

# Reduced-Complexity ML Detection and Capacity-Optimized Training for Spatial Modulation Systems

Rakshith Rajashekar, *Student Member, IEEE*, K.V.S. Hari, *Senior Member, IEEE*, and L. Hanzo, *Fellow, IEEE*

**Abstract**—Spatial Modulation (SM) is a recently developed low-complexity Multiple-Input Multiple-Output scheme that jointly uses antenna indices and a conventional signal set to convey information. It has been shown that the Maximum-Likelihood (ML) detector of an SM system involves joint detection of the transmit antenna index and of the transmitted symbol, hence, the ML search complexity grows linearly with the number of transmit antennas and the size of the signal set. To circumvent the problem, we show that the ML search complexity of an SM system may be rendered independent of the constellation size, provided that the signal set employed is a square- or a rectangular-QAM. Furthermore, we derive bounds for the capacity of the SM system and derive the optimal power allocation between the data and the training sequences by maximizing the worst-case capacity bound of the SM system operating with imperfect channel state information. We show, with the aid of our simulation results, that the proposed detector is ML-optimal, despite its lowest complexity amongst the existing detectors. Furthermore, we show that employing the proposed optimal power allocation provides a substantial gain in terms of the SM system’s capacity as well as signal-to-noise ratio compared to its equal-power-allocation counterpart. Finally, we compare the performance of the SM system to that of the conventional Multiple-Input Multiple-Output (MIMO) system and show that the SM system is capable of outperforming the conventional MIMO system by a significant margin, when both the systems are employing optimal power splitting.

**Index Terms**—Spatial modulation, ML decoding, computational complexity, training, and channel estimation.

## I. INTRODUCTION

Spatial Modulation (SM) [1]-[4] is a recently developed low-complexity Multiple-Input Multiple-Output (MIMO) scheme, where the information bitstream is divided into blocks of  $\log_2(N_t M)$  bits, where  $N_t$  is the number of transmit antennas and  $M$  is the size of the classic signal set. In each such block,  $\log_2(M)$  bits select a symbol  $s$  from an  $M$ -ary signal set (such as  $M$ -QAM or -PSK), and  $\log_2(N_t)$  bits select an antenna out of  $N_t$  transmit antennas for the transmission of the symbol  $s$ , effectively conveying  $(\log_2 M + \log_2 N_t)$  bits. The throughput achieved by this scheme is  $R = \log_2(N_t M)$

bits per channel use (bpcu). Thus, the SM scheme achieves a throughput increase of  $\log_2 N_t$  bits over single-antenna systems. In order to achieve higher throughputs, either  $N_t$  or  $M$ , or both have to be increased, which renders this scheme suitable for both low and for moderately high bandwidth efficiencies. One of the important benefits of this scheme is that it is free from Inter-Antenna Interference (IAI), which is achieved by activating only a single transmit antenna in any symbol duration.

In [5], it was shown that the Maximum Likelihood (ML) detector of the SM system involves joint detection of both the transmit antenna index and of the transmitted symbol. Thus, the ML search complexity in the SM system grows as  $N_t M$ . Recently, various sub-optimal low-complexity detectors were proposed for the SM system in [6]-[9]. Furthermore, specially tailored Sphere Decoding (SD) schemes [10]-[13] were proposed for SM systems, which were termed as Transmitter-centric SD (Tx-SD) [14], Receiver-centric SD (Rx-SD) [15] and Combined SD (C-SD) schemes [14]. The Tx-SD arrangement of [14] was proposed for SM systems having  $N_t \leq N_r$ , where  $N_r$  is the number of receive antennas. It was shown in [14] that the Rx-SD is suitable for SM systems with a large  $N_r$ , while the C-SD is suitable for systems using a large  $N_t$  or  $M$ , or both. In [17], the SD of [15] was further developed for SM systems having  $N_r < N_t \leq 2N_r - 1$ , and a generalized ML-optimal detector was proposed for SM systems having arbitrary  $N_t$ . Though all the detectors mentioned above aim at achieving near-ML performance, their complexities grow with the size of the signal constellation used.

In practical scenarios the Channel State Information at the Receiver (CSIR) is imperfect due to imperfections in the channel estimation process. Some recent studies on the performance of SM systems operating with imperfect CSIR can be found in [18]-[21]. However in [18], no solution was proposed to reduce the performance degradation imposed. An iterative channel-estimation/data detection algorithm was proposed in [19] that aims for reducing the performance degradation due to imperfect CSIR, but, this algorithm incurs a significant computational complexity and data detection delay at the receiver. In [20], [21], the performance of space shift keying (a special case of SM) operating with imperfect CSIR was studied by considering various training periods.

Against this background the following are the novel contributions of this paper:

- 1) We show that for a square- or a rectangular-QAM signal set, the ML search complexity of an SM system becomes

Rakshith Rajashekar and K.V.S. Hari are with the Department of Electrical Communication Engineering, Indian Institute of Science, India (e-mail: {rakshithmr, hari}@ece.iisc.ernet.in).

L. Hanzo is with the School of ECS, University of Southampton, UK (e-mail: lh@ecs.soton.ac.uk).

The financial support of the DST, India and of the EPSRC, UK under the auspices of the India-UK Advanced Technology Center (IU-ATC) is gratefully acknowledged. L. Hanzo would like to thank the European Research Council for their support in terms of the Advanced Fellow Grant.

*independent* of the constellation size and hence, grows only with  $N_t$  (Section III). This is achieved by employing a simple *hard-limiting* operation, and the resultant detector is termed as the Hard-Limiter based ML (HL-ML) detector. Furthermore, we compare the computational complexity of the proposed HL-ML detector to that of the various existing sub-optimal detectors [6]-[9] as well as of SDs [14]-[17] and show that the proposed detector has the lowest computational complexity among these detectors. Since the proposed detector imposes a constant complexity which is independent of the constellation size, it proves very beneficial for SM systems that adapt modulation order at the transmitter depending on the channel conditions. An additional benefit of our proposed detector is that it remains optimal even when the receiver operates with imperfect channel estimates.

- 2) Note that the algorithm proposed in [19] for SM systems operating with imperfect CSIR incurs a significant computational complexity and naturally increasing the fraction of pilot/training symbols [20] for improving the quality of the Channel Estimate (CE) reduces the effective data throughput. Our objective is to improve the system performance without increasing the receiver complexity or increasing the fraction of training symbols. This is achieved by optimally splitting the transmission power between training and data blocks so that the Mutual Information (MI) of the SM system operating with imperfect CSIR is maximized (Section IV). An important advantage of our approach is that the power sharing strategy does not depend on the instantaneous channel realization, nor does it require any feedback information to be sent to the transmitter from the receiver<sup>1</sup>.

Note that the statement in 1) is not restricted to SM systems alone, but is applicable to any ICI-free system, such as the Space-Time Shift Keying scheme of [24]. The hard-limiter can be used in the ML detection of conventional Space-Time Block Codes (STBC) as well. However the reduction in the ML decoding complexity is marginal due to multiple symbols encoded in the STBC. Further details on this can be found in [25].

*Notations:* Boldface uppercase letters represent matrices. Boldface lowercase letters represent vectors, while  $\text{tr}(\cdot)$  represents the trace of a matrix,  $\|\cdot\|$  denotes the two-norm of a vector;  $\langle \cdot, \cdot \rangle$  represents the inner product of two vectors;  $\Re(\cdot)$  and  $\Im(\cdot)$  are the real and imaginary parts of a complex-valued quantity;  $|\cdot|$  is the magnitude of a complex quantity, or the cardinality of a given set;  $\lfloor \cdot \rfloor$  indicates the operation of rounding a real number to the nearest integer;  $\mathbf{H}([a : b], :)$  represents a matrix with rows  $a, a + 1, \dots, b - 1, b$  of  $\mathbf{H}$  and  $\mathbf{H}(:, [a : b])$  denotes a matrix with columns  $a, (a + 1), \dots, (b - 1), b$  of

<sup>1</sup>In [22], [23], the power allocation for transmit antennas for data transmission phase was proposed for space shift keying system, where the scheme in [23] needs partial channel state information at the transmitter. Both the schemes of [22], [23], are assumed to have perfect CSIR. To the best of our knowledge, power allocation between training and data blocks in the SM system operating with imperfect CSIR has not been reported in the existing literature.

$\mathbf{H}$ ;  $\mathbb{R}$  and  $\mathbb{C}$  represent the field of real and complex numbers, respectively;  $\mathcal{CN}(\mu, \sigma^2)$  denotes a complex Gaussian random variable with mean  $\mu$  and variance  $\sigma^2$ . If  $S_1$  and  $S_2$  are two sets such that  $S_2 \subset S_1$ , then  $X = S_1 \setminus S_2$  represents a set containing all the elements of  $S_1$  that are not present in  $S_2$ .  $N - \text{PAM} = \{-N + 1, -N + 3, \dots, -1, 1, \dots, N - 3, N - 1\}$ , where  $N$  is a power of two. With a slight abuse of notation, both random variables and their instantiations are represented by lower case letters.

## II. SM SYSTEM MODEL

An SM system having  $N_t$  transmit as well as  $N_r$  receive antennas and communicating over a quasi-static, frequency-flat fading channel, can be modeled as

$$\mathbf{y} = \mathbf{H}\mathbf{x}_{l,s} + \mathbf{n}, \quad (1)$$

where  $\mathbf{x}_{l,s} \in \mathbb{C}^{N_t}$  is the transmitted vector, which is of the form

$$\mathbf{x}_{l,s} = \underbrace{[0, \dots, 0]_{l-1}}_{l-1}, \underbrace{[s, 0, \dots, 0]_{N_t-l}}_{N_t-l} \in \mathbb{C}^{N_t}, \quad (2)$$

where  $l \in L = \{i\}_{i=1}^{N_t}$  and  $s$  is a complex symbol from the signal set  $S$  with  $|S| = M$ ,  $\mathbf{y} \in \mathbb{C}^{N_r}$  is the received vector,  $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_{N_t}] \in \mathbb{C}^{N_r \times N_t}$  is the channel matrix,  $\mathbf{n} \in \mathbb{C}^{N_r}$  is the noise vector. The entries of the channel matrix and the noise vector are from  $\mathcal{CN}(0,1)$  and  $\mathcal{CN}(0, \sigma^2)$  such that  $\sigma^2 = \frac{\gamma}{\rho}$ , respectively, where  $\rho$  is the average signal-to-noise ratio (SNR) at each receive antenna and  $\gamma = \frac{\sum_{i=1}^M |s_i|^2}{M}$  is the average energy of the signal set  $S$ . Throughout this paper we assume  $S$  to be a square- or a rectangular-QAM constellation. Assuming perfect channel state information (CSI) and ML detection at the receiver, we have

$$(\hat{l}, \hat{s})_{ML} = \arg \min_{l \in L, s \in S} \|\mathbf{y} - \mathbf{H}\mathbf{x}_{l,s}\|_2^2. \quad (3)$$

## III. ML SEARCH COMPLEXITY REDUCTION IN SM SYSTEMS EMPLOYING A LATTICE CONSTELLATION

*Definition 1:* Given an optimization problem, the number of evaluations of the optimization metric is defined as the *order of complexity* for the algorithm.

*Definition 2:* The number of real-valued multiplications involved in solving an optimization problem is defined as the *computational complexity*<sup>2</sup>.

It is straightforward from (3) that the order of complexity for the ML detector is  $MN_t$ , since  $|S| = M$  and  $|L| = N_t$ , and its *computational complexity* is  $6N_r N_t M$  which increases linearly with  $M$ . Note that for a given  $(l, s)$ ,  $\|\mathbf{y} - \mathbf{h}_{l,s}\|^2 = \sum_{i=1}^{N_r} |y_i - h_{i,l}s|^2$  takes  $6N_r$  real-valued multiplications, since  $h_{i,l}s$  takes 4 real-valued multiplications and  $|y_i - h_{i,l}s|^2 = [\Re(y_i - h_{i,l}s)]^2 + [\Im(y_i - h_{i,l}s)]^2$  takes 2 real-valued multiplications.

<sup>2</sup>Given an optimization problem, if  $\mathcal{C}$  is the computational complexity in evaluating the optimization metric once, then the overall computational complexity is given by  $[\text{order of complexity} \times \mathcal{C}]$ . Thus, minimizing the order minimizes the overall computational complexity.

Before presenting our proposed low-complexity ML detector, we provide a brief overview of the sub-optimal and optimal detectors conceived for the SM system in the existing literature and compare them to that of the proposed detector. Note that the comparison of various detectors based on their computational complexity alone is difficult, thus for the ease of comparison the order of complexity is also provided for the various detectors. Appendix A provides the details of the computational complexity of various detectors listed in column four of Table I, including that of the proposed HL-ML detector. It is clear from Table I that our proposed detector is the only ML-optimal detector whose computational complexity does not increase with the size of the constellation. The proposed HL-ML detector is presented in the proof of the following proposition.

*Proposition 1:* If  $S$  is a lattice constellation such as a square or a rectangular QAM which can be viewed as a Cartesian product of two PAM signal sets, say  $S_1 = N_1$ -PAM and  $S_2 = N_2$ -PAM, then the *order of complexity* for ML detector in an SM system having  $N_t$  transmit antennas becomes *independent* of the constellation size. In other words, the ML search complexity in (3) reduces from  $N_t N_1 N_2$  to  $N_t$ .

*Proof:* Considering (3), we have

$$(\hat{l}, \hat{s})_{ML} = \arg \min_{l \in L, s \in S} \|\mathbf{y} - \mathbf{H}\mathbf{x}_{l,s}\|_2^2, \quad (4)$$

$$= \arg \min_{l \in L, s \in S} \|\mathbf{y} - \mathbf{h}_l s\|_2^2 \Leftrightarrow \arg \min_{l \in L, s \in S} |\hat{y}_l - s|_2^2, \quad (5)$$

where  $\hat{y}_l = \frac{\mathbf{h}_l^H \mathbf{y}}{\|\mathbf{h}_l\|^2}$ . The equivalence in (5) follows from (A.55) of [26]. Thus, the ML solution can be equivalently written as

$$\begin{aligned} (\hat{l}, \hat{s})_{ML} &= \arg \min_{l \in L, s \in S} |\hat{y}_l - s|_2^2, \\ &= \arg \min_{l \in L} \left[ \min_{s_I \in N_1 - \text{PAM}, s_Q \in N_2 - \text{PAM}} \{(\Re(\hat{y}_l) - s_I)^2 + (\Im(\hat{y}_l) - s_Q)^2\} \right]. \end{aligned}$$

In what follows, we show that the complexity of the inner optimization problem given by  $\min_{s \in S} |\hat{y}_l - s|_2^2$  is a constant and does not depend on the size of  $S$ . Thus, the overall complexity of the proposed algorithm depends only on the variable  $l$ .

Since  $s_I$  and  $s_Q$  are from orthogonal dimensions, we have

$$\begin{aligned} (\hat{l}, \hat{s})_{ML} &= \arg \min_{l \in L} \left[ \min_{s_I \in N_1 - \text{PAM}} (\Re(\hat{y}_l) - s_I)^2 \right. \\ &\quad \left. + \min_{s_Q \in N_2 - \text{PAM}} (\Im(\hat{y}_l) - s_Q)^2 \right]. \end{aligned}$$

Now consider the optimization problem  $\Re(\hat{s}_l) = \min_{s_I \in N_1 - \text{PAM}} (\Re(\hat{y}_l) - s_I)^2$ . Since the symbols of  $N_1 - \text{PAM}$  are given by  $\{2i + 1\}_{i=\frac{N_1-1}{2}}$ ,  $\Re(\hat{s}_l)$  can be expressed

as follows:

$$\begin{aligned} \Re(\hat{y}_l) < -N_1 + 2 &\Rightarrow \Re(\hat{s}_l) = -N_1 + 1 \\ -N_1 + 2 \leq \Re(\hat{y}_l) < -N_1 + 4 &\Rightarrow \Re(\hat{s}_l) = -N_1 + 3 \\ &\vdots \\ -2 \leq \Re(\hat{y}_l) < 0 &\Rightarrow \Re(\hat{s}_l) = -1 \\ 0 \leq \Re(\hat{y}_l) < 2 &\Rightarrow \Re(\hat{s}_l) = 1 \\ &\vdots \\ N_1 - 4 \leq \Re(\hat{y}_l) < N_1 - 2 &\Rightarrow \Re(\hat{s}_l) = N_1 - 3 \\ N_1 - 2 \leq \Re(\hat{y}_l) &\Rightarrow \Re(\hat{s}_l) = N_1 - 1. \end{aligned}$$

However,  $\Re(\hat{s}_l)$  can be obtained directly without employing the above set of comparisons. Assuming that the active transmit antenna index is  $l$ , we have  $\hat{y}_l = (2j+1) + \hat{n}_l$  for some  $j \in \{-\frac{N_1}{2}, -\frac{N_1}{2} + 1, \dots, \frac{N_1}{2} - 2, \frac{N_1}{2} - 1\}$ , where  $\hat{n}_l = \frac{\mathbf{h}_l^H \mathbf{n}}{\|\mathbf{h}_l\|^2}$ . Then consider  $2 \left\lfloor \frac{\Re(\hat{y}_l) + 1}{2} \right\rfloor - 1$ , which simplifies to  $2(j+k) - 1$ , where  $k = \left\lfloor \frac{\Re(\hat{n}_l)}{2} \right\rfloor$ . Note that  $2(j+k) - 1$  is of the form  $2z - 1$  for  $z \in \mathbb{Z}$ , that is closest to  $\Re(\hat{y}_l)$ . But the symbols of  $N_1 - \text{PAM}$  range from  $-N_1 + 1$  to  $N_1 - 1$ . Therefore we have to ensure that  $2 \left\lfloor \frac{\Re(\hat{y}_l) + 1}{2} \right\rfloor - 1$  lies in the  $N_1 - \text{PAM}$  range, which is achieved based on the following comparisons:

$$\max \left( 2 \left\lfloor \frac{\Re(\hat{y}_l) + 1}{2} \right\rfloor - 1, -N_1 + 1 \right), \quad (6)$$

which ensures that  $2 \left\lfloor \frac{\Re(\hat{y}_l) + 1}{2} \right\rfloor - 1$  is less than or equal to  $(-N_1 + 1)$ , and

$$\min \left[ \max \left( 2 \left\lfloor \frac{u_1 + 1}{2} \right\rfloor - 1, -N_1 + 1 \right), N_1 - 1 \right], \quad (7)$$

ensures that  $2 \left\lfloor \frac{\Re(\hat{y}_l) + 1}{2} \right\rfloor - 1$  is greater than or equal to  $(N_1 - 1)$ .

Proceeding along similar lines for  $\Im(\hat{s}_l) = \min_{s_Q \in N_2 - \text{PAM}} (\Im(\hat{y}_l) - s_Q)^2$ , we obtain

$$\begin{aligned} \Re(\hat{s}_l)_{ML} &= \min \left[ \max \left( 2 \left\lfloor \frac{u_1 + 1}{2} \right\rfloor - 1, -N_1 + 1 \right), N_1 - 1 \right], \\ \Im(\hat{s}_l)_{ML} &= \min \left[ \max \left( 2 \left\lfloor \frac{u_2 + 1}{2} \right\rfloor - 1, -N_2 + 1 \right), N_2 - 1 \right], \end{aligned} \quad (8)$$

where  $u_1 = \Re(\hat{y}_l)$  and  $u_2 = \Im(\hat{y}_l)$ , yielding  