that when the CSI is imperfect, CSI-Free-GSD-MISO outperforms MLD. This indicates that the proposed CSI-free SD approach inherently has a good robustness to CSI uncertainty. Moreover, since there is no need to estimate CSI for the CSI-free SD approach, it is very appealing in wireless communications. As a by-product, the performance of the CSI-free SD approach is not affected by the CSI estimate algorithms.

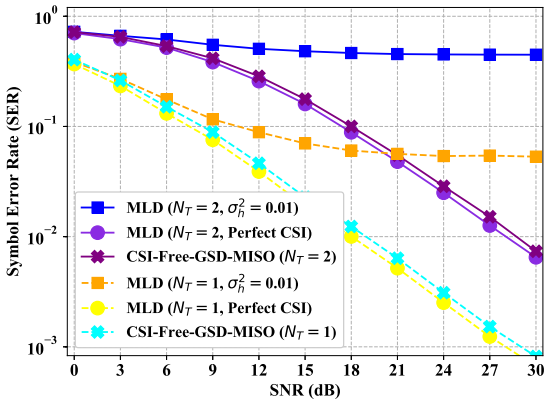

Fig. 7. The SER performance (of different SD algorithms) vs SNR: $U = 8$, $N_{\mathrm{R}} = 1$, $L = 2000$, UCM and 16QAM modulation.

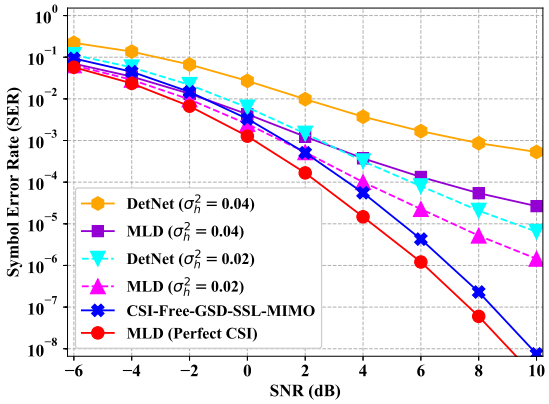

Fig. 8. The SER performance of different SD algorithms: $U = 8$, $N_{\mathrm{T}} = N_{\mathrm{R}} = 16$, CCM ($\rho = 0.5$) and BPSK modulation.

Next, we further evaluate different SD algorithms for the correlated MIMO channels. The SER performance of different SD algorithms is shown in Fig. 8. It is observed that as the SNR value increases, the SER performance of the three SD algorithms becomes better and better. It is not surprising that MLD with perfect CSI achieves the best performance among the three detection algorithms. However, when the CSI is imperfect (even if $\sigma_{\Delta h}^2$ is relatively small, e.g., $\sigma_{\Delta h}^2 = 0.02$), CSI-Free-GSD-SSL-MIMO outperforms MLD. In general, the MLD approach offers the limit performance (i.e., a lower bound) among various SD algorithms that utilize CSI. Hence, it is not surprising that MLD outperforms DetNet in the two cases.

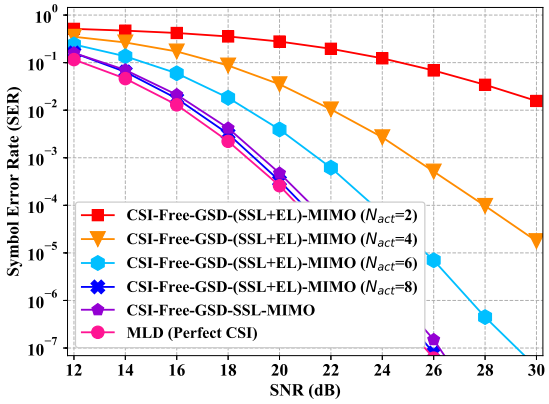

Fig. 9. The SER performance of the two CSI-free geometric SD algorithms: : $U = 10$, $N_{\mathrm{T}} = 4$, $N_{\mathrm{R}} = 8$, CCM ($\rho = 0.5$) and 16QAM modulation. N_{act} denotes the number of active receiving antennas.

The SER performance of the two CSI-free geometric SD algorithms, i.e., CSI-Free-GSD-SSL-MIMO and CSI-Free-GSD-(SSL+EL)-MIMO, is demonstrated in Fig. 9. It can be seen that CSI-Free-GSD-(SSL+EL)-MIMO with all receiving antennas active outperforms CSI-Free-GSD-SSL-MIMO with the same number of receiving antennas. The reason for this is that CSI-Free-GSD-(SSL+EL)-MIMO first extracts cluster information from the received signals of each receiving antenna independently and then combines the extracted cluster information to generate a decision criterion. Due to the independent processing, a base learner who can perform better plays a more important role in SD. In contrast, CSI-Free-GSD-SSL-MIMO coequally treats the cluster information extracted from all received signals, which weakens the role of more useful received signals. Moreover, since CSI-Free-GSD-(SSL+EL)-MIMO is designed based on EL, it can be implemented in parallel and has lower computational complexity.

It is also observed from Fig. 9 that when the number of active receiving antennas is two, the SER performance of CSI-Free-GSD-(SSL+EL)-MIMO is not very satisfactory. However, when the number of active antennas is greater than three (i.e., $N_{\mathrm{act}} \geq 4$), much better SER performance can be achieved. The reason for this is that more receiving antennas provide larger diversity gains, which is desired in practice. It is also seen that the performance improvement thanks to the increase of active antennas becomes small, which implies that the performance loss due to the use of less antennas is negligible. Apparently, the use of less active receiving antennas means lower computational complexity and less consumptions on computing and storage resources.

Besides, the EL-based SD approach offers sufficient flexibility, e.g., it enables the modular design. Typically, it is very convenient to add or reduce active receiving antennas, which facilitates the incorporation of receiving antenna selection techniques and SD.

