Received September 29, 2020, accepted October 9, 2020, date of publication October 14, 2020, date of current version October 27, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3030932

A FFT-Like MIMO Detection Algorithm

PENG DU[1,2], YUAN ZHANG[2], (Member, IEEE), AND TEER BA[2], (Member, IEEE)

[1]College of Automation & College of Artificial Intelligence, Nanjing University of Posts and Telecommunications, Nanjing 210023, China
[2]National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China

Corresponding author: Yuan Zhang (y.zhang@seu.edu.cn)

This work was supported by the National Key Research and Development Program of China under Grant 2018YFB1800800.

ABSTRACT This paper studies the multiple-input-multiple-output (MIMO) detection problem. Existing works model the MIMO detection as a tree or factor graph. This work adds a new member called the clique graph model to the MIMO detection graph model family so that MIMO detection can be transformed into finding the maximum clique in the constructed graph. Then, a clique based MIMO detection algorithm is developed, which uses an architecture similar to the Fast Fourier Transform (FFT) and can be executed in parallel naturally. The execution time of the algorithm can be proportional to $\log_2 N_t$ where N_t is the transmit antenna number. Simulation results are provided to evaluate the bit error rate (BER) performance of the proposed algorithm.

INDEX TERMS Detection, graph model, MIMO, parallel processing.

I. INTRODUCTION

Multiple-input-multiple-output (MIMO) detection has been extensively studied and many different methods have been proposed in the literature [1], [2]. Among them, we briefly review the two types of MIMO detection methods related to this work, namely the sphere decoding algorithm and the message passing algorithm.

The first is the sphere decoding (SD) algorithm, the key step of which is to express the MIMO detection problem as a tree model [3]. The early form of the SD algorithm has high complexity [4], [5]. After that, many strategies were proposed to reduce the complexity. Among them, there are two strategies that are widely accepted. The first is the best-K strategy [6], that is, after each processing, only the best K items are retained for the next round; the second is to apply Dijkstra's algorithm to search for the solution [7], [8].

The second is the message passing (MP) algorithm, the key step of which is to express the MIMO detection problem as a factor graph model [9], [10]. The early form of the MP algorithm also has high complexity. In order to reduce complexity, a large class of approximate MP (AMP) algorithms for merging messages are proposed in the literature. In the first AMP algorithm, an Onsager item is included to ensure the performance of the algorithm [11], [14]. Afterwards, starting from the original AMP algorithm, many variants that can further improve performance were proposed. Among them, including the VAMP (vector AMP) algorithm [15] and the OAMP (orthogonal AMP) algorithm [16], [17], which are actually equivalent to each other.

These methods (or equivalent, graph models) have their own characteristics in terms of obtaining high-quality solutions and paying the cost of complexity. In this work, a new member called the *clique graph model* will be added to the MIMO detection graph model family. In this graph model, each vertex represents each possible value of each variable, and whether there is an edge between any two vertices represents whether the combination of two corresponding values is feasible. Then, the vertices corresponding to the true solution will form a maximum clique in the constructed graph with a high probability. In this way, MIMO detection is actually to find the maximum clique in the constructed graph. Based on the above observations, we have also developed a clique-based MIMO detection algorithm. The algorithm uses an architecture similar to the Fast Fourier Transform (FFT) to decompose the task of computing a N_t-point clique into subtasks of computing two $(N_t/2)$-point cliques, and so on, until only the computation of the 2-point clique is required, where N_t represents the number of transmit antennas. The complexity analysis shows that the amount of computation required by the algorithm is proportional to the number of transmit antennas N_t. Further, due to the natural parallel

The associate editor coordinating the review of this manuscript and approving it for publication was Cong Pu.

TABLE 1. Summary of Notations.

notation	description
N_{t}	the number of transmitting antennas
N_{r}	the number of receiving antennas
M	the number of points in the constellation
S	$S = \{s_1, s_2, ..., s_M\}$, the set of points in the constellation
$\mathbf{x}$	$\mathbf{x} = [x_1, x_2, ..., x_{N_{\mathrm{t}}}]^{\mathrm{T}}$, the transmitted signals after constellation mapping
$\mathbf{H}$	the channel matrix of size $N_{\mathrm{r}} \times N_{\mathrm{t}}$
$\mathbf{z}$	the noise of size $N_{\mathrm{r}} \times 1$
$\mathbf{y}$	$\mathbf{y} = [y_1, y_2, ..., y_{N_{\mathrm{r}}}]^{\mathrm{T}}$, the received signals
v_{im}	vertex v_{im} represents the value of x_i is s_m
$\mathbf{Q}, \mathbf{R}$	perform QR decomposition on the matrix $\mathbf{H}$ to obtain $\mathbf{H} = \mathbf{QR}$
G	the constructed graph
I	a subset $I \subseteq \{1, 2, \cdots, N_{\mathrm{t}}\}$
G_I	the subgraph of G composed of all vertices $\{v_{i\cdot}\}_{i\in I}$ and corresponding edges $(v_{i\cdot}, v_{i'\cdot}\}_{i,i'\in I}$
$\mathbf{x}_I$	$\mathbf{x}_I = \{x_i\}_{i\in I}$
$\mathbf{x}_I[h]$	the value of $\mathbf{x}_I$ in the hth round
K_h	the number of items retained in the hth round

execution structure, the actual execution time of the algorithm can only be proportional to the base-2 logarithm of the number of transmit antennas (i.e., $\log_2 N_{\mathrm{t}}$).

The contributions of this work are summarized as follows.

- Firstly, a new graph model of the MIMO detection problem is established and the MIMO detection is expressed as a maximum clique problem.
- Secondly, a clique based MIMO detection algorithm is developed. The algorithm uses an architecture similar to the FFT, and can be executed in parallel naturally. The execution time of the algorithm can only be proportional to $\log_2 N_{\mathrm{t}}$ where N_{t} is the antenna number. Compared with the SD algorithm, the computation steps of the proposed algorithm are fixed; compared with the AMP algorithm, the proposed algorithm does not need to compute the matrix inverse.
- Thirdly, simulation experiments are carried out to verify that the bit error rate (BER) performance of the proposed algorithm is better than the existing algorithms.

Additionally, as a step toward a detection algorithm suitable for communication systems with a large number of antennas, this work focuses on the basic uncoded MIMO system. In modern MIMO communication systems, space-time-frequency (STF) coding is a very important way to use multiple antennas [18]–[21]. We will study how to extend the proposed FFT-like detection algorithm to the STF coded MIMO system in our future work.

The rest of this paper is organized as follows. Section II introduces the existing MIMO detection models. Section III presents the new clique graph model. Section IV develops a clique based MIMO detection algorithm. Section V evaluates the BER performance of the proposed algorithm. Section VI gives concluding remarks. The main notations used in this paper are summarized in Table 1.

