Received March 16, 2020, accepted March 28, 2020, date of publication April 1, 2020, date of current version April 15, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2984817

Low-Complexity Differential Spatial Modulation Schemes for a Lot of Antennas

RUEY-YI WEI, (Senior Member, IEEE)
Department of Communication Engineering, National Central University, Taoyuan, Taiwan
e-mail: rywei@ncu.edu.tw

This work was supported by the Ministry of Science and Technology, Taiwan, under Grant MOST108-2221-E-008-019-MY2.

ABSTRACT Spatial modulation (SM) uses a single antenna for transmission each time to avoid synchronization and interferences among transmitter antennas, and differential SM (DSM) is a kind of SM technique that does not need pilot symbols and channel estimation. The original noncoherent detection for DSM is very complicated, so various low-complexity detectors have been proposed. However, they are still too complicated to be applied to DSM with many transmitter antennas. In this paper, low-complexity DSM schemes for dozens or hundreds of transmitter antennas are designed. Two bit-mapping schemes are proposed: one is group arrangement for maximizing data rates, and the other is full mapping so that DSM can be easily detected at the receiver. A low-complexity detection algorithm that is less complicated than all existing fixed-complexity detectors is proposed as well. In addition, two detection strategies for the proposed bit mapping methods are proposed. Simulation results show that the proposed low-complexity DSM schemes have satisfactory error performance.

INDEX TERMS Differential detection, differential encoding, massive MIMO, spatial modulation.

I. INTRODUCTION

Wireless communication systems equipped with multiple antennas can achieve high spectral efficiency. Spatial modulation (SM) [1]–[4] is a kind of multi-antenna technique which uses a single antenna for transmission each time so it avoids synchronization and interferences among transmitter antennas. By selecting indices of antennas, SM can transmit additional data bits without increasing radio-frequency circuits and power consumption.

Coherent SM needs pilot symbols for channel estimation. When the channel varies rapidly, pilot symbols should be transmitted frequently so the rate loss is significant. For such channels, differential SM (DSM) [5]–[18] with differential detection is more suitable than coherent SM. Different from SM, DSM is modulated block-by-block. Conventionally, DSM is viewed as a special case of differential space-time modulation (DSTM) [12], [19] which tests all data matrices one-by-one at the noncoherent maximum-likelihood (ML) receiver. A differential detector utilizes the previous symbol (block) as a reference symbol (block) to detect the current symbol (block). In DSTM and DSM systems with N_T transmitter antennas, each transmitted block consists of N_T time slots so the reference block contains sufficient information of channel coefficients. More specifically, each antenna is activated exactly once during each transmitted block of DSM, so all channel coefficients of N_T transmitter antennas can be estimated from the previous block. For a fixed transmission rate, because the number of $N_T \times N_T$ data matrices increases exponentially with N_T, the ML detector is very complicated if N_T is not small. To reduce the complexity of the receiver, several detectors were proposed in [7]–[9], but their complexity also grows exponentially with N_T. Consequently, they were simulated for $N_T \leq 8$ only.

Massive multiple-input-multiple-output (MIMO) systems where the base station is equipped with dozens or hundreds of antennas have attracted much attention in recent years. For noncoherent downlink massive MIMO systems, only two schemes [10] and [11] were proposed. Unlike regular DSM uses $N_T \times N_T$ square space-time matrices, in the DSM scheme proposed in [11], the number of time slots per transmitted block is $T < N_T$, resulting in $N_T \times T$ rectangular space-time matrices. Because the reference block does not have enough information of channel coefficients, conventional two-block differential detection [12] cannot be used. In fact, for a specific transmitter antenna, it always has a nonzero probability that this antenna is not activated during finite transmitted blocks. Hence, multiple-block differential detection [13]

The associate editor coordinating the review of this manuscript and approving it for publication was Hayder Al-Hraishawi.

still cannot be used. Consequently, only recursive decision-feedback differential detection [14] which uses all previously received blocks with a forgetting factor can be applied to the transmitter in [11]. Since the signals received a long time ago are still utilized, the scheme in [11] is not suitable for time-varying channels though it is simple enough for large N_T.

To the best of my knowledge, the detection algorithm proposed in [10] is the simplest detector for regular DSM. The detector we proposed in [9] is not simple enough because antenna-index matrices, which are antenna indices of all symbols in a block, are still detected one-by-one. In [10], instead of detecting indices of all received symbols jointly, the authors proposed to detect them individually. All individually-detected values form a temporarily decided antenna-index matrix which is very likely illegitimate, e.g., containing some identical antenna indices (note that each antenna is activated only once). If the temporary matrix is invalid, some legitimate antenna-index matrices modified from the temporary matrix are tested. Among all tested legitimate matrices, the one whose probability is the highest is chosen as the finally-determined antenna-index matrix. However, the number of tested matrices is not fixed and likely very large for $N_T \gg 1$, so only $N_T = 6$ and 16 were simulated in [10]. Detailed discussion of [10] can be found later in Sec. III.

In this paper, the proposed detection algorithm also detects the received symbols individually like [10]. However, the values of antenna indices are determined successively, and a later detected index value cannot be the same as previously decided values. The detection order is determined according to their reliability, and a more reliable symbol is detected earlier than a less reliable symbol. By doing so, the temporarily-decided antenna-index matrix does not have the same antenna indices, so it is very likely legitimate. If it is not, the proposed modification method is also simple. By the proposed simple fixed-complexity detection, simulation results of square DSM with $N_T = 256$ can be obtained. Compared with the noncoherent ML detection, the proposed algorithm is much less complicated, at the price of higher error probability which is only slightly higher for $N_T = 4, 6$ and 8 according to simulation results.

On the other hand, the bit mapping method proposed in [5] and used in [7] and [10] is complicated for $N_T \gg 1$. To my best knowledge, bit mapping designed for large N_T such as $N_T = 16$ has not been proposed yet. In this paper, new mapping schemes for a lot of transmitter antennas are proposed also. There are independently-mapping symbols and mapping groups in the proposed scheme. An algorithm of group arrangement for maximizing data rate is proposed. Besides, a full mapping method to simplify detection is proposed as well.

Coded DSM schemes for increasing transmit diversity have been proposed in [15], [16], [18]. For M-ary PSK (phase-shift keying) with a fixed value of M, coded DSM schemes have less data rate than the original DSM which maximizes transmission rate. Note that this paper only considers the original DSM, so coded DSM is beyond the scope of this paper.

The main results and contributions of this paper are summarized as follows.

1) The differential encoding of DSM is performed by multiplication of $N_T \times N_T$ matrices which is not easy for $N_T \gg 1$. This paper indicates that differential encoding of DSM by matrix multiplication is equivalent to differentially encoding of N_T DPSK (differential phase-shift keying) symbols whose reference symbols are in the previous block. DSM transmits additional data bits by selecting different reference orders.
2) For a lot of transmitter antennas, a low-complexity bit mapping technique and a group arrangement algorithm to maximize data rates are proposed as well.
3) A low-complexity detection algorithm for DSM which is less complicated than all existing fixed-complexity detectors is proposed. Simulation results indicate that error performance of the proposed detection is close to that of the noncoherent ML detection.
4) To simplify the detection complexity, a new full-mapping scheme and two detection strategies are proposed. Using the full-mapping scheme and symbol-wise mapping, the complexity of the receiver is extremely low so error performance can be simulated for downlink massive MIMO systems.

The remainder of the paper is organized as follows. In Sec. II, background knowledge of DSM in [5] and [9] is briefly reviewed. In Sec. III, a new multi-group bit mapping scheme and a new detection algorithm for DSM are proposed. Group arrangement for bit mapping, full-mapping DSM and detection strategies are proposed in Sec. IV. Finally, Sec. V concludes this paper.

Notation: $(.)^T$ and $\|.\|$ denote the transpose and the Frobenius norm of a matrix, respectively. diag$\{.\}$ represents the operation from a row vector to a diagonal matrix. $\lfloor\rfloor$ denotes the floor function. $\mathcal{CN}(0, \sigma^2)$ denotes the zero-mean, σ^2-variance, complex Gaussian distribution.

II. REVIEW OF DSM

Consider a communication system with N_T transmitter antennas and N_R receiver antennas. The channels between antenna pairs are Rayleigh-fading and independent of each other. If the symbol s is transmitted via the mth antenna, the constellation vector is represented by a column-vector $[0, \cdots, 0, s, 0, \cdots, 0]^T$ where the only nonzero entry s is the mth element. Each block of DSM contains N_T time slots, and all N_T column-vectors form an $N_T \times N_T$ matrix of transmitted signals $\mathbf{S}(t)$ for the tth block which satisfies two restrictions: (1) At each time slot, only one antenna is activated, i.e., there is only one nonzero entry in each column of $\mathbf{S}(t)$. (2) In each block, each transmitter antenna is activated exactly once,

i.e., there is only one nonzero entry in each row of $\mathbf{S}(t)$. For the tth block, the $N_R \times N_T$ matrix of received signals is

$$\mathbf{Y}(t) = \mathbf{H}(t)\mathbf{S}(t) + \mathbf{N}(t) \tag{1}$$

where $\mathbf{H}(t)$ is the $N_R \times N_T$ matrix of channel coefficients whose entries are $\mathcal{CN}(0,1)$, and $\mathbf{N}(t)$ is the $N_R \times N_T$ matrix of AWGN with $\mathcal{CN}(0, N_0)$ entries.