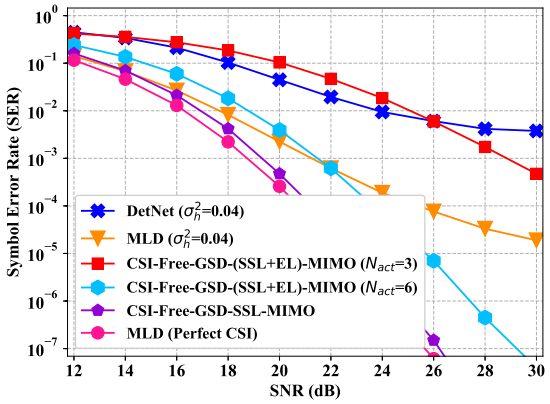

Fig. 10. The SER performance of different SD algorithms: $U = 10$, $N_{\mathrm{T}} = 4$, $N_{\mathrm{R}} = 8$, CCM ($\rho = 0.5$) and 16QAM modulation.

The SER performance of different SD algorithms is provided in Fig. 10. It can be observed that CSI-Free-GSD-(SSL+EL)-MIMO with only three active receiving antennas (i.e., $N_{\mathrm{act}} = 3$) can achieve similar SER performance with DetNet, even though the estimated channel matrix $\hat{\mathbf{H}}$ used by DetNet is only slightly contaminated by the noise, i.e., $\sigma_{\Delta h}^2 = 0.04$. Moreover, when the SNR value is greater than 22dB, CSI-Free-GSD-(SSL+EL)-MIMO with $N_{\mathrm{act}} = 6$ active receiving antennas outperforms the (optimal) MLD algorithm. Along with the observation that CSI-Free-GSD-SSL-MIMO surpasses MLD with imperfect CSI, these observations fully demonstrate the appealing advantages of the proposed CSI-free SD approach.

VI. Discussion: Limitations and Advantages

In this section, we compare and analyze our approach with existing MIMO detection design methodologies, including (conventional) model-based design methodology (e.g., [9]–[12]), deep learning (data-driven) based design methodology (e.g., [17]–[19], [46]), and hybrid model-based deep learning design methodology (e.g., [34], [47]).

Subject to multiple factors, our algorithms temporarily are mainly applicable to small-scale MIMO settings, while the conventional model-based MIMO detectors (e.g., classical linear or iteration-based detectors) can achieve good performance for large-scale MIMO settings. Because

of the powerful modeling and representation abilities, deep-learning based receivers can exploit latent or unknown gains (e.g., due to joint processing or data symbols [46]) or complex structure information to improve performance of interest in complicated systems or environments. The model-based deep learning designs can relax the requirement of "big data" and meanwhile enjoy the advantages of model-based methods. But, our methods fail to reap these advantages.

However, our algorithms enjoy the following salient advantages. First, because CSI estimate is not required in our algorithms, they are not affected by the quality of estimated CSI and/or the performance of CSI estimation algorithm. Second, they have very low computational complexity. For example, there is no need to even invoke the clustering operation in Algorithm 2, which thus significantly reduces the complexity. Finally, when incorporating EL our method is very flexible, e.g., it can be implemented modularly and in parallel. These features make our algorithms very appealing in short packet based communications, in particular, like Internet-of-Things (IoT) [48].

It has been recognized that IoT calls for a new type of communication technologies [48]–[50]. First and foremost, the traffics in IoT applications are typically in the form of short packets [48]. Moreover, there are often stringent requirements on latency and reliability in mission-critical IoT applications, such as factory automation and vehicle-to-everything (V2X) communications. The requirements of low latency, low power consumption, high reliability as well as short packet transmission from IoT applications make our algorithms preferable over the existing techniques:

- Comparison with model-based detectors: Our algorithms offer better robustness and higher transmission efficiency, since they are not affected by CSI estimate. Moreover, lower SER and higher reliability can be achieved, by exploiting received signals of effective data.
- Comparison with deep learning based receivers: Thanks to using the lightweight ML methods (e.g., the clustering algorithm), the size of needed training samples is very low, which yields low computational complexity and low sample complexity.
- Comparison with hybrid model-based deep learning detectors: As a hybrid model and data based design, our methods enjoy the benefits of low complexities, which yield better real-time. Hence, our algorithms are suitable for IoT applications, in which stringent requirements (e.g., low latency, low power consumption and high reliability) are often imposed.

Thanks to the above advantages, our algorithms are also competitive in the other small-scale MIMO cases, if both low complexity and high performance are desired.

VII. Conclusion

To avoid performance loss of SD due to the use of inevitable inaccurate CSI and approach the limit performance offered by the optimal MLD algorithm, in this paper we leveraged SSL to design a CSI-free SD approach. First, we proposed an efficient CSI-free geometric SD algorithm for the MISO channels. We proved that the designed algorithm can approach the optimal MLD algorithm. Then, we incorporated SSL and EL to design a parallelizable CSI-free SD algorithm, so as to enable parallel implementation and reduce complexity. The algorithm designed based on SSL and EL is sufficiently flexible and thus enables modular design. Finally, simulation results confirmed the superiority of our proposal. In particular, the proposed algorithms can approach the (optimal) MLD algorithm with perfect CSI, but outperforms it for imperfect CSI.

The current algorithms are mainly applicable to small-scale MIMO settings (e.g., $N_{\mathrm{T}} \leq 16$), since they are subject to the size of available data samples, tolerant time-delay, computing resource, and so on. It is our future work to develop efficient algorithms for more challenging large-scale MIMO settings. To simplify design and reduce complexity, we chose the naive voting ensemble method to design the EL-based SD algorithm. A more appropriate method may be to learn the ensemble or aggregation procedure, e.g., via the mixture-of-experts techniques.

Appendix A
The k-means Clustering Algorithm

The k-means algorithm is a typical unsupervised learning algorithm. Given a dataset $\mathcal{D} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N\}$, the goal of the k-means algorithm is to partition it into m clusters such that each point in a cluster is similar to points from its own cluster than with points from some other cluster. Towards this end, we define prototype vectors $\boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_m$ and indicator variables r_{ij} which is 1 if and only if $\boldsymbol{x}_i$ is assigned to cluster j (and 0 otherwise). To cluster the dataset $\mathcal{D}$, we try to minimize the following distortion measure which minimizes the distance of each point from the prototype vector:

$$J(\{r_{ij}\}, \{\boldsymbol{\mu}_j\}) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{m} r_{ij} \|\boldsymbol{x}_i - \boldsymbol{\mu}_j\|^2. \tag{23}$$

Unfortunately, it is challenging to obtain an optimal solution of problem (23) for large (and even moderate) N and m. An efficient sub-optimal algorithm to tackle problem (23) is the alternating iterative method summarized in Algorithm 4. A typical convergence criterion for