II. EXISTING MIMO DETECTION MODELS

Consider a MIMO system equipped with N_{t} transmit antennas and N_{r} receive antennas. Without loss of generality, assume that N_{r} is not less than N_{t}. Suppose that M-ary modulation is used, and the set of points in the constellation is $S = \{s_1, s_2, \ldots, s_M\}$. Let $\mathbf{x} = [x_1, x_2, \ldots, x_{N_{\mathrm{t}}}]^{\mathrm{T}}$ denote the transmitted signals after constellation mapping. That is, we have $\mathbf{x} \in S^{N_{\mathrm{t}}}$. Let $\mathbf{H}$ denote the channel matrix of size $N_{\mathrm{r}} \times N_{\mathrm{t}}$ and $\mathbf{z} \sim \mathrm{CN}(\mathbf{0}, \sigma_{\mathrm{z}}^2\mathbf{I})$ denote the noise signal where CN represents circularly symmetric complex Gaussian distribution. Then we have,

$$\mathbf{y} = \mathbf{Hx} + \mathbf{z}, \tag{1}$$

where $\mathbf{y} = [y_1, y_2, \ldots, y_{N_{\mathrm{r}}}]^{\mathrm{T}}$ representing the received signals. Assume that the receiver has perfectly obtained the information of the channel matrix $\mathbf{H}$ and the noise variance σ_{z}^2. The task of MIMO detection is to infer the value of $\mathbf{x}$ based on the value of $\mathbf{y}$.

Generally speaking, MIMO detection is to solve the following optimization problems,

$$\min_{\mathbf{x}\in S^{N_{\mathrm{t}}}} ||\mathbf{y} - \mathbf{Hx}||^2. \tag{2}$$

This is an integer optimization problem, and its solution space may be very huge when the modulation dimension and the number of antennas are not small.

To solve this integer optimization problem, the simple idea is to relax the integer constraint in (2), so that various descent direction methods can be used to find the optimal relaxation solution, and then rounded as the detection result. The typical representative of this idea is the MMSE or ZF algorithm and many other algorithms. Although this relaxation idea has low complexity, the gap between the obtained relaxed solution and the optimal solution is usually large, resulting in a high bit error rate.

In order to solve (2) better, the more advanced idea is to keep the nature of integer programming unchanged and express the optimization problem as a graph. Two types of graph models have been proposed for solving (2) in the literature. The first type is the tree model, and its typical representative is the sphere decoding algorithm. The second type is the factor graph model, and its typical representative

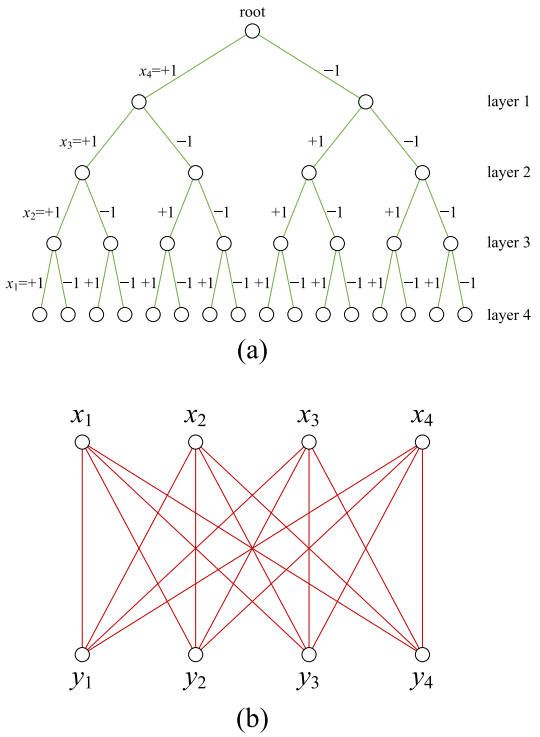

FIGURE 1. Existing graph models of MIMO detection for a simple 4 × 4 MIMO system with $S = \{+1, -1\}$: (a) the tree model, (b) the factor graph model.

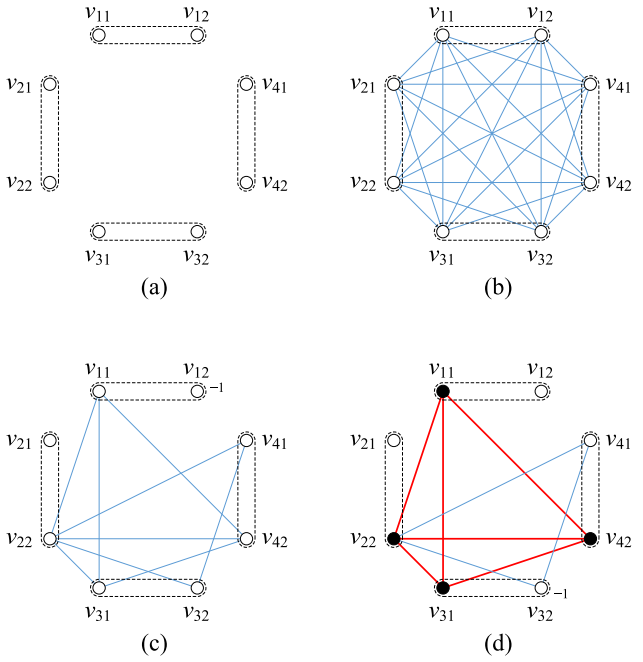

FIGURE 2. The clique graph model of MIMO detection for a simple 4 × 4 MIMO system with $S = \{+1, -1\}$: (a) the vertices, (b) all possible edges, (c) the constructed graph model, (d) the maximum clique.

is the message passing algorithm. Fig. 1(a) and (b) illustrates an example of the tree and factor graph model, respectively.

III. A NEW GRAPH MODEL

The above-mentioned MIMO detection graph models have their own characteristics in terms of obtaining high-quality solutions and paying the cost of complexity. In this work, a new member called the *clique graph model* will be added to the MIMO detection graph model family. Fig. 2(c) illustrates an example of this new graph model. In the following, we will introduce the definition and construction process of the vertices and edges in the clique graph model, respectively.

A. THE VERTICES

In the MIMO system of (1), there are N_t unknown variables $\{x_i, 1 \le i \le N_t\}$. Each variable x_i can only take its value in the constellation point set S containing M elements. Therefore, we define a total of NM vertices $\{v_{im}, 1 \le i \le N_t, 1 \le m \le M\}$, where each vertex v_{im} represents that the value of x_i is s_m.

For example, consider the MIMO detection for a simple 4×4 MIMO system with $S = \{+1, -1\}$. Thus, there will be a total of $4 \times 2 = 8$ vertices $\{v_{im}, 1 \le i \le 4, 1 \le m \le 2\}$ in the graph. Among them, vertex v_{11} and v_{12} represents that the value of x_1 is $+1$ and -1 respectively, and so on. Fig. 2(a) illustrates the vertices for this example.

