

Differentially-encoded rectangular spatial modulation approaches the performance of its coherent counterpart

Lixia Xiao, *Member, IEEE*, Pei Xiao, *Senior Member, IEEE*, Hang Ruan, Naoki Ishikawa, *Member, IEEE*, Lei Lu, Yue Xiao, and Lajos Hanzo, *Fellow, IEEE*

Abstract—A simplified rectangular differential spatial modulation (S-RDSM) scheme is conceived for massive multiple-input multiple-output (MIMO) systems dispensing with the channel state information (CSI). In the proposed S-RDSM scheme, the information bits are first mapped to a conventional SM symbol and then rectangular differential encoding is invoked between a pair of SM symbols. Then a non-coherent detector relying on a forgetting factor is developed, which requires no CSI at the receiver. Explicitly, a low-complexity hard limited maximum likelihood (HL-ML) detector is conceived for our generalized S-RDSM scheme, which is characterized by our theoretical analysis. Furthermore, we derive the optimal forgetting factor in closed form, which is capable of significantly reducing the complexity of the associated optimization. Finally, the upper bounds of the average bit error probability (ABEP) are derived using the moment generating function (MGF), and are validated by our simulation results. Both the theoretical and simulation results have shown that the proposed S-RDSM system outperforms the existing non-coherent schemes, despite operating at 10% of the benchmarker’s complexity, whilst approaching the performance of its coherent SM counterpart at a comparable complexity.

Index Terms—Rectangular Differential Spatial Modulation, Spatial Modulation, Average Bit Error Probability, Maximum Likelihood.

I. INTRODUCTION

LARGE-SCALE multiple-input multiple-output (MIMO) schemes [1]-[3] are capable of significantly increasing the capacity, hence they are deemed to constitute a key technology for next-generation wireless systems. In massive MIMO, the

transmitter and/or the receiver employs numerous antennas, which require numerous radio frequency (RF) chains, hence imposing a substantial implementation cost, energy consumption and signal processing complexity as well as pilot overhead. In order to simplify the massive MIMO structure, sparse RF chain based spatial modulation (SM) aided MIMO systems have been developed in [5]-[14], where the information bits are conveyed both by the activated transmit antenna (TA) indices as well as by the classic amplitude phase modulation (APM) symbols. Hence SM constitutes a promising low-cost massive MIMO solution for next generation wireless communications [5]-[14].

However, the benefits of the massive SM-MIMO have been mainly exploited under the idealized simplifying assumption of having perfect channel state information (CSI), even though relying on a tolerable pilot overhead is challenging. In order to address these challenges, differential spatial modulation (DSM) techniques [15]-[35], which rely on a single RF transmit structure dispensing with CSI knowledge have been developed. As for the DSM system, there is a pair of classic categories: square DSM (SDSM) [15]-[31] and rectangular DSM (RDSM) [32]-[35].

Specifically, for the SDSM system, one out of Q spreading matrices (SPMs) is activated to disperse N_t symbols to N_t TAs via N_t time instants. The design of SPM plays an important role in striking a trade-off between the transmit rate and the performance of the SDSM system. Specifically, the popular SPM designs include but are not limited to: 1) full SPM (FSPM) design [15]-[29] and 2) full-diversity (FD) SPM design [30]-[31]. For a FSPM-SDSM system, there is a total of $Q = 2^{\lfloor \log_2(N_t!) \rfloor}$ legitimate SPMs, where $\lfloor \cdot \rfloor$ is the floor operator. Each $N_t \times N_t$ -element full rank SPM has a single nonzero element in each column. As N_t increases, the value of Q increases exponentially, which makes bit-to-symbol mapping, demapping and signal detection challenging. It has been demonstrated in [24], [25] that the FSPM-SDSM structure is capable of supporting at most $N_t = 32$ TAs. In order to reduce the number of SPMs, FD-SDSM has been developed in [30]-[31], where there is only a total of $Q = N_t$ legitimate SPMs for bit-to-symbol mapping, and each SPM only transmits a single APM symbol via N_t time slots. Hence the transmission rate is significantly reduced compared to its coherent SM counterpart. Consequently, the known SDSM systems are not suitable for large-scale MIMO configurations.

To circumvent the aforementioned problem, RDSM systems

L. Xiao, P. Xiao and H. Ruan are with 5G Innovation Center of University of Surrey, Surrey, GU27XH, UK. (e-mail: lixiaoxiao@hust.edu.cn, p.xiao@surrey.ac.uk, hr648@outlook.com).

L. Xiao is also with the Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan 430074, China.

N. Ishikawa is with Hiroshima City University, Hiroshima 731-3194, Japan. (e-mail: naoki@ishikawa.cc).

L. Lu is with Wireless Network RAN Research Department Shanghai Huawei Technologies. (e-mail: kevin.lu@huawei.com).

Y. Xiao is with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China.

L. Hanzo is with the school of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (email: l-h@ecs.soton.ac.uk).

L. Xiao and P. Xiao would like to acknowledge the financial support from Shanghai Huawei Technologies Co. Ltd.

L. Hanzo would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council projects EP/N004558/1, EP/P034284/1, EP/P034284/1, EP/P003990/1 (COALESCE), of the Royal Society’s Global Challenges Research Fund Grant as well as of the European Research Council’s Advanced Fellow Grant QuantCom.

are developed in [32]-[35] based on the concept of FD-SDSM. Specifically, the information bits are firstly mapped to a full-diversity SPM, then they are converted to a non-square matrix after differential coding. Hence, one out of $Q = \log_2(N_t)$ SPMs is activated to convey a single symbol via T time instants. Furthermore, the transmission rate of the RDSM using $T = 1$ is comparable to that of its coherent SM counterpart. As a result, RDSM becomes a promising candidate for large-scale MIMO communications dispensing with CSI.

In the existing RDSM systems, non-coherent detection relies on the previous estimated symbols, hence tends to suffer from error propagation. To improve the performance of RDSM, a new non-coherent detector relying on a forgetting factor is designed in [32]-[35]. Specifically, in [32]-[33], the forgetting factor is chosen by minimizing the mean square error of the reference signal \hat{Y} . In [34], the forgetting factor is determined through maximizing the effective signal to noise ratio (SNR). However, these design methods require time-consuming exhaustive search, where a large number of random variables are generated. To address this limitation, in [35], an adaptive forgetting factor is proposed, which does not require off-line optimization. However, it only relies on the current received symbols. Thus, this adaptive forgetting factor typically result in a locally optimal solution, rather than the global optimum. On the other hand, the locally optimal nature of this forgetting factor and the resultant error propagation of RDSM render the derivation of average bit error probability (ABEP) bound challenging. Moreover, the exhaustive search based maximum likelihood (ML) detector employed in the RDSM of [32]-[35] imposes a high complexity.

Against this background, the new contributions of this paper are as follows:

- 1) We propose a simplified RDSM (S-RDSM) system for massive MIMO downlink communications. Specifically, the information bits are first mapped to a conventional SM symbol. Then rectangular differential encoding is employed relying on a pair of adjacent SM symbols, where the complexity of the bit-to-symbol mapping and demapping is comparable to that of its coherent SM counterpart.
- 2) We employed the hard limited ML (HL-ML) detector for the new generalized S-RDSM system in our theoretical analysis. The complexity of our HL-ML based S-RDSM is independent of the constellation size L , and it is a linear function of N_t for any value of T . Our simulation results indicate that our solution is capable of operating at 10% complexity of its ML counterpart in the massive MIMO configurations considered.
- 3) We demonstrate that the transmit signal cardinality of the proposed S-RDSM system is the same as that of its coherent SM counterpart and it is much lower than that of the existing RDSM schemes [33]-[35]. Furthermore, we show that the error propagation encountered by the proposed S-RDSM system is mainly owing to a pair of consecutive blocks. Based on this observation, the approximate ABEP upper bound of our proposed S-RDSM system is derived using the moment generating

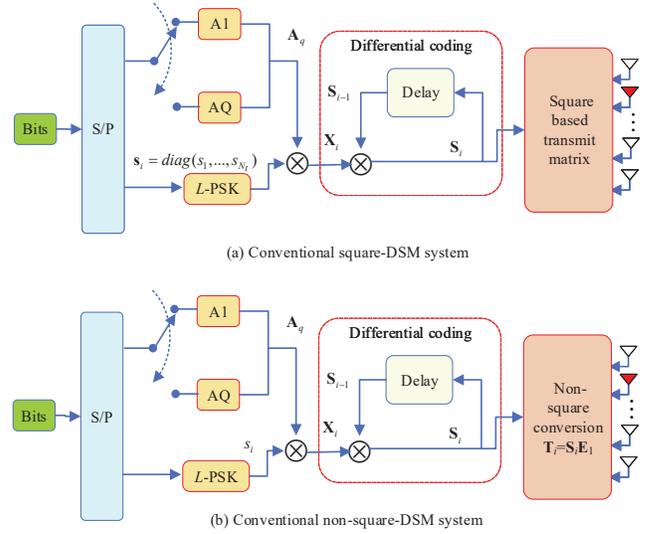


Fig. 1. System model of conventional DSM systems.

function (MGF) and its accuracy is confirmed by our simulation results.

- 4) We propose a new forgetting factor, which relies on all the bits of the entire transmission block and optimize it by deriving a closed-form expression, which is capable of significantly reducing the complexity of the forgetting factor optimization.

The remainder of this paper is organized as follows. Section II provides a rudimentary review of the conventional DSM systems. In Section III, the system model and HL-ML detector of our proposed S-RDSM system are introduced. In Section IV, the ABEP upper bound of our S-RDSM system is derived. Section V presents our simulation results. Finally, Section VI concludes this paper.

Notation: $\|\cdot\|^2$ denotes the Frobenious norms of a matrix. $|\cdot|$ represents the cardinality of a set. $(\cdot)^T$, $(\cdot)^*$ and $(\cdot)^H$ stand for the transpose, the conjugated, and the Hermitian transpose of a vector/matrix. $\langle x \rangle$ returns the value of x mod N_t . $\lfloor \cdot \rfloor$ is the floor operation. $diag[\cdot]$ returns a square diagonal matrix with the elements of vector. \mathbf{I}_{N_t} is the $(N_t \times N_t)$ -element identity matrix. $x \gg y$ and $x \ll y$ indicate that x is much greater and much smaller than y , respectively. $\mathbf{E}(\mathbf{x})$ denotes the expectation operator of \mathbf{x} .

II. REVIEW OF CONVENTIONAL DSM SYSTEMS

Let us consider a DSM system having N_t TAs and N_r receiver antennas (RAs). The system models of the square DSM and non-square DSM systems are depicted in Fig. 1.

A. Conventional square DSM systems

In the SDSM system, the information bits are conveyed by the activated SPM \mathbf{A}_q $q = 1, \dots, Q$ as well as by the APM symbols via N_t time slots. As shown in Fig. 1, the information bit segment of length B is partitioned into two parts: 1) $B_1 = \log_2(Q)$ bits are mapped to one of the SPMs \mathbf{A}_q ; and 2) B_2 bits are mapped to L -PSK symbols s_1, \dots, s_{N_t}

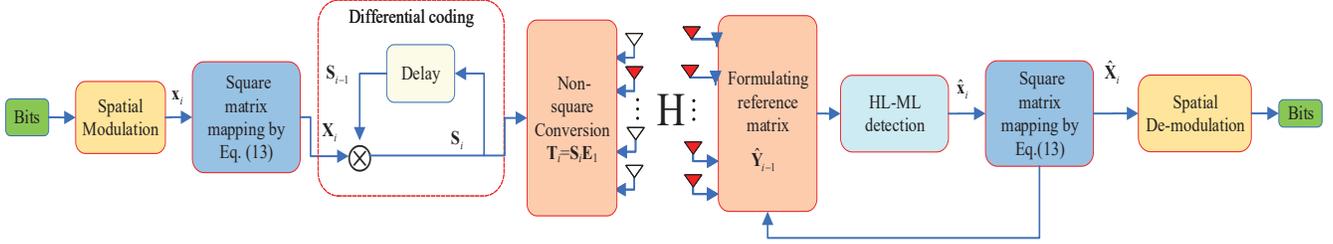


Fig. 2. System model of the proposed S-RDSM system.

that are transmitted by the activated SPM \mathbf{A}_q . As a result, the i -th signal block is expressed as

$$\mathbf{X}_i = \mathbf{A}_q \text{diag}[s_1, s_2, \dots, s_{N_t}]. \quad (1)$$

Differential encoding is employed between a pair of adjacent blocks as

$$\mathbf{S}_i = \mathbf{S}_{i-1} \mathbf{X}_i, \quad (2)$$

where $\mathbf{S}_0 = \mathbf{I}_{N_t}$.

Let us denote that $\tilde{\mathbf{H}}_i \in \mathbb{C}^{N_r \times N_t}$ and $\tilde{\mathbf{N}}_i \in \mathbb{C}^{N_r \times N_t}$ the channel and the noise matrices, whose elements obey the complex Gaussian distributions $\mathcal{CN}(0, 1)$ and $\mathcal{CN}(0, \sigma^2)$. The i -th block's received signal $\mathbf{Y}_i \in \mathbb{C}^{N_r \times N_t}$ can be expressed as

$$\mathbf{Y}_i = \tilde{\mathbf{H}}_i \mathbf{S}_{i-1} \mathbf{X}_i + \tilde{\mathbf{N}}_i. \quad (3)$$

Assuming that the channel coefficients remain near-constant over W time blocks, i.e., we have $\tilde{\mathbf{H}}_1 \approx \tilde{\mathbf{H}}_W$, the ML detector of SDSM can be formulated as

$$\hat{\mathbf{X}}_i = \arg \min_{\mathbf{X}_i \in \mathbb{X}} \|\mathbf{Y}_i - \mathbf{Y}_{i-1} \mathbf{X}_i\|^2, \quad (4)$$

where \mathbb{X} is the set of SDSM symbols.

The main difference of SDSM system lies in the design of \mathbf{A}_q . Based on the design of \mathbf{A}_q , there are two popular types of SDSM systems, namely FSPM-SDSM and FD-SDSM, which are introduced as follows.

1) Full SPM based SDSM system:

In the FSPM-SDSM system, there is a total of $N_t!$ SPMs but only $Q = 2^{\lfloor \log_2(N_t!) \rfloor}$ SPMs are required for conveying N_t symbols. For each $(N_t \times N_t)$ -element full-rank SPM, there is only a single nonzero element in each column. Hence, the transmission rate of the FSPM-SDSM scheme is $R_{\text{FSPM}} = \lfloor 2^{\lfloor \log_2(N_t!) \rfloor} + N_t \log_2(L)/N_t \rfloor$ bpcu, where bpcu denotes bits per channel use. For a large value of N_t , Q becomes excessive, which makes finding the list bit-to-symbol mapping and demapping challenging.

2) Full diversity based SDSM system:

In the FD-SDSM system, we have a total of $Q = N_t$ SPMs as

$$\{\mathbf{A}_1, \dots, \mathbf{A}_{N_t}\} = \{\mathbf{I}_{N_t}, \mathbf{M}, \mathbf{M}^2, \dots, \mathbf{M}^{N_t-1}\}, \quad (5)$$

where the matrix \mathbf{M} is defined as

$$\mathbf{M} = \begin{bmatrix} 0 & 0 & \dots & 0 & u \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}, \quad (6)$$

where u is designed differently in [30] and [31], respectively. In the FD-SDSM system, each activated SPM only transmits a single L -PSK symbol via N_t time slots. The transmission rate of the FD-SDSM system is $R_{\text{FD}} = \log_2(N_t L)/N_t$, which is low for the massive MIMO antenna configuration.

B. Conventional non-square DSM systems

Again in the RDSM systems, the information bits are conveyed by the activated SPMs \mathbf{A}_q $q = 1, \dots, Q$ as well as by the APM symbols via T ($1 \leq T \leq N_t$) time slots. The SPM is selected based on (5) and (6) with $u = e^{j2\pi/L}$. As shown in Fig.1 (b), $B_1 = \log_2(N_t)$ bits are spread by one of the SPMs \mathbf{A}_q and $B_2 = \log_2(L)$ bits are modulated to an L -PSK symbols s_i as $\mathbf{X}_i = \mathbf{A}_q s_i$. Then differential coding is employed according to $\mathbf{S}_i = \mathbf{S}_{i-1} \mathbf{X}_i$, where $\mathbf{S}_0 = [\hat{\mathbf{E}}_1, \dots, \hat{\mathbf{E}}_k, \dots, \hat{\mathbf{E}}_{N_t/T}]$, which can be obtained by [34]. Finally, the encoded transmit signal is converted to a non-square matrix form formulated as

$$\mathbf{T}_i = \mathbf{S}_i \hat{\mathbf{E}}_1, \quad (7)$$

where $\hat{\mathbf{E}}_1 \in \mathbb{C}^{N_t \times T}$ is designed differently in [34]. The transmission rate of the RDSM system is $R_{\text{CN}} = \log_2(N_t L)/T$ bpcu.