Algorithm 4: The k-means Clustering Algorithm

1: **input:** sample points $\mathcal{D} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N\}$; m - number of total clusters

2: **initialize** mean vectors $\{\boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_m\}$ (e.g., choose m points within $\mathcal{D}$ randomly)

3: **repeat**

(a) **initialize** cluster sets: let $\mathcal{D}_j = \emptyset\,(1 \le j \le m)$

(b) **for** $i = 1, 2, \cdots, N$: (update cluster sets)

1) compute distance between $\boldsymbol{x}_i$ and each point $\boldsymbol{\mu}_j$: $d_{ij} = \|\boldsymbol{x}_i - \boldsymbol{\mu}_j\|_2$

2) classify: $\mathcal{D}_{\lambda_j} = \mathcal{D}_{\lambda_j} \bigcup \{\boldsymbol{x}_i\}$ with λ_j given by $\lambda_j = \arg\min_{j \in \{1,\cdots,m\}} d_{ij}$

(c) **for** $j = 1, 2 \cdots, m$: (update mean vectors)

$\boldsymbol{\mu}_j \leftarrow \mathrm{card}(\mathcal{D}_j)^{-1} \sum_{\boldsymbol{x} \in \mathcal{D}_j} \boldsymbol{x}$

until some convergence criterion is met

4: **output:** mean vectors $\{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \cdots, \boldsymbol{\mu}_m\}$

Algorithm 4 is that the mean vectors do not change. The output of the k-means algorithm is m mean vectors, based on which the Euclidean space is divided into m VCs. Most importantly, the method to divide the Euclidean space is in essence the Voronoi tessellation method, which enables the proposed SD algorithms to approach the MLD algorithm with perfect CSI.

APPENDIX B
PROOF OF LEMMA 1

First, we rigorously prove that the regions generated by the MLD algorithm are VCs, i.e., the MLD algorithm generates a VT. In fact, for an arbitrary transmitted symbol vector $\mathbf{c}_j \in \mathcal{C}$, the likelihood function is given by [51]

$$L(y|p, \mathbf{h}, \mathbf{c}_j) = \frac{1}{\pi} \exp\big(-|y - \sqrt{p}\mathbf{h}^{\mathrm{H}}\mathbf{c}_j|^2\big). \tag{24}$$

Let y be an arbitrary received signal and $\hat{\mathbf{c}}(y)$ be an estimate of the constellation point of y. Then, according to the principle of maximum likelihood detection, $\hat{\mathbf{c}}(y) = \mathbf{c}_j$ if and only if

$$L(y|p, \mathbf{h}, \mathbf{c}_j) > L(y|p, \mathbf{h}, \mathbf{c}_k), \quad (\forall\, k \neq j). \tag{25}$$

The inequality in (25) is equivalent to

$$|y - \sqrt{p}\mathbf{h}^{\mathrm{H}}\mathbf{c}_j| < |y - \sqrt{p}\mathbf{h}^{\mathrm{H}}\mathbf{c}_k|, \quad (\forall\, k \neq j). \tag{26}$$

According to the definition of VC, $\mathcal{V}(\sqrt{p}\mathbf{h}^{\mathrm{H}}\mathbf{c}_j)$ is given by

$$\mathcal{V}(\sqrt{p}\mathbf{h}^{\mathrm{H}}\mathbf{c}_j) = \big\{y \,|\, |y - \sqrt{p}\mathbf{h}^{\mathrm{H}}\mathbf{c}_j| < |y - \sqrt{p}\mathbf{h}^{\mathrm{H}}\mathbf{c}_k|, \forall\, k \neq j\big\}.$$

It is observed that: 1) $\hat{\mathbf{c}}(y) = \mathbf{c}_j$ if and only if $y \in \mathcal{V}(\mathbf{c}_j^{\text{MLD}})$; and 2) the complex plane $\mathbb{C}$ is divided into M VCs $\{\mathcal{V}(\sqrt{p}\mathbf{h}^{\text{H}}\mathbf{c}_j)\}$, or equivalently, the MLD algorithm generates a VT.

Next, we concentrate on the second part of this lemma. [7] Let $\mathcal{A}_j = \{y \,|\, |y - \sqrt{p}\mathbf{h}^{\text{H}}\mathbf{c}_j| \leq \sqrt{p}d_{\min}\} \subset \mathcal{V}(\mathbf{c}_j^{\text{MLD}})$. The proof mainly consists of three steps. The first step is to show that for an arbitrary symbol vector $\mathbf{c}_j \in \mathcal{C}$, the K_j received signals $\mathcal{S}_j = \{y_{j,1}, y_{j,2}, \cdots, y_{j,K_j}\}$ lie in $\mathcal{A}_j$ with a high probability. Because $y_{j,i} = \sqrt{p}\mathbf{h}^{\text{H}}\mathbf{c}_j + w_i$ and $w_i \sim \mathcal{CN}(0,1)$, the (joint) probability density function of an arbitrary received signal $y_{j,i} \in \mathcal{S}_j$ with respect to its real part and imaginary part can be expressed as [51]

$$p(y_{\text{R}}, y_{\text{I}}) = \frac{1}{\pi}\exp\Big(-(y_{\text{R}} - y'_{\text{R}})^2 - (x_{\text{I}} - x'_{\text{I}})^2\Big), \tag{27}$$

where y_{R}, y_{I}, y'_{R} and y'_{I} represent the real part of $y_{j,i}$, the imaginary part of $y_{j,i}$, the real part of $\sqrt{p}\mathbf{h}^{\text{H}}\mathbf{c}_j$ and the imaginary part of $\sqrt{p}\mathbf{h}^{\text{H}}\mathbf{c}_j$, respectively.