B. THE EDGES

Before giving the definition of the edge in the graph, we need to introduce some notations. First, we perform QR decomposition on the matrix $\mathbf{H}$ to obtain $\mathbf{H} = \mathbf{QR}$ and calculate $\mathbf{y}' = \mathbf{Q}^T\mathbf{y}$. It should be noted that we do not actually need to perform a complete QR decomposition. This is because we only need the second row from back to front of the matrix $\mathbf{R}$ and vector $\mathbf{y}'$ to get a constraint involving only two variables as,

$$|y'_{N_t-1} - r_{N_t-1,N_t-1}x_{N_t-1} - r_{N_t-1,N_t}x_{N_t}| \le a\sigma_z, \qquad (3)$$

where a is a parameter which can be set to for example 3. We call such a constraint a two-dimensional constraint for the two variables x_{N_t-1} and x_{N_t}. Furthermore, by adjusting the order of the variables, we can use this method to get the two-dimensional constraint for any two variables x_i and $x_{i'}$ with $i \ne i'$. Now we give the definition of the edges in the graph. For any two vertices v_{im} and $v_{i'm'}$ in the graph, there is a two-dimensional partial solution $(x_i = s_m, x_{i'} = s_{m'})$. If this two-dimensional partial solution satisfies the corresponding two-dimensional constraint, we will call this two-dimensional partial solution feasible, and use an edge $(v_{im}, v_{i'm'})$ to connect the corresponding two vertices in the graph.

Continue to consider the previous example of a simple 4×4 MIMO system with $S = \{+1, -1\}$. First of all, according to the definition of edges, we know that there is no edge between vertices with the same user index; but it is possible to have an edge between vertices with different user index. Fig. 2(b) illustrates all possible edges for this example. Of all these possible edges, only some are feasible. Assume the transmitted signals are $\mathbf{x} = [+1, -1, +1, -1]^T$, the channel

matrix is,

$$\mathbf{H} = \begin{bmatrix} 1.0933 & -1.2141 & -0.7697 & -1.0891 \\ 1.1093 & -1.1135 & 0.3714 & 0.0326 \\ -0.8637 & -0.0068 & -0.2256 & 0.5525 \\ 0.0774 & 1.5326 & 1.1174 & 1.1006 \end{bmatrix}, \quad (4)$$

the noise variance $\sigma_z^2 = 0.1$ and the randomly generated noise is $\mathbf{z} = [0.4883, 0.0272, -0.4717, -0.2347]^{\mathrm{T}}$, and the received signals are $\mathbf{x} = [3.1151, 2.5888, -2.1066, -1.6733]^{\mathrm{T}}$. Substituting these numbers, we can calculate that only nine sets of two-dimensional partial solutions are feasible: $(x_1 = +1, x_2 = -1)$, $(x_1 = +1, x_3 = +1)$, $(x_1 = +1, x_4 = -1)$, $(x_2 = -1, x_3 = +1)$, $(x_2 = -1, x_3 = -1)$, $(x_2 = -1, x_4 = +1)$, $(x_2 = -1, x_4 = -1)$, $(x_3 = +1, x_4 = -1)$, and $(x_3 = -1, x_4 = +1)$. Accordingly, we can only keep nine edges in the graph: (v_{11}, v_{22}), (v_{11}, v_{31}), (v_{11}, v_{42}), (v_{22}, v_{31}), (v_{22}, v_{32}), (v_{22}, v_{41}), (v_{22}, v_{42}), (v_{31}, v_{42}), and (v_{32}, v_{41}). Fig. 2(c) illustrates the constructed graph model for this example.

IV. THE CLIQUE BASED MIMO DETECTION ALGORITHM

In this section, the clique graph model based MIMO detection algorithm is developed.

A. THE PRINCIPLE

How to use the constructed new graph model for MIMO detection? The answer to this question depends on the following two observations.

We first introduce the first observation. Suppose the true solution of (2) is $\{x_i = s_{m_i^*}\}_{1\le i\le N_t}$, which corresponds to N_t vertices $\{v_{im_i^*}\}_{1\le i\le N_t}$ in the graph model. Further, this true solution can be decomposed into $N_t(N_t-1)/2$ two-dimensional true partial solutions $\{x_i = s_{m_i^*}, x_{i'} = s_{m_{i'}^*}\}_{1\le i\ne i'\le N_t}$, which correspond to the positions of $N_t(N_t-1)/2$ edges $\{(v_{im_i^*}, v_{i'm_{i'}^*})\}_{1\le i\ne i'\le N_t}$ in the graph model. As introduced in the previous section, when constructing the graph model, many edges between pairs of vertices are deleted due to infeasibility and do not exist. However, for the $N_t(N_t-1)/2$ edges derived from the true solution, since this solution is true, we can think that these $N_t(N_t-1)/2$ edges exist in the constructed graph with a high probability. Applying the concept of clique in graph theory, we can say that the N_t vertices corresponding to the true solution should form a clique in the graph model. According to the definition of the graph, it can be seen that the clique in the constructed graph includes at most N_t vertices. Therefore, we can say that *the N_t vertices corresponding to the true solution should form a maximum clique in the graph model*. In this way, MIMO detection is actually to find the maximum clique in the constructed graph model. This is the first observation that our proposed MIMO detection algorithm relies on.

The second observation is a further extension of the first observation. We only consider a subset of the true solution $\{x_{i_1} = s_{m_{i_1}^*}, \cdots, x_{i_L} = s_{m_{i_L}^*}\}$, which corresponds to L vertices in the graph model. Similarly, this subset of the true solution can be decomposed into $L(L-1)/2$ two-dimensional true partial solutions, which correspond to the positions of $L(L-1)/2$ edges in the graph model. Also since this solution is true, we can think that these $L(L-1)/2$ edges exist in the constructed graph with a high probability. Using graph theory terms, we can say that *the L vertices corresponding to any subset of L elements of the true solution should form a clique in the graph*. This is the second observation that our proposed MIMO detection algorithm relies on.

Still consider the previous example of a simple 4×4 MIMO system with $S = \{+1, -1\}$. In this example, we have assumed the true solution is $\{x_1 = +1, x_2 = -1, x_3 = +1, x_4 = -1\}$, which corresponds to four vertices $\{v_{11}, v_{22}, v_{31}, v_{42}\}$ in the graph, as shown by the four black solid circles in Fig. 2(d). This true solution can be decomposed into six two-dimensional true partial solutions which correspond to the positions of six edges (v_{11}, v_{22}), (v_{11}, v_{31}), (v_{11}, v_{42}), (v_{22}, v_{31}), (v_{22}, v_{42}), and (v_{31}, v_{42}) in the graph model, as shown by the six red lines in Fig. 2(d). It can be seen that these four vertices form the maximum clique in the graph. Further, the subsets of the true solution also form smaller cliques in the graph. For example, the vertices $\{v_{11}, v_{22}, v_{31}\}$ also form a clique in the graph.

B. THE MIMO DETECTION ALGORITHM

In the previous section, we have modeled MIMO detection as the maximum clique problem. Many maximum clique algorithms have been proposed in the field of combinatorial optimization. However, those algorithms are all for general graphs. This work will proposes a new algorithm framework to find the maximum clique in the constructed graph specifically for the MIMO detection problem.

First introduce some notations. Let G be the constructed graph. Since the graph G is derived from N_t variables $\{x_i\}_{1\le i\le N_t}$, we also call G a N_t-dimensional graph. Given a subset $I = \{i_1, \cdots, i_L\} \subseteq \{1, 2, \cdots, N_t\}$, let G_I be the subgraph of G composed of all vertices $\{v_{i\cdot}\}_{i\in I}$ and corresponding edges $(v_{i\cdot}, v_{i'\cdot})_{i,i'\in I}$ associated with the subset I, and call G_I a L-dimensional subgraph of G. In addition, we call a clique containing n vertices a n-point clique.