Let us denote that $\mathbf{H}_i \in \mathbb{C}^{N_r \times N_t}$ and $\mathbf{N}_i \in \mathbb{C}^{N_r \times T}$ the channel and the noise matrices, whose elements obey the complex Gaussian distributions $\mathcal{CN}(0, 1)$ and $\mathcal{CN}(0, \sigma^2)$. The i -th block's received signal $\mathbf{Y}_i \in \mathbb{C}^{N_r \times T}$ can be expressed as

$$\mathbf{Y}_i = \mathbf{H}_i \mathbf{S}_{i-1} \mathbf{X}_i \hat{\mathbf{E}}_1 + \mathbf{N}_i. \quad (8)$$

Assuming that the channel coefficients remain near-constant over W time blocks, i.e., we have $\mathbf{H}_1 \approx \mathbf{H}_W$, the detector of RDSM in [34] is expressed as

$$\hat{\mathbf{X}}_i = \arg \min_{\mathbf{X}_i \in \mathbb{X}_R} \left\| \mathbf{Y}_i - \hat{\mathbf{Y}}_{i-1} \mathbf{X}_i \right\|^2, \quad (9)$$

where \mathbb{X}_R is the set of RDSM symbols and the value of $\hat{\mathbf{Y}}_{i-1}$ is defined by

$$\hat{\mathbf{Y}}(i) = \begin{cases} \sum_{k=1}^{N_t/T} \mathbf{Y}(k) \hat{\mathbf{E}}_k^H, & \text{if } i = N_t/T, \\ \mathbf{Y}(i) \hat{\mathbf{E}}^{(1-\alpha)} + \hat{\mathbf{Y}}(i-1) \hat{\mathbf{X}}(i) \hat{\mathbf{E}}^{(\alpha)}, & \end{cases} \quad (10)$$

with

$$\begin{cases} \hat{\mathbf{E}}^{(\alpha)} = \alpha \hat{\mathbf{E}}_1 \hat{\mathbf{E}}_1^H + \sum_{k=2}^{N_t/T} \hat{\mathbf{E}}_k \hat{\mathbf{E}}_k^H, \\ \hat{\mathbf{E}}^{(1-\alpha)} = (1-\alpha) \hat{\mathbf{E}}_1^H. \end{cases} \quad (11)$$

Due to the complex design of $(\hat{\mathbf{E}}_1, \dots, \hat{\mathbf{E}}_k, \dots, \hat{\mathbf{E}}_{N_t/T})$ and the introduction of α , it is difficult to derive the ABEP and the low-complexity detector for the conventional RDSM system.

III. PROPOSED SIMPLIFIED RECTANGULAR DSM SYSTEM

A. System model

In this section, a novel simplified RDSM system termed as S-RDSM is proposed. The system model is shown in Fig. 2. Specifically, $\log 2(N_t) + \log 2(L)$ bits are first mapped to a conventional SM symbol [6] \mathbf{x}_i as

$$\mathbf{x}_i = \underbrace{[0, \dots, 0]_{q_i-1}}_{q_i-1}, \underbrace{[s_i, 0, \dots, 0]_{N_t-q_i}}_{N_t-q_i}^T, \quad (12)$$

where q_i is the activated TA and s_i is the mapped L -PSK symbol. Then \mathbf{x}_i is mapped to a square matrix by

$$\mathbf{X}_i = \begin{bmatrix} \vdots & \vdots & \cdots & s_i & \cdots & 0 \\ s_i & 0 & \cdots & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots & \cdots & s_i \\ 0 & \cdots & s_i & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{C}^{N_t \times N_t}. \quad (13)$$

$\begin{matrix} \uparrow & & \uparrow & \uparrow & & \uparrow \\ q_i & & N_t & 1 & & q_i - 1 \end{matrix}$

Next, differential coding is employed according to

$$\mathbf{S}_i = \mathbf{S}_{i-1} \mathbf{X}_i, \quad (14)$$

where $\mathbf{S}_0 = \mathbf{I}_{N_t}$. Finally, the encoded S-RDSM transmit symbol is formulated as

$$\mathbf{T}_i = \mathbf{S}_{i-1} \mathbf{X}_i \mathbf{E}_1 = \mathbf{S}_{i-1} \mathbf{X}_i^{[1:T]}, \quad (15)$$

where \mathbf{E}_1 and $\mathbf{X}_i^{[1:T]}$ are the first T columns of \mathbf{I}_{N_t} and \mathbf{X}_i , respectively. The transmission rate of the proposed S-RDSM system is $R_{\text{ER}} = \log_2(N_t L)/T$ bpcu.

B. ML detector with the forgetting factor α

At the receiver, the reference symbols are defined as the 0-th block signal and the corresponding received signal $\mathbf{Y}_0 \in \mathbb{C}^{N_r \times N_t}$ is formulated as:

$$\mathbf{Y}_0 = \mathbf{H}_0 \mathbf{I}_{N_t} + \mathbf{N}_0, \quad (16)$$

and the i -th ($i > 0$) block's received signal $\mathbf{Y}_i \in \mathbb{C}^{N_r \times T}$ can be expressed as

$$\mathbf{Y}_i = \mathbf{H}_i \mathbf{S}_{i-1} \mathbf{X}_i \mathbf{E}_1 + \mathbf{N}_i, \quad (17)$$

where $\mathbf{H}_0 \in \mathbb{C}^{N_r \times N_t}$ vs $\mathbf{H}_i \in \mathbb{C}^{N_r \times N_t}$, and $\mathbf{N}_0 \in \mathbb{C}^{N_r \times N_t}$ vs $\mathbf{N}_i \in \mathbb{C}^{N_r \times T}$ denote the channel and the noise matrices, whose elements obey the complex Gaussian distributions $\mathcal{CN}(0, 1)$ and $\mathcal{CN}(0, \sigma^2)$. Assuming that the channel coefficients remain near constant over W time blocks, i.e., we have $\mathbf{H}_1 \approx \mathbf{H}_W \approx \mathbf{H}_0$, the received signal \mathbf{Y}_i can be represented by

$$\mathbf{Y}_i = \mathbf{H}_i \mathbf{X}_1 \cdots \mathbf{X}_{i-1} \mathbf{X}_i \mathbf{E}_1 + \mathbf{N}_i. \quad (18)$$

Hence, the ML detector without the forgetting factor α is expressed as

$$\hat{\mathbf{x}}_i = \arg \min_{\mathbf{x}_i \in \mathcal{X}} \left\| \mathbf{Y}_i - \tilde{\mathbf{Y}}_{i-1} \mathbf{X}_i \mathbf{E}_1 \right\|^2, \quad (19)$$

where \mathcal{X} is the set of SM symbol, $\tilde{\mathbf{Y}}_{i-1} = \mathbf{Y}_0 \mathbf{X}_1 \cdots \mathbf{X}_{i-1} \mathbf{X}_i$ and \mathbf{X}_i is obtained by \mathbf{x}_i via (13). Due to the non-square structure, the performance of (19) can be further improved by introducing the forgetting factor α as

$$\hat{\mathbf{x}}_i = \arg \min_{\mathbf{x}_i \in \mathcal{X}} \left\| \mathbf{Y}_i - \hat{\mathbf{Y}}_{i-1} \mathbf{X}_i \mathbf{E}_1 \right\|^2, \quad (20)$$

where $\hat{\mathbf{Y}}_i$ is defined as

$$\hat{\mathbf{Y}}_i = \begin{cases} \mathbf{Y}_0, & \text{if } i = 0, \\ \mathbf{Y}_i \mathbf{E}^{(1-\alpha)} + \hat{\mathbf{Y}}_{i-1} \hat{\mathbf{x}}_i \mathbf{E}^\alpha, & \text{if } i > 0, \end{cases} \quad (21)$$

with $\mathbf{E}^{(1-\alpha)} \in \mathbb{C}^{T \times N_t}$ and $\mathbf{E}^\alpha \in \mathbb{C}^{N_t \times N_t}$ given by

$$\begin{aligned} \mathbf{E}^{(1-\alpha)} &= [(1-\alpha) \mathbf{I}_T, \mathbf{O}_T, \dots, \mathbf{O}_T], \\ \mathbf{E}^\alpha &= \begin{bmatrix} \alpha \mathbf{I}_T & \mathbf{O}_T & \cdots & \mathbf{O}_T \\ \mathbf{O}_T & \mathbf{I}_T & \cdots & \mathbf{O}_T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{O}_T & \mathbf{O}_T & \cdots & \mathbf{I}_T \end{bmatrix}, \end{aligned} \quad (22)$$

where $\alpha \in [0, 1]$, \mathbf{O}_T is the $(T \times T)$ -element zero matrix. It is plausible that (20) is the same as (19) for the case of $\alpha = 1$.

Furthermore, the ML detector with α of (20) can also be represented by

$$\begin{aligned} (\hat{q}_i, \hat{s}_i) &= \arg \min_{\forall \mathbf{X}} \left[\left\| \mathbf{Y}_i - \hat{\mathbf{Y}}_{i-1} \mathbf{X} \mathbf{E}_1 \right\|^2 \right] \\ &= \arg \min_{\forall (q, s)} \left[\sum_{\tau=1}^T \left\| \mathbf{Y}_i^\tau - \hat{\mathbf{Y}}_{i-1}^{(q+\tau-1)} s \right\|^2 \right] \\ &= \arg \min_{\forall (q, s)} \left[\sum_{\tau=1}^T \left\| \hat{\mathbf{Y}}_{i-1}^{(q+\tau-1)} s \right\|^2 - 2\Re \left(\sum_{\tau=1}^T (\mathbf{Y}_i^\tau)^H \hat{\mathbf{Y}}_{i-1}^{(q+\tau-1)} s \right) \right] \\ &= \arg \min_{\forall (q, s)} \left[\sum_{\tau=1}^T \left\| \hat{\mathbf{Y}}_{i-1}^{(q+\tau-1)} \right\|^2 - 2\Re \left[\sum_{\tau=1}^T (\mathbf{Y}_i^\tau)^H \hat{\mathbf{Y}}_{i-1}^{(q+\tau-1)} s \right] \right]. \end{aligned} \quad (23)$$

C. Low-complexity HL-ML detector

In this section, our low-complexity HL-ML detector is developed, which detects the activated SPM index and APM symbol separately. For the conventional ML detector of (23), each antenna index corresponds to L constellation symbols and we find the optimal one from $N_t L$ indices. In the proposed HL-ML detector, we first obtain the optimal constellation symbol for each antenna index by match filter (MF). Then we estimate the final result from N_t ones, which can significantly reduce the complexity.

TABLE I
ADVANTAGES OVER EXISTING DSM SCHEMES.

	[10] 2013	[11,12,14] 2014	[17] 2015	[19] 2015	[20] 2015	[13] 2016	[15] 2016	[16] 2016	[18] 2017	[25-26] 2017	[21-24] 2018 2019	[27-30] 2017- 2019	Proposed
Single or two RF chain	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Finite cardinality	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓
Low complexity				✓	✓				✓				✓
High throughput							✓					✓	✓
High diversity			✓						✓	✓	✓	✓	✓
Large scale												✓	✓

Specifically, we first obtain the optimal constellation symbol \tilde{s}_q ($q = 1, \dots, N_t$) for each antenna index as

$$\tilde{s}_q = \mathbb{D} \left(\sum_{\tau=1}^T (\hat{\mathbf{Y}}_{i-1}^{(q+\tau-1)})^H \mathbf{Y}_i^\tau / \sum_{\tau=1}^T \left\| \hat{\mathbf{Y}}_{i-1}^{(q+\tau-1)} \right\|^2 \right), \quad (24)$$

where $\mathbb{D}(\cdot)$ denotes the digital demodulation function. Then the activated SPM index is estimated by

$$\hat{q}_i = \arg \min_{q=1, \dots, N_t} \left[\sum_{\tau=1}^T \left\| \hat{\mathbf{Y}}_{i-1}^{(q+\tau-1)} \right\|^2 - 2\Re \left[\sum_{\tau=1}^T (\mathbf{Y}_i^\tau)^H \hat{\mathbf{Y}}_{i-1}^{(q+\tau-1)} \tilde{s}_q \right] \right]. \quad (25)$$

After obtaining the optimal index \hat{q}_i , the symbol can be finally acquired by

$$\hat{s}_i = \mathbb{D} \left(\sum_{\tau=1}^T (\hat{\mathbf{Y}}_{i-1}^{(\hat{q}_i+\tau-1)})^H \mathbf{Y}_i^\tau / \sum_{\tau=1}^T \left\| \hat{\mathbf{Y}}_{i-1}^{(\hat{q}_i+\tau-1)} \right\|^2 \right). \quad (26)$$

We note that (25) and (26) have been proven to be the same as (23) in Appendix A. According to [23], the complexities of the ML and HL-ML detectors are quantified in terms of the number of real-valued floating point (flop) operations. For specific matrices $\mathbf{A} \in \mathbb{C}^{m \times n}$, $\mathbf{B} \in \mathbb{C}^{n \times p}$, $\mathbf{c} \in \mathbb{C}^{n \times 1}$ and $\mathbf{d} \in \mathbb{C}^{n \times 1}$, the operations of \mathbf{AB} , $\|\mathbf{c}\|_F^2$ and $\mathbf{c} \pm \mathbf{d}$ require $8mnp - 2mp$, $4n - 1$, and $2n$ flops, respectively [12]. The complexity of ML and HL-ML detectors can be expressed as

$$\begin{aligned} C_{\text{ML}} &= (\underbrace{6N_r T}_{\hat{\mathbf{Y}}_{i-1} \mathbf{X}_i \mathbf{E}_1} + \underbrace{2N_r T}_{\mathbf{Y}_i - \hat{\mathbf{Y}}_{i-1} \mathbf{X}_i \mathbf{E}_1} + \underbrace{4N_r T}_{\|\cdot\|^2}) N_t L, \\ C_{\text{HL-ML}} &= [(4N_r - 1)T + (8N_r T - 2T) + \underbrace{2}_{\hat{s}_i} + 2] N_t \\ &\quad \underbrace{\sum_{\tau=1}^T \left\| \hat{\mathbf{Y}}_{i-1}^{(q+\tau)} \right\|^2}_{\sum_{\tau=1}^T (\mathbf{Y}_i^\tau)^H \hat{\mathbf{Y}}_{i-1}^{(q+\tau)}} \\ &= 12N_r T N_t - 3T N_t + 4N_t. \end{aligned} \quad (27)$$

The resultant complexity reduction ratio can be formulated as:

$$r_c = (C_{\text{ML}} - C_{\text{HL-ML}}) / C_{\text{ML}} = \frac{12N_r T L - 12N_r T + 3T - 4}{12N_r T L} \approx 1 - \frac{1}{L}. \quad (28)$$

According to (28), it is plausible that the proposed low-complexity detector is capable of reducing the complexity to about 25%, 12.5%, 6.25% and 3.125% of that of the ML detector for QPSK, 8-PSK, 16-PSK and 32-PSK, respectively.

TABLE II
COMPARISONS WITH EXISTING DSM SYSTEMS.

Scheme	Transmission rate	Cardinality of transmit signal	Detection complexity order
FAM -SDSM [17]	$\frac{2^{\lfloor \log_2(N_t!) \rfloor}}{N_t + \log_2(L)}$	$N_t!$	$O(N_t!)$
FD -SDSM [30]	$\frac{\log_2(N_t L)}{N_t}$	$N_t L^2$	$O(N_t L)$
RDSM [34] [35]	$\log_2(L N_t) / T$	$N_t L^2$	$O(N_t L)$
STBC -RDSM [33]	$\log_2(L^2 N_t / 2) / 2$	$\gg N_t L$	$O(N_t)$
Proposed S-RDSM	$\log_2(L N_t) / T$	$N_t L$	$O(N_t)$

D. Advantages over the existing DSM schemes

Table I and II compares the proposed S-RDSM scheme to the existing DSM schemes. In the proposed S-RDSM scheme, we have $u = 1$ of (6), which is the main difference with respect to the conventional RDSM scheme of [34]. As observed from Tables I and II, the advantages of the proposed S-RDSM represented by its simple structure may be summarized as follows.