For each received signal $y_{j,i}$, the probability $\mathbb{P}(y_{j,i} \in \mathcal{A}_j)$ can be calculated as

$$\mathbb{P}(y_{j,i} \in \mathcal{A}_j) = \int_{\mathcal{A}_j} \frac{1}{\pi}\exp\Big(-(y_{\text{R}} - y'_{\text{R}})^2 - (x_{\text{I}} - x'_{\text{I}})^2\Big)dy_{\text{R}}dy_{\text{I}} = 1 - \exp\big(-pd_{\min}^2\big). \tag{28}$$

Then, the probability $\mathbb{P}(\mathcal{S}_j \subset \mathcal{A}_j)$ can be calculated as

$$\begin{aligned}\mathbb{P}(\mathcal{S}_j \subset \mathcal{A}_j) \overset{(1)}{=} \mathbb{P}(y_{j,1} \in \mathcal{A}_j)\mathbb{P}(y_{j,2} \in \mathcal{A}_j)\cdots\mathbb{P}(y_{j,K_j} \in \mathcal{A}_j) =& \Big(1 - \exp\big(-pd_{\min}^2\big)\Big)^{K_j} \overset{(2)}{\geq} \\ & 1 - K_j\exp\big(-pd_{\min}^2\big),\end{aligned} \tag{29}$$

where (1) is due to the conditional independence (i.e., the received signals are independent given the CSI and transmitted symbol vector), and (2) is due to the inequality $(1-x)^n \geq 1 - nx$ for $x \in (0,1)$ and for an arbitrary positive integer n. It is seen that if $pd_{\min}^2$ is sufficiently large, $\mathbb{P}(\mathcal{S}_j \subset \mathcal{A}_j)$ approaches 1.

The second step is to prove that $|w_{\mathbf{c}_j} - w_{\mathbf{c}_j}^{\text{MLD}}| \leq \varepsilon$ holds with a high probability. Note that the centroid $w_{\mathbf{c}_j}$ calculated as $w_{\mathbf{c}_j} = K_j^{-1}\sum_{i=1}^{K_j} y_{j,i}$ is distributed as

$$w_{\mathbf{c}_j} \sim \mathcal{CN}(\sqrt{p}\mathbf{h}^{\text{H}}\mathbf{c}_j, 1/K_j). \tag{30}$$

Similar to the calculation of probability $\mathbb{P}(y_{j,i} \in \mathcal{A}_j)$ in (28), the probability $\mathbb{P}(|w_{\mathbf{c}_j} - w_{\mathbf{c}_j}^{\text{MLD}}| \leq \varepsilon)$ can be calculated as

$$\begin{aligned}\mathbb{P}(|w_{\mathbf{c}_j} - w_{\mathbf{c}_j}^{\text{MLD}}| \leq \varepsilon) =& \frac{K_j}{\pi}\int_{|w_{\mathbf{c}_j} - w_{\mathbf{c}_j}^{\text{MLD}}| \leq \varepsilon} \exp\Big(-K_j|w_{\mathbf{c}_j} - w_{\mathbf{c}_j}^{\text{MLD}}|^2\Big)dw_{\mathbf{c}_j} \\ =& 1 - \exp\big(-K_j\varepsilon^2\big).\end{aligned} \tag{31}$$

[7] It is implicitly assumed that the k-means algorithm can generate the correct clusters. This conclusion has been verified by a large number of simulation experiments and can also be proved rigidly if $\sqrt{p}d_{\min}$ is sufficiently large.

The final step is to bound the probability $\mathbb{P}(\max_j |w_{\mathbf{c}_j} - w_{\mathbf{c}_j}^{\text{MLD}}| \leq \varepsilon)$. In fact, the probability $\mathbb{P}(\max_j |w_{\mathbf{c}_j} - w_{\mathbf{c}_j}^{\text{MLD}}| \leq \varepsilon)$ can be lower bounded by

$$\mathbb{P}(\max_j |w_{\mathbf{c}_j} - w'_{\mathbf{c}_j}| \leq \varepsilon) \geq \prod_{j=1}^{M} \mathbb{P}(\mathcal{S}_j \subset \mathcal{A}_j)\mathbb{P}(|w_{\mathbf{c}_j} - w_{\mathbf{c}_j}^{\text{MLD}}| \leq \varepsilon)$$

$$= \prod_{j=1}^{M} \Big(1 - K_j \exp\big(-pd_{\text{min}}^2\big)\Big)\Big(1 - \exp\big(-K_j\varepsilon^2\big)\Big). \tag{32}$$

By induction, it can be shown that for real numbers $\{a_i \in (0,1)\}$, the following inequality holds

$$\prod_{i=1}^{M}(1-a_i) \geq 1 - \sum_{i=1}^{M} a_i. \tag{33}$$

According to inequality (33), the probability $\mathbb{P}(\max_j |w_{\mathbf{c}_j} - w_{\mathbf{c}_j}^{\text{MLD}}| \leq \varepsilon)$ can be further lower bounded by

$$\mathbb{P}(\max_j |w_{\mathbf{c}_j} - w_{\mathbf{c}_j}^{\text{MLD}}| \leq \varepsilon) \geq 1 - \sum_{j=1}^{M} K_j \exp\big(-pd_{\text{min}}^2\big) - \sum_{j=1}^{M} \exp\big(-K_j\varepsilon^2\big), \tag{34}$$

which completes the proof.

APPENDIX C
PROOF OF THEOREM 2

Without loss of generality, we consider VC $\mathcal{R}_j$ and assume that symbol vector $\mathbf{c}_j \in \mathcal{C}$ is associated to $\mathcal{R}_j$. The received signal $y_u \in \mathcal{R}_j$ is given by

$$y_u = \sqrt{p}\mathbf{h}^{\text{H}}\mathbf{c}_j + w_u. \tag{35}$$

The centroid $w_{\mathbf{c}_j} = U^{-1}\sum_{y_u \in \mathcal{Y}_{\mathbf{c}_j}} y_u$ is distributed as

$$w_{\mathbf{c}_j} \sim \mathcal{CN}(\sqrt{p}\mathbf{h}^{\text{H}}\mathbf{c}_j, 1/U). \tag{36}$$

Let $x_{\text{R}} = \text{Re}(w_{\mathbf{c}_j})$ and $x_{\text{I}} = \text{Im}(w_{\mathbf{c}_j})$ represent the real part and imaginary part of $w_{\mathbf{c}_j}$, respectively. $x'_{\text{R}} = \text{Re}(\sqrt{p}\mathbf{h}^{\text{H}}\mathbf{c}_j)$ and $x'_{\text{I}} = \text{Im}(\sqrt{p}\mathbf{h}^{\text{H}}\mathbf{c}_j)$ are defined similarly. The probability density function of pair $(x_{\text{R}}, x_{\text{I}})$ is given by

$$p(x_{\text{R}}, x_{\text{I}}) = \frac{U}{\pi}\exp\Big(-U(x_{\text{R}} - x'_{\text{R}})^2 - U(x_{\text{I}} - x'_{\text{I}})^2\Big). \tag{37}$$