The idea of the proposed algorithm is conveniently illustrated by considering the special case of N_t an integer power of 2, i.e., $N_t = 2^p$. Since N_t is divisible by two, we can consider computing the maximum clique by separating all variables $\{x_i\}_{1\le i\le N_t}$ into two $(N_t/2)$-elements subsets consisting of the first half of the variables $\mathbf{x}_I = \{x_i\}_{i\in I}$ where $I = \{1, \cdots, N_t/2\}$ and the second half of the variables $\mathbf{x}_J = \{x_i\}_{i\in J}$ where $J = \{N_t/2+1, \cdots, N_t\}$. Fig. 3 shows this decomposition.

To understand the significance of Fig. 3 as an organizing principle for the proposed clique based MIMO detection, it is helpful to consider the computation of $(N_t/2)$-point cliques in $(N_t/2)$-dimensional subgraph as a black box. Take the first black box as an example, which focuses on the subgraph G_I. The input of this black box is the initial value of all variables associated with I, denoted as $\mathbf{x}_I[0]$. The initial value of each variable may be any one from s_1 to s_M. Thus, for

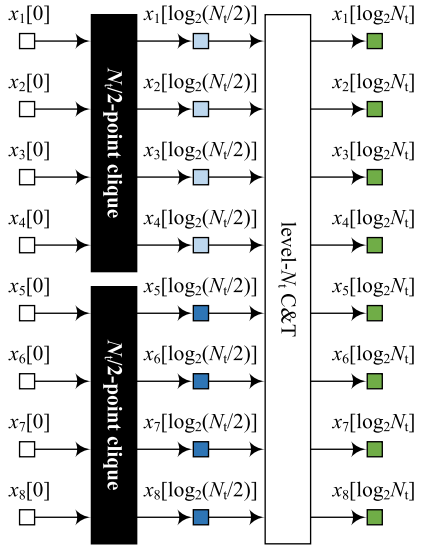

FIGURE 3. Flow graph of the decomposition of an N_t-point clique computation into two ($N_t/2$)-point clique computations with $N_t = 8$, where "C&T" represents "combination and test".

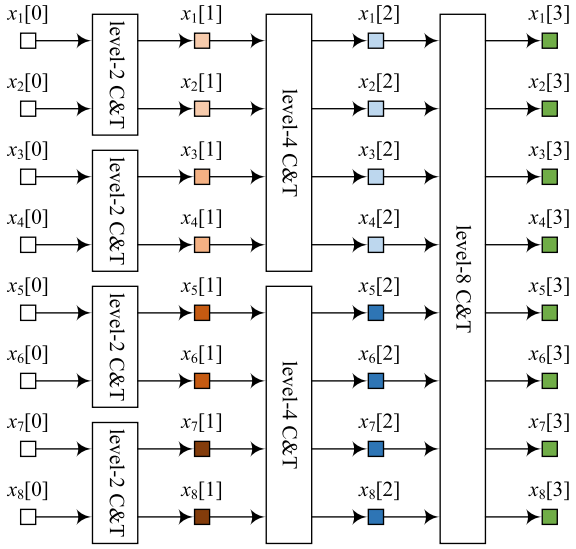

FIGURE 4. Flow graph of complete decomposition of an 8-point clique computation, where "C&T" represents "combination and test".

a $(N_t/2)$-dimensional subgraph G_I, there are $M^{N_t/2}$ possible values for $\mathbf{x}_I[0]$. The output of the black box is all possible $(N_t/2)$-point cliques in the $(N_t/2)$-dimensional subgraph G_I, denoted as $\mathbf{x}_I[\log_2(N_t/2)]$, where the index $\log_2(N_t/2)$ is for the convenience of subsequent expansion.

We temporarily put aside the working mechanism of the black box, but first focus on how to deal with the two $(N_t/2)$-point cliques generated by the two black boxes, and get the final N_t-point clique. This involves two steps.

- *The Combination Step:* The first step is to combine to generate a N_t-dimensional solution. Given two $(N_t/2)$-point cliques $\mathbf{x}_I[\log_2(N_t/2)]$ and $\mathbf{x}_J[\log_2(N_t/2)]$, since $I \cap J = \emptyset$, they can be combined directly. That is, let $\mathbf{x}_{I\cup J}[\log_2 N_t]$ denote the solution obtained after the combination, then we have,

$$\mathbf{x}_{I\cup J}[\log_2 N_t] = \mathbf{x}_I[\log_2(N_t/2)] \cup \mathbf{x}_J[\log_2(N_t/2)]. \quad (5)$$

  In particular, since $I \cup J$ is the full set, we abbreviate $\mathbf{x}_{I\cup J}[\log_2 N_t]$ as $\mathbf{x}[\log_2 N_t]$ here.
- *The Test Step:* It should be noted that $\mathbf{x}[\log_2 N_t]$ is not necessarily a N_t-point clique (that is, $\mathbf{x}[\log_2 N_t]$ is not necessarily the true solution). So the second step is to test whether $\mathbf{x}[\log_2 N_t]$ is a N_t-point clique. Since the graph has been constructed, we can test whether $\mathbf{x}[\log_2 N_t]$ is a N_t-point clique only according to the definition of clique. Only the solution that passes the test can be retained, otherwise the solution must be discarded.

In this way, we break the original N_t-point clique computation into two $(N_t/2)$-point clique computations. For convenience, we refer to these two steps collectively as the level-N_t C&T operation, as illustrated in Fig. 3.

If $N_t/2$ is even, as it is when N_t is equal to a power of 2, then we can consider computing each of the $(N_t/2)$-point clique by breaking it into computing two $(N_t/4)$-point cliques, which would then be combined and tested, that is, the level-$(N_t/2)$ C&T operation, to yield the $(N_t/2)$-point clique. Inserting the computation of $(N_t/2)$-point clique into the flow graph of Fig. 3, we obtain the flow graph of computing the N_t-point clique which can be obtained by computing four $(N_t/4)$-point cliques. Consequently, we would proceed by decomposing the $(N_t/4)$-point clique computations into $(N_t/8)$-point clique computations and continue until we were left with only 2-point clique computations. This requires $\log_2 N_t$ stages of computation. For the 8-point clique that we have been using as an illustration, the complete computation flow graph is shown in Fig. 4.

C. THE PRACTICAL VERSION

In the previous section, we proposed a clique-based MIMO detection algorithm. This algorithm uses a framework similar to Fast Fourier Transform (FFT), and can be expected to greatly increase the execution speed. However, when the modulation dimension and the number of antennas are large, the complexity of the algorithm is still high. For example, consider a 256×256 MIMO system and 16QAM modulation, i.e., $N_t = 256$ and $M = 16$. At this time, there will be $N_t \times M = 4096$ vertices in the graph model. Further, since there may be edges between any two vertices v_{im} and $v_{i'm'}$, at most $4096 \times (4096 - 16)/2 = 8,355,840$ computations are required to determine which edges are in the graph model. Obviously, such a high level of complexity is impractical. Therefore, we need to further improve the clique based MIMO detection algorithm. The basic idea of improvement is that it is not necessary to generate a complete graph model in advance, but it is sufficient to generate the partial information of the graph when needed.