- 1) **High throughput:** The throughput of the S-RDSM scheme using $T = 1$ is comparable to that of coherent SM, which is much higher than that of the existing SDSM systems [15]-[31].
- 2) **Finite cardinality:** The cardinality of the proposed S-RDSM's transmit signal is $|\mathbf{S}_i| = |\mathbf{X}_1 \mathbf{X}_2 \cdots \mathbf{X}_i| = N_t L$, which is much lower than that of the existing DSM schemes. As a result, the unbounded differential constellation size issues are further mitigated. The cardinality of the encoded signal of different RDSM schemes is calculated in detail as follows.

The main difference between the proposed S-RDSM scheme and the RDSM scheme of [34] is that u of (6) is equal to 1. According to (6), we have

$$\begin{aligned} \mathbf{M}^1 &= \begin{bmatrix} 0 \cdots 0 u \\ 1 \cdots 0 0 \\ \vdots \\ 0 \cdots 1 0 \end{bmatrix}, \dots, \mathbf{M}^{N_t} = \begin{bmatrix} u 0 \cdots 0 \\ 0 u \cdots 0 \\ \vdots \\ 0 0 \cdots u \end{bmatrix} \\ \vdots \\ \mathbf{M}^{LN_t} &= \begin{bmatrix} u^L 0 \cdots 0 \\ 0 u^L \cdots 0 \\ \vdots \\ 0 0 \cdots u^L \end{bmatrix}, \mathbf{M}^{LN_t+1} = \begin{bmatrix} 0 0 0 u^{L+1} \\ u^L 0 0 0 \\ \vdots \\ 0 0 u^L 0 \end{bmatrix}. \end{aligned} \quad (29)$$

Since $u = e^{j2\pi/L}$ is designed in [34], we have $u^L = 1$. The cardinality of the transmit signal of [34] is $|\mathbf{S}_i| = |\mathbf{M}^i||\mathcal{S}| = N_t L^2$, where $|\mathcal{S}|$ is the cardinality of the L -PSK symbol.

In the proposed S-RDSM system, we have $u = 1$. The cardinality of the corresponding transmit signal is $|\mathbf{S}_i|_p = N_t L$.

- 3) **Simplified transmit structure:** More importantly, as a benefit of the finite cardinality and special structure, the error propagation in the proposed S-RDSM system is mainly imposed by a pair of adjacent blocks, which is helpful for our ABEP performance analysis and it will be exploited in the next section.

IV. SIMPLIFIED FORGETTING FACTOR OPTIMIZATION

A. Conventional optimization

Assuming that no error is encountered by the previous $(i-1)$ blocks, the reference signal can be formulated as:

$$\tilde{\mathbf{Y}}_{i-1} = \mathbf{H}\mathbf{S}_{i-1} + \tilde{\mathbf{N}}_{i-1}, \quad (30)$$

with $\tilde{\mathbf{N}}_{i-1} = \mathbf{N}_{i-1}\mathbf{E}^{(1-\alpha)} + \tilde{\mathbf{N}}_{i-2}\tilde{\mathbf{X}}_{i-1}\mathbf{E}^\alpha$, where $\tilde{\mathbf{N}}_0 = \mathbf{N}_0$, which is demonstrated in Appendix B. The forgetting factor can be optimized according to:

$$\hat{\alpha} = \arg \min_{\alpha} \sum_{i=1}^{W/T} \mathbf{E}(\|\tilde{\mathbf{N}}_i\|^2) \approx \arg \min_{\alpha} \sum_{i=1}^{W/T} \mathbf{E}(\|\tilde{\mathbf{N}}_i\|^2), \quad (31)$$

with

$$\tilde{\mathbf{N}}_i = \mathbf{N}_i\mathbf{E}^{(1-\alpha)} + \tilde{\mathbf{N}}_{i-1}\mathbf{M}_1\mathbf{E}^\alpha, \quad (32)$$

where $\tilde{\mathbf{N}}_0 = \mathbf{N}_0$ and \mathbf{M}_1 is the \mathbf{M} of (6) with $u = 1$.

B. Proposed optimization

According to (31), the optimal forgetting factors have to be obtained through simulation, where a large number of random variables are generated. To reduce the complexity of the optimization, the closed form of (31) is derived in this section. Specifically, (32) can be further represented by

$$\begin{aligned} \tilde{\mathbf{N}}_i &= \mathbf{N}_i\mathbf{E}^{(1-\alpha)} + \tilde{\mathbf{N}}_{i-1}\mathbf{M}_1\mathbf{E}^\alpha \\ &= \mathbf{N}_i\mathbf{E}^{(1-\alpha)} + \mathbf{N}_{i-1}\mathbf{E}^{(1-\alpha)}\mathbf{M}_1\mathbf{E}^\alpha + \dots + \\ &\mathbf{N}_1\mathbf{E}^{(1-\alpha)}(\mathbf{M}_1\mathbf{E}^\alpha)^{i-1} + \tilde{\mathbf{N}}_0(\mathbf{M}_1\mathbf{E}^\alpha)^i \\ &= [\tilde{\mathbf{N}}_i^1, \tilde{\mathbf{N}}_i^2, \dots, \tilde{\mathbf{N}}_i^{N_t}], \end{aligned} \quad (33)$$

where we have

$$\tilde{\mathbf{N}}_i^u \sim \mathcal{CN}(0, (\sum_{j=0}^{i-1} \|\Psi_j^u\|^2 + \|\Upsilon_i^u\|^2)\sigma^2), u = 1, \dots, N_t, \quad (34)$$

with

$$\begin{aligned} \Psi_i &= \Psi_{i-1}\mathbf{M}_1\mathbf{E}^\alpha, \\ \Upsilon_i &= \Upsilon_{i-1}\mathbf{M}_1\mathbf{E}^\alpha, \end{aligned} \quad (35)$$

where $\Psi_0 = \mathbf{E}^{(1-\alpha)}$, $\Upsilon_0 = \mathbf{I}_{N_t}$, Ψ_j^u and Υ_i^u denote the u -th column of Ψ_j and Υ_i , respectively. The average noise variance of the i -th block is expressed as:

$$\sigma_i^2 = \frac{\sum_{u=1}^{N_t} \left(\sum_{j=0}^{i-1} \|\Psi_j^u\|^2 + \|\Upsilon_i^u\|^2 \right)}{N_t} \sigma^2, \quad (36)$$

which plays an important role in the ABEP analysis. Then $\sum_{i=1}^{W/T} \|\tilde{\mathbf{N}}_i\|^2$ in (31) can be represented by

$$\varphi = \sum_{i=1}^{W/T} \|\tilde{\mathbf{N}}_i\|^2 = \sum_{i=1}^{W/T} \sum_{u=1}^{N_t} \|\tilde{\mathbf{N}}_i^u\|^2 = \sum_{i=1}^{W/T} \sum_{u=1}^{N_t} \varphi_i^u, \quad (37)$$

where $\varphi_i^u = \|\tilde{\mathbf{N}}_i^u\|^2$. Hence, the mean value of (37) can be expressed as

$$\mathbf{E} \left(\sum_{i=1}^{W/T} \|\tilde{\mathbf{N}}_i\|^2 \right) = \mathbf{E}(\varphi) = \frac{\partial M_\varphi(s)}{\partial s} \Big|_{s=0}, \quad (38)$$

where $M_\varphi(s) = \int_{\varphi} e^{\varphi s} f_\varphi(\varphi) d\varphi$. Since we have

$$\frac{2\varphi_i^u}{\left(\sum_{j=0}^{i-1} \|\Psi_j^u\|^2 + \|\Upsilon_i^u\|^2 \right) \sigma^2} \sim \chi^2(2N_r), \quad (39)$$

the MGF of φ_i^u can be further obtained by

$$M_{\varphi_i^u}(s) = \left(1 - \left(\sum_{j=0}^{i-1} \|\Psi_j^u\|^2 + \|\Upsilon_i^u\|^2 \right) \sigma^2 s \right)^{-N_r}. \quad (40)$$

Then, the MGF of φ is given by

$$M_\varphi(s) = \prod_{i=1}^{W/T} \prod_{u=1}^{N_t} \left(1 - \left(\sum_{j=0}^{i-1} \|\Psi_j^u\|^2 + \|\Upsilon_i^u\|^2 \right) \sigma^2 s \right)^{-N_r}. \quad (41)$$

As a result, (38) is represented by

$$\frac{\partial M_\varphi(s)}{\partial s} \Big|_{s=0} = \sum_{i=1}^{W/T} \sum_{u=1}^{N_t} \left(\sum_{j=0}^{i-1} \|\Psi_j^u\|^2 + \|\Upsilon_i^u\|^2 \right) \sigma^2 N_r. \quad (42)$$

Finally, (31) is simplified according to:

$$\begin{aligned} \hat{\alpha} &= \arg \min_{\alpha} \mathbf{E} \left(\sum_{i=1}^{W/T} \|\tilde{\mathbf{N}}_i\|^2 \right) \\ &= \arg \min_{\alpha} \left\{ \sum_{i=1}^{W/T} \sum_{u=1}^{N_t} \left[\left(\sum_{j=0}^{i-1} \|\Psi_j^u\|^2 + \|\Upsilon_i^u\|^2 \right) \sigma^2 N_r \right] \right\} \\ &= \arg \min_{\alpha} \left\{ \sum_{i=1}^{W/T} \sum_{u=1}^{N_t} \left[\left(\sum_{j=0}^{i-1} \|\Psi_j^u\|^2 + \|\Upsilon_i^u\|^2 \right) \right] \right\} \\ &= \arg \min_{\alpha} \left\{ \sum_{i=1}^{W/T} \left[\|\Upsilon_i\|^2 + \sum_{j=0}^{i-1} \|\Psi_j\|^2 \right] \right\}. \end{aligned} \quad (43)$$

It is plausible that the forgetting factor optimization is independent of both N_r as well as of σ^2 and can be expressed in closed form without relying on any simulations.

C. Effective SNR analysis

According to (36), the average noise variance of W/T blocks is expressed as

$$\bar{\sigma}_v^2 = \underbrace{\frac{\sum_{i=1}^{W/T} \left[\|\Upsilon_i\|^2 + \sum_{j=0}^{i-1} \|\Psi_j\|^2 \right]}{W/T}}_{\beta} \sigma^2, \quad (44)$$

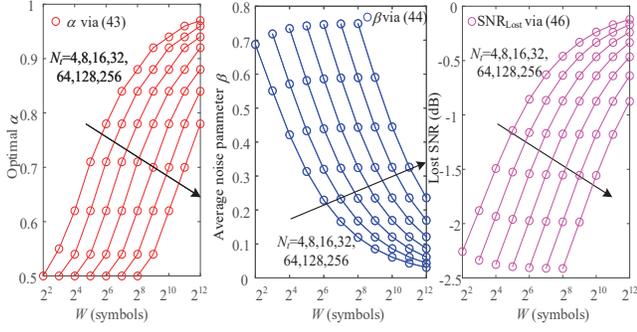


Fig. 3. Theoretical forgetting factor design and effective SNR analysis for $T = 1$, which is suitable for all the SNRs.

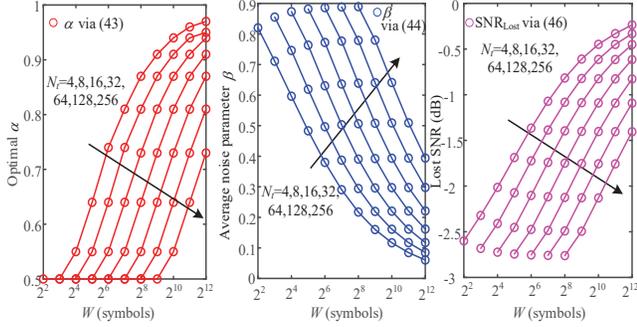


Fig. 4. Theoretical forgetting factor design and effective SNR analysis for $T = 2$, which is suitable for all the SNRs.

which is used for performance prediction. Moreover, the specific effective SNR analysis can be derived based on the noise variance of (36). In particular, the SNR-loss of the i -th block due to non-coherent detection can be formulated as:

$$\begin{aligned} \text{SNR}_i^P &= \frac{1}{\sigma^2 + \sigma_i^2} = \frac{1}{1 + \frac{\sum_{u=1}^{N_t} \sum_{j=0}^{i-1} \|\Psi_j^u\|^2 + \|\Upsilon_i^u\|^2}{N_t}} \text{SNR}_i^{\text{Coherent}} \\ &= \text{SNR}_i^{\text{Coherent}} \underbrace{-10 \log_{10} \left[1 + \frac{\sum_{u=1}^{N_t} \sum_{j=0}^{i-1} \|\Psi_j^u\|^2 + \|\Upsilon_i^u\|^2}{N_t} \right]}_{\text{Lost SNR (dB)}}. \end{aligned} \quad (45)$$

As a result, the average SNR-loss for W/T blocks is expressed as

$$\text{SNR}_{\text{Lost}} = \frac{\sum_{i=1}^{W/T} -10 \log_{10} \left[1 + \frac{\sum_{u=1}^{N_t} \left(\sum_{j=0}^{i-1} \|\Psi_j^u\|^2 + \|\Upsilon_i^u\|^2 \right)}{N_t} \right]}{W/T}. \quad (46)$$

In order to gain further insights, Figs. 3 and 4 present the optimal forgetting factor via (43), the average noise variance β of (44) and the average SNR-loss of (46) in the case of $T = 1$ and $T = 2$, respectively. As observed from Fig. 3, the value of β decreases as W increases and N_t/W decreases, hence resulting in reduced performance loss. Specifically, the value of β may be reduced to around 0 for $N_t = 4$ and $W = 2^{12}$, which indicates that non-coherent detection did not impose any extra noise and it is capable of approaching the performance of its coherent counterpart. Similar trends can be observed in the case of $T = 2$ in Fig. 4.

V. ABEP ANALYSIS OF S-RDSM SYSTEMS

In this section, the ABEP bound of the proposed S-RDSM scheme is derived. In the proposed S-RDSM system, non-coherent detection relies on the previous estimated data, which may result in error propagation. As a result, the analysis of error propagation is pivotal in the derivation of the ABEP bound of the proposed S-RDSM system. Assuming that the transmit signals of the i -th, the $(i+1)$ -st and the $(i+2)$ -nd block are \mathbf{x}_i^t and \mathbf{x}_{i+1}^m , \mathbf{x}_{i+2}^p , respectively, and the corresponding estimated signals are $\hat{\mathbf{x}}_i^j$, $\hat{\mathbf{x}}_{i+1}^k$ and $\hat{\mathbf{x}}_{i+2}^r$, respectively, the error propagation and the ABEP analysis are discussed in the sequel.