The probability that each symbol vector $\mathbf{c}_j \in \mathcal{C}$ is correctly associated to VC $\mathcal{R}_j$ is calculated as

$$\begin{aligned}\mathbb{P}(\mathbf{c}_j \in \mathcal{R}_j) =& \frac{U}{\pi}\int_{\mathcal{R}_j} \exp\big(-U(x_{\mathrm{R}}-x'_{\mathrm{R}})^2 - U(x_{\mathrm{I}}-x'_{\mathrm{I}})^2\big)dx_{\mathrm{R}}dx_{\mathrm{I}}\\ \geq& \frac{U}{\pi}\int_{\mathcal{A}_j} \exp\big(-U(x_{\mathrm{R}}-x'_{\mathrm{R}})^2 - U(x_{\mathrm{I}}-x'_{\mathrm{I}})^2\big)dx_{\mathrm{R}}dx_{\mathrm{I}}\\ =& \frac{U}{\pi}\int_{x_{\mathrm{R}}^2+x_{\mathrm{I}}^2\leq pd_{\min}^2} \exp\big(-Ux_{\mathrm{R}}^2 - Ux_{\mathrm{I}}^2\big)dx_{\mathrm{R}}dx_{\mathrm{I}} = 1-\exp\big(-pUd_{\min}^2\big),\end{aligned} \tag{38}$$

where $\mathcal{A}_j = \{(s,t)\,|\,(s-x'_{\mathrm{R}})^2+(t-x'_{\mathrm{I}})^2 \leq pd_{\min}^2\}$ is a disk in $\mathbb{C}$ with center $(x'_{\mathrm{R}}, x'_{\mathrm{I}})$ and radius $r=\sqrt{p}d_{\min}$.

The probability that all symbol vectors are correctly associated to the corresponding VCs can be lower bounded by

$$\begin{aligned}&\mathbb{P}(\mathbf{c}_1 \in \mathcal{R}_1, \mathbf{c}_2 \in \mathcal{R}_2, \cdots, \mathbf{c}_M \in \mathcal{R}_M) = 1-\mathbb{P}(\mathbf{c}_1 \notin \mathcal{R}_1 \text{ or } \mathbf{c}_2 \notin \mathcal{R}_2 \text{ or } \cdots \text{ or } \mathbf{c}_M \notin \mathcal{R}_M)\\ \geq& 1-\sum_{\mathbf{c}_i\in\mathcal{C}}\mathbb{P}(\mathbf{c}_i \notin \mathcal{R}_i) \overset{(1)}{\geq} 1-\sum_{\mathbf{c}_i\in\mathcal{C}}\exp\big(-pUd_{\min}^2\big) \overset{(2)}{=} 1-M\exp\big(-pUV_0^2\big),\end{aligned} \tag{39}$$

where (1) is due to the inequality in (38) and (2) is due to the assumption $d_{\min} \geq V_0$. The remaining of the derivation is straightforward, which is omitted.

REFERENCES

[1] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, 2014.

[2] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[3] S. Yang and L. Hanzo, "Fifty years of MIMO detection: The road to large-scale MIMOs," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 1941–1988, 2015.

[4] Z. Guo and P. Nilsson, "Algorithm and implementation of the K-best sphere decoding for MIMO detection," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 491–503, 2006.

[5] S. Wu, L. Kuang, Z. Ni, J. Lu, D. Huang, and Q. Guo, "Low-complexity iterative detection for large-scale multiuser MIMO-OFDM systems using approximate message passing," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 902–915, 2014.

[6] J. Cspedes, P. M. Olmos, M. Snchez-Fernndez, and F. Perez-Cruz, "Expectation propagation detection for high-order high-dimensional MIMO systems," *IEEE Trans. Commun.*, vol. 62, no. 8, pp. 2840–2849, 2014.

[7] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, 2010.

[8] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2011.

[9] J. Qian and C. Masouros, "Multipair relaying with space-constrained large-scale MIMO arrays: Spectral and energy efficiency analysis with incomplete CSI," *IEEE OJ ComSoc.*, vol. 2, pp. 2357–2371, 2021.

[10] J. Qian, C. Masouros, and A. Garcia-Rodriguez, "Partial CSI acquisition for size-constrained massive MIMO systems with user mobility," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 9016–9020, 2018.

[11] A. Garcia-Rodriguez and C. Masouros, "Exploiting the increasing correlation of space constrained massive MIMO for CSI relaxation," *IEEE Trans. Commun.*, vol. 64, no. 4, pp. 1572–1587, 2016.

[12] C. Masouros, M. Sellathurai, and T. Ratnarajah, "Large-scale MIMO transmitters in fixed physical spaces: The effect of transmit correlation and mutual coupling," *IEEE Trans. Commun.*, vol. 61, no. 7, pp. 2794–2804, 2013.

[13] H. Sun, X. Chen, Q. Shi, M. Hong, X. Fu, and N. D. Sidiropoulos, "Learning to optimize: Training deep neural networks for interference management," *IEEE Trans. Signal Process.*, vol. 66, no. 20, pp. 5438–5453, 2018.

[14] W. Cui, K. Shen, and W. Yu, "Spatial deep learning for wireless scheduling," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 6, pp. 1248–1261, 2019.

[15] F. Liang, C. Shen, W. Yu, and F. Wu, "Towards optimal power control via ensembling deep neural networks," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1760–1776, 2020.

[16] W. Lee, M. Kim, and D. Cho, "Deep power control: Transmit power control scheme based on convolutional neural network," *IEEE Commun. Lett.*, vol. 22, no. 6, pp. 1276–1279, 2018.

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

[18] N. Samuel, T. Diskin, and A. Wiesel, "Deep MIMO detection," in *2017 IEEE 18th International Workshop on SPAWC*, 2017, pp. 1–5.

[19] M. Mohammadkarimi, M. Mehrabi, M. Ardakani, and Y. Jing, "Deep learning-based sphere decoding," *IEEE Trans. Wireless Commun.*, vol. 18, no. 9, pp. 4368–4378, 2019.