We put the pseudo-code description of the practical version of clique based MIMO detection algorithm in **Algorithm** 1. The algorithm requires a total of $\log_2 N_t$ rounds of

Algorithm 1 The Clique Based MIMO Detection

1: Initialize $x_i[0]$ which can be any one in S;
2: **for** $h = 1 : \log_2 N_\mathrm{t}$ **do**
3: **for** $l = 1 : (N_\mathrm{t}/2^h)$ **do**
4: Let $I = \{(l-1)2^h + 1, \cdots, (l-1)2^h + 2^{h-1}\}$;
5: Let $J = \{(l-1)2^h + 2^{h-1} + 1, \cdots, (l-1)2^h + 2^h\}$;
6: **for** all possible $\mathbf{x}_I[h-1]$ and $\mathbf{x}_J[h-1]$ **do**
7: Combine $\mathbf{x}_{I\cup J}[h] = \mathbf{x}_I[h-1] \cup \mathbf{x}_J[h-1]$;
8: Compute the metric of $\mathbf{x}_{I\cup J}[h]$ by (6);
9: **end for**
10: Select K_h $\mathbf{x}_{I\cup J}[h]$s with the smallest metric;
11: **end for**
12: **end for**
13: Select $\mathbf{x}[\log_2 N_\mathrm{t}]$ with the smallest metric as the output;

computation (see Line 2). At the hth round, the algorithm performs a total of $N_\mathrm{t}/2^h$ level-(2^h) C&T operations (see Line 3). In the lth operation of the hth round, the algorithm first selects two (2^h-1)-point cliques $\mathbf{x}_I[h-1]$ and $\mathbf{x}_J[h-1]$ from the previous round and combines them into a new (2^h)-dimensional partial solution $\mathbf{x}_{I\cup J}[h]$ (see Lines 7), and then tests whether it is a clique (see Lines 8). This test needs to use the partial information of the graph. Many different test methods can be designed. This work considers the following test method.

We need to adjust the order of the variables in $\mathbf{x} = [x_1, x_2, \cdots, x_{N_\mathrm{t}}]^\mathrm{T}$ so that $\mathbf{x}$ becomes $[\mathbf{x}\backslash\mathbf{x}_{I\cup J}, \mathbf{x}_{I\cup J}]$. Let the variable obtained after adjusting the order be $\mathbf{x}'$. Correspondingly, the same adjustment is performed on the channel matrix $\mathbf{H}$, and the adjusted channel matrix is denoted as $\mathbf{H}'$. Then, we perform QR decomposition on the matrix $\mathbf{H}'$ to obtain $\mathbf{H}' = \mathbf{Q}'\mathbf{R}'$ and calculate $\mathbf{y}' = (\mathbf{Q}')^\mathrm{T}\mathbf{y}$. It should be noted that we do not actually need to perform a complete QR decomposition. This is because we only need a sub-matrix in the matrix $\mathbf{R}'$, whose position in $\mathbf{R}'$ is from the $(N_\mathrm{t} - |I \cup J| + 1)$th to the N_tth row and from the $(N_\mathrm{t} - |I \cup J| + 1)$th to the N_tth column. Let this square sub-matrix be $\mathbf{R}''$. Correspondingly, denote the sub-vector formed by the $(N_\mathrm{t} - |I \cup J| + 1)$th to the N_tth elements of $\mathbf{y}'$ as $\mathbf{y}''$. Next, we calculate the metric associated with this (2^h)-dimensional partial solution $\mathbf{x}_{I\cup J}[h]$ as,

$$\zeta(\mathbf{x}_{I\cup J}[h]) \triangleq ||\mathbf{y}'' - \mathbf{R}''\mathbf{x}_{I\cup J}[h]||^2. \tag{6}$$

Sort all possible $\mathbf{x}_{I\cup J}[h]$s in ascending order according to their metric values, and then keep the first K_h $\mathbf{x}_{I\cup J}[h]$s with the smallest metric to enter the next round, where K_h is a parameter (see Lines 10).

After all $\log_2 N_\mathrm{t}$ rounds of computations are completed, if there is more than one output, choose the one with the smallest metric as the final output (see Lines 13).

D. THE COMPLEXITY ANALYSIS

Finally, we make an approximate analysis of the complexity of the proposed MIMO detection algorithm. The algorithm consists of a total of $\log_2 N_\mathrm{t}$ rounds. At the hth round, the algorithm consists of a total of $N_\mathrm{t}/2^h$ sub-rounds. In each sub-round, approximately K_h^2 new (2^h)-dimensional partial solution $\mathbf{x}_{I\cup J}[h]$s will be generated and tested. So in each sub-round the amount of performed computation are approximately proportional to K_h^2. Therefore, the amount of the computation of the whole algorithm is approximately proportional to,

$$\sum_{h=1}^{\log_2 N_\mathrm{t}} \sum_{l=1}^{\frac{N_\mathrm{t}}{2^h}} K_h^2 \le (N_\mathrm{t} - 1) \cdot (\max_h K_h)^2. \tag{7}$$

This shows that the complexity of the proposed MIMO detection algorithm is approximately proportional to the number of transmit antennas N_t.

Furthermore, it can be noted that the proposed MIMO detection algorithm has a natural parallel implementation architecture. Actually, in Fig. 3, the two black boxes are independent of each other and can be executed in parallel. In this way, taking the level 2 processing in Fig. 4 as an example, it can be found that all four level-2 processing modules are independent of each other and can be executed in parallel, and so on. Thus, if the factors that can be executed in parallel are taken into account, the execution time complexity of the proposed MIMO detection algorithm will be approximately proportional to,

$$\sum_{h=1}^{\log_2 N_\mathrm{t}} K_h^2 \le \log_2 N_\mathrm{t} \cdot (\max_h K_h)^2. \tag{8}$$

This shows that the execution time complexity of the proposed MIMO detection algorithm can be approximately proportional to the base-2 logarithm of the number of transmit antennas N_t.

V. PERFORMANCE EVALUATION

In this section, we will provide the results of simulation evaluation of the performance of the proposed clique-based MIMO detection algorithm and compare it with other existing MIMO detection algorithms.

A. SIMULATION SETTINGS

In the simulations conducted in this work, a 256×256 MIMO system is considered, and a 64QAM or 16QAM constellation is used. The Kronecker channel model is used and the channel matrix $\mathbf{H}$ is expressed as $\mathbf{H} = \mathbf{C}_\mathrm{R}^{1/2}\tilde{\mathbf{H}}\mathbf{C}_\mathrm{T}^{1/2}$, where $\tilde{\mathbf{H}}$ consists of independently identically complex Gaussian distributed elements with zeros mean and unit variance, and $\mathbf{C}_\mathrm{R}$ and $\mathbf{C}_\mathrm{T}$ are the receive and transmit correlation matrices, respectively. For both $\mathbf{C}_\mathrm{R}$ and $\mathbf{C}_\mathrm{T}$, the value of the (i, j)th element is set to $\alpha^{|i-j|}$ where $\alpha \in [0, 1)$ is the correlation coefficient. Given signal-to-noise ratio (SNR) ρ, noise is generated by independently identically complex Gaussian distributed random variables with zeros mean and variance $1/\rho$.