The number of permutating the antennas index in each block is $N_T!$, and Q permutations are used where Q is a power of two. If Q achieves the maximum value, i.e., $Q = 2^{\lfloor \log_2 N_T! \rfloor}$, the DSM is called **full rate**. The signal constellation for data symbols is M-ary PSK where $M = 2^b$ and b is an integer. There are totally $\log_2 Q + N_T b$ data bits at each block, so the spectral efficiency is $R = \frac{\log_2 Q}{N_T} + b$ bits/s/Hz. For the tth block, $\log_2 Q$ bits determine an antenna-index matrix $\mathbf{A}(t) \in \mathcal{A} = \{\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_Q\}$ and $N_T b$ bits decide N_T data symbols $\mathbf{x}(t) = [x_1(t), x_2(t), \cdots, x_{N_T}(t)]$. All antenna-index matrices $\mathbf{A}_q$ where $q \in \{1, 2, \cdots, Q\}$ have entries 0 or 1 and there is only one nonzero entry in each row and column. Note that complex-valued antenna-index matrices proposed in [6], [9] are not considered in the paper because they should be detected one by one and thus cannot be detected efficiently. An $N_T \times N_T$ data matrix $\mathbf{X}(t)$ is calculated by

$$\mathbf{X}(t) = \text{diag}\{\mathbf{x}(t)\}\mathbf{A}(t). \tag{2}$$

Because $\mathbf{X}(t)$ is a unitary matrix, DSM can be viewed as a special case of DSTM. Therefore, differential encoding and differential detection of DSTM is applied to DSM. At the transmitter, $\mathbf{S}(t)$ is determined by

$$\mathbf{S}(t) = \mathbf{S}(t-1)\mathbf{X}(t). \tag{3}$$

The initial reference matrix $\mathbf{X}(0)$ is the identity matrix, so the transmitted matrix $\mathbf{S}(t)$ is unitary and satisfies the two restrictions. At the receiver, the noncoherent ML detection is

$$\hat{\mathbf{X}}(t) = \arg\min_{\tilde{\mathbf{X}} \in \mathcal{X}} \|\mathbf{Y}(t) - \mathbf{Y}(t-1)\tilde{\mathbf{X}}\|^2 \tag{4}$$

where $\mathcal{X}$ denotes the set of all possible values of $\mathbf{X}(t)$. To obtain $\hat{\mathbf{X}}(t)$, the receiver has to try all $Q \times M^{N_T}$ elements in $\mathcal{X}$. If N_T is not small, differential encoding and detection is not easy.

In [9], given an antenna-index matrix, we indicate that differential detection of DSM is equivalent to differentially detecting data symbols separately. For $\mathbf{A}(t)$ and $\mathbf{A}_q$ where $q \in \{1, 2, \cdots, Q\}$, define $\mathbf{p} = (p^{(1)}, p^{(2)}, \cdots, p^{(N_T)})$ and $\mathbf{p}_q = (p_q^{(1)}, p_q^{(2)}, \cdots, p_q^{(N_T)})$ where $p^{(k)}$ and $p_q^{(k)} \in \{1, 2, \cdots, N_T\}$ represent the position of 1 in the kth column of $\mathbf{A}(t)$ and $\mathbf{A}_q$, respectively, for $k \in \{1, 2, \cdots, N_T\}$. Let $s_k(t)$ denote the transmitted symbol in the kth time slot of the tth block $\mathbf{S}(t)$. According to the differential encoding in (3), in the kth time slot of the tth block, the activated antenna is the antenna used in the $p^{(k)}$th time slot of the $t-1$th block, the reference symbol is $s_{p^{(k)}}(t-1)$, and the transmitted data symbol is $x_{p^{(k)}}(t)$. That is, $s_k(t) = s_{p^{(k)}}(t-1)x_{p^{(k)}}(t)$. Therefore, the receiver can use $y_{ip^{(k)}}(t-1)$ and $y_{ik}(t)$ only to detect $x_{p^{(k)}}(t)$ where $i \in \{1, 2, \cdots, N_R\}$ and $y_{ij}(t)$ denotes the entry in the ith row and the jth column of $\mathbf{Y}(t)$.

Example 1: Assume $N_T = 4$, $N_R = 1$, and

$$\mathbf{A}(t) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

whose $\mathbf{p} = (3142)$, so

$$\mathbf{X}(t) = \begin{pmatrix} 0 & x_1(t) & 0 & 0 \\ 0 & 0 & 0 & x_2(t) \\ x_3(t) & 0 & 0 & 0 \\ 0 & 0 & x_4(t) & 0 \end{pmatrix}.$$

If $\mathbf{S}(t-1)$ is

$$\begin{pmatrix} 0 & s_2(t-1) & 0 & 0 \\ 0 & 0 & 0 & s_4(t-1) \\ 0 & 0 & s_3(t-1) & 0 \\ s_1(t-1) & 0 & 0 & 0 \end{pmatrix},$$

then in

$$\mathbf{S}(t) = \begin{pmatrix} 0 & 0 & 0 & s_4(t) \\ 0 & 0 & s_3(t) & 0 \\ s_1(t) & 0 & 0 & 0 \\ 0 & s_2(t) & 0 & 0 \end{pmatrix},$$

we have $s_1(t) = s_3(t-1)x_3(t)$, $s_2(t) = s_1(t-1)x_1(t)$, $s_3(t) = s_4(t-1)x_4(t)$ and $s_4(t) = s_2(t-1)x_2(t)$. For simplicity, $y_j(t)$ is used instead of $y_{1j}(t)$ $\forall j \in \{1, 2, \cdots, N_T\}$. For this $\mathbf{A}(t)$, $\mathbf{Y}(t) - \mathbf{Y}(t-1)\tilde{\mathbf{X}}$ in (4) with $\tilde{\mathbf{X}} = \mathbf{X}(t)$ is $[y_1(t) - y_3(t-1)x_3(t), y_2(t) - y_1(t-1)x_1(t), y_3(t) - y_4(t-1)x_4(t), y_4(t) - y_2(t-1)x_2(t)]$, and the minimization in (4) can be done separately, e.g., $\hat{x}_3(t) = \arg\min_{\tilde{x}} |y_1(t) - y_3(t-1)\tilde{x}|^2$. Therefore, $[y_3(t-1), y_1(t)]$, $[y_1(t-1), y_2(t)]$, $[y_4(t-1), y_3(t)]$ and $[y_2(t-1), y_4(t)]$ are utilized for detecting $x_3(t)$, $x_1(t)$, $x_4(t)$ and $x_2(t)$, respectively.

The low-complexity ML detector proposed in [9] is described as follows. At the receiver, $\forall q \in \{1, 2, \cdots, Q\}$, the determined data symbols of $\mathbf{A}_q$, $\hat{\mathbf{x}}_q(t) = [\hat{x}_1^{(q)}(t), \hat{x}_2^{(q)}(t), \cdots, \hat{x}_{N_T}^{(q)}(t)]$, are obtained by

$$\hat{x}_{p_q^{(k)}}^{(q)}(t) = \arg\min_{\tilde{x}} \sum_{i=1}^{N_R} |y_{ik}(t) - y_{ip_q^{(k)}}(t-1)\tilde{x}|^2 \tag{5}$$

and the metric of $\mathbf{A}_q$ is

$$m_q(t) = \sum_{k=1}^{N_T} \sum_{i=1}^{N_R} |y_{ik}(t) - y_{ip_q^{(k)}}(t-1)\hat{x}_{p_q^{(k)}}^{(q)}(t)|^2. \tag{6}$$

The detected value of $\mathbf{A}(t)$ is $\mathbf{A}_{\hat{q}}$ satisfying

$$\mathbf{A}_{\hat{q}} = \arg\min_{\mathbf{A}_q \in \mathcal{A}} m_q(t) \tag{7}$$

and the detected value of $\mathbf{x}(t)$ is $\hat{\mathbf{x}}_{\hat{q}}(t)$. The minimization of (5) and (7) needs to try $Q \times (MN_T + 1)$ times, so the complexity reduction compared with the one-by-one ML detector in (4) is $1 - \frac{MN_T+1}{M^{N_T}}$.

Two methods for the mapping from $\log_2 Q$ data bits to $\mathbf{p}$ were proposed in [5]. The first method is looking-up a table which is too huge if N_T is not small, so the second method is used in [7] and [10] which consists of two steps:

Step 1 Map $\log_2 Q$ data bits to $\mathbf{q} = (q_1, q_2, \cdots, q_{N_T-1})$ where $q_k \in \{0, 1, 2, \cdots, N_T - k\}$ $\forall k = 1, 2, \cdots, N_T - 1$.

Step 2 Convert $\mathbf{q}$ to $\mathbf{p}$. Define an ordered list $\Theta = \{1, 2, \cdots, N_T\}$. The value of $p^{(1)}$ is the value of the $q_1 + 1$th element in Θ and then this element is removed from Θ. After that, the value of $p^{(2)}$ is the value of the $q_2 + 1$th element in Θ and then this element is removed from Θ, and so on. Finally, $p^{(N_T)}$ is the only element in Θ.

In Step 1 in [5], an integer m ($0 \le m < Q$) is formed by $\log_2 Q$ data bits first. Then $\mathbf{q}$ is determined based on the equation

$$m = q_1(N_T - 1)! + q_2(N_T - 2)! + \cdots + q_{N_T-1}1!. \quad (8)$$

The algorithm first finds the value of q_1 which is the maximum value satisfying $q_1(N_T - 1)! \le m$, and then finds the value of q_2 which is the maximum value satisfying $q_2(N_T - 2)! \le m - q_1(N_T - 1)!$, and so on. Because $p^{(N_T)}$ is the only element in Θ in Step 2, there is no data bit mapped to q_{N_T} in Step 1. In this paper, this two-step mapping is used, but new mapping methods instead of (8) will be proposed in the following sections.

III. THE PROPOSED FULL-RATE DSM

A. BIT MAPPING

The multiplication of $N_T \times N_T$ matrices in (3) is not easy if $N_T \gg 1$. In fact, the transmitter can be realized without matrix multiplication. For each block, $\log_2 Q$ data bits determine $\mathbf{p} = (p^{(1)}, p^{(2)}, \cdots, p^{(N_T)}) \in \{\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_Q\}$. After that, differential encoding of DPSK whose reference symbols are the symbols of the $t - 1$th block is done N_T times independently. The order of the reference symbols is specified by $\mathbf{p}$. At time slot $k = 1, 2, \cdots, N_T$ of the tth block, using the $p^{(k)}$th symbol of the $t - 1$th block as the reference symbol, the data symbol is differentially encoded. Unlike coherent SM which selects orders of antenna activation, DSM selects orders of reference symbols. Note that by reference order $\mathbf{p}$ only, one cannot know which transmitter antenna is used. For instance, consider Example 1 whose $\mathbf{p}$ is (3142). For the first time slot of the tth block, the used transmitter antenna is the antenna used by the third symbol of the $t - 1$th block, but we do not know whether it is the first or second antenna.

The complexity of the ML detection proposed in [9] is proportional to the number of reference order Q. If N_T is large, then Q is extremely huge and thus the decoding is too complicated to be realized. The high complexity is due to the joint detection of $p^{(1)}, p^{(2)}, \cdots, p^{(N_T)}$ in (7). To further simplify the detection complexity, we should first separately detect $p^{(1)}, p^{(2)}, \cdots, p^{(N_T)}$ and then consider the relation of $p^{(1)}, p^{(2)}, \cdots, p^{(N_T)}$, like [10]. There are two limits on the relation between $p^{(1)}, p^{(2)}, \cdots, p^{(N_T)}$. The first one is essential to DSM.

Limit 1: $p^{(i)} \ne p^{(j)} \forall i \ne j$.

For $N_T > 2$, Q is always less than $N_T!$ because $N_T!$ is not a power of two. Hence, some patterns of $(p^{(1)}, p^{(2)}, \cdots, p^{(N_T)})$ are not used by the mapping from data bits to $\mathbf{p}$.

Limit 2: For $N_T > 2$, there are $N_T! - Q$ invalid permutations for $\mathbf{p}$.

For example, consider $N_T = 4$ for which there are $4! = 24$ permutations for $\mathbf{p}$. Because only $\lfloor \log_2 24 \rfloor = 4$ data bits are used for choosing $\mathbf{p}$, there are $24 - 16 = 8$ permutations are not used and thus are invalid permutations for $\mathbf{p}$.

In the bit mapping scheme proposed in [5] and introduced in Sec. II, the integer m is formed by $\log_2 Q$ data bits, so if $N_T \gg 1$, the value of m is too large and the mapping to $\mathbf{q}$ by (8) is very complicated. In such a case, the mapping should be divided into several mapping groups and some independent symbols. For clarity of presentation, $q'_{N_T+1-k} = q_k$ is used instead of q_k which has $N_T + 1 - k$ possible values. For instance, $\{q'_8, q'_7, q'_6\}$ is used instead of $\{q_1, q_2, q_3\}$ for $N_T = 8$. For q'_{2^l} corresponding to $p_{N_T+1-2^l}$ where l is a positive integer, all 2^l values can be mapped by l data bits. Hence, the bit mapping for q'_{2^l} can be done independently. A group is composed of multiple elements, say $\{q'_{\ell_1}, q'_{\ell_2}, \cdots, q'_{\ell_n}\}$ where $\ell_1 > \ell_2 > \cdots > \ell_n$ are not powers of two. To maximize the data rate, $\ell_1, \ell_2, \cdots, \ell_n$ are not necessarily consecutive integers in general. However, the mapping method in [5] cannot be used directly if $\ell_1, \ell_2, \cdots, \ell_n$ are not consecutive integers.

A new mapping method for a group is proposed as follows. The proposed mapping consists of the same two steps in Section II, but the mapping from data bits to $\mathbf{q}$ is different from (8). An integer m_{ℓ_1}, formed by $\lfloor \log_2(\ell_1 \times \ell_2 \times \cdots \times \ell_n) \rfloor$ data bits, is mapped to $\{q'_{\ell_1}, q'_{\ell_2}, \cdots, q'_{\ell_n}\}$ by using $n - 1$ division. $\forall k \in \{1, 2, \cdots, n - 1\}$, m_{ℓ_k} divided by ℓ_k gives out a quotient of $m_{\ell_{k+1}}$ with a remainder of $q'_{\ell_k} \forall k \in \{1, 2, \cdots, n - 1\}$. Finally $q'_{\ell_n} = m_{\ell_n}$. The equation of m_{ℓ_1} and $\{q'_{\ell_1}, q'_{\ell_2}, \cdots, q'_{\ell_n}\}$ is

$$m_{\ell_1} = q'_{\ell_1} + q'_{\ell_2} \times \ell_1 + q'_{\ell_3} \times \ell_1 \times \ell_2 + q'_{\ell_4} \times \ell_1 \times \ell_2 \times \ell_3 + \cdots + q'_{\ell_n} \times \ell_1 \times \ell_2 \times \cdots \times \ell_{n-1}. \quad (9)$$

B. A NEW LOW-COMPLEXITY DETECTION ALGORITHM

In [10], the values of $p^{(k)}$ are detected individually, so the temporarily detected value of $\mathbf{p}$, denoted by $\hat{\mathbf{p}}$, is likely to be invalid. To satisfy Limit 1, the algorithm in [10] replaces the repeated values among the elements of $\hat{\mathbf{p}}$ by unused values, and all possible replacements are tried. However, the number of identical values of $p^{(k)}$ may differ, so the complexity is not fixed and perhaps very high for $N_T \gg 1$. For Limit 2, if the detected value of $p^{(k)}$ is invalid, the value of $p^{(k)}$ is exchanged to a value of $p^{(k')}$ where $k' > k$. Nevertheless, it is possible that all possible exchanges are still invalid, but [10] does not discuss such a situation.

In this paper, by deciding the elements of $\hat{\mathbf{p}}$ one-by-one according to their reliability, the proposed detection algorithm always obeys Limit 1. A more reliable symbol is detected before a less reliable symbol. Once $p^{(k)}$ is detected to be $\hat{p}^{(k)}$, $\hat{p}^{(k)}$ is removed from the candidate list, denoted by Φ, for the remaining undetected $p^{(k')}$. Hence, the elements of $\hat{\mathbf{p}}$ are never identical.

Set the initial value of the candidate list $\Phi = \{1, 2, \cdots, N_T\}$. The proposed detection algorithm consists of the following seven steps of which explanations are given later.

Step 1 For $k, l = 1, 2, \cdots, N_T$, compute

$$\hat{x}_k^l = \arg\min_{\tilde{x}} \sum_{i=1}^{N_R} |y_{ik}(t) - y_{il}(t-1)\tilde{x}|^2 \quad (10)$$

and

$$\eta_k^l = \sum_{i=1}^{N_R} |y_{ik}(t) - y_{il}(t-1)\hat{x}_k^l|^2. \quad (11)$$

Step 2 For $k = 1, 2, \cdots, N_T$, sort all $\eta_k^1, \eta_k^2, \cdots, \eta_k^{N_T}$ in ascending order. The sorted metrics are denoted by $\eta_k^{l_k(1)} \le \eta_k^{l_k(2)} \le \cdots \le \eta_k^{l_k(N_T)}$ where $l_k(1), l_k(2), \cdots, l_k(N_T) \in \{1, 2, \cdots, N_T\}$. Set $i = 1$.

Step 3 Find the value of $k \in \Phi$, denoted by $\hat{k}$, which has the maximum value of $\eta_k^{l_k(2)} - \eta_k^{l_k(1)}$. That is

$$\hat{k} = \arg\max_{k \in \Phi} \eta_k^{l_k(2)} - \eta_k^{l_k(1)}. \quad (12)$$

The detected value of $p^{(\hat{k})}$ is $\hat{p}^{(\hat{k})} = l_{\hat{k}}(1)$. Remove $\hat{k}$ from Φ.

Step 4 For all elements in Φ, say k, find the value of n satisfying $l_k(n) = \hat{p}^{(\hat{k})}$. Delete $\eta_k^{l_k(n)}$ and move the following sorted metric $\eta_k^{l_k(n+1)}, \cdots, \eta_k^{l_k(N_T+1-i)}$ up. In other words, $\forall j = n, n+1, \cdots, N_T - i$, the updated value of $l_k(j)$ is the original value of $l_k(j+1)$.

Step 5 If $i < N_T$, add 1 to i and repeat Step 3 and Step 4; otherwise, go to Step 6.

Step 6 Check whether $\hat{\mathbf{p}} = (\hat{p}^{(1)}, \hat{p}^{(2)}, \cdots, \hat{p}^{(N_T)})$ is a valid reference order. If it is invalid, modify $\hat{\mathbf{p}}$ so that it becomes legitimate.

Step 7 Convert $\hat{\mathbf{p}}$ to data bits. For $k = 1, 2, \cdots, N_T$, the detected value of $x_k(t)$ is $\hat{x}_k(t) = \hat{x}_k^{\hat{p}^{(k)}}$, and recover b data bits accordingly.

The flow chart of the proposed algorithm is illustrated in Fig. 1. In Step 1, based on the assumption that $s_k(t)$ is differentially encoded on $s_l(t-1)$, the detected data symbol is $\hat{x}_k^l$ and the metric is η_k^l. After the sorting of $\eta_k^1, \eta_k^2, \cdots, \eta_k^{N_T}$ in Step 2, the value of k in Φ which maximizes the metric difference in (12) is chosen in Step 3. The reason why the comparison is based on the metric difference $\eta_k^{l_k(2)} - \eta_k^{l_k(1)}$ instead of the metric $\eta_k^{l_k(1)}$ is explained in the following lemma.

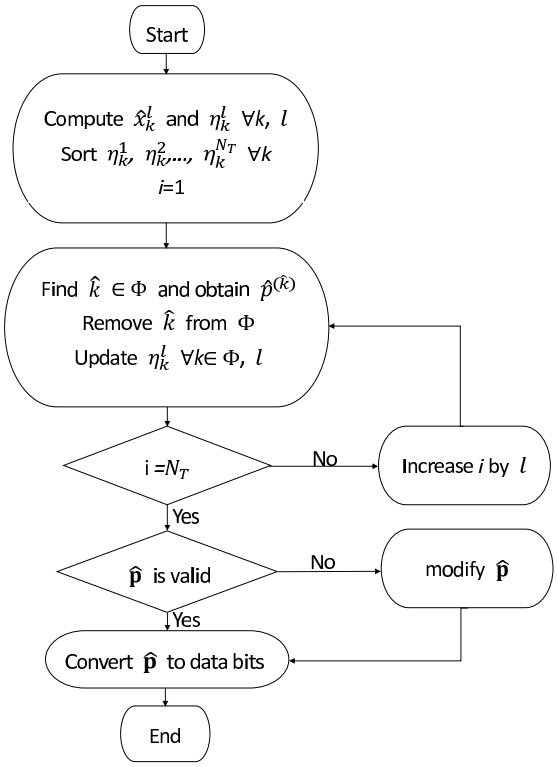


FIGURE 1. The flow chart of the proposed detector.

Lemma 1: Consider $\tilde{\mathbf{p}} = (l_1(1), l_2(1), \cdots, l_{N_T}(1))$ with $l_1(1) = l_2(1)$. There are two possible modifications of $\tilde{\mathbf{p}}$ to obey Limit 1: replacing $l_2(1)$ by $l_2(2)$ which results in $\tilde{\mathbf{p}}_1 = (l_1(1), l_2(2), l_3(1), \cdots, l_{N_T}(1))$ or replacing $l_1(1)$ by $l_1(2)$ which results in $\tilde{\mathbf{p}}_2 = (l_1(2), l_2(1), l_3(1), \cdots, l_{N_T}(1))$. If $\eta_1^{l_1(2)} - \eta_1^{l_1(1)} > \eta_2^{l_2(2)} - \eta_2^{l_2(1)}$, then $\tilde{\mathbf{p}}_1$ is more likely than $\tilde{\mathbf{p}}_2$.

Proof: The metric of the invalid $\tilde{\mathbf{p}}$ is $\eta = \sum_{k=1}^{N_T} \eta_k^{l_k(1)}$, and the metrics of $\tilde{\mathbf{p}}_1$ and $\tilde{\mathbf{p}}_2$ are $\eta_1 = \eta - \eta_2^{l_2(1)} + \eta_2^{l_2(2)}$ and $\eta_1 = \eta - \eta_1^{l_1(1)} + \eta_1^{l_1(2)}$, respectively. If $\eta_1^{l_1(2)} - \eta_1^{l_1(1)} > \eta_2^{l_2(2)} - \eta_2^{l_2(1)}$, then we have $\eta_1 < \eta_2$ which means that $\tilde{\mathbf{p}}_1$ is more likely than $\tilde{\mathbf{p}}_2$. □

The lemma indicates that we should choose $p^{(1)} = l_1(1)$ instead of $p^{(2)} = l_2(1)$. Therefore, the priority of $l_1(1), l_2(1), \cdots, l_{N_T}(1)$ should depend on the metric difference $\eta_k^{l_k(2)} - \eta_k^{l_k(1)}$ where $k \in \{1, 2, \cdots, N_T\}$.

When $p^{(\hat{k})}$ is decided, all $p^{(k)}$ where $k \neq \hat{k}$ cannot be the same value of $p^{(\hat{k})}$, so we delete the corresponding metric in Step 4. The element $\hat{k}$ is removed from Φ because Φ is the set of the indices of the remaining undetected symbols. After Step 3 and Step 4 are done N_T times iteratively, a temporarily decided value of $\mathbf{p}$ satisfying Limit 1, $\hat{\mathbf{p}}$, is obtained. In Step 6, to check whether $\hat{\mathbf{p}}$ satisfies Limit 2, usually $\hat{\mathbf{p}}$ is converted to $\hat{\mathbf{q}} = (\hat{q}_1, \hat{q}_2, \cdots, \hat{q}_{N_T})$ for $N_T > 4$. If $\hat{\mathbf{p}}$ is illegitimate, the modification of $\hat{\mathbf{p}}$ is necessary. A modification rule for one mapping group is proposed in Sec. III.D. In the final step, the conversion from $\hat{\mathbf{p}}$ to data bits is the inverse function of bit mapping.

C. COMPARISON BETWEEN THE PROPOSED DETECTION AND EXISTING DETECTION TECHNIQUES

First consider the complexity issue. Definitions of complexity in [7], [8] and [10] are all different. In [7], the search complexity contains a term $N_T 2^{N_T-1}$, so it grows exponentially with N_T. In [8], an example of $N_T = 4$ and $N_R = 3$ shows the complexity reduction compared with the one-by-one ML detector is only between 44% and 70% for different SNRs, which is worse than the detector of [9] reviewed in Sec. II. Hence, the complexities of both detectors proposed in [7] and [8] are not low enough for large N_T. In [10], the complexity is linear with N_T^2, but it varies as discussed at the beginning of Sec. III.B. In the proposed algorithm, the main complexity is the computation of the Euclidean distance in (10) in Step 1, which is also linear with N_T^2, and the sorting in Step 2. Notice that $(\hat{x}_1^{\ell_1(1)}, \hat{x}_2^{\ell_2(1)}, \cdots, \hat{x}_{N_T}^{\ell_{N_T}(1)})$ of the proposed detector is the temporarily-decided antenna-order in [10], and the number of computation in (10) is also needed in [10]. However, the Euclidean distance in (10) is transformed into a different form (7) in [10], but the so-called HL-ML detector is valid for QAM only. Unlike [10], the complexity of the proposed detector is fixed except Step 6. In fact, Step 6 is simple and will be omitted for the full-mapping DSM in the next section.

Then consider the issue of error performance. In [7], [8] and [10], there is no mathematical analysis of error probabilities, and only computer simulation results are provided. The gap between two curves of a low-complexity detector and the ML detector is an indicator of the error performance of the low-complexity detector. In this paper, error performance of the proposed schemes are also justified by computer simulations.

D. A MODIFICATION RULE AND EXAMPLES FOR DSM WITH A MAPPING GROUP

Consider DSM which consists of a mapping group $\{q'_{\ell_1}, q'_{\ell_2}, \cdots, q'_3\}$ where $q'_3 = q_{N_T-2}$ and several independently-mapping symbols q'_{2^l}. Note that such DSM is always full rate. Let $q'_{3,\max}$ denote the maximum value of q'_3, i.e., the value of $q'_3 = q'_{\ell_n}$ in (9) where m_{ℓ_1} is formed by $\lfloor \log_2(\ell_1 \times \ell_2 \times \cdots \times 3) \rfloor$ data bits $11\cdots1$. Apparently, possible values of q'_3 are either $\{0, 1, 2\}$ or $\{0, 1\}$, so $q'_{3,\max}$ is either 2 or 1.

For the DSM described above, the modification in Step 6 is listed as follows. If $\hat{\mathbf{p}}$ is invalid, switch the values of $\hat{p}^{(N_T-2)}$ and $\hat{p}^{(k')}$ where k' denotes the latest $\hat{k}$ except $N_T - 2$ in Step 3. If $q'_{3,\max} = 1$, then check whether the updated $\hat{\mathbf{p}}$ is a valid reference order. If it is still invalid, undo the switch and switch the values of $\hat{p}^{(N_T-2)}$ and $\hat{p}^{(k'')}$ where k'' denote the second latest $\hat{k}$ except $N_T - 2$ in Step 3.

For simplicity, if $\hat{\mathbf{p}}$ is an invalid reference order, the modification changes the value of $\hat{p}^{(N_T-2)}$. Changing other values of $\hat{p}^{(k)}$ where $k \neq N_T - 2$ may make $\hat{\mathbf{p}}$ valid, but it is more complicated. Hence, in Step 6, $\hat{p}^{(N_T-2)}$ and $\hat{p}^{(k')}$ which is the most unlikely among $\hat{p}^{(1)}, \hat{p}^{(2)}, \cdots, \hat{p}^{(N_T-3)}, \hat{p}^{(N_T-1)}, \hat{p}^{(N_T)}$ is exchanged. If $q'_{3,\max} = 2$, there is only one possible invalid value of $\hat{p}^{(N_T-2)}$ corresponding to $\hat{q}_{N_T-2} = 2$; hence, a new value of $\hat{p}^{(N_T-2)}$ always makes $\hat{\mathbf{p}}$ valid. However, if $q'_{3,\max} = 1$, $\hat{q}_{N_T-2} = 2$ is always invalid, and $\hat{q}_{N_T-2} = 1$ is possibly invalid (it depends on the value of $\hat{\mathbf{p}}$). In such a case, there are two possible invalid values of $\hat{p}^{(N_T-2)}$ corresponding to $\hat{q}_{N_T-2} = 1$ and 2, so the updated $p^{(\hat{k})}$ should be checked again; once $p^{(\hat{k})}$ is still invalid, $\hat{p}^{(N_T-2)} = \hat{p}^{(k'')}$ will always make $\hat{\mathbf{p}}$ valid.

Examples of DSM for $N_T = 4, 6, 8$ are given below to demonstrate Step 6 of the proposed detection algorithm, and to show gaps between BER curves of the proposed algorithm and BER curves of the ML detection. For $N_T > 8$, DSM with a mapping group is complicated, and the ML detection cannot be realized. Because they are full rate, their spectral efficiency is $R = \frac{\log_2 Q}{N_T} + b$ bits/s/Hz.

Example 2 ($N_T = 4$ and $Q = 16$): Because there are two independently-mapping symbols $q'_4 = q_1$ and $q'_2 = q_3$ and one remaining symbol $q'_3 = q_2$ in the mapping group, the mapping equation in (9) is not used. Therefore, we have $q'_4 \in \{0, 1, 2, 3\}$, $q'_3 \in \{0, 1\}$, $q'_2 \in \{0, 1\}$ and $q'_1 = q_4 = 1$. In other words, at the transmitter, two data bits choose four possible values of $p^{(1)} \in \{1, 2, 3, 4\}$, one data bit selects two possible values of $p^{(2)} \in \{1, 2, 3\} \neq p^{(1)}$, and the last data bit decides two possible values of $p^{(3)} \in \{1, 2, 3, 4\} \neq p^{(1)}$ and $\neq p^{(2)}$. Consequently, we have $q'_{3,\max} = 1$, so $\hat{q}_2 = \hat{q}'_3 = 2$ is invalid. Because there is only one symbol in the mapping group, $\hat{q}_2 = 1$ is always legitimate. Invalid reference orders for Limit 2 either have $\hat{p}^{(2)} = 4$ or $(\hat{p}^{(1)}, \hat{p}^{(2)}) = (4, 3)$. Notice that both invalid cases of $\hat{\mathbf{p}}$ correspond to $\hat{q}_2 = 2$.

Example 3 ($N_T = 6$ and $Q = 512$): There are two independently-mapping symbols, q'_4 and q'_2, and a mapping group $\{q'_6, q'_5, q'_3\}$. Among total $6 \times 5 \times 3 = 90$ patterns of $\{q'_3, q'_5, q'_6\}$, valid 64 patterns are $(0, 0, 0), (0, 0, 1), \cdots, (0, 4, 5), (1, 0, 0), (1, 0, 1), \cdots, (1, 4, 5), (2, 0, 0), (2, 0, 1), (2, 0, 2), (2, 0, 3)$; while invalid 26 patterns are $(2, 0, 4), (2, 0, 5), \cdots, (2, 4, 5)$. In Step 6, $\hat{\mathbf{p}}$ is a valid permutation order if and only if $\hat{q}'_6 + \hat{q}'_5 \times 6 + \hat{q}'_3 \times 30 < 64$. Apparently, $q'_{3,\max} = 2$ and there is only a possible invalid value of $\hat{q}_4 = \hat{q}'_3$ which is 2. If $\hat{\mathbf{p}}$ is illegitimate, the proposed switch in Step 6 makes $\hat{q}'_3 < 2$ and thus the updated $\hat{\mathbf{p}}$ is valid.

Example 4 ($N_T = 8$ and $Q = 2^{15}$): There are three independently-mapping symbols q'_8, q'_4 and q'_2, and a mapping group $\{q'_7, q'_6, q'_5, q'_3\}$. Among total $7 \times 6 \times 5 \times 3 = 630$ patterns of (q'_3, q'_5, q'_6, q'_7), valid 512 patterns are $(0, 0, 0, 0), (0, 0, 0, 1), \cdots, (0, 4, 5, 6), (1, 0, 0, 0), (1, 0, 0, 1), \cdots, (1, 4, 5, 6), (2, 0, 0, 0), (2, 0, 0, 1), \cdots, (2, 2, 1, 0)$; while invalid 118 patterns are $(2, 2, 1, 1), (2, 2, 1, 2), \cdots, (2, 4, 5, 6)$. In Step 6, $\hat{\mathbf{p}}$ is a valid permutation order if and only if $\hat{q}'_7 + \hat{q}'_6 \times 7 + \hat{q}'_5 \times 42 + \hat{q}'_3 \times 210 < 512$. Similar to Example 3, this example has $q'_{3,\max} = 2$ and a possible invalid value of $\hat{q}_4 = \hat{q}'_3$ is 2. If $\hat{\mathbf{p}}$ is invalid, the proposed switch in Step 6 let $\hat{q}'_3$ has a new value, so the updated $\hat{\mathbf{p}}$ is legitimate.

Throughout the paper, quasi-static Rayleigh fading is used in computer simulations. Simulation results of QPSK with

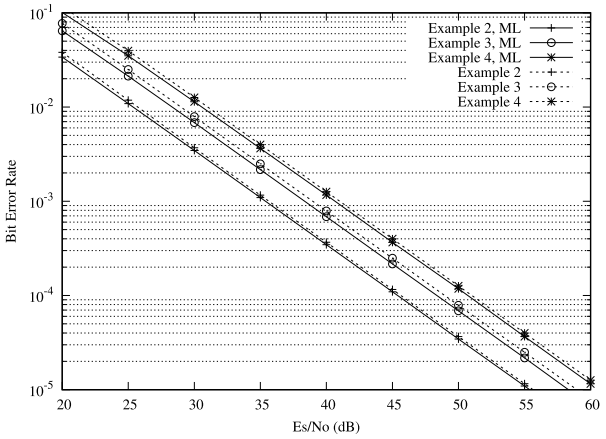


FIGURE 2. Simulation results of Examples 2-4 with $M = 4$ and $N_R = 1$.

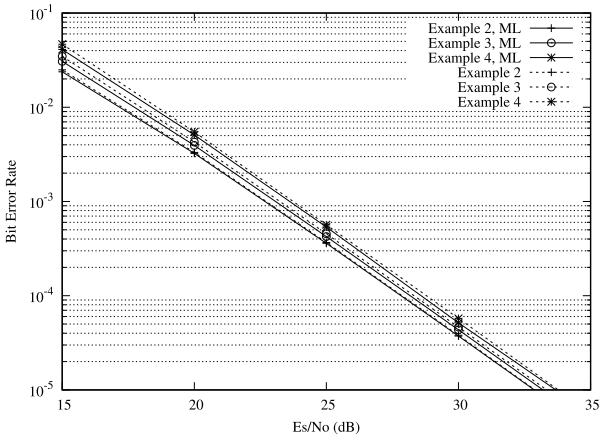


FIGURE 3. Simulation results of Examples 2-4 with $M = 8$ and $N_R = 2$.

$N_R = 1$ and 8PSK with $N_R = 2$ are shown in Fig. 2 and Fig. 3, respectively, where solid curves represent the proposed detection algorithm and dashed curves denote the ML detection. It can be observed that the gap between the ML detection and the proposed detection is very small. Notice that for $N_R = 1$, DPSK with $M = 8$ performs better than DSM with $M = 4$ and $N_T = 3$ or 4 [5]. However, as N_T and M increase, the data rate of DSM increases, so the advantage of DSM over DPSK becomes more obvious. For example, in Example 4 whose $N_T = 8$, the DSM using 8PSK has data rate 4.875 bit/s/Hz, but DPSK with similar data rate is 32-DPSK whose BER is very high.

IV. PROPOSED GROUP ARRANGEMENT, FULL-MAPPING DSM AND DETECTION STRATEGIES

A. GROUP ARRANGEMENT

Because q'_k has k possible values, its information is $\log_2 k$ bits. Consequently, total information of the antenna indices in a block is $\sum_{k=1}^{N_T} \log_2 k = \log_2 N_T!$ bits. It means that for full-rate DSM, the number of data bits deciding the reference order $\mathbf{p}$ is $\lfloor \log_2 N_T! \rfloor$ and the unused information is $r = \log_2 N_T! - \lfloor \log_2 N_T! \rfloor$ bit. Similarly, for a group $\{q'_{\ell_1}, q'_{\ell_2}, \cdots, q'_{\ell_n}\}$, the unused information is $\sum_{i=1}^{n} \log_2 \ell_i - \lfloor \sum_{i=1}^{n} \log_2 \ell_i \rfloor$ bit.

For $N_T \gg 1$, a mapping group is impractical because m_{ℓ_1} in (9) is too large. For such a case, in addition to independent-mapping symbols q'_{2^ℓ}, multiple mapping groups are necessary. If the sum of unused information of all groups does not exceed r, the resulting DSM is full-rate DSM. Arrangement steps for two mapping groups of full-rate DSM are listed as follows.

Step 1 Define a set $\Omega = \{3, 4, 5, \cdots, N_T\}$, and remove $4, 8, \cdots, 2^{\lfloor \log_2 N_T \rfloor}$ from Ω. The resulting Ω does not contain indices of independent-mapping symbols. Let a denote the value of the product of all elements in Ω, so $\lfloor \log_2 a \rfloor$ denotes the number of total data bits of both groups and $r = \log_2 a - \lfloor \log_2 a \rfloor$ denotes the unused information in Ω.

Step 2 Divide Ω into two groups $G_1 = \{\ell_1, \ell_2, \cdots, \ell_n\}$ and $G_2 = \{\ell'_1, \ell'_2, \cdots, \ell'_{n'}\}$ where $\ell_i < \ell'_j$, $\ell_i < \ell_{i+1}$, $\ell'_j < \ell'_{j+1}$ $\forall i, j$, and n satisfies $\prod_{i=1}^{n} \ell_i < \sqrt{a}$ and $\ell'_1 \times \prod_{i=1}^{n} \ell_i \geq \sqrt{a}$.

Step 3 Check if the inequation

$$\sum_{i=1}^{n} \log_2 \ell_i - \lfloor \sum_{i=1}^{n} \log_2 \ell_i \rfloor \leq r \tag{13}$$

is true; if not, switch ℓ_i and ℓ'_j until the inequality holds. If all tries fail, switch two (or more) elements in G_1 and two (or more) elements in G_2 until the inequality holds. The two groups $\{q'_{\ell_1}, q'_{\ell_2}, \cdots, q'_{\ell_n}\}$ and $\{q'_{\ell'_1}, q'_{\ell'_2}, \cdots, q'_{\ell'_{n'}}\}$ are the desired groups where $\{\ell_1, \ell_2, \cdots, \ell_n\}$ and $\{\ell'_1, \ell'_2, \cdots, \ell'_{n'}\}$ are the updated G_1 and G_2, i.e., G_1 and G_2 after switch if necessary, respectively.

In order to roughly divide Ω into two balanced subsets, i.e., the product of all elements in G_1 is close to the product of all elements in G_2, $\prod_{i=1}^{n} \ell_i < \sqrt{a}$ and $\ell'_1 \times \prod_{i=1}^{n} \ell_i \geq \sqrt{a}$ are required in Step 2. The product of all elements in G_1 is less than $\sqrt{a}$ because it is likely to increase the product in G_1 by the switch in Step 3. Because the value r is the unused information for full-rate DSM, the unused information in G_1 $\sum_{i=1}^{n} \log_2 \ell_i - \lfloor \sum_{i=1}^{n} \log_2 \ell_i \rfloor$ bit cannot be larger than r. If it fails, the exchange between elements in G_1 and G_2 is needed. For simplicity, first one-element switch is tried and a possible switch order is $(\ell_n, \ell'_1), (\ell_n, \ell'_2), \cdots, (\ell_n, \ell'_{n'}), (\ell_{n-1}, \ell'_1), (\ell_{n-1}, \ell'_2), \cdots$. Usually, the one-element switch can be successful. The following lemma proves that the inequation (13) guarantees full-rate DSM.

Lemma 2: The two groups $\{q'_{\ell_1}, q'_{\ell_2}, \cdots, q'_{\ell_n}\}$ and $\{q'_{\ell'_1}, q'_{\ell'_2}, \cdots, q'_{\ell'_{n'}}\}$ which satisfy (13) always achieve full-rate DSM, i.e.,

$$\lfloor \log_2 \prod_{i=1}^{n} \ell_i \rfloor + \lfloor \log_2 \prod_{j=1}^{n'} \ell'_j \rfloor = \lfloor \log_2 a \rfloor \tag{14}$$

Proof: Defining $x = \sum_{i=1}^{n} \log_2 \ell_i$ and $y = \sum_{j=1}^{n'} \log_2 \ell'_j$, $\log_2 a$ can be written as $\log_2 a = x + y$ because of $a = \prod_{i=1}^{n} \ell_i \times \prod_{j=1}^{n'} \ell'_j$. Let r_1 and r_2 denote $x - \lfloor x \rfloor$ and $y - \lfloor y \rfloor$, respectively. Hence, the decimal part of $x + y$ is $r = x + y - \lfloor x+y \rfloor = \begin{cases} r_1 + r_2 & \text{if } r_1 + r_2 < 1 \\ r_1 + r_2 - 1 & \text{otherwise} \end{cases}$. Because of $r_1 > r_1 + r_2 - 1$, inequation (13) which is $x - \lfloor x \rfloor \le x + y - \lfloor x+y \rfloor$ implies $r = r_1 + r_2$ which is $x - \lfloor x \rfloor + y - \lfloor y \rfloor = x + y - \lfloor x+y \rfloor$. Consequently, the equation becomes $\lfloor x \rfloor + \lfloor y \rfloor = \lfloor x + y \rfloor$ which is (14). □

The following is an example of the proposed arrangement.

Example 5 ($N_T = 16$ and $Q = 2^{44}$): The DSM has four independent-mapping symbols q'_{16}, q'_8, q'_4 and q'_2 which ten data bits are mapped to. In Step 1, Ω has parameters $\log_2 a = 34.25$ and $r = 0.25$. Two groups in Step 2 are $G_1 = \{3, 5, 6, 7, 9, 10\}$ and $G_2 = \{11, 12, 13, 14, 15\}$. The inequation (13) fails due to $\sum_{i=1}^{6} \log_2 \ell_i = 15.79$ which implies that the unused information in G_1 is 0.79 larger than 0.25. The first switch makes $G_1 = \{3, 5, 6, 7, 9, 11\}$ which has $\sum_{i=1}^{6} \log_2 \ell_i = 15.93$, so it still cannot satisfy (13). The second switch makes $G_1 = \{3, 5, 6, 7, 9, 12\}$ which has $\sum_{i=1}^{6} \log_2 \ell_i = 16.05$, so it is successful. Hence, there are four independently-mapping symbols q'_{16}, q'_8, q'_4 and q'_2, and two mapping groups $\{q'_{12}, q'_9, q'_7, q'_6, q'_5, q'_3\}$ and $\{q'_{15}, q'_{14}, q'_{13}, q'_{11}, q'_{10}\}$. For $\{q'_{12}, q'_9, q'_7, q'_6, q'_5, q'_3\}$, there are $2^{16} = 65536$ valid reference orders among total $12 \times 9 \times 7 \times 6 \times 5 \times 3 = 68040$ permutations; while for $\{q'_{15}, q'_{14}, q'_{13}, q'_{11}, q'_{10}\}$, there are $2^{18} = 262144$ valid reference orders among total $15 \times 14 \times 13 \times 11 \times 10 = 300300$ permutations.

The proposed two-group arrangement can be generalized for more groups. For k-group arrangement, groups $G_1, G_2, \cdots, G_k$ should have product close to $\sqrt[k]{a}$ in Step 2, and switch elements to ensure the sum of the unused information in all groups does not exceed r. An example of three-group arrangement is given as follows.

Example 6 ($N_T = 20$ and $Q = 2^{61}$): In Step 1, Ω has parameters $\log_2 a = 51.0774$ and $r = 0.0774$ only. Three groups in Step 2 are $G_1 = \{3, 5, 6, 7, 9, 10\}$, $G_2 = \{11, 12, 13, 14\}$ and $G_3 = \{15, 17, 18, 19, 20\}$. The switch between G_1 and G_2 makes $G_1 = \{3, 5, 6, 7, 9, 12\}$ as Example 5, and the unused information in G_1 is 0.0541 which means that the unused information in G_2 cannot exceed $0.0774 - 0.0541 = 0.0233$. Because $G_2 = \{10, 11, 13, 14\}$ has $\sum_{i=1}^{4} \log_2 \ell'_i = 14.2892$, the switch between G_2 and G_3 is necessary. All one-element exchanges fail, and then a two-element exchange which makes $G_2 = \{10, 11, 15, 20\}$ and $G_3 = \{13, 14, 17, 18, 19\}$ is successful, which have $\sum_{i=1}^{4} \log_2 \ell'_i = 15.0102$ and $\sum_{i=1}^{5} \log_2 \ell''_i = 20.0131$, respectively. Hence, there are four independently-mapping symbols q'_{16}, q'_8, q'_4 and q'_2, and three mapping groups $\{q'_{12}, q'_9, q'_7, q'_6, q'_5, q'_3\}$, $\{q'_{20}, q'_{15}, q'_{11}, q'_{10}\}$ and $\{q'_{19}, q'_{18}, q'_{17}, q'_{14}, q'_{13}\}$.

Unlike Examples 2, 3 and 4, the above examples have more than one mapping groups, so the detection algorithm in Sec. III.D cannot be directly applied to the two examples.

B. FULL-MAPPING DSM AND DETECTION ALGORITHMS

For multiple mapping groups, if $\hat{\mathbf{p}}$ is invalid in Step 6, to find a valid $\hat{\mathbf{p}}$ efficiently and reasonably is a difficult problem. To solve this problem, we propose a new idea to satisfy Limit 2: all $N_T!$ permutations are valid reference orders, i.e., $N_T! - Q$ originally invalid permutations for $\mathbf{p}$ are also mapped by data bits. For instance, in Example 2 whose N_T is 4, in addition to sixteen valid patterns of $\{q'_1, q'_2, q'_3, q'_4\}$, (0, 0, 0, 0), (0, 0, 0, 1), (0, 0, 0, 2), (0, 0, 0, 3), (0, 0, 1, 0), (0, 0, 1, 1), (0, 0, 1, 2), (0, 0, 1, 3), (0, 1, 0, 0), (0, 1, 0, 1), (0, 1, 0, 2), (0, 1, 0, 3), (0, 1, 1, 0), (0, 1, 1, 1), (0, 1, 1, 2), (0, 1, 1, 3), eight illegitimate patterns of $\{q'_1, q'_2, q'_3, q'_4\}$, (0, 0, 2, 0), (0, 0, 2, 1), (0, 0, 2, 2), (0, 0, 2, 3), (0, 1, 2, 0), (0, 1, 2, 1), (0, 1, 2, 2), (0, 1, 2, 3) are also mapped by data bits at the transmitter. Because there are no invalid reference orders, Step 6 is omitted. DSM which maps data bits to all $N_T!$ reference orders is called **full-mapping DSM**.

The full mapping is accomplished by expanding the ranges of all mapping groups and independently-mapping symbols. For a group $\{q'_{\ell_1}, q'_{\ell_2}, \cdots, q'_{\ell_n}\}$ or a symbol q'_{ℓ_1} which can be viewed as a group with $n = 1$, there are $d = \lfloor \log_2(\ell_1 \times \ell_2 \times \cdots \times \ell_n) \rfloor$ data bits. For the original DSM, among all $\prod_{i=1}^{n} \ell_i$ permutations in the group, 2^d legitimate permutations are mapped by d data bits, and $\prod_{i=1}^{n} \ell_i - 2^d < 2^d$ permutations are invalid. For the proposed full mapping, $\prod_{i=1}^{n} \ell_i - 2^d$ (among 2^d) valid permutations and $\prod_{i=1}^{n} \ell_i - 2^d$ invalid permutations are mapped by the same d data bits one by one. First choose a data bit in another mapping group or symbol, say b_0. For these $\prod_{i=1}^{n} \ell_i - 2^d$ chosen valid permutations, they are the same if $b_0 = 0$, but they would become the originally-invalid permutations one by one if $b_0 = 1$. Steps of the proposed full mapping are listed below. First d data bits form an integer m'_{ℓ_1}. If $b_0 = 1$ and $m'_{\ell_1} < \ell_1 \times \ell_2 \times \cdots \times \ell_n - 2^d$, then $m_{\ell_1} = m'_{\ell_1} + 2^d$; otherwise, $m_{\ell_1} = m'_{\ell_1}$. By doing so, the value of m_{ℓ_1} is $0 \le m_{\ell_1} \le \prod_{i=1}^{n} \ell_i - 1$. Mapping m_{ℓ_1} to $\{q'_{\ell_1}, q'_{\ell_2}, \cdots, q'_{\ell_n}\}$ is the same as what we propose in Sec. III.A, i.e., using (9). By the proposed mapping, if $m'_{\ell_1} < \ell_1 \times \ell_2 \times \cdots \times \ell_n - 2^d$, then $b_0 = 0$ and $b_0 = 1$ have different mapping.

For the proposed full-mapping, there are two detection strategies: ignoring the information of b_0 or not. The former is simple because the detected d data bits are simply the rightmost d bits in the binary representation of $\hat{m}_{\ell_1}$ which denotes the detected value of m_{ℓ_1} converted from $\hat{\mathbf{q}}$ in the last detection step. The $d + 1$-th bit from the right which may contain the information of b_0 is discarded. Examples 2-6 with $M = 4$ and $N_R = 1$ using full-mapping and the simple detection are simulated and presented in Fig. 4. For $N_T = 4$, 6 and 8, the redundancy ratio $(N_T! - Q)/N_T!$ is 1/3, 13/45 and 59/315, respectively. For a larger value of N_T, the curve of full-mapping is closer to the curve of the original mapping in Fig. 2 because the ratio of the redundancy to the number of permutations becomes smaller.

The latter strategy is to use the information of b_0. To find the improvement of the latter over the former, we

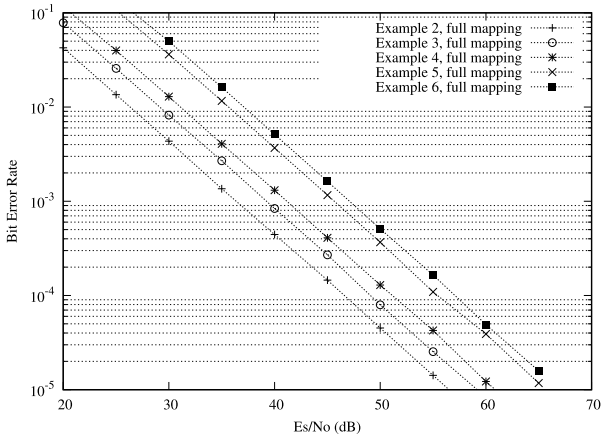


FIGURE 4. Simulation results of Examples 2-6 using full-mapping with $M = 4$ and $N_R = 1$.

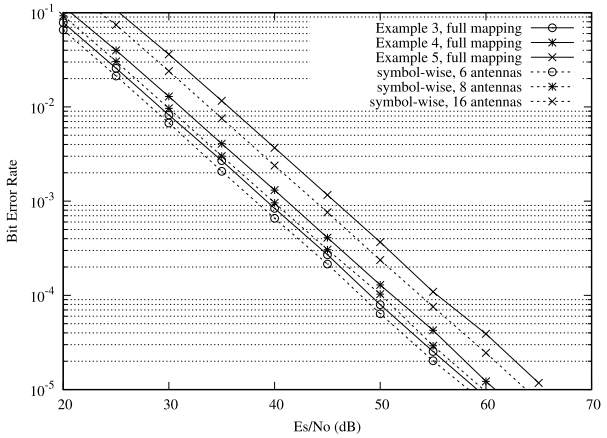


FIGURE 5. Simulation results of full-mapping DSM with $M = 4$ and $N_R = 1$.

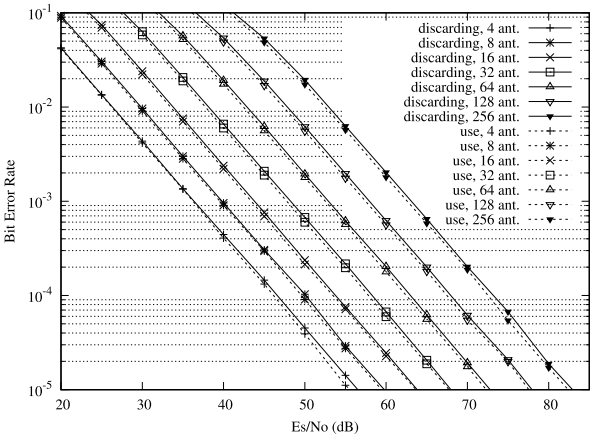


FIGURE 6. Simulation results of full-mapping reduced-rate DSM different detection strategies with $M = 4$ and $N_R = 1$.

propose a simple symbol-wise full-mapping scheme. In this scheme, all symbols are independently mapped, i.e., $N_T - 1$ independently-mapping symbols for N_T antennas. Let $b_\ell^1, b_\ell^2, \cdots, b_\ell^{\lfloor \log_2 \ell \rfloor}$ denote the $\lfloor \log_2 \ell \rfloor$ data bits corresponding to symbol q'_ℓ where $\ell \in \{N_T, N_T - 1, \cdots, 2\}$. First $b_\ell^1, b_\ell^2, \cdots, b_\ell^{\lfloor \log_2 \ell \rfloor}$ form the integer m'_ℓ. If $b_{\ell-1}^1 = 1$ and $m'_\ell < \ell - 2^{\lfloor \log_2 \ell \rfloor}$, then add $2^{\lfloor \log_2 \ell \rfloor}$ to m'_ℓ. Take $\ell = 7$ as an example. Two bits b_7^1 and b_7^2 are the two data bits corresponding to symbol q'_7. First b_7^1 and b_7^2 form the integer m'_7 ($0 \le m'_7 \le 3$). If $b_6^1 = 1$ and $m'_7 = 0$, 1 or 2, then add 4 to m'_7. Therefore, q'_7 is full-mapping due to $0 \le m'_7 \le 6$. Obviously, for $N_T \ge 6$, DSM using the simple mapping cannot achieve full rate. For instance, $N_T = 8$ has $Q = 256$ and $N_T = 16$ has $Q = 2^{13}$, so they have one and two bits loss compared with Example 4 and Example 5, respectively.

Fig. 5 compares simulation results of symbol-wise full-mapping DSM with those of group-wise full-mapping DSM in Example 3, 4 and 5. With the same N_T, symbol-wise DSM outperforms group-wise DSM because the former has less codewords and thus a smaller error coefficient than the latter.

Symbol-wise full-mapping DSM has higher redundancy ratio than group-wise full-mapping DSM, so discarding the information of $b_{\ell-1}^1$ when detecting q'_ℓ is somewhat wasteful. Hence, a technique to simply use the information of $b_{\ell-1}^1$ is proposed as follows. Let the detection order of $\hat{p}^{(1)}, \hat{p}^{(2)}, \cdots, \hat{p}^{(N_T)}$ in Step 3 be $o_{N_T}, o_{N_T-1}, \cdots, o_1$. In the final detection step, if $\hat{q}'_\ell < \ell - 2^{\lfloor \log_2 \ell \rfloor}$ or $\hat{q}'_\ell > 2^{\lfloor \log_2 \ell \rfloor} - 1$, then the temporarily detected value of $b_{\ell-1}^1$, denoted by $\bar{b}_{\ell-1}^1$, is the $\lfloor \log_2 \ell \rfloor + 1$-th bit from the right in the binary representation of $\hat{q}'_\ell$. Let $\tilde{b}_{\ell-1}^1$ denote the value of $b_{\ell-1}^1$ converted from $\hat{q}'_{\ell-1}$. If $\bar{b}_{\ell-1}^1$ is different from $\tilde{b}_{\ell-1}^1$, then the finally detected value of $b_{\ell-1}^1$, denoted by $\hat{b}_{\ell-1}^1$, is determined according to the detection order in Step 3, i.e.,

$$\hat{b}_{\ell-1}^1 = \begin{cases} \bar{b}_{\ell-1}^1, & \text{if } o_\ell < o_{\ell-1} \\ \tilde{b}_{\ell-1}^1, & \text{if } o_\ell > o_{\ell-1}. \end{cases} \tag{15}$$

Simulation results of symbol-wise full-mapping DSM with $N_T \le 256$ using different detection strategies are shown in Fig. 6 where curves labeled by "discarding" represent discarding the information of the bit belonging to another group and curves labeled by "use" denotes using the information of the bit belonging to another group by (15). With the same N_T, the latter provides a little gain over the former.

V. CONCLUSION

In this paper, low-complexity DSM schemes for a lot of transmitter antennas have been designed. A low-complexity detection algorithm whose error performance is close to the error performance of the noncoherent ML detection is proposed. The algorithm is less complicated than all existing fixed-complexity detectors. In addition, two techniques of mapping data bits to reference orders which are suitable for a lot of transmitter antennas are proposed: one is the group arrangement algorithm for full-rate DSM, and the other is the full-mapping for simple detectors. Two types of full mapping are proposed: group-wise and symbol-wise. The group-wise full-mapping is proposed for multi-group full-rate DSM so

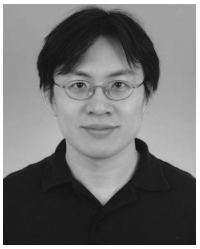
that the proposed low-complexity detector can be applied to it, and the symbol-wise full-mapping is proposed to further decrease the complexity so it is suitable for downlink massive-MIMO systems. Simulation results confirm that all proposed low-complexity DSM schemes have satisfactory error performance.

ACKNOWLEDGMENT

The author thanks Mr. P.-M. Tsai for helping with the group arrangement.

REFERENCES

[1] R. Mesleh, H. Haas, S. Sinanovic, C. Ahn, and S. Yun, "Spatial modulation," *IEEE Trans. Veh. Technol.*, vol. 57, no. 4, pp. 2228–2242, Jul. 2008.
[2] J. Jeganathan, A. Ghrayeb, and L. Szczecinski, "Spatial modulation: Optimal detection and performance analysis," *IEEE Commun. Lett.*, vol. 12, no. 8, pp. 545–547, Aug. 2008.
[3] J. Jeganathan, A. Ghrayeb, L. Szczecinski, and A. Ceron, "Space shift keying modulation for MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3692–3703, Jul. 2009.
[4] M. Renzo, H. Haas, and P. Grant, "Spatial modulation for multiple-antenna wireless systems: A survey," *IEEE Commun. Mag.*, vol. 49, no. 12, pp. 182–191, Dec. 2011.
[5] Y. Bian, X. Cheng, M. Wen, L. Yang, H. V. Poor, and B. Jiao, "Differential spatial modulation," *IEEE Trans. Veh. Technol.*, vol. 64, no. 7, pp. 3262–3268, Jul. 2015.
[6] N. Ishikawa and S. Sugiura, "Unified differential spatial modulation," *IEEE Wireless Commun. Lett.*, vol. 3, no. 4, pp. 337–340, Aug. 2014.
[7] M. Wen, X. Cheng, Y. Bian, and H. V. Poor, "A low-complexity near-ML differential spatial modulation detector," *IEEE Signal Process. Lett.*, vol. 22, no. 11, pp. 1834–1838, Nov. 2015.
[8] Z. Li, X. Cheng, S. Han, M. Wen, L.-Q. Yang, and B. Jiao, "A low-complexity optimal sphere decoder for differential spatial modulation," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, San Diego, CA, USA, Dec. 2015, pp. 1–6.
[9] R.-Y. Wei and T.-Y. Lin, "Low-complexity differential spatial modulation," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 356–359, Apr. 2019.
[10] L. Xiao, P. Yang, X. Lei, Y. Xiao, S. Fan, S. Li, and W. Xiang, "A low-complexity detection scheme for differential spatial modulation," *IEEE Commun. Lett.*, vol. 19, no. 9, pp. 1516–1519, Sep. 2015.
[11] N. Ishikawa and S. Sugiura, "Rectangular differential spatial modulation for open-loop noncoherent massive-MIMO downlink," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1908–1920, Mar. 2017.
[12] B. M. Hochwald and W. Sweldens, "Differential unitary space-time modulation," *IEEE Trans. Commun.*, vol. 48, no. 12, pp. 2041–2052, Dec. 2000.
[13] C. Gao, A. M. Haimovich, and D. Lao, "Multiple-symbol differential detection for MPSK space-time block codes: Decision metric and performance analysis," *IEEE Trans. Commun.*, vol. 54, no. 8, pp. 1502–1510, Aug. 2006.
[14] H. Leib, "Data-aided noncoherent demodulation of DPSK," *IEEE Trans. Commun.*, vol. 43, nos. 2–4, pp. 722–725, Feb. 1995.
[15] R. Rajashekar, N. Ishikawa, S. Sugiura, K. V. S. Hari, and L. Hanzo, "Full-diversity dispersion matrices from algebraic field extensions for differential spatial modulation," *IEEE Trans. Veh. Technol.*, vol. 66, no. 1, pp. 385–394, Jan. 2017.
[16] R. Rajashekar, C. Xu, N. Ishikawa, S. Sugiura, K. V. S. Hari, and L. Hanzo, "Algebraic differential spatial modulation is capable of approaching the performance of its coherent counterpart," *IEEE Trans. Commun.*, vol. 65, no. 10, pp. 4260–4273, Oct. 2017.
[17] C. Xu, R. Rajashekar, N. Ishikawa, S. Sugiura, and L. Hanzo, "Single-RF index shift keying aided differential space–time block coding," *IEEE Trans. Signal Process.*, vol. 66, no. 3, pp. 773–788, Feb. 2018.
[18] C. Xu, P. Zhang, R. Rajashekar, N. Ishikawa, S. Sugiura, L. Wang, and L. Hanzo, "Finite-cardinality single-RF differential space-time modulation for improving the diversity-throughput tradeoff," *IEEE Trans. Commun.*, vol. 67, no. 1, pp. 318–335, Jan. 2019.
[19] B. L. Hughes, "Differential space-time modulation," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2567–2578, Nov. 2000.



RUEY-YI WEI (Senior Member, IEEE) received the B.S. degree in electronics engineering from National Chiao-Tung University, Hsinchu, Taiwan, in 1993, and the Ph.D. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, in 1998.

From 1998 to 2000, he was with private companies in Taiwan and focused on the modem design. Since August 2000, he has joined the faculty of National Central University, where he is currently a Professor with the Department of Communication Engineering. From July 2013 to July 2014, he was a Visiting Scholar with the University of Washington, Seattle. His research interests include the areas of coded modulation, noncoherent detection, and MIMO systems.

Dr. Wei received the Best Paper Award for young scholars from the IEEE IT/COM Society Taipei/Tainan Chapter, in 2007. Since 2011, he has been an Associate Editor of the IEEE COMMUNICATIONS SURVEYS AND TUTORIALS.