[20] J. Sun, Y. Zhang, J. Xue, and Z. Xu, "Learning to search for MIMO detection," *IEEE Trans. Wireless Commun.*, vol. 19, no. 11, pp. 7571–7584, 2020.

[21] S. Park, H. Jang, O. Simeone, and J. Kang, "Learning to demodulate from few pilots via offline and online meta-learning," *IEEE Trans. Signal Process.*, vol. 69, pp. 226–239, 2021.

[22] L. Sun, Y. Wang, A. L. Swindlehurst, and X. Tang, "Generative-adversarial-network enabled signal detection for communication systems with unknown channel models," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 47–60, 2021.

[23] N. Shlezinger, R. Fu, and Y. C. Eldar, "DeepSIC: Deep soft interference cancellation for multiuser MIMO detection," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1349–1362, 2021.

[24] N. Shlezinger, N. Farsad, Y. C. Eldar, and A. J. Goldsmith, "Viterbinet: A deep learning based viterbi algorithm for symbol detection," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3319–3331, 2020.

[25] Y.-D. Huang, P. P. Liang, Q. Zhang, and Y.-C. Liang, "A machine learning approach to MIMO communications," in *2018 IEEE ICC*, 2018, pp. 1–6.

[26] Q. Zhang, P. P. Liang, Y.-D. Huang, Y. Pei, and Y.-C. Liang, "Label-assisted transmission for short packet communications: A machine learning approach," *IEEE Trans. Vehi. Technol.*, vol. 67, no. 9, pp. 8846–8859, 2018.

[27] S. Takabe, M. Imanishi, T. Wadayama, and K. Hayashi, "Deep learning-aided projected gradient detector for massive overloaded MIMO channels," in *2019 IEEE ICC*, 2019, pp. 1–6.

[28] M. Shao and W.-K. Ma, "Binary MIMO detection via homotopy optimization and its deep adaptation," *IEEE Trans. Signal Process.*, vol. 69, pp. 781–796, 2021.

[29] Q. Chen, S. Zhang, S. Xu, and S. Cao, "Efficient MIMO detection with imperfect channel knowledge - a deep learning approach," in *2019 IEEE WCNC*, 2019, pp. 1–6.

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

[31] A. Mohammad, C. Masouros, and Y. Andreopoulos, "Complexity-scalable neural-network-based MIMO detection with learnable weight scaling," *IEEE Trans. Commun.*, vol. 68, no. 10, pp. 6101–6113, 2020.

[32] J. Zhang, Y. He, Y.-W. Li, C.-K. Wen, and S. Jin, "Meta learning-based MIMO detectors: Design, simulation, and experimental test," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1122–1137, 2021.

[33] X. Tan, W. Xu, K. Sun, Y. Xu, Y. Beery, X. You, and C. Zhang, "Improving massive MIMO message passing detectors with deep neural network," *IEEE Trans. Veh. Technol.*, vol. 69, no. 2, pp. 1267–1280, 2020.

[34] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Model-driven deep learning for MIMO detection," *IEEE Trans. Signal Process.*, vol. 68, pp. 1702–1715, 2020.

[35] J. Liao, J. Zhao, F. Gao, and G. Y. Li, "A model-driven deep learning method for massive MIMO detection," *IEEE Commun. Lett.*, vol. 24, no. 8, pp. 1724–1728, 2020.

[36] X. Jin and H.-N. Kim, "Parallel deep learning detection network in the MIMO channel," *IEEE Commun. Lett.*, vol. 24, no. 1, pp. 126–130, 2020.

[37] D. Shirase, T. Takahashi, S. Ibi, K. Muraoka, N. Ishii, and S. Sampei, "Deep unfolding-aided gaussian belief propagation for correlated large MIMO detection," in *GLOBECOM 2020*, 2020, pp. 1–6.

[38] Y. Wei, M.-M. Zhao, M. Hong, M.-J. Zhao, and M. Lei, "Learned conjugate gradient descent network for massive MIMO detection," *IEEE Trans. Signal Process.*, vol. 68, pp. 6336–6349, 2020.

[39] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *IEEE Signal Process. Mag.*, vol. 38, no. 2, pp. 18–44, 2021.

[40] M. Haenggi, "Stochastic geometry for wireless networks," *Cambridge University Press*, 2012.

[41] R. Xu and D. Wunsch, "Survey of clustering algorithms," *IEEE Trans. Neural Netw.*, vol. 16, no. 3, pp. 645–678, 2005.

[42] Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. 28, no. 1, pp. 84–95, 1980.

[43] K. P. Murphy, *Probabilistic machine learning: An introduction*. MIT press, 2022.

[44] Z.-H. Zhou, *Ensemble Methods: Foundations and Algorithms*, 1st ed., ser. Chapman & Hall/CRC Data Mining and Knowledge Discovery Serie. Chapman and Hall/CRC, 2012.

[45] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7112–7139, 2014.

[46] M. Honkala, D. Korpi, and J. M. J. Huttunen, "DeepRx: Fully convolutional deep learning receiver," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3925–3940, 2021.

[47] N. Shlezinger, J. Whang, Y. C. Eldar, and A. G. Dimakis, "Model-based deep learning," *CoRR*, vol. abs/2012.08405, 2020. [Online]. Available: https://arxiv.org/pdf/2012.08405.pdf

[48] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultrareliable, and low-latency wireless communication with short packets," *Proceedings of the IEEE*, vol. 104, no. 9, pp. 1711–1726, 2016.

[49] C. Boyer and S. Roy, "Backscatter communication and RFID: Coding, energy, and MIMO analysis," *IEEE Trans. Commun.*, vol. 62, no. 3, pp. 770–785, 2014.

[50] G. Yang, Q. Zhang, and Y.-C. Liang, "Cooperative ambient backscatter communications for green Internet-of-Things," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 1116–1130, 2018.

[51] J. G. Proakis, *Digital Communications*. New York, NY, USA: McGraw-Hill, 1995.

Responses to Reviewers' Comments for Manuscript TCOM-TPS-21-1382.R1

CSI-Free Geometric Symbol Detection via Semi-supervised Learning and Ensemble Learning

Jianjun Zhang, Christos Masouros and Yongming Huang

August 27, 2022

Dear Editor,

We would like to thank you for handling the review process of our paper. We are also indebted to you and the reviewers for the helpful comments. According to your suggestions, we have updated the original manuscript and submitted a revised version.