Three types of MIMO detection algorithms are considered and compared. The first type is the MIMO detection algorithm based on the message passing (i.e. the factor graph model). The complexity of the full version of this type of

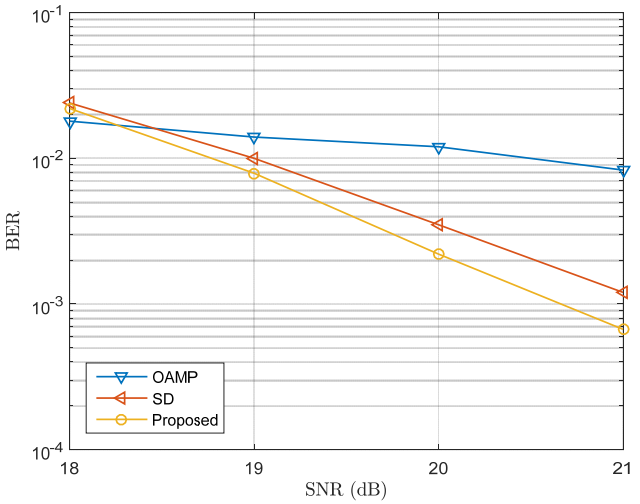

FIGURE 5. The BER performance of the proposed MIMO detection algorithm for 256 × 256 MIMO with 64QAM and $\alpha = 0.5$.

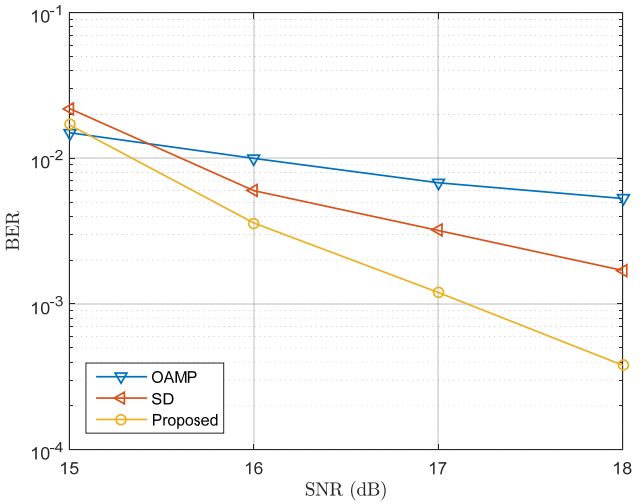

FIGURE 6. The BER performance of the proposed MIMO detection algorithm for 256 × 256 MIMO with 64QAM and $\alpha = 0$.

MIMO detection algorithm is too high. In order to get the results within an acceptable length of time, we chose the recently proposed OAMP detection algorithm [16], [17]. This algorithm is an approximate message passing algorithm, and its complexity is mainly reflected in the need to compute the inverse of the matrix derived from the channel matrix multiple times. According to the simulation configuration used in this work, the algorithm needs to compute the inverse of the 256 × 256 matrix multiple times. The second type is the MIMO detection algorithm based on the sphere decoding (i.e., the tree graph model). Similarly, in order to get the results within an acceptable length of time, we choose a version which uses Dijkstra's algorithm and keeps only the best K for each processing [7], [8]. The third type is the proposed clique based MIMO detection algorithm (i.e., the clique graph model).

Since the antenna configuration considered in the simulation is 256×256, the proposed algorithm requires $\log_2 256 = 8$ stages of calculation. The parameter K_h required in each stage is set to the same as 50. Correspondingly, the parameter K required in the sphere decoding detection algorithm is also set to 50. It is assumed that the channel estimation is ideal. In addition, this work focuses on the uncoded system, so the simulation result is the raw bit error rate that does not include any decoding operation. Given the parameter configuration, the simulation experiment is repeated one hundred times and then averaged as the final result.

B. SIMULATION RESULTS

In the first set of simulation experiments, we set the parameter α in the channel model to 0.5 and use the 64QAM modulation. We plot the simulation results of three different types of MIMO detection algorithms in Fig. 5. Observing the curves in the figure, it can be seen that the sphere decoding and the proposed algorithm have a similar descending law, which is faster than the OAMP algorithm. This is mainly due to the different basic ideas of these algorithms. For example, for the OAMP algorithm, it seems impossible to achieve the BER performance of 10^{-3} under the considered configuration. But for sphere decoding and the proposed algorithm, in order to achieve the BER performance of 10^{-3}, the required SNR is approximately 21.4dB and 20.6dB respectively, which is about 1dB difference. Moreover, the proposed algorithm also has advantages in terms of complexity. Specifically, first, the proposed algorithm does not need to compute the matrix inverse; second, a large number of computations in the proposed algorithm can naturally be executed in parallel, which helps to shorten the execution time; third, the computation steps in the proposed algorithm are fixed and can be known in advance. Therefore, the proposed algorithm is more suitable for MIMO detection when the number of antennas is large and the modulation order is high.

In the second set of simulation experiments, we set the parameter α in the channel model to 0 and still use the 64QAM modulation. We plot the simulation results of three different types of MIMO detection algorithms in Fig. 6. Observing the curves in the figure, it can be seen that the BER curve of the sphere decoding algorithm declines faster than the OAMP algorithm, and the BER curve of the proposed algorithm declines faster than the sphere decoding algorithm. Moreover, compared with the previous set of simulation experiments, since the correlation coefficient α is reduced to 0, the SNR required for the BER performance to reach 10^{-3} is also reduced. For example, for the sphere decoding algorithm and the proposed algorithm, in order to achieve the BER performance of 10^{-3}, the required SNR is approximately 19dB and 17.2dB, a decrease of approximately 2.4dB and 3.4dB, respectively. The required SNR reduction of the proposed algorithm is more than that of the sphere decoding algorithm.

In the third set of simulation experiments, we set the parameter α in the channel model to 0.5 and use the 16QAM modulation. We plot the simulation results of three different types of MIMO detection algorithms in Fig. 7. Observing

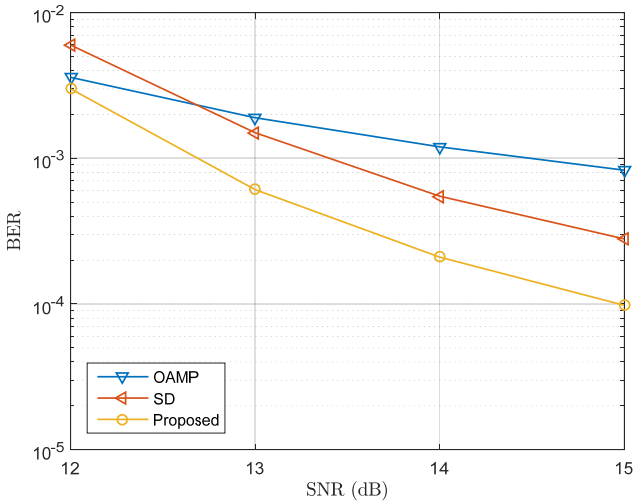

FIGURE 7. The BER performance of the proposed MIMO detection algorithm for 256 × 256 MIMO with 16QAM and $\alpha = 0.5$.

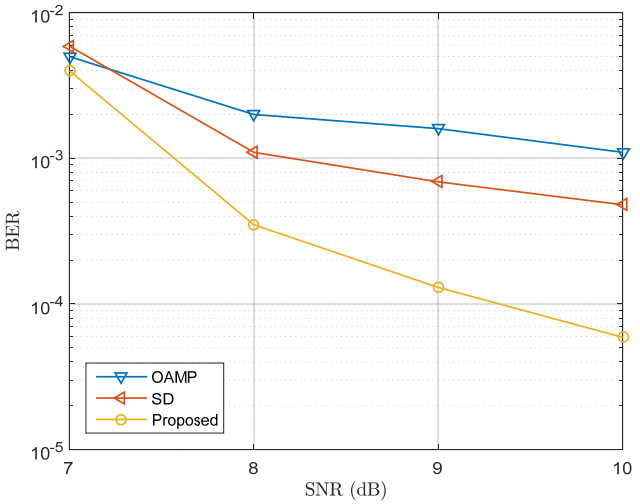

FIGURE 8. The BER performance of the proposed MIMO detection algorithm for 256 × 256 MIMO system with 16QAM and $\alpha = 0$.

the curves in the figure, it can be seen that among all three types of algorithms, the BER curve of the proposed algorithm has the fastest decline, not to mention the advantages of the proposed algorithm in shortening the execution time. Moreover, compared with the modulation using 64QAM, since the constellation becomes sparse, the SNR required for the BER performance to reach 10^{-3} is also reduced. For example, for the OAMP algorithm, the sphere decoding algorithm, and the proposed algorithm, in order to achieve the BER performance of 10^{-3}, the required SNR is approximately 14.4dB, 13.4dB, and 12.7dB, respectively. Compared with the 64QAM modulation, the required SNR of the proposed algorithm is reduced by about 7.9dB.

In the fourth set of simulation experiments, we set the parameter α in the channel model to 0 and still use the 16QAM modulation. We plot the simulation results of three different types of MIMO detection algorithms in Fig. 8. Observing the curves in the figure, it can be seen that the BER performance gap of the three types of algorithms still exist. For example, for the OAMP algorithm, the sphere decoding algorithm, and the proposed algorithm, to achieve a BER performance of 10^{-3}, the required SNR is approximately 10dB, 8.3dB, and 7.7dB, respectively. Compared with the case where the correlation coefficient is 0.5, the performance advantage of the proposed algorithm over the OAMP algorithm is increased from 1.7dB to 2.3dB, while the performance advantage over the sphere decoding algorithm remains about 0.6dB. Therefore, the proposed algorithm is more suitable for MIMO detection when the number of antennas is large.

In the next three sets of experiments, we all set the parameter α in the channel model to 0. In the fifth set of simulation experiments, we consider the situation where the number of transmitting and receiving antennas are not equal, that is, consider a 128 × 256 MIMO system with 64QAM modulation. We plot the simulation results of three different types of MIMO detection algorithms in Fig. 9. Observing the

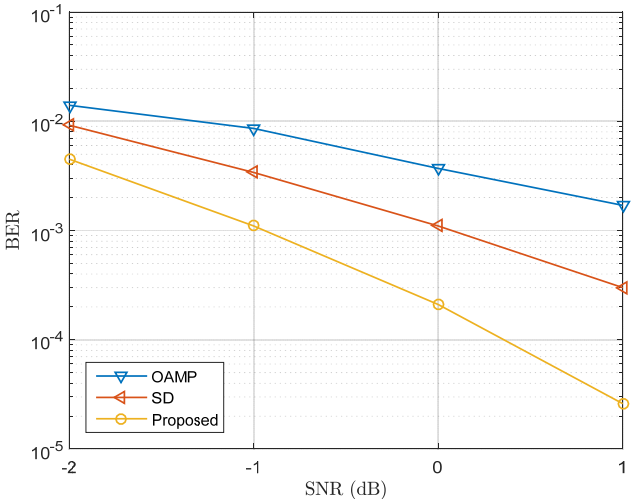

FIGURE 9. The BER performance of the proposed MIMO detection algorithm for 128 × 256 MIMO system with 64QAM and $\alpha = 0$.

curves in the figure, it can be seen that the BER performance of the proposed algorithm is still better than the OAMP algorithm and the sphere decoding algorithm. For example, to achieve a BER performance of 10^{-3}, the required SNR is approximately 2dB, 0dB, and -1dB, respectively. Therefore, the proposed algorithm is also suitable for MIMO detection when the number of transmitting and receiving antennas are not equal.

In the sixth set of simulation experiments, we evaluate the effect of channel estimation error on BER performance. In the simulations, we consider a 256 × 256 MIMO system with 16QAM modulation and assume that the channel matrix used by the receiver is $\hat{\mathbf{H}} = \mathbf{H} + \delta\mathbf{e}$, where δ is the coefficient and $\mathbf{e} \sim \mathrm{CN}(\mathbf{0}, \sigma_z^2\mathbf{I})$ representing the channel estimation error. We plot the simulation results of three different types of MIMO detection algorithms with $\delta = 0.01$ in Fig. 10. Observing the curves in the figure, it can be seen that when there is channel estimation errors, the BER performance of

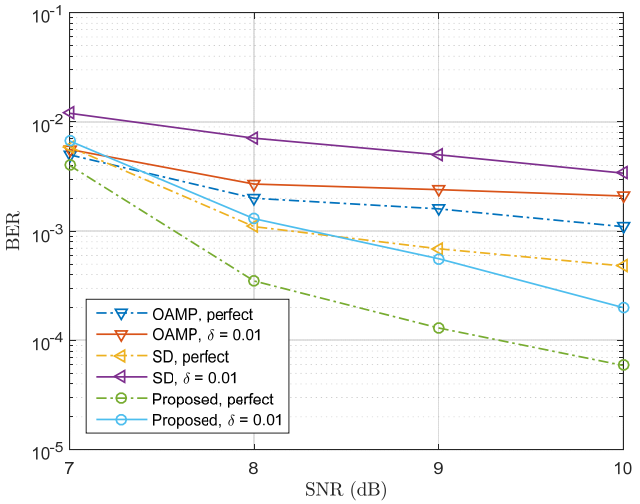

FIGURE 10. Impact of the channel estimation errors for 256 × 256 MIMO with 16QAM and $\alpha = 0$.

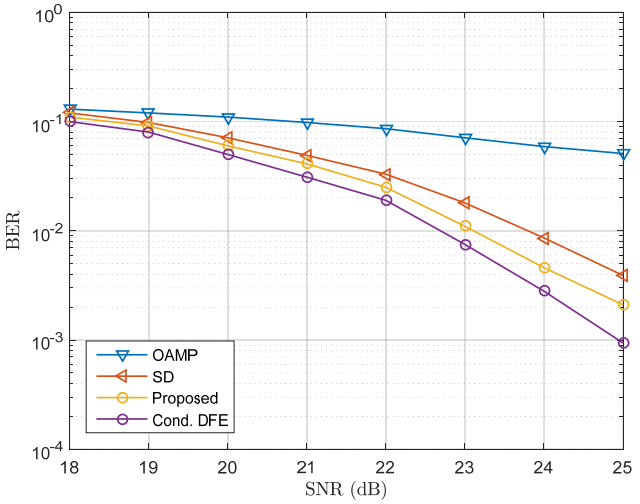

FIGURE 11. The comparison of BER performance for 8 × 8 MIMO with 256QAM and $\alpha = 0$.

each detection algorithm has deteriorated. For example, when the SNR is 10dB, the BER of the proposed algorithm changes from 5.9×10^{-5} to 2.0×10^{-4}, while the OAMP algorithm changes from 1.1×10^{-3} to 2.1×10^{-3}, and the sphere decoding algorithm changes from 4.8×10^{-4} to 3.4×10^{-3}. Comparing these simulation results, it can be concluded that the proposed algorithm has certain robustness to channel estimation errors.

In the final set of simulation experiments, we compare the proposed algorithm with the detection algorithm recently proposed in the literature. In [22], the authors proposed a detection algorithm that combines conditional optimization with DFE, which can approach the performance of maximum likelihood detection. However, when considering a 256×256 MIMO system, the computational complexity of the conditional DFE algorithm is too high to output simulation results within an acceptable length of time. Therefore, we consider a low-dimensional 8 × 8 MIMO with 256QAM the same as [22], and then plot the simulation results in Fig. 11. Observing the curves in the figure, it can be seen that the BER performance of the proposed algorithm is close to that of the quasi-optimal conditional DFE algorithm. For example, to achieve a BER performance of 10^{-2}, the required SNR of the proposed detection algorithm and the conditional DFE algorithm is approximately 23.1dB and 22.7dB, respectively. The difference between the two is 0.4dB. However, the proposed algorithm can handle high-dimensional 256 × 256 MIMO systems, while the conditional DFE algorithm can only handle low-dimensional MIMO systems. It can also be seen that The BER performance of the other two algorithms is also worse than the proposed algorithm and the conditional DFE algorithm under this configuration. Compared with the conditional DFE algorithm, the proposed algorithm has similar BER performance but can handle situations with a large number of antennas, and is more practical for MIMO detection.

VI. CONCLUSION AND FUTURE WORK

This paper studied the MIMO detection problem. Firstly, a new graph model of the MIMO detection problem was established and MIMO detection was transformed into finding the maximum clique in the graph. Then, a clique based MIMO detection algorithm was presented which uses an architecture similar to the FFT. The execution time of the algorithm can be proportional to $\log_2 N_t$. Simulation results showed that the proposed algorithm can achieve good BER performance.

As a step toward a detection algorithm suitable for communication systems with a large number of antennas, this work focused on the basic uncoded MIMO system. In modern MIMO communication systems, space-time-frequency (STF) coding is a very important way to utilize multiple antennas. In future work, we will study how to extend the proposed FFT-like detection algorithm to the STF coded MIMO system with a large number of antennas.

ACKNOWLEDGMENT

The authors thank reviewers for pointing out future research directions to them.

REFERENCES

[1] 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, Sep. 2015.

[2] M. A. Albreem, M. Juntti, and S. Shahabuddin, "Massive MIMO detection techniques: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3109–3132, Aug. 2019.

[3] B. M. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389–399, Mar. 2003.

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

[5] J. Jalden and B. Ottersten, "On the complexity of sphere decoding in digital communications," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1474–1484, Apr. 2005.

[6] 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, Mar. 2006.
[7] T. Fukatani, R. Matsumoto, and T. Uyematsu, "Two methods for decreasing the computational complexity of the MIMO ML decoder," *IEICE Trans. Fundamentals Electron., Commun. Comput. Sci.*, vol. E87-A, no. 10, pp. 2571–2576, Oct. 2004.
[8] R. Y. Chang and W.-H. Chung, "Best-first tree search with probabilistic node ordering for MIMO detection: Generalization and performance-complexity tradeoff," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 780–789, Feb. 2012.
[9] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.
[10] H. Loeliger, "An introduction to factor graphs," *IEEE Signal Process. Mag.*, vol. 21, no. 1, pp. 28–41, Jan. 2004.
[11] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 45, pp. 18914–18919, Nov. 2009.
[12] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: I. Motivation and construction," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Cairo, Egypt, Jan. 2010, pp. 1–5.
[13] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: II. Analysis and validation," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Cairo, Egypt, Jan. 2010, pp. 1–5.
[14] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 764–785, Feb. 2011.
[15] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," *IEEE Trans. Signal Process.*, vol. 65, no. 10, pp. 6664–6684, Oct. 2019.
[16] J. Ma and L. Ping, "Orthogonal AMP," *IEEE Access*, vol. 5, pp. 2020–2033, 2017.
[17] J. Ma, L. Liu, X. Yuan, and L. Ping, "On orthogonal AMP in coded linear vector systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 12, pp. 5658–5672, Dec. 2019.
[18] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.
[19] A. F. Molisch, M. Z. Win, and J. H. Winters, "Space-time-frequency (STF) coding for MIMO-OFDM systems," *IEEE Commun. Lett.*, vol. 6, no. 9, pp. 370–372, Sep. 2002.
[20] W. Su, Z. Safar, and K. J. R. Liu, "Towards maximum achievable diversity in space, time, and frequency: Performance analysis and code design," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1847–1857, Jul. 2005.
[21] D. Darsena, G. Gelli, L. Paura, and F. Verde, "Blind channel shortening for Space-Time-Frequency block coded MIMO-OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 3, pp. 1022–1033, Mar. 2012.
[22] M. K. Izadinasab and O. Damen, "Bridging the gap between MMSE-DFE and optimal detection of MIMO systems," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 220–231, Jan. 2020.

PENG DU received the Ph.D. degree from Southeast University, Nanjing, China, in 2004. He was a Technical Staff Member with Lucent Technologies, Nanjing. Since 2008, he has been with the Nanjing University of Posts and Telecommunications, Nanjing. His research interests include wireless communication system design and digital signal processing.

YUAN ZHANG (Member, IEEE) received the Ph.D. degree from Southeast University, Nanjing, China, in 2004. Since 2005, he has been with the National Mobile Communications Research Laboratory, Southeast University, as a Faculty Member. His research interests include wireless communications, networking, and computing. He was a co-recipient of the Best Paper Awards from the IEEE ICC in 2014 and the IEEE ICCC in 2019.

TEER BA (Member, IEEE) received the bachelor's degree from Southeast University, Nanjing, China, in 2000. He is currently an Associate Researcher with the National Communication Research Laboratory, Southeast University. His research interests include terrestrial and satellite communication system design.