A. Error propagation analysis

Assuming that no error occurs at the i -th and $(i+1)$ -st block, according to (30), the reference signals $\hat{\mathbf{Y}}_i^t$ and $\hat{\mathbf{Y}}_{i+1}^m$ having no error are formulated as

$$\begin{aligned} \hat{\mathbf{Y}}_i^t &= \mathbf{Y}_i \mathbf{E}^{(1-\alpha)} + \hat{\mathbf{Y}}_{i-1} \mathbf{X}_i^t \mathbf{E}^\alpha \\ &= \mathbf{H} \mathbf{S}_{i-1} (\mathbf{X}_i^t \mathbf{E}_1 \mathbf{E}^{(1-\alpha)} + \mathbf{X}_i^t \mathbf{E}^\alpha) + \mathbf{N}_i \mathbf{E}^{(1-\alpha)} + \hat{\mathbf{N}}_{i-1} \mathbf{X}_i^t \mathbf{E}^\alpha \\ &= \mathbf{H} \mathbf{S}_i + \mathbf{N}_i \mathbf{E}^{(1-\alpha)} + \hat{\mathbf{N}}_{i-1} \mathbf{X}_i^t \mathbf{E}^\alpha, \\ \hat{\mathbf{Y}}_{i+1}^m &= \mathbf{Y}_{i+1} \mathbf{E}^{(1-\alpha)} + \hat{\mathbf{Y}}_i^t \mathbf{X}_{i+1}^m \mathbf{E}^\alpha \\ &= \mathbf{Y}_{i+1} \mathbf{E}^{(1-\alpha)} + \mathbf{Y}_i \mathbf{E}^{(1-\alpha)} \mathbf{X}_{i+1}^m \mathbf{E}^\alpha + \hat{\mathbf{Y}}_{i-1} \mathbf{X}_i^t \mathbf{E}^\alpha \mathbf{X}_{i+1}^m \mathbf{E}^\alpha \\ &= \mathbf{H} \mathbf{S}_{i+1} + \mathbf{N}_{i+1} \mathbf{E}^{(1-\alpha)} + \hat{\mathbf{N}}_i \mathbf{X}_{i+1}^m \mathbf{E}^\alpha. \end{aligned} \quad (47)$$

When an error occurs during a block signal's detection, the reference signals will include erroneous results, which results in error propagation. Assuming that the erroneous result starts at the i -th block, the error propagations of both the $(i+1)$ -st block, the $(i+2)$ -nd blocks are analyzed.

1) Error occurs at the i -th block:

Assuming that the error occurs at the i -th block and the transmit signal \mathbf{x}_i^t is estimated to $\hat{\mathbf{x}}_i^j$, the probability $P(\mathbf{x}_i^t \rightarrow \hat{\mathbf{x}}_i^j)$ of this pairwise error probability (PEP) event can be calculated as

$$\begin{aligned} P(\mathbf{x}_i^t \rightarrow \hat{\mathbf{x}}_i^j | \mathbf{H}) &= P(\|\mathbf{Y}_i - \hat{\mathbf{Y}}_{i-1} \mathbf{X}_i^t \mathbf{E}_1\|^2 \geq \|\mathbf{Y}_i - \hat{\mathbf{Y}}_{i-1} \mathbf{X}_i^j \mathbf{E}_1\|^2) \\ &= P(\|\mathbf{N}_i - \bar{\mathbf{N}}_i^t\|^2 \geq \|\mathbf{H} \mathbf{S}_{i-1} (\mathbf{X}_i^t \mathbf{E}_1 - \mathbf{X}_i^j \mathbf{E}_1) + \mathbf{N}_i - \bar{\mathbf{N}}_i^j\|^2) \\ &\approx Q \left(\sqrt{\frac{\|\mathbf{H} \mathbf{S}_{i-1} (\mathbf{X}_i^t \mathbf{E}_1 - \mathbf{X}_i^j \mathbf{E}_1)\|^2}{2(\sigma^2 + \sigma_i^2)}} \right), \end{aligned} \quad (48)$$

where we have $\bar{\mathbf{N}}_i^t = \hat{\mathbf{N}}_{i-1} \mathbf{X}_i^t \mathbf{E}_1$, $\bar{\mathbf{N}}_i^j = \hat{\mathbf{N}}_{i-1} \mathbf{X}_i^j \mathbf{E}_1$ and σ_i^2 can be obtained via (36). According to (86)-(94) of Appendix C, we have

$$\begin{aligned} P(\mathbf{x}_i^t \rightarrow \hat{\mathbf{x}}_i^j) &= \left[\frac{1}{2} - \frac{1}{2} \sqrt{\frac{c_1}{1+c_1}} \right] \sum_{k=0}^{TN_r-1} \binom{TN_r-1+k}{k} \left[\frac{1}{2} + \frac{1}{2} \sqrt{\frac{c_1}{1+c_1}} \right]^k, \end{aligned} \quad (49)$$

where $c_1 = |\mathbf{x}_i^t - \hat{\mathbf{x}}_i^j|^2 / 4(\sigma^2 + \sigma_i^2)$.

Since the i -th block's signal \mathbf{X}_i^t is estimated as $\hat{\mathbf{X}}_i^j$, the reference signals $\hat{\mathbf{Y}}_i^j$ having one error can be expressed as

$$\begin{aligned} \hat{\mathbf{Y}}_i^j &= \mathbf{Y}_i \mathbf{E}^{(1-\alpha)} + \hat{\mathbf{Y}}_{i-1} \hat{\mathbf{X}}_i^j \mathbf{E}^\alpha \\ &= (\mathbf{H} \mathbf{S}_i \mathbf{E}_1 + \mathbf{N}_i) \mathbf{E}^{(1-\alpha)} + (\mathbf{H} \mathbf{S}_{i-1} + \hat{\mathbf{N}}_{i-1}) \hat{\mathbf{X}}_i^j \mathbf{E}^\alpha \\ &= \mathbf{H} \mathbf{S}_{i-1} (\mathbf{X}_i^t \mathbf{E}_1 \mathbf{E}^{(1-\alpha)} + \mathbf{X}_i^j \mathbf{E}^\alpha) + \mathbf{N}_i \mathbf{E}^{(1-\alpha)} + \hat{\mathbf{N}}_{i-1} \mathbf{X}_i^j \mathbf{E}^\alpha. \end{aligned} \quad (50)$$

This erroneous reference signals will give rise to error propagations during the ensuing signal detection.

2) *Error propagation of the $(i+1)$ -st block:*

For the detection of the $(i+1)$ -st block with no error propagation, there are only two outcomes: 1) \mathbf{x}_{i+1}^m has the smallest Euclidean distance (ED); 2) \mathbf{x}_{i+1}^u ($\mathbf{x}_{i+1}^u \neq \mathbf{x}_{i+1}^m$) has the smallest ED. When the erroneous reference signals are employed for signal detection, there are also two error events corresponding to these two outcomes. However, at high SNRs, both $P(\mathbf{x}_i^t \rightarrow \mathbf{x}_i^j) \ll 10^{-3}$ and $P(\mathbf{x}_{i+1}^m \rightarrow \mathbf{x}_{i+1}^u) \ll 10^{-3}$ hold true, hence we have $P(\mathbf{x}_{i+1}^m \rightarrow \mathbf{x}_{i+1}^m) \approx 1$ and $\sum_{u \neq m}^{2^B} P(\mathbf{x}_{i+1}^m \rightarrow \mathbf{x}_{i+1}^u) \ll 10^{-3}$. Then it is readily seen that

$$\underbrace{P(\mathbf{x}_i^t \rightarrow \mathbf{x}_i^j) \sum_{u \neq m}^{2^B} P(\mathbf{x}_{i+1}^m \rightarrow \mathbf{x}_{i+1}^u)}_{\ll 10^{-3} P(\mathbf{x}_i^t \rightarrow \mathbf{x}_i^j)} \ll \underbrace{P(\mathbf{x}_i^t \rightarrow \mathbf{x}_i^j) P(\mathbf{x}_{i+1}^m \rightarrow \mathbf{x}_{i+1}^m)}_{P(\mathbf{x}_i^t \rightarrow \mathbf{x}_i^j)}. \quad (51)$$

Hence, the error propagation imposed by the events of $\mathbf{x}_i^j \neq \mathbf{x}_i^t$ and $\mathbf{x}_{i+1}^u \neq \mathbf{x}_{i+1}^m$ can be neglected. As a result, we mainly consider the error propagation caused by the events of $\mathbf{x}_i^j \neq \mathbf{x}_i^t$ and $\mathbf{x}_{i+1}^u = \mathbf{x}_{i+1}^m$. In this case, we originally have

$$\left\| \mathbf{Y}_{i+1} - \hat{\mathbf{Y}}_i^t \mathbf{X}_{i+1}^m \mathbf{E}_1 \right\|^2 = \min_{\mathbf{x}_{i+1} \in \mathcal{X}} \left\| \mathbf{Y}_{i+1} - \hat{\mathbf{Y}}_i^t \mathbf{X}_{i+1} \mathbf{E}_1 \right\|^2. \quad (52)$$

Since \mathbf{x}_i^t has been estimated to \mathbf{x}_i^j , the detection of the $(i+1)$ -st block signal becomes

$$\left\| \mathbf{Y}_{i+1} - \hat{\mathbf{Y}}_i^j \mathbf{X}_{i+1}^k \mathbf{E}_1 \right\|^2 = \min_{\mathbf{x}_{i+1} \in \mathcal{X}} \left\| \mathbf{Y}_{i+1} - \hat{\mathbf{Y}}_i^j \mathbf{X}_{i+1} \mathbf{E}_1 \right\|^2. \quad (53)$$

Specifically, the right parts of (52) and (53) can be further represented by

$$\begin{aligned} \mathbf{Y}_{i+1} - \hat{\mathbf{Y}}_i^t \mathbf{X}_{i+1} \mathbf{E}_1 &= \mathbf{H} \mathbf{S}_i \mathbf{X}_{i+1}^m \mathbf{E}_1 + \mathbf{N}_{i+1} \\ &- (\mathbf{H} \mathbf{S}_i + \mathbf{N}_i \mathbf{E}^{(1-\alpha)} + \hat{\mathbf{N}}_{i-1} \mathbf{X}_i^t \mathbf{E}^\alpha) \mathbf{X}_{i+1} \mathbf{E}_1 \\ &= \mathbf{H} \mathbf{S}_{i-1} (\mathbf{X}_i^t \mathbf{X}_{i+1}^m \mathbf{E}_1 - \mathbf{X}_i^t \mathbf{X}_{i+1} \mathbf{E}_1) + \\ &\underbrace{\mathbf{N}_{i+1} - \mathbf{N}_i \mathbf{E}^{(1-\alpha)} \mathbf{X}_{i+1} \mathbf{E}_1 - \hat{\mathbf{N}}_{i-1} \mathbf{X}_i^t \mathbf{E}^\alpha \mathbf{X}_{i+1} \mathbf{E}_1}_{\mathbf{N}_{i+1}(\alpha)}, \end{aligned} \quad (54)$$

and

$$\begin{aligned} \mathbf{Y}_{i+1} - \hat{\mathbf{Y}}_i^j \mathbf{X}_{i+1} \mathbf{E}_1 &= \mathbf{H} \mathbf{S}_i \mathbf{X}_{i+1} \mathbf{E}_1 + \mathbf{N}_{i+1} \\ &- (\mathbf{H} \mathbf{S}_i + \mathbf{N}_i \mathbf{E}^{(1-\alpha)} + \hat{\mathbf{N}}_{i-1} \mathbf{X}_i^j \mathbf{E}^\alpha) \mathbf{X}_{i+1} \mathbf{E}_1 \\ &= \mathbf{H} \mathbf{S}_{i-1} (\mathbf{X}_i^t \mathbf{X}_{i+1}^m \mathbf{E}_1 - \mathbf{X}_i^j \mathbf{X}_{i+1} \mathbf{E}_1) + \\ &\underbrace{\mathbf{N}_{i+1} - \mathbf{N}_i \mathbf{E}^{(1-\alpha)} \mathbf{X}_{i+1} \mathbf{E}_1 - \hat{\mathbf{N}}_{i-1} \mathbf{X}_i^j \mathbf{E}^\alpha \mathbf{X}_{i+1} \mathbf{E}_1}_{\mathbf{N}_{i+1,e}(\alpha)}. \end{aligned} \quad (55)$$

Since the variances of $\mathbf{N}_{i+1}(\alpha)$ and $\mathbf{N}_{i+1,e}(\alpha)$ are similar, we have

$$\left\| \mathbf{Y}_{i+1} - \hat{\mathbf{Y}}_i^t \mathbf{X}_{i+1}^m \mathbf{E}_1 \right\|^2 \approx \left\| \mathbf{Y}_{i+1} - \hat{\mathbf{Y}}_i^j \mathbf{X}_{i+1}^k \mathbf{E}_1 \right\|^2. \quad (56)$$

Then it becomes plausible that $\mathbf{X}_i^t \mathbf{X}_{i+1}^m = \mathbf{X}_i^j \mathbf{X}_{i+1}^k$ and $\mathbf{X}_i^t \mathbf{x}_{i+1}^m = \mathbf{X}_i^j \mathbf{x}_{i+1}^k$. Specifically, assuming that (q_i^t, s_i^t) , (q_{i+1}^m, s_{i+1}^m) , (q_i^j, s_i^j) and (q_{i+1}^k, s_{i+1}^k) are the SPM index and APM symbol of \mathbf{x}_i^t , \mathbf{x}_{i+1}^m , \mathbf{x}_i^j and \mathbf{x}_{i+1}^k , both the index $l_{i+1}^{j,k}$ and the symbol $s_{i+1}^{j,k}$ of $\mathbf{X}_i^j \mathbf{x}_{i+1}^k$, as well as the index $l_{i+1}^{t,m}$ and

the symbol $s_{i+1}^{t,m}$ of $\mathbf{X}_i^t \mathbf{x}_{i+1}^m$ can be obtained by

$$\begin{aligned} l_{i+1}^{j,k} &= \begin{cases} q_i^j + q_{i+1}^k - 1, & \text{if } q_i^j + q_{i+1}^k - 1 \leq N_t, \\ q_i^j + q_{i+1}^k - 1 - N_t, & \text{else,} \end{cases} \\ l_{i+1}^{t,m} &= \begin{cases} q_i^t + q_{i+1}^m - 1, & \text{if } q_i^t + q_{i+1}^m - 1 \leq N_t, \\ q_i^t + q_{i+1}^m - 1 - N_t, & \text{else,} \end{cases} \\ s_{i+1}^{j,k} &= s_i^j s_{i+1}^k, \quad s_{i+1}^{t,m} = s_i^t s_{i+1}^m. \end{aligned} \quad (57)$$

Since $\mathbf{X}_i^j \mathbf{x}_{i+1}^k = \mathbf{X}_i^t \mathbf{x}_{i+1}^m$ holds true, we have

$$l_{i+1}^{j,k} = l_{i+1}^{t,m}, \quad s_{i+1}^{j,k} = s_{i+1}^{t,m}. \quad (58)$$

As a result, for a given transmit signal \mathbf{x}_i^t and \mathbf{x}_{i+1}^m as well as for the erroneous result \mathbf{x}_i^j , the $(i+1)$ -st block signal will be definitely detected as

$$\begin{aligned} q_{i+1}^k &= \begin{cases} q_i^t + q_{i+1}^m - q_i^j + N_t, & \text{if } q_i^t + q_{i+1}^m - q_i^j < 1, \\ q_i^t + q_{i+1}^m - q_i^j, & \text{if } 1 \leq q_i^t + q_{i+1}^m - q_i^j \leq N_t, \\ q_i^t + q_{i+1}^m - q_i^j - N_t, & \text{if } N_t < q_i^t + q_{i+1}^m - q_i^j, \end{cases} \\ s_{i+1}^k &= s_i^{j*} s_{i+1}^m. \end{aligned} \quad (59)$$

3) *Error propagation of the $(i+2)$ -nd block:*

For the detection of the $(i+2)$ -nd block without error propagation, there are also two outcomes: 1) \mathbf{x}_{i+2}^p has the smallest ED; 2) \mathbf{x}_{i+1}^u ($\mathbf{x}_{i+2}^u \neq \mathbf{x}_{i+2}^p$) has the smallest ED, which results in different error propagations. Due to the same reason as illustrated in the $(i+1)$ -st block's analysis, the error propagation is mainly caused by the events of $\mathbf{x}_i^j \neq \mathbf{x}_i^t$, $\mathbf{x}_{i+1}^m \neq \mathbf{x}_{i+1}^k$ and $\mathbf{x}_{i+2}^u = \mathbf{x}_{i+2}^p$.

Based on the erroneous estimated results \mathbf{x}_i^j and \mathbf{x}_{i+1}^k , the reference signal used for the $(i+2)$ -nd block's detection is updated via (21) as

$$\begin{aligned} \hat{\mathbf{Y}}_{i+1}^k &= \mathbf{Y}_{i+1} \mathbf{E}^{(1-\alpha)} + \hat{\mathbf{Y}}_i^j \mathbf{X}_{i+1}^k \mathbf{E}^\alpha \\ &= \mathbf{Y}_{i+1} \mathbf{E}^{(1-\alpha)} + (\mathbf{Y}_i \mathbf{E}^{(1-\alpha)} + \hat{\mathbf{Y}}_{i-1} \mathbf{X}_i^j \mathbf{E}^\alpha) \mathbf{X}_{i+1}^k \mathbf{E}^\alpha \\ &= \mathbf{Y}_{i+1} \mathbf{E}^{(1-\alpha)} + \mathbf{Y}_i \mathbf{E}^{(1-\alpha)} \mathbf{X}_{i+1}^k \mathbf{E}^\alpha + \hat{\mathbf{Y}}_{i-1} \mathbf{X}_i^j \mathbf{E}^\alpha \mathbf{X}_{i+1}^k \mathbf{E}^\alpha \\ &= \mathbf{H} \mathbf{S}_{i+1} \mathbf{E}_1 \mathbf{E}^{(1-\alpha)} + \mathbf{H} \mathbf{S}_i \mathbf{E}_1 \mathbf{E}^{(1-\alpha)} \mathbf{X}_{i+1}^k \mathbf{E}^\alpha + \\ &\mathbf{H} \mathbf{S}_{i-1} \mathbf{X}_i^t \mathbf{E}^\alpha \mathbf{X}_{i+1}^m \mathbf{E}^\alpha + \mathbf{N}_{i+1} \mathbf{E}^{(1-\alpha)} + \mathbf{N}_i \mathbf{E}^{(1-\alpha)} \mathbf{X}_{i+1}^k \mathbf{E}^\alpha \\ &+ \hat{\mathbf{N}}_{i-1} \mathbf{X}_i^t \mathbf{E}^\alpha \mathbf{X}_{i+1}^m \mathbf{E}^\alpha \\ &= \mathbf{H} \mathbf{S}_i (\mathbf{X}_{i+1}^m \mathbf{E}_1 \mathbf{E}^{(1-\alpha)} + \mathbf{E}_1 \mathbf{E}^{(1-\alpha)} \mathbf{X}_{i+1}^k \mathbf{E}^\alpha + \mathbf{E}^\alpha \mathbf{X}_{i+1}^m \mathbf{E}^\alpha) \\ &= \mathbf{H} \mathbf{S}_{i+1} + \mathbf{H} \mathbf{S}_i [\mathbf{E}_1 \mathbf{E}^{(1-\alpha)} (\mathbf{X}_{i+1}^k \mathbf{E}^\alpha - \mathbf{X}_{i+1}^m \mathbf{E}^\alpha)] \\ &+ \mathbf{N}_{i+1} \mathbf{E}^{(1-\alpha)} + \mathbf{N}_i \mathbf{E}^{(1-\alpha)} \mathbf{X}_{i+1}^k \mathbf{E}^\alpha + \hat{\mathbf{N}}_{i-1} \mathbf{X}_i^t \mathbf{E}^\alpha \mathbf{X}_{i+1}^m \mathbf{E}^\alpha. \end{aligned} \quad (60)$$

For the case of $\alpha = 1$, we have $\mathbf{H} \mathbf{S}_i [\mathbf{E}_1 \mathbf{E}^{(1-\alpha)} (\mathbf{X}_{i+1}^k \mathbf{E}^\alpha - \mathbf{X}_{i+1}^m \mathbf{E}^\alpha)] = \mathbf{N}_i \mathbf{E}^{(1-\alpha)} \mathbf{X}_{i+1}^k \mathbf{E}^\alpha = \mathbf{O}_{N_r \times N_t}$. Then it becomes plausible that we have $\hat{\mathbf{Y}}_{i+1}^k = \hat{\mathbf{Y}}_{i+1}^m$ and hence no error is encountered by the reference signal $\hat{\mathbf{Y}}_{i+1}^k$.

For the case of $\alpha \neq 1$, since \mathbf{x}_i^t has been estimated to \mathbf{x}_i^j , \mathbf{x}_{i+1}^m has been estimated to \mathbf{x}_{i+1}^k , the detection of the $(i+2)$ -nd block becomes

$$\left\| \mathbf{Y}_{i+2} - \hat{\mathbf{Y}}_{i+1}^k \mathbf{X}_{i+2}^p \mathbf{E}_1 \right\|^2 = \min_{\mathbf{x}_{i+2} \in \mathcal{X}} \left\| \mathbf{Y}_{i+2} - \hat{\mathbf{Y}}_{i+1}^k \mathbf{X}_{i+2} \mathbf{E}_1 \right\|^2. \quad (61)$$

Specifically, the values of $\mathbf{Y}_{i+2} - \hat{\mathbf{Y}}_{i+1}^k \mathbf{X}_{i+2}^p \mathbf{E}_1$ and

$\mathbf{Y}_{i+2} - \hat{\mathbf{Y}}_{i+1}^k \mathbf{X}_{i+2}^r \mathbf{E}_1$ can be further represented by

$$\begin{aligned} & \mathbf{Y}_{i+2} - \hat{\mathbf{Y}}_{i+1}^k \mathbf{X}_{i+2}^p \mathbf{E}_1 \\ &= \mathbf{H} \mathbf{S}_i [\mathbf{E}_1 \mathbf{E}^{(1-\alpha)} (\mathbf{X}_{i+1}^k \mathbf{E}^\alpha - \mathbf{X}_{i+1}^m \mathbf{E}^\alpha)] \mathbf{X}_{i+2}^p \mathbf{E}_1 + \mathbf{N}_{i+2} \\ & \quad - \underbrace{(\mathbf{N}_{i+1} \mathbf{E}^{(1-\alpha)} + \mathbf{N}_i \mathbf{E}^{(1-\alpha)} \mathbf{X}_{i+1}^k \mathbf{E}^\alpha + \hat{\mathbf{N}}_{i-1} \mathbf{X}_i^t \mathbf{E}^\alpha \mathbf{X}_{i+1}^m \mathbf{E}^\alpha) \mathbf{X}_{i+2}^p \mathbf{E}_1}_{\mathbf{N}_{i+2,e,p}(\alpha)}. \end{aligned} \quad (62)$$

$$\begin{aligned} & \mathbf{Y}_{i+2} - \hat{\mathbf{Y}}_{i+1}^k \mathbf{X}_{i+2}^r \mathbf{E}_1 \\ &= \mathbf{H} \mathbf{S}_{i+2} \mathbf{E}_1 - \mathbf{H} \mathbf{S}_{i+1} \mathbf{X}_{i+2}^r \mathbf{E}_1 - \\ & \quad \mathbf{H} \mathbf{S}_i [\mathbf{E}_1 \mathbf{E}^{(1-\alpha)} (\mathbf{X}_{i+1}^k \mathbf{E}^\alpha - \mathbf{X}_{i+1}^m \mathbf{E}^\alpha)] \mathbf{X}_{i+2}^r \mathbf{E}_1 + \mathbf{N}_{i+2} \\ & \quad - \underbrace{(\mathbf{N}_{i+1} \mathbf{E}^{(1-\alpha)} + \mathbf{N}_i \mathbf{E}^{(1-\alpha)} \mathbf{X}_{i+1}^k \mathbf{E}^\alpha + \hat{\mathbf{N}}_{i-1} \mathbf{X}_i^t \mathbf{E}^\alpha \mathbf{X}_{i+1}^m \mathbf{E}^\alpha) \mathbf{X}_{i+2}^r \mathbf{E}_1}_{\mathbf{N}_{i+2,e,r}(\alpha)}. \end{aligned} \quad (63)$$

Since the variances of $\mathbf{N}_{i+2,e,p}(\alpha)$ and $\mathbf{N}_{i+2,e,r}(\alpha)$ are close to σ_{i+2}^2 , and the average variances of $\mathbf{N}_{i+2,r}^{\text{error}}$ and $\mathbf{N}_{i+2,p}^{\text{error}}$ are similar, we have

$$\begin{aligned} & P(\mathbf{X}_{i+2}^p \rightarrow \mathbf{X}_{i+2}^r | \hat{\mathbf{Y}}_{i+1}^k) \\ &= P(\|\mathbf{Y}_{i+2} - \hat{\mathbf{Y}}_{i+1}^k \mathbf{X}_{i+2}^r \mathbf{E}_1\|^2) \leq P(\|\mathbf{Y}_{i+2} - \hat{\mathbf{Y}}_{i+1}^k \mathbf{X}_{i+2}^p \mathbf{E}_1\|^2) \\ &= Q\left(\sqrt{\frac{\|\mathbf{H} \mathbf{S}_{i+1} (\mathbf{X}_{i+2}^p \mathbf{E}_1 - \mathbf{X}_{i+2}^r \mathbf{E}_1)\|^2}{2(\sigma^2 + V(\mathbf{N}_{i+2}^{\text{error}}) + \sigma_{i+2}^2)}}\right), \end{aligned} \quad (64)$$

where $V(\mathbf{N}_{i+2}^{\text{error}})$ denotes the variance of $\mathbf{N}_{i+2,r}^{\text{error}}$ or $\mathbf{N}_{i+2,p}^{\text{error}}$. Upon considering $T = 1$ and $M = 4$ for example, the value of $V(\mathbf{N}_{i+2}^{\text{error}})$ belongs to the set

$$V(\mathbf{N}_{i+2}^{\text{error}}) = (0, \alpha_1^2 \alpha^2, \alpha_1^2, 2\alpha_1^2 \alpha^2, 2\alpha_1^2, 4\alpha_1^2 \alpha^2, 4\alpha_1^2), \quad (65)$$

with $\alpha_1 = 1 - \alpha$. The corresponding probabilities can be obtained via (98) of Appendix D as

$$\begin{aligned} P(V(\mathbf{N}_{i+2}^{\text{error}}) = \alpha_1^2 \alpha^2) &= \frac{(N_t - 1) 2M^2}{(N_t M)^3 N_t}, \\ P(V(\mathbf{N}_{i+2}^{\text{error}}) = \alpha_1^2) &= \frac{(N_t - 1)^2 2M^2}{(N_t M)^3 N_t}, \\ P(V(\mathbf{N}_{i+2}^{\text{error}}) = 2\alpha_1^2 \alpha^2) &= \frac{2M^2}{(N_t M)^3}, \\ P(V(\mathbf{N}_{i+2}^{\text{error}}) = 2\alpha_1^2) &= \frac{2(N_t - 1)M^2}{(N_t M)^3}, \\ P(V(\mathbf{N}_{i+2}^{\text{error}}) = 4\alpha_1^2 \alpha^2) &= \frac{M^2}{(N_t M)^3}, \\ P(V(\mathbf{N}_{i+2}^{\text{error}}) = 4\alpha_1^2) &= \frac{(N_t - 1)M^2}{(N_t M)^3}. \end{aligned} \quad (66)$$

The average value of $V(\mathbf{N}_{i+2}^{\text{error}})$ can be obtained by (65)-(66). To provide further insights, Table III presents the value of $V(\mathbf{N}_{i+2}^{\text{error}})$ for $T = 1$ and $M = 4$. Observe from Table III that the probability of $V(\mathbf{N}_{i+2}^{\text{error}}) = \alpha_1^2 = (1 - \alpha)^2$ is the largest and decreases as N_t increases. Since α is chosen close to 1, the value of $V(\mathbf{N}_{i+2}^{\text{error}}) = \alpha_1^2$ is small. To be more specific, Table IV presents the average value of $V(\mathbf{N}_{i+2}^{\text{error}})$ for $\alpha = 0.8$ and $M = 4$. Observe from Table IV that the average value of $V(\mathbf{N}_{i+2}^{\text{error}})$ is close to zero in the context of large-scale MIMO setups. As a result, we have

$$\begin{aligned} & P(\mathbf{X}_i^t \rightarrow \mathbf{X}_i^j, \mathbf{X}_{i+2}^p \rightarrow \mathbf{X}_{i+2}^r) \\ &= P(\mathbf{X}_{i+2}^p \rightarrow \mathbf{X}_{i+2}^r | \hat{\mathbf{Y}}_{i+1}^k) P(\mathbf{X}_i^t \rightarrow \mathbf{X}_i^j) \\ &\approx P(\mathbf{X}_{i+2}^p \rightarrow \mathbf{X}_{i+2}^r) P(\mathbf{X}_i^t \rightarrow \mathbf{X}_i^j). \end{aligned} \quad (67)$$

Since $P(\mathbf{X}_{i+2}^p \rightarrow \mathbf{X}_{i+2}^r) \ll 10^{-3}$ at high SNRs, we have $P(\mathbf{X}_i^t \rightarrow \mathbf{X}_i^j, \mathbf{X}_{i+2}^p \rightarrow \mathbf{X}_{i+2}^r) \ll P(\mathbf{X}_i^t \rightarrow \mathbf{X}_i^j)$. It can

TABLE III
THE VALUE OF $V(\mathbf{N}_{i+2}^{\text{error}})$ OF (65) AND ITS CORRESPONDING PROBABILITY OF (66) FOR $T = 1$ AND $M = 4$.

$\frac{P(V(\mathbf{N}_{i+2}^{\text{error}}))}{N_t}$	0	$\alpha_1^2 \alpha^2$	α_1^2	$2\alpha_1^2 \alpha^2$	$2\alpha_1^2$	$4\alpha_1^2 \alpha^2$	$4\alpha_1^2$
$N_t = 4$	0.58	0.09	0.28	7e-3	0.02	4e-3	0.01
$N_t = 8$	0.77	0.03	0.19	1e-2	7e-3	5e-4	3e-3
$N_t = 16$	0.88	7e-3	0.11	1e-4	2e-3	1e-4	9e-4

TABLE IV
AVERAGE VALUE OF $V(\mathbf{N}_{i+2}^{\text{error}})$ FOR $\alpha = 0.8$ AND $M = 4$ CALCULATING BY (65) AND (66).

T	N_t	4	8	16
$T = 1$		0.0182	0.0095	0.0049
$T = 2$		0.0328	0.0182	0.0095

be concluded that the error propagation of the $(i + 2)$ -nd block can be neglected. Hence the error propagation is mainly encountered by a pair of adjacent blocks.

B. Approximate upper bound of the S-RDSM system

According to our error propagation analysis, the ABEP of the $(i + 1)$ -st block is mainly associated with the i -th block's signal detection, which has two estimated results at the i -th block: 1) $\mathbf{x}_i^j = \mathbf{x}_i^t$; 2) $\mathbf{x}_i^j \neq \mathbf{x}_i^t$. Hence the approximate upper ABEP bound of the $(i + 1)$ -st block can be formulated as:

$$\begin{aligned} P_{b,i+1}^U &= \frac{1}{2^B} \sum_{t=1}^{2^B} \underbrace{P_{b,\mathbf{x}_i^t \rightarrow \mathbf{x}_i^t}^{i+1} P(\mathbf{x}_i^t \rightarrow \mathbf{x}_i^t)}_{\text{ABEP of } \mathbf{x}_i^j = \mathbf{x}_i^t} + \\ & \quad \frac{1}{2^B} \sum_{t=1}^{2^B} \sum_{j=1}^{2^B} \underbrace{P_{b,\mathbf{x}_i^t \rightarrow \mathbf{x}_i^j}^{i+1} P(\mathbf{x}_i^t \rightarrow \mathbf{x}_i^j)}_{\text{ABEP of } \mathbf{x}_i^j \neq \mathbf{x}_i^t}, \end{aligned} \quad (68)$$

where $P_{b,\mathbf{x}_i^t \rightarrow \mathbf{x}_i^t}^{i+1}$ and $P_{b,\mathbf{x}_i^t \rightarrow \mathbf{x}_i^j}^{i+1}$ denote the ABEP of the conditions of $\mathbf{x}_i^j = \mathbf{x}_i^t$ and $\mathbf{x}_i^j \neq \mathbf{x}_i^t$ at the $(i + 1)$ -st block, respectively.

1) Calculation of $P(\mathbf{x}_i^t \rightarrow \mathbf{x}_i^t)$:

The probability of $P(\mathbf{x}_i^t \rightarrow \mathbf{x}_i^t)$ is obtained by

$$P(\mathbf{x}_i^t \rightarrow \mathbf{x}_i^t) = \begin{cases} 0, & \text{if } \sum_{j=1, j \neq t}^{2^B} P(\mathbf{x}_i^t \rightarrow \mathbf{x}_i^j) > 1, \\ 1 - \sum_{j=1, j \neq t}^{2^B} P(\mathbf{x}_i^t \rightarrow \mathbf{x}_i^j), & \text{else.} \end{cases} \quad (69)$$

2) Calculation of $P_{b,\mathbf{x}_i^t \rightarrow \mathbf{x}_i^j}^{i+1}$:

In this case, we can ignore the effect of error propagation, hence the value of $P_{b,\mathbf{x}_i^t \rightarrow \mathbf{x}_i^j}^{i+1}$ can be formulated as:

$$P_{b,\mathbf{x}_i^t \rightarrow \mathbf{x}_i^j}^{i+1} = \frac{1}{B 2^B} \sum_{t=1}^{2^B} \sum_{j=1, j \neq t}^{2^B} P(\mathbf{x}_{i+1}^m \rightarrow \mathbf{x}_{i+1}^k) d(\mathbf{x}_{i+1}^m, \mathbf{x}_{i+1}^k), \quad (70)$$

where $d(\mathbf{x}_{i+1}^m, \mathbf{x}_{i+1}^k)$ is the Hamming distance (HD) of \mathbf{x}_{i+1}^m and \mathbf{x}_{i+1}^k , $P(\mathbf{x}_{i+1}^m \rightarrow \mathbf{x}_{i+1}^k)$ denotes the PEP, which can be obtained by (49) with $c_1 = |\mathbf{x}_{i+1}^m - \mathbf{x}_{i+1}^k|^2 / 4(\sigma^2 + \sigma_{i+1}^2)$.

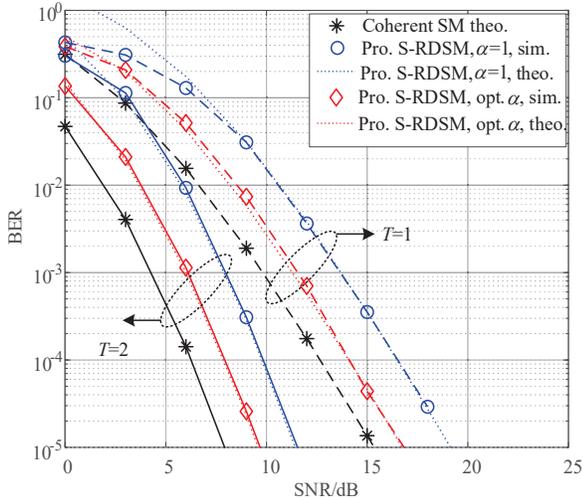


Fig. 5. Theoretical ABEP results of the proposed S-RDSM system with $N_t = 4, N_r = 4, W = 256$, which are calculated by (73).

3) Calculation of $P_{b, \mathbf{x}_i^t \rightarrow \mathbf{x}_i^j}^{i+1}$: For the case of the event of $\mathbf{x}_i^t \rightarrow \mathbf{x}_i^j$, the value of $P_{b, \mathbf{x}_i^t \rightarrow \mathbf{x}_i^j}^{i+1}$ can be formulated as:

$$P_{b, \mathbf{x}_i^t \rightarrow \mathbf{x}_i^j}^{i+1} = \frac{1}{B2^B} \sum_{m=1}^{2^B} \sum_{k=1}^{2^B} P(\mathbf{x}_{i+1}^m \rightarrow \mathbf{x}_{i+1}^k | \mathbf{x}_i^t \rightarrow \mathbf{x}_i^j) d(\mathbf{x}_{i+1}^m, \mathbf{x}_{i+1}^k), \quad (71)$$

where $P(\mathbf{x}_{i+1}^m \rightarrow \mathbf{x}_{i+1}^k | \mathbf{x}_i^t \rightarrow \mathbf{x}_i^j)$ is the PEP event under the condition of $\mathbf{x}_i^j \neq \mathbf{x}_i^t$, which is close to 1 based on our error propagation analysis and $d(\mathbf{x}_{i+1}^m, \mathbf{x}_{i+1}^k)$ can be obtained by (59). By substituting (69), (70), (71) and (49) into (68), we can derive the approximate upper bound of the ABEP as

$$P_{b, i+1}^U = \frac{1}{2^B} \sum_{t=1}^{2^B} \left[\underbrace{P_{b, \mathbf{x}_i^t \rightarrow \mathbf{x}_i^t}^{i+1}}_{\text{by (70)}} \underbrace{P(\mathbf{x}_i^t \rightarrow \mathbf{x}_i^t)}_{\text{by (69)}} + \sum_{j \neq t} \underbrace{P(\mathbf{x}_i^t \rightarrow \mathbf{x}_i^j)}_{\text{by (49)}} \right] \cdot \left[\frac{1}{B2^B} \sum_{m=1}^{2^B} \underbrace{d[(q_{i+1}^k, s_{i+1}^k), (q_{i+1}^m, s_{i+1}^m)]}_{\text{by (59)}} \right]. \quad (72)$$

As a result, the average approximate upper ABEP bound for W/T blocks is obtained by

$$P_b^U = \frac{\sum_{i=1}^{W/T} P_{b, i+1}^U}{W/T}. \quad (73)$$

VI. SIMULATION RESULTS

In this section, the performance of the proposed S-RDSM, of the existing RDSM and of coherent SM systems are compared in flat Rayleigh fading channels. In all the comparisons, the HL-ML detector is employed for the proposed S-RDSM system and the ML detector is used for the other systems.

Figs. 5-8 portrays the theoretical ABEP upper bounds of our proposed S-RDSM systems in conjunction with different

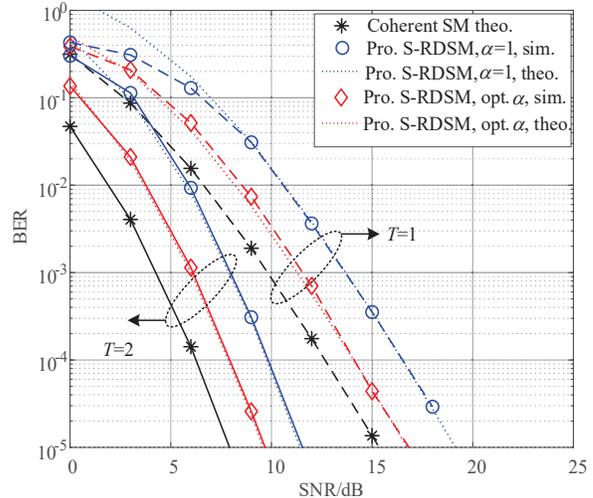


Fig. 6. Theoretical ABEP results of the proposed S-RDSM system with $N_t = 8, N_r = 4, W = 256$, which are calculated by (73).

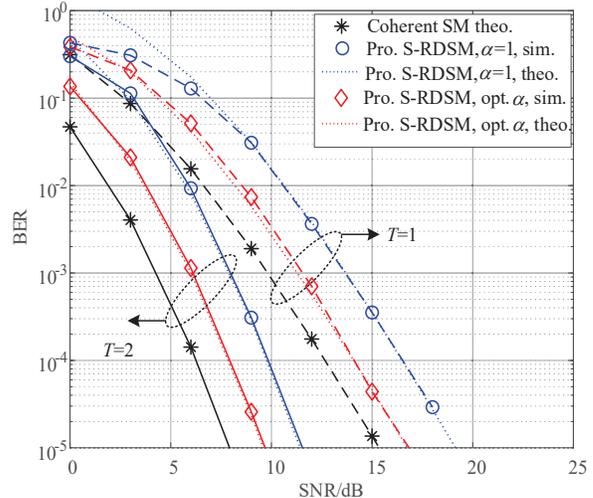


Fig. 7. Theoretical ABEP results of the proposed S-RDSM system with $N_t = 16, N_r = 4, W = 256$, which are calculated by (73).

number of TAs $N_t = 4, 8, 16, 64$ at $W = 256, N_r = 4$. The factors of $\alpha = 0.88, 0.84, 0.78, 0.71, 0.62$ are used for $N_t = 4, 8, 16, 64$ at $T = 1$, while $\alpha = 0.87, 0.81, 0.73, 0.64, 0.55$ are used for $N_t = 4, 8, 16, 64$ at $T = 2$. Moreover, the theoretical coherent SM results having an identical throughput are added as benchmarks.

Observe from Figs. 5-8 that the upper bounds derived for the proposed S-RDSM systems operating both with and without a forgetting factor become very tight upon increasing the SNR values. Furthermore, the proposed S-RDSM system using the optimized forgetting factor is capable of providing significant performance gains over its counterpart operating without a forgetting factor. As the ratio N_t/W decreases, the performance advantage of using the optimal factor becomes more dominant.

Figs. 9 and 10 compare the performance of the proposed S-RDSM system to that of the existing RDSM systems and coherent SM systems at 8 bpcu and 9 bpcu, respectively. Specifically, in Fig. 8, $N_t = 64, N_r = 4, T = 1, L = 4$

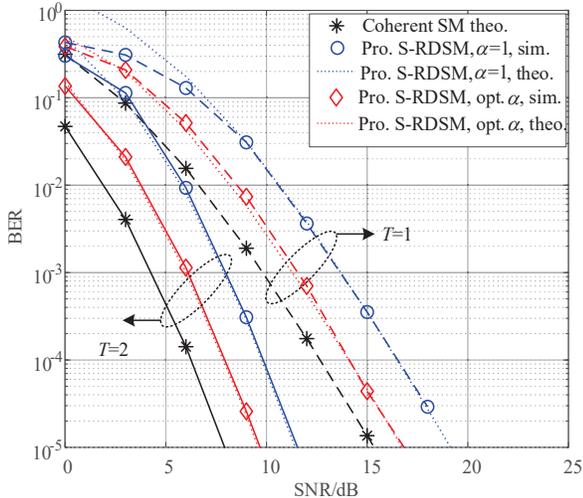


Fig. 8. Theoretical ABEP results of the proposed S-RDSM system with $N_t = 64, N_r = 4, W = 256$, which are calculated by (73).

and $W = 256, 512, 1024, 2048$ are used for the proposed S-RDSM system, while $N_t = 64, N_r = 4, T = 1, L = 4$ and $W = 256$ are used in the conventional RDSM [34] systems; $\alpha = 0.62, 0.7, 0.78, 0.84$ are employed for $W = 256, 512, 1024, 2048$; $N_t = 64, N_r = 4, L = 4$ are employed for the coherent SM system. In Fig. 9, $N_t = 128, N_r = 4, T = 1, L = 4$ are employed for both the proposed S-RDSM and the conventional RDSM systems, while $N_t = 128, N_r = 4, L = 4$ are employed for its coherent SM counterpart. $\alpha = 0.7$ and $\alpha = 0.8$ are selected for $W = 1024$ and 2048 , respectively. Furthermore, the coherent SM relying on realistic estimated channel has been added for comparison. According to [37], the estimated channel matrix can be expressed as

$$\hat{\mathbf{H}} = \mathbf{H} + \mathbf{H}_e, \quad (74)$$

where $\mathbf{H}_e \in \mathbb{C}^{N_r \times N_t}$, whose elements obey the complex Gaussian distributions $\mathcal{CN}(0, \sigma_e^2)$.

It is observed from Figs. 9-10 that the performance of the proposed S-RDSM system is comparable to that of conventional RDSM systems, despite its simplified structure. As predicted in Fig. 3(b), the performance of the proposed S-RDSM system improves, as the value of W increases. This is because as W increases, the average noise variance imposed by the non-coherent detector is reduced. Furthermore, it is also seen from Figs. 9-10 that the proposed S-RDSM system is capable of providing a 1.5-dB performance gain over the coherent SM with $\sigma_e^2 = \sigma^2$.

VII. CONCLUSIONS

Simplified rectangular differential spatial modulation was conceived for massive MIMO downlink communication without the knowledge of CSI. The proposed S-RDSM system retains all the advantages of its coherent SM counterpart, including its high transmit rate, low-complexity bit-to-symbol mapping and signal detection. The approximate ABEP upper bounds of the proposed S-RDSM system have been derived and validated by the simulation results. Furthermore, we proposed a novel forgetting factor optimization method relying

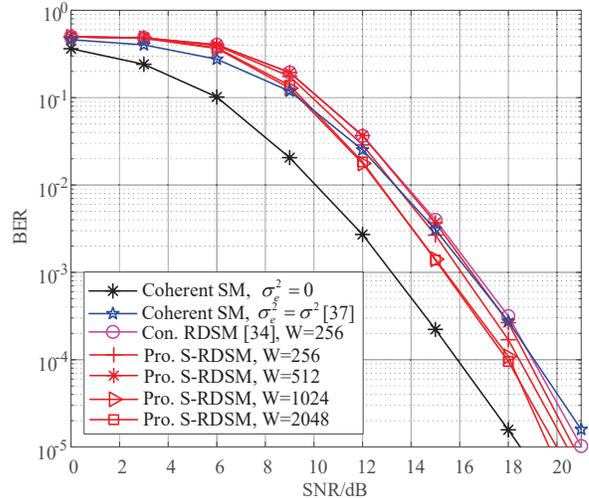


Fig. 9. Performance comparisons of the proposed S-RDSM systems to the existing RDSM and coherent SM systems having $N_t = 64$ at 8 bpcu.

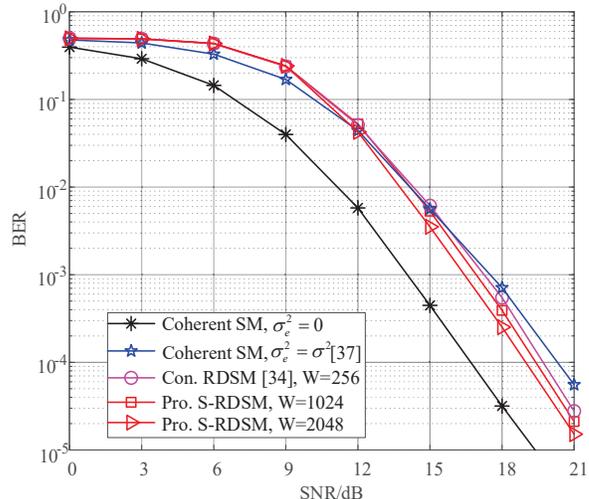


Fig. 10. Performance comparisons of the proposed S-RDSM systems to the existing RDSM and coherent SM systems having $N_t = 128$ at 9 bpcu.

on a closed-form expression instead of simulations, as seen in the literature. Both the simulation and theoretical results have demonstrated that the proposed S-RDSM system using the optimized forgetting factor is capable of approaching the performance of its coherent counterpart, but without CSI knowledge, which is very attractive for future massive MIMO communications.

APPENDIX A PROOF OF HL-ML DETECTOR

According to (23), (25) and (26), it becomes plausible that the main difference between the ML detector and the HL-ML detector is the way of obtaining \hat{s}_i . If we can prove that the

following expression

$$\begin{aligned} \tilde{s}_q &= \arg \min_{\forall s} \left[\sum_{\tau=1}^T \left\| \hat{\mathbf{Y}}_{i-1}^{(q+\tau-1)} \right\|^2 - 2\Re \left(\sum_{\tau=1}^T (\mathbf{Y}_i^\tau)^H \hat{\mathbf{Y}}_{i-1}^{(q+\tau-1)} s \right) \right] \\ &= \arg \max_{\forall s} \underbrace{\Re \left(\sum_{\tau=1}^T (\mathbf{Y}_i^\tau)^H \hat{\mathbf{Y}}_{i-1}^{(q+\tau-1)} \right) / \sum_{\tau=1}^T \left\| \hat{\mathbf{Y}}_{i-1}^{(q+\tau-1)} \right\|^2}_{a+jb} s, \end{aligned} \quad (75)$$

holds true, then we can conclude that the outputs of the HL-ML detector and the ML detector are identical.

Specifically, assuming $\tilde{s}_q = \mathbb{D}(a - jb) = c + jd$, $s = c_1 + jd_1 \neq \tilde{s}_q$ and $c^2 + d^2 = c_1^2 + d_1^2 = 1$, we have

$$|a - jb - \tilde{s}_q|^2 < |a - jb - s|^2, \quad (76)$$

hence we obtain

$$\begin{aligned} (a - c)^2 + (b + d)^2 &< (a - c_1)^2 + (b + d_1)^2 \\ \Downarrow \\ -2ac + c^2 + 2bd + d^2 &< -2ac_1 + c_1^2 + 2bd + d_1^2 \\ \Downarrow \\ ac - bd &> ac_1 - bd_1 \\ \Downarrow \\ \Re[(a + jb)(c + jd)] &> \Re[(a + jb)(c_1 + jd_1)]. \end{aligned} \quad (77)$$

As a result, to satisfy (75), we should have

$$\begin{aligned} \tilde{s}_q &= \mathbb{D}(a - jb) \\ &= \mathbb{D} \left[\left(\sum_{\tau=1}^T (\mathbf{Y}_i^\tau)^H \hat{\mathbf{Y}}_{i-1}^{(q+\tau-1)} / \sum_{\tau=1}^T \left\| \hat{\mathbf{Y}}_{i-1}^{(q+\tau-1)} \right\|^2 \right)^* \right] \\ &= \mathbb{D} \left[\sum_{\tau=1}^T (\hat{\mathbf{Y}}_{i-1}^{(q+\tau-1)})^H \mathbf{Y}_i^\tau / \sum_{\tau=1}^T \left\| \hat{\mathbf{Y}}_{i-1}^{(q+\tau-1)} \right\|^2 \right]. \end{aligned} \quad (78)$$

According to the above analysis, when \tilde{s}_q is defined by (24), then (25) and (26) yield the same performance as the ML detector of (20).

APPENDIX B PROOF OF EQ. (30)

$$\begin{aligned} \hat{\mathbf{Y}}_{i-1} &= \mathbf{Y}_{i-1} \mathbf{E}^{(1-\alpha)} + \hat{\mathbf{Y}}_{i-2} \hat{\mathbf{X}}_{i-1} \mathbf{E}^\alpha \\ &= \mathbf{Y}_{i-1} \mathbf{E}^{(1-\alpha)} + (\mathbf{Y}_{i-2} \mathbf{E}^{(1-\alpha)} + \hat{\mathbf{Y}}_{i-3} \hat{\mathbf{X}}_{i-2} \mathbf{E}^\alpha) \hat{\mathbf{X}}_{i-1} \mathbf{E}^\alpha \\ &= \mathbf{Y}_{i-1} \mathbf{E}^{(1-\alpha)} + \mathbf{Y}_{i-2} \mathbf{E}^{(1-\alpha)} \hat{\mathbf{X}}_{i-1} \mathbf{E}^\alpha \\ &\quad + \mathbf{Y}_{i-3} \mathbf{E}^{(1-\alpha)} \hat{\mathbf{X}}_{i-2} \mathbf{E}^\alpha \hat{\mathbf{X}}_{i-1} \mathbf{E}^\alpha \\ &\quad + \dots + \mathbf{Y}_1 \mathbf{E}^{(1-\alpha)} \hat{\mathbf{X}}_2 \mathbf{E}^\alpha \dots \hat{\mathbf{X}}_{i-1} \mathbf{E}^\alpha + \hat{\mathbf{Y}}_0 \hat{\mathbf{X}}_1 \mathbf{E}^\alpha \dots \hat{\mathbf{X}}_{i-1} \mathbf{E}^\alpha \\ &= \mathbf{H}(\mathbf{S}_{i-1} \mathbf{E}_1 \mathbf{E}^{(1-\alpha)} + \mathbf{S}_{i-2} \mathbf{E}_1 \mathbf{E}^{(1-\alpha)} \hat{\mathbf{X}}_{i-1} \mathbf{E}^\alpha + \dots + \\ &\quad \mathbf{S}_1 \mathbf{E}_1 \mathbf{E}^{(1-\alpha)} \hat{\mathbf{X}}_2 \mathbf{E}^\alpha \dots \hat{\mathbf{X}}_{i-1} \mathbf{E}^\alpha + \hat{\mathbf{X}}_1 \mathbf{E}^\alpha \hat{\mathbf{X}}_2 \mathbf{E}^\alpha \dots \hat{\mathbf{X}}_{i-1} \mathbf{E}^\alpha) + \hat{\mathbf{N}}_{i-1} \end{aligned} \quad (79)$$

with

$$\begin{aligned} \hat{\mathbf{N}}_{i-1} &= \mathbf{N}_{i-1} \mathbf{E}^{(1-\alpha)} + \mathbf{N}_{i-2} \mathbf{E}^{(1-\alpha)} \hat{\mathbf{X}}_{i-1} \mathbf{E}^\alpha + \dots + \\ &\quad \mathbf{N}_1 \mathbf{E}^{(1-\alpha)} \hat{\mathbf{X}}_2 \mathbf{E}^\alpha \dots \hat{\mathbf{X}}_{i-1} \mathbf{E}^\alpha + \mathbf{N}_0 \hat{\mathbf{X}}_1 \mathbf{E}^\alpha \hat{\mathbf{X}}_2 \mathbf{E}^\alpha \dots \hat{\mathbf{X}}_{i-1} \mathbf{E}^\alpha \end{aligned} \quad (80)$$

Assuming that no error is encountered by the previous $(i - 1)$ blocks, since $\mathbf{E}_1 \mathbf{E}^{(1-\alpha)} + \mathbf{E}^\alpha = \mathbf{I}_{N_i}$, we have

$$\begin{aligned} &\mathbf{S}_{i-1} \mathbf{E}_1 \mathbf{E}^{(1-\alpha)} + \mathbf{S}_{i-2} \mathbf{E}_1 \mathbf{E}^{(1-\alpha)} \hat{\mathbf{X}}_{i-1} \mathbf{E}^\alpha + \dots + \\ &\mathbf{S}_1 \mathbf{E}_1 \mathbf{E}^{(1-\alpha)} \hat{\mathbf{X}}_2 \mathbf{E}^\alpha \dots \hat{\mathbf{X}}_{i-1} \mathbf{E}^\alpha + \hat{\mathbf{X}}_1 \mathbf{E}^\alpha \hat{\mathbf{X}}_2 \mathbf{E}^\alpha \dots \hat{\mathbf{X}}_{i-1} \mathbf{E}^\alpha \\ &= \mathbf{S}_{i-1} \mathbf{E}_1 \mathbf{E}^{(1-\alpha)} + \mathbf{S}_{i-2} \mathbf{E}_1 \mathbf{E}^{(1-\alpha)} \hat{\mathbf{X}}_{i-1} \mathbf{E}^\alpha + \dots + \\ &\mathbf{S}_1 \hat{\mathbf{X}}_2 \mathbf{E}^\alpha \dots \hat{\mathbf{X}}_{i-1} \mathbf{E}^\alpha \\ &= \mathbf{S}_{i-1} \mathbf{E}_1 \mathbf{E}^{(1-\alpha)} + \mathbf{S}_{i-2} \hat{\mathbf{X}}_{i-1} \mathbf{E}^\alpha \\ &= \mathbf{S}_{i-2} (\hat{\mathbf{X}}_{i-1} \mathbf{E}_1 \mathbf{E}^{(1-\alpha)} + \hat{\mathbf{X}}_{i-1} \mathbf{E}^\alpha) \\ &= \mathbf{S}_{i-2} \hat{\mathbf{X}}_{i-1} \\ &= \mathbf{S}_{i-1}. \end{aligned} \quad (81)$$

Hence, Eq. (71) can be simplified by (30).

APPENDIX C PROOF OF EQ. (49)

According to (48), the PEP event of $P(\mathbf{x}_i^t \rightarrow \mathbf{x}_i^j | \mathbf{H}_i)$ can be expressed as

$$\begin{aligned} &P(\mathbf{x}_i^t \rightarrow \mathbf{x}_i^j | \mathbf{H}_i) \\ &= P \left(\left\| \mathbf{Y}_i - \hat{\mathbf{Y}}_{i-1} \mathbf{X}_i^{t,[1:T]} \right\|^2 \geq \left\| \mathbf{Y}_i - \hat{\mathbf{Y}}_{i-1} \mathbf{X}_i^{j,[1:T]} \right\|^2 \right) \\ &= P \left(\underbrace{\left\| \mathbf{N}_i - \tilde{\mathbf{N}}_i^t \right\|^2}_{\Theta_{i,t}} \geq \underbrace{\left\| \mathbf{N}_i - \tilde{\mathbf{N}}_i^j - \mathbf{H}_i \mathbf{S}_{i-1} (\mathbf{X}_i^{t,[1:T]} - \mathbf{X}_i^{j,[1:T]}) \right\|^2}_{\Delta} \right) \\ &= P \left(\sum_{\tau=1}^T \left\| \Theta_{i,t}^\tau \right\|^2 \geq \sum_{\tau=1}^T \left\| \Theta_{i,j}^\tau - \Delta^\tau \right\|^2 \right), \end{aligned} \quad (82)$$

where $\tilde{\mathbf{N}}_i^t = \hat{\mathbf{N}}_{i-1} \mathbf{X}_i^{t,[1:T]}$, $\tilde{\mathbf{N}}_i^j = \hat{\mathbf{N}}_{i-1} \mathbf{X}_i^{j,[1:T]}$, Δ^τ , $\Theta_{i,t}^\tau$, $\Theta_{i,j}^\tau$ are the τ column of Δ , $\Theta_{i,t}$, $\Theta_{i,j}$. Since we have $\Theta_{i,t} \in \mathcal{CN}(0, \sigma^2 + \sigma_i^2)$, $\Theta_{i,j} \in \mathcal{CN}(0, \sigma^2 + \sigma_i^2)$ at high SNRs, (82) can be further represented by

$$P(\mathbf{x}_i^t \rightarrow \mathbf{x}_i^j | \mathbf{H}_i) = P \left(\sum_{\tau=1}^T 2\Re \left[(\Theta_{i,t}^\tau)^H \Delta^\tau \right] \geq \sum_{\tau=1}^T \|\Delta^\tau\|^2 \right), \quad (83)$$

with

$$2\Re \left[(\Theta_{i,t}^\tau)^H \Delta^\tau \right] \in \mathcal{N}(0, 2(\sigma^2 + \sigma_i^2)) \|\Delta^\tau\|^2. \quad (84)$$

Hence, we can obtain that

$$\sum_{\tau=1}^T 2\Re \left[(\Theta_{i,t}^\tau)^H \Delta^\tau \right] \in \mathcal{N}(0, 2(\sigma^2 + \sigma_i^2) \sum_{\tau=1}^T \|\Delta^\tau\|^2), \quad (85)$$

and $P(\mathbf{x}_i^t \rightarrow \mathbf{x}_i^j | \mathbf{H}_i)$ can be expressed as

$$P(\mathbf{x}_i^t \rightarrow \mathbf{x}_i^j | \mathbf{H}_i) \approx Q \left(\sqrt{\sum_{\tau=1}^T \|\Delta^\tau\|^2 / 2(\sigma^2 + \sigma_i^2)} \right), \quad (86)$$

where we have $Q(x) = \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \exp \left(-\frac{x^2}{2 \sin^2 \theta} \right) d\theta$. As a result, the PEP value can be expressed as

$$\begin{aligned} P(\mathbf{x}_i^t \rightarrow \mathbf{x}_i^j) &= \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \int_{\gamma} f_\gamma(\gamma) \exp \left(-\frac{\gamma}{4(\sigma^2 + \sigma_i^2) \sin^2 \theta} \right) d\theta \\ &= \frac{1}{\pi} \int_0^{\frac{\pi}{2}} M_\gamma \left(-\frac{1}{4(\sigma^2 + \sigma_i^2) \sin^2 \theta} \right) d\theta, \end{aligned} \quad (87)$$

where $\gamma = \left\| \mathbf{H}_i \mathbf{S}_{i-1} (\mathbf{x}_i^t - \mathbf{x}_i^j) \right\|^2 = \sum_{\tau=1}^T \|\Delta^\tau\|^2$ and

$$\begin{aligned} M_\gamma(s) &= \int e^{\gamma s} f_\gamma(\gamma) d\gamma = \int_{\Delta} e^{\sum_{\tau=1}^T \|\Delta^\tau\|^2 s} f_\Delta(\Delta) d\Delta \\ &= \int e^{\|\Delta^1\|^2 s} f_{\Delta^1}(\Delta^1) d\Delta^1 \dots \int e^{\|\Delta^T\|^2 s} f_{\Delta^T}(\Delta^T) d\Delta^T \\ &= \prod_{\tau=1}^T M_{\|\Delta^\tau\|^2 s}(s). \end{aligned} \quad (88)$$

Moreover, the value of $\|\Delta^\tau\|^2$ can be further expressed as

$$\begin{aligned}\gamma^\tau &= \|\Delta^\tau\|^2 = \left\| \mathbf{H}_i \mathbf{S}_{i-1} (\mathbf{X}_i^{t, [\tau]} - \mathbf{X}_i^{j, [\tau]}) \right\|^2 \\ &= \left\| \mathbf{h}_{l_i^{t, \tau}} s_{i-1}^t s_i^t - \mathbf{h}_{l_i^{j, \tau}} s_{i-1}^j s_i^j \right\|^2 \\ &= \underbrace{\sum_{r=1}^{N_r} \left| h_{l_i^{t, \tau}}^r s_{i-1}^t s_i^t - h_{l_i^{j, \tau}}^r s_{i-1}^j s_i^j \right|^2}_{\gamma_r^\tau}.\end{aligned}\quad (89)$$

According to (57), the index of the τ -th column of \mathbf{X}_i^t and \mathbf{X}_i^j can be expressed as

$$\begin{aligned}l_i^{t, \tau} &= \begin{cases} q_{i-1} + q_i^t + \tau - 2, & \text{if } q_{i-1} + q_i^t + \tau - 2 \leq N_t; \\ q_{i-1} + q_i^t + \tau - 2 - N_t, & \text{if } q_{i-1} + q_i^t + \tau - 2 > N_t, \end{cases} \\ l_i^{j, \tau} &= \begin{cases} q_{i-1} + q_i^j + \tau - 2, & \text{if } q_{i-1} + q_i^j + \tau - 2 \leq N_t; \\ q_{i-1} + q_i^j + \tau - 2 - N_t, & \text{if } q_{i-1} + q_i^j + \tau - 2 > N_t, \end{cases}\end{aligned}\quad (90)$$

where q_{i-1} and s_{i-1} are the SPM index and APM symbol of the first column of \mathbf{S}_{i-1} . Hence, γ_r^τ can be further expressed as

$$\gamma_r^\tau = \begin{cases} |s_{i-1}|^2 |s_i^t - s_i^j|^2 |h_{l_i^{t, \tau}}^r|^2, & \text{if } q_i^t = q_i^j, \\ |s_{i-1}|^2 |h_{l_i^{t, \tau}}^r s_i^t - h_{l_i^{j, \tau}}^r s_i^j|^2, & \text{if } q_i^t \neq q_i^j. \end{cases}\quad (91)$$

Since $|h_{l_i^{t, \tau}}^r|^2 \sim \text{EXP}(1)$ and $|h_{l_i^{t, \tau}}^r s_i^t - h_{l_i^{j, \tau}}^r s_i^j|^2 \sim \chi^2(2)$ follow the exponentially and chi-square distributions, respectively, while the MGF of γ_r^τ is given by

$$\begin{aligned}M_{\gamma_r^\tau}(s) &= \begin{cases} (1 - |s_i^t - s_i^j|^2 s)^{-1}, & \text{if } q_i^t = q_i^j, \\ (1 - 2s)^{-1}, & \text{if } q_i^t \neq q_i^j. \end{cases} \\ &= (1 - |\mathbf{x}_i^t - \mathbf{x}_i^j|^2 s)^{-1}.\end{aligned}\quad (92)$$

According to (88) and (92), the MGF of γ is obtained by

$$M_\gamma(s) = \prod_{\tau=1}^T \prod_{r=1}^{N_r} M_{\gamma_r^\tau}(s) = (1 - |\mathbf{x}_i^t - \mathbf{x}_i^j|^2 s)^{-TN_r}.\quad (93)$$

According to (87), the value of $P(\mathbf{x}_i^t \rightarrow \mathbf{x}_i^j)$ can be finally formulated as

$$\begin{aligned}P(\mathbf{x}_i^t \rightarrow \mathbf{x}_i^j) &= \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \int_\gamma f_\gamma(\gamma) \exp\left(-\frac{\gamma}{4(\sigma^2 + \sigma_c^2) \sin^2 \theta}\right) d\theta \\ &= \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \left(1 + \frac{\|\mathbf{x}_i^t - \mathbf{x}_i^j\|^2}{4(\sigma^2 + \sigma_c^2)}\right)^{-TN_r} d\theta \\ &= \left[\frac{1}{2} - \frac{1}{2} \sqrt{\frac{c}{1+c}}\right] \sum_{k=0}^{TN_r-1} \binom{TN_r-1+k}{k} \left[\frac{1}{2} + \frac{1}{2} \sqrt{\frac{c}{1+c}}\right]^k,\end{aligned}\quad (94)$$

where $c = \|\mathbf{x}_i^t - \mathbf{x}_i^j\|^2 / 4(\sigma^2 + \sigma_c^2)$.

APPENDIX D PROOF OF (66)

In this section, we demonstrated that the value of $\mathbf{H}\mathbf{S}_i[\mathbf{E}_1 \mathbf{E}^{(1-\alpha)}(\mathbf{X}_{i+1}^k \mathbf{E}^\alpha - \mathbf{X}_{i+1}^m \mathbf{E}^\alpha)]\mathbf{X}_{i+2} \mathbf{E}_1$ is always a zero matrix or a matrix having small values. Assuming that

$\tilde{\mathbf{Y}}_i = \mathbf{H}\mathbf{S}_i \mathbf{E}_1$, whose elements obey $\mathcal{CN}(0, 1)$ we have

$$\begin{aligned}\tilde{\mathbf{Y}}_i \mathbf{E}^{(1-\alpha)} \mathbf{X}_{i+1}^m \mathbf{E}^\alpha \mathbf{X}_{i+2} \mathbf{E}_1 &= \left[(1-\alpha) \tilde{\mathbf{Y}}_i^1, \dots, (1-\alpha) \tilde{\mathbf{Y}}_i^T, \mathbf{0}, \dots, \mathbf{0} \right] \mathbf{X}_{i+1}^m \mathbf{E}^\alpha \mathbf{X}_{i+2} \mathbf{E}_1 \\ &= \left[(1-\alpha) s_{i+1}^m \tilde{\mathbf{Y}}_i^{q_i^m+1}, \dots, (1-\alpha) s_{i+1}^m \tilde{\mathbf{Y}}_i^T, \mathbf{0}, \dots, \mathbf{0}, \right. \\ &\quad \left. (1-\alpha) s_{i+1}^m \tilde{\mathbf{Y}}_i^1, \dots, (1-\alpha) s_{i+1}^m \tilde{\mathbf{Y}}_i^{q_i^m+1-1} \right] \mathbf{E}^\alpha \mathbf{X}_{i+2} \mathbf{E}_1, \\ \tilde{\mathbf{Y}}_i \mathbf{E}^{(1-\alpha)} \mathbf{X}_{i+1}^k \mathbf{E}^\alpha \mathbf{X}_{i+2} \mathbf{E}_1 &= \left[(1-\alpha) \tilde{\mathbf{Y}}_i^1, \dots, (1-\alpha) \tilde{\mathbf{Y}}_i^T, \mathbf{0}, \dots, \mathbf{0} \right] \mathbf{X}_{i+1}^k \mathbf{E}^\alpha \mathbf{X}_{i+2} \mathbf{E}_1 \\ &= \left[(1-\alpha) s_{i+1}^k \tilde{\mathbf{Y}}_i^{q_i^k+1}, \dots, (1-\alpha) s_{i+1}^k \tilde{\mathbf{Y}}_i^T, \mathbf{0}, \dots, \mathbf{0}, \right. \\ &\quad \left. (1-\alpha) s_{i+1}^k \tilde{\mathbf{Y}}_i^1, \dots, (1-\alpha) s_{i+1}^k \tilde{\mathbf{Y}}_i^{q_i^k+1-1} \right] \mathbf{E}^\alpha \mathbf{X}_{i+2} \mathbf{E}_1.\end{aligned}\quad (95)$$

To achieve a high throughput, we usually have $T = 1$, so that we have

$$\begin{aligned}V(\mathbf{N}_{i+2}^{\text{error}}) &= V[\tilde{\mathbf{Y}}_i \mathbf{E}^{(1-\alpha)}(\mathbf{X}_{i+1}^m - \mathbf{X}_{i+1}^k) \mathbf{E}^\alpha \mathbf{X}_{i+2} \mathbf{E}_1] \\ &= \begin{cases} V(\mathbf{Y}_\Delta), & \text{if } q_{i+1}^k = q_{i+1}^m, s_{i+1}^k \neq s_{i+1}^m, \\ V_m + V_k, & \text{if } q_{i+1}^k \neq q_{i+1}^m, \end{cases}\end{aligned}\quad (96)$$

with

$$\begin{aligned}\mathbf{Y}_\Delta &= \left[\dots (1-\alpha) \Lambda_{i+1} \tilde{\mathbf{Y}}_i^{q_i^k+1}, \dots, \mathbf{0} \right] \mathbf{E}^\alpha \mathbf{X}_{i+2} \mathbf{E}_1 \\ V(\mathbf{Y}_\Delta) &= \begin{cases} (1-\alpha)^2 \alpha^2 \Lambda_{i+1}^2 \mathbf{1}, & \text{if } q_{i+1}^k = q_{i+1}^m = 1, \\ (1-\alpha)^2 \Lambda_{i+1}^2 \mathbf{1}, & \text{if } q_{i+1}^k \neq 1, \langle q_{i+1}^k + q_{i+1}^m - 1 \rangle = 1, \\ \mathbf{0}_{N_r \times 1}, & \text{else,} \end{cases} \\ V_m &= V(\tilde{\mathbf{Y}}_i \mathbf{E}^{(1-\alpha)} \mathbf{X}_{i+1}^m \mathbf{E}^\alpha \mathbf{X}_{i+2} \mathbf{E}_1) \\ &= \begin{cases} (1-\alpha)^2 \alpha^2 \mathbf{1}, & \text{if } q_{i+1}^m = q_{i+1}^r = 1 \\ (1-\alpha)^2 \mathbf{1}, & \text{if } q_{i+1}^m \neq 1, \langle q_{i+1}^m + q_{i+1}^r - 1 \rangle = 1 \\ \mathbf{0}, & \text{else} \end{cases} \\ V_k &= V(\tilde{\mathbf{Y}}_i \mathbf{E}^{(1-\alpha)} \mathbf{X}_{i+1}^k \mathbf{E}^\alpha \mathbf{X}_{i+2} \mathbf{E}_1) \\ &= \begin{cases} (1-\alpha)^2 \alpha^2 \mathbf{1}, & \text{if } q_{i+1}^k = q_{i+1}^r = 1, \\ (1-\alpha)^2 \mathbf{1}, & \text{if } q_{i+1}^k \neq 1, \langle q_{i+1}^k + q_{i+1}^r - 1 \rangle = 1, \\ \mathbf{0}, & \text{else,} \end{cases}\end{aligned}\quad (97)$$

where $V(x)$ represents the variance of x and $\Lambda_{i+1} = s_{i+1}^m - s_{i+1}^k$.

$$\begin{aligned}V(\mathbf{N}_{i+2}^{\text{error}}) &= \begin{cases} (1-\alpha)^2 \alpha^2 \Lambda_{i+1}^2 \mathbf{1}, & \text{if } q_{i+1}^k = q_{i+1}^m = q_{i+1}^r = 1, \\ (1-\alpha)^2 \Lambda_{i+1}^2 \mathbf{1}, & \text{if } q_{i+1}^k = q_{i+1}^m \neq 1, \langle q_{i+1}^k + q_{i+1}^m - 1 \rangle = 1, \\ (1-\alpha)^2 \alpha^2 \mathbf{1}, & \text{if } q_{i+1}^k = 1, q_{i+1}^m = 1, q_{i+1}^r \neq 1, \\ (1-\alpha)^2 \alpha^2 \mathbf{1}, & \text{if } q_{i+1}^k = 1, q_{i+1}^m \neq 1, q_{i+1}^r = 1, \\ (1-\alpha)^2 \mathbf{1}, & \text{if } q_{i+1}^k \neq 1, q_{i+1}^m \neq q_{i+1}^r, \langle q_{i+1}^k + q_{i+1}^r - 1 \rangle = 1 \\ \mathbf{0}, & \text{else.} \end{cases}\end{aligned}\quad (98)$$

According to (98), we can obtain the corresponding probabilities by (66).

REFERENCES

- [1] M. Di Renzo, H. Haas, A. Ghayeb, S. Sugiura, and L. Hanzo, "Spatial modulation for generalized MIMO: challenges, opportunities and implementation," *Proceedings of the IEEE*, vol. 102, no. 1, pp. 56-103, Jan. 2014.
- [2] A. Chockalingam and B. S. Rajan, *Large MIMO systems*. Cambridge, U.K.: Cambridge Univ. Press, 2014.
- [3] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186-195, Feb. 2014.

- [4] R. Mesleh, H. Haas, S. Sinanovic, C. W. Ahn, and S. Yun, "Spatial modulation," *IEEE Trans. Veh. Technol.*, vol. 57, no. 4, pp. 2228-2241, Jul. 2008.
- [5] M. Wen, B. Zheng, K. J. Kim, M. D. Renzo, T. A. Tsiftsis, K. Chen and N. A. Dahir, "A survey on spatial modulation in emerging wireless systems: Research progresses and applications," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 9, pp. 1949-1972, Sep. 2019.
- [6] J. Li, M. Wen, X. Cheng, Y. Yan, S. Song and M. Lee, "Generalized precoding-aided quadrature spatial modulation," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1881-1886, Feb. 2017.
- [7] J. Li, Y. Peng, Y. Yan, X. Jiang, H. Hai and M. Zukerman, "Cognitive radio network assisted by OFDM with index modulation," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 1106-1110, Jan. 2020.
- [8] P. Yang, Y. Xiao, Y. L. Guan, K. V. S. Hari, A. Chockalingam, S. Sugiura, H. Haas, M. Renzo, C. Masouros, Z. Liu, L. Xiao, S. Li and L. Hanzo, "Single-carrier spatial modulation: A promising design for large-scale broadband antenna systems," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1687-1716, Feb. 2016.
- [9] S. Wang, Y. Li, M. Zhang, and J. Wang, "Energy-efficient and low-complexity uplink transceiver for massive spatial modulation MIMO," *IEEE Trans. Veh. Technol.*, vol. 64, no. 10, pp. 4617-4632, Oct. 2015.
- [10] L. He, J. Wang, J. Song, and L. Hanzo, "On the multi-user, multi-cell massive spatial modulation uplink: How many antennas for each user?," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1437-1451, Dec. 2016.
- [11] P. Patcharamanepakorn, S. Wu, C. Wang, E. M. Aggoune, M. M. Alwakel, X. Ge, and M. Di. Renzo, "Spectral, energy, and economic efficiency of 5G multicell massive MIMO systems with generalized spatial modulation", *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 9715-9731, Dec. 2016.
- [12] L. Xiao, Y. Xiao, C. Xu, X. Lei, P. Yang, S. Li and L. Hanzo, "Compressive sensing assisted spatial multiplexing aided spatial modulation," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 794-807, Feb. 2018.
- [13] L. Xiao, P. Xiao, Y. Xiao, H. Hass, A. Mohamed, and L. Hanzo, "Compressive sensing assisted generalized quadrature spatial modulation for massive MIMO," *IEEE Trans. Commun.* vol. 67, no. 7, pp. 4795-4810, May 2019.
- [14] L. Xiao, P. Xiao, Z. Liu, W. Yu, H. Hass and L. Hanzo, "A compressive sensing assisted massive SM-VBLAST system: Error probability and capacity analysis," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1990-2005, Mar. 2020.
- [15] Y. Bian, M. Wen, X. Cheng, H. V. Poor, and B. Jiao, "A differential scheme for spatial modulation," in *Proc. 2013 IEEE Globecom. Conf.*, Atlanta, USA, pp. 3925-3930, Dec. 2013.
- [16] M. Wen, Z. Ding, X. Cheng, Y. Bian, H. V. Poor, and B. Jiao, "Performance analysis of differential spatial modulation with two transmit antennas," *IEEE Commun. Lett.*, vol. 18, no. 3, pp. 475-478, Mar. 2014.
- [17] 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, Aug. 2014.
- [18] J. Li, M. Wen, X. Cheng, Y. Yan, S. Song, and M. H. Lee, "Differential spatial modulation with gray coded antenna activation order," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1100-1103, Jun. 2016.
- [19] N. Ishikawa and S. Sugiura, "Unified differential spatial modulation," *IEEE Wireless Commun. Lett.*, vol. 3, no. 4, pp. 337-340, Feb. 2014.
- [20] J. Liu, L. Dan, P. Yang, L. Xiao, F. Yu and Y. Xiao, "High-rate APSK-aided differential spatial modulation: design method and performance analysis," *IEEE Commun. Lett.*, vol. 21, no. 1, pp. 168-171, Sep. 2016.
- [21] M. Zhang, M. Wen, X. Cheng, and L. Yang, "A dual-hop virtual MIMO architecture based on hybrid differential spatial modulation," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6356-6370, Sept. 2016.
- [22] W. Zhang, Q. Yin, and H. Deng, "Differential full diversity spatial modulation and its performance analysis with two transmit antennas," *IEEE Commun. Lett.*, vol. 19, no. 4, pp. 677-680, April. 2015.
- [23] L. Xiao, Y. Xiao, P. Yang, J. Liu, S. Li and W. Xiang, "Space-time block coded differential spatial modulation," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 8821-8834, Oct. 2017.
- [24] M. Wen, X. Cheng, Y. Bian, and H. V. Poor, "A low-complexity near-ML differential spatial modulation detector," *IEEE Signal Proc. Lett.*, vol. 22, no. 11, pp. 1834-1838, Nov. 2015.
- [25] 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.
- [26] 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.
- [27] 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.
- [28] C. Xu, T. Bai, J. Zhang, R. Rajashekar, R. G. Maunder, Z. Wang and L. Hanzo, "Adaptive coherent/non-Coherent spatial modulation aided unmanned aircraft systems," *IEEE Wireless Commun.*, vol. 26, no. 4, pp. 170-177, May, 2019.
- [29] C. Xu, J. Zhang, T. Bai, P. Botsinis, R. G. Maunder, R. Zhang and L. Hanzo, "Adaptive coherent/non-coherent single/multiple-antenna aided Channel coded ground-to-air aeronautical communication," *IEEE Trans. Commun.*, vol. 67 no. 2, pp. 1099-1116, Feb. 2019.
- [30] R. Rajashekar, C. Xu, N. Ishikaw, S. Sugiura, K.V.S. Hari, and L. Hanzo, "Full-diversity dispersion matrices from algebraic field extensions for differential spatial modulation," *IEEE Trans. Commun.*, vol. 66, no. 1, pp. 385-394, Jan. 2017.
- [31] R. Rajashekar, C. Xu, N. Ishikaw, 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.
- [32] N. Ishikawa, and S. Sugiura, "Rectangular differential spatial modulation for open-loop noncoherent massive-MIMO downlink," *IEEE Trans. Wireless Commun.*, vol. 67, no. 9, pp. 6586-6597, Sep. 2019.
- [33] C. Wu, Y. Xiao, L. Xiao, P. Yang, X. Lei, "Space-time block coded rectangular differential spatial modulation: System design and performance analysis", *IEEE Trans. Commun.*, vol. 16, no. 3, pp. 1908-1920, Mar. 2017.
- [34] N. Ishikawa, R. Rajashekar, C. Xu, S. Sugiura, and L. Hanzo, "Differential space-time coding dispending with channel estimation approaches the performance of its coherent counterpart in the open-loop massive MIMO-OFDM downlink," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6190 - 6204, 2018.
- [35] N. Ishikawa, R. Rajashekar, C. Xu, M. El-Hajjar, S. Sugiura, L. L. Yang, and L. Hanzo, "Differential-detection aided large-scale generalized spatial modulation is capable of operating in high-mobility millimeter-wave channels," *IEEE J. Sel. Top. Signal Process.*, vol. 13, no. 6, pp. 1360-1374, Oct. 2019.
- [36] R. Rajashekar, K.V.S. Hari and L. Hanzo, "Reduced-complexity ML detection and capacity-optimized training for spatial modulation systems," *IEEE Trans. Commun.*, vol. 62, no. 1, pp. 112-125, Jan. 2014.
- [37] R. Mesleh, O. S. Badarneh, A. Younis and H. Haas, "Performance analysis of spatial modulation and space-shift keying with imperfect channel estimation over generalized $\eta - \mu$ Fading Channels," *IEEE Trans. Veh. Technol.*, vol. 64, no. 1, pp. 88-96, Jan. 2015.