The most significant revision in our manuscript is that we have added a new section (i.e., Section VI) to discuss the limitations of our algorithms and the gains of our techniques when compared to other MIMO detectors.

In this revision, all of the comments raised by the reviewers have been addressed. To enhance legibility of this response letter, the reviewers' comments are typeset in *italic font* and our responses are written in plain font.

Yours Sincerely,

Jianjun Zhang, Christos Masouros and Yongming Huang

IEEE Transactions on Communications
Paper ID: TCOM-TPS-21-1382
Authors' Response to Editor

We would like to thank you for your insightful suggestions and comments, which have helped us improve the quality of our paper. We have revised our paper incorporating all your suggestions and comments.

Comment 1:

1. All reviewers agree that the paper has significantly improved and is now in a rather good shape. However, Reviewer 1 raises an important issue: the paper is lacking a section that clearly identifies the limitations of the proposed methods and a discussion what exactly could be the gains of this technique compared to other MIMO detectors. Please add such a discussion as a new section to the paper.

Response:

We are very grateful to you for your helpful suggestion. In the revised manuscript, we have added a new section (i.e., Section VI), in which we have discussed the limitations of our methods and the gains of our techniques compared to other MIMO detectors. We would like to refer you to Section VI of the revised paper or the response to Comment 1 of Reviewer 1 for the details.

Comment 2:

2. Please also check the inconsistency by Reviewer 3.

Response:

We very much appreciate your kindly reminder. We have checked the inconsistency raised by Reviewer 3 and updated the manuscript. We would like to refer you to the response to Comment 1 of Reviewer 3 for the details.

IEEE Transactions on Communications
Paper ID: TCOM-TPS-21-1382
Authors' Response to Reviewer 1

We would like to thank you for your insightful suggestions and comments, which have helped us improve the quality of our paper. We have revised our paper incorporating all your suggestions and comments.

Comment 1:

1. The authors have made a considerable effort to address the comments raised in the previous round. This is appreciated. The main finding though is that the proposed MIMO detection algorithm has a good deal of limitations, and it is not clear exactly in which settings is the proposed method expected to be beneficial, and what are the exact benefits in these settings (complexity, performance, robustness, etc). Therefore, for the paper to be suitable for publication, I ask the authors to clearly identify the settings in which their method is preferable over existing techniques. Please clarify what are the exact gains in these settings of using relatively old machine learning techniques (LVQs) over model-based MIMO detectors (such as SD), deep learning based MIMO receivers (e.g., [R1]), and hybrid model-based deep learning techniques (see [R2]).

[R1] Honkala, Mikko, Dani Korpi, and Janne MJ Huttunen. "DeepRx: Fully convolutional deep learning receiver." IEEE Transactions on Wireless Communications 20.6 (2021): 3925-3940.

[R2] Shlezinger, Nir, et al. "Model-based deep learning." arXiv preprint arXiv:2012.08405 (2020).

Response:

We are very grateful to you for your insightful comment. We also thank you very much for bringing these papers to our attention. Before identifying the setting in which our algorithms are preferable, we would like to highlight the advantages of our algorithms.

The advantages of our algorithms are three-fold:

- First, our algorithms have very low computational complexity. For example, there is no need to even invoke the clustering operation, which significantly reduces the complexity.
- Second, CSI is not required in our methods, which makes our algorithms unaffected by the quality of estimated CSI and/or the performance of CSI estimation algorithm.
- Finally, our approach is flexible, e.g., it can be implemented modularly and in parallel.

A typical setting in which our algorithms are preferable over the existing techniques is short packet communications (SPC) [R3], like Internet-of-Things (IoT). It has been recognized that IoT calls for a new type of communication technologies [R4], [R5]. First and foremost, the traffics in IoT applications are typically in the form of short packets [R3]. Moreover, there are stringent requirements on

latency and reliability in mission-critical IoT applications, e.g., factory automation, remote surgery and vehicle-to-everything (V2X) communications. Therefore, it is of utmost importance to design low latency, low power consumption and high reliability SPC transmission for IoT applications.

[R3] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultrareliable, and low-latency wireless communication with short packets," Proceedings of the IEEE, vol. 104, no. 9, pp. 1711-1726, 2016.

[R4] C. Boyer and S. Roy, "Backscatter communication and RFID: Coding, energy, and MIMO analysis," IEEE Trans. Commun., vol. 62, no. 3, pp. 770-785, 2014.

[R5] G. Yang, Q. Zhang, and Y.-C. Liang, "Cooperative ambient backscatter communications for green internet-of-things," IEEE Internet Things J., vol. 5, no. 2, pp. 1116-1130, 2018.

We would like to highlight that our algorithms are particularly well-suited for the SPC-based IoT applications. In fact, the conventional pilot-assisted transmission (no matter whether it is a model-based MIMO detector, deep learning based MIMO receiver, or hybrid model-based deep learning detector) requires significant overhead to obtain accurate CSI for further symbol detection, which thereby reduces the transmission efficiency. Besides, we would like to further identify and clarify the benefits of our algorithms from the following three aspects in detail:

- Comparison with model-based MIMO detectors: (1) Our algorithms offers better robustness [1] and larger (data) transmission efficiency, since CSI estimation is not required in our algorithms, and thus they are not affected by the quality of estimated CSI and/or the performance of CSI estimation algorithm. (2) Our algorithms can efficiently use received signals of effective data to assist symbol recovery, which decreases symbol error rate (SER) and thus improves reliability.

- Comparison with deep learning based MIMO receivers: The deep learning based MIMO receivers (e.g., [R1]) typically require a huge number of training samples to train a deep neural network. In contrast, due to the use of lightweight machine learning method, the size of training samples needed by our algorithms is very low, which is even less than that of pilots used by the model-based MIMO detectors. Therefore, an appealing advantage of our algorithms is the low computational complexity.

- Comparison with hybrid model-based deep learning techniques: First, we would like to clarify that strictly speaking, our algorithms also belong to the hybrid model and data design category. [2] However, compared with the hybrid model-based deep learning methods, our algorithms enjoy the benefits of low complexities (including computation, sample and implementation complexities), which yields better real-time. As a result, our algorithms are very suitable for IoT applications, which requires low latency, low power consumption and high reliability.

Due to the above advantages, we would also like to mention that our algorithms are competitive in the other small-scale MIMO settings, where both low complexity and high performance are desired.

[1] Both theoretical analysis and simulation results show that our algorithms can approach the performance of the optimal maximum likelihood detection algorithm with perfect CSI, and outperforms it when the CSI is imperfect.

[2] In fact, the clustering operation (which belongs to machine learning category) is invoked to estimate the centroids of received signals, which are then used by model-based detectors.

According to your suggestion, we have added a new section (i.e., Section VI) to cite the papers and discuss the limitations and gains our algorithms by comparing with the other detectors/receivers. We have updated our manuscript as follows:

The 1-st Paragraph of Section VI (Page 24):

In this section, we compare and analyze our approach with existing MIMO detection design methodologies, including (conventional) model-based design methodology (e.g., [9]-[12]), deep learning (data-driven) based design methodology (e.g., [17]-[19], [46]), and hybrid model-based deep learning design methodology (e.g., [34], [47]).

The 2-nd Paragraph of Section VI (Page 25):

Subject to multiple factors, our algorithms temporarily are mainly applicable to small-scale MIMO settings, while the conventional model-based MIMO detectors (e.g., classical linear or iteration-based detectors) can achieve good performance for large-scale MIMO settings. Because of the powerful modeling and representation abilities, deep-learning based receivers can exploit latent or unknown gains (e.g., due to joint processing or data symbols [46]) or complex structure information to improve performance of interest in complicated systems or environments. The model-based deep learning designs can relax the requirement of "big data" and meanwhile enjoy the advantages of model-based methods. But, our methods fail to reap these advantages.

The 3-rd Paragraph of Section VI (Page 25):

However, our algorithms enjoy the following salient advantages. First, because CSI estimate is not required in our algorithms, they are not affected by the quality of estimated CSI and/or the performance of CSI estimation algorithm. Second, they have very low computational complexity. For example, there is no need to even invoke the clustering operation in Algorithm 2, which thus significantly reduces the complexity. Finally, when incorporating EL our method is very flexible, e.g., it can be implemented modularly and in parallel. These features make our algorithms very appealing in short packet based communications, in particular, like Internet-of-Things (IoT) [48].

The 4-th Paragraph of Section VI (Page 25):

It has been recognized that IoT calls for a new type of communication technologies [48]-[50]. First and foremost, the traffics in IoT applications are typically in the form of short packets [48]. Moreover, there are often stringent requirements on latency and reliability in mission-critical IoT applications, such as factory automation and vehicle-to-everything (V2X) communications. The requirements of low latency, low power consumption, high reliability as well as short packet transmission from IoT applications make our algorithms preferable over the existing techniques:

- Comparison with model-based detectors: Our algorithms offer better robustness and higher transmission efficiency, since they are not affected by CSI estimate. Moreover, lower SER and higher reliability can be achieved, by exploiting received signals of effective data.
- Comparison with deep learning based receivers: Thanks to using the lightweight ML methods (e.g., the clustering algorithm), the size of needed training samples is very low, which yields low computational complexity and low sample complexity.
- Comparison with hybrid model-based deep learning detectors: As a hybrid model and data based design, our methods enjoy the benefits of low complexities, which yield better real-time. Hence, our algorithms are suitable for IoT applications, in which stringent requirements (e.g., low latency, low power consumption and high reliability) are often imposed.

Thanks to the above advantages, our algorithms are also competitive in the other small-scale MIMO cases, if both low complexity and high performance are desired.

IEEE Transactions on Communications
Paper ID: TCOM-TPS-21-1382
Authors' Response to Reviewer 2

We would like to thank you for your insightful suggestions and comments, which have helped us improve the quality of our paper. We have revised our paper incorporating all your suggestions and comments.

Comment 1:

1, All my concerns were properly addressed, thus I recommend the manuscript for publication.

Response:

We very much appreciate your detailed review and positive affirmation.

IEEE Transactions on Communications
Paper ID: TCOM-TPS-21-1382
Authors' Response to Reviewer 3

We would like to thank you for your insightful suggestions and comments, which have helped us improve the quality of our paper. We have revised our paper incorporating all your suggestions and comments.

Comment 1:

1. The reviewer would like to disagree. The reviewer consents with the fact that you don't need to consider time explicitly, but this is not what was meant.

It is kind of nit-picky, but: If Y would be a set, then $\{1,2,1\} = \{1,2\}$ *which is not what you want to have. You want to have an order (namely time), such that* $\{1,2,1\}! = \{1,1,2\}! = \{1,2\}$.

Response:

We very much appreciate your detailed review and precious suggestion. We are also very grateful to you for patiently reminding us this issue once again. The received signals are given by

$$y_i = \sqrt{p}\mathbf{h}^{\mathrm{H}}\mathbf{s}_i + w_i,\ (i = 1, 2, 3, \cdots, T), \tag{1}$$

where $\{w_i\}$ are independently and identically distributed (i.i.d.) as $\mathcal{CN}(0,1)$, i.e., $w_i \sim \mathcal{CN}(0,1)$. The received signals are collected into set $\mathcal{Y} = \{y_1, y_2, \cdots, y_T\}$. We would like to highlight that as a continuous random variable, each complex Gaussian random variable w_i takes values in $(-\infty, +\infty)$, and thus the probability that two (or more) received signals (e.g., y_i and y_j) are equal is zero. More importantly, even if two (or more) received signals are equal in an extreme case, they should not be merged, since each of them corresponds to a transmitted symbol vector. Moreover, these received signals should be recorded chronologically (so as to correctly recover the corresponding symbols), which naturally introduces an order. For these reasons, we can safely ignore the time component, which does not cause any misunderstanding, and meanwhile makes our expressions more concisely.

Following your comment, we have updated the manuscript as follows:

Footnote 1 (Page 7):

Here, we ignore the mathematical rigor of the definition of the set, i.e., a set should not contain repetitive elements. In fact, the probability that two (or more) received signals (e.g., y_i and y_j) are equal is zero due to the continuous distribution of $\{w_i\}$. Even if they are equal they should not be merged, because each received signal corresponds to a transmitted symbol vector. The elements within $\mathcal{Y}$ should also be recorded chronologically, so as to recover the transmitted symbols correctly, which naturally introduces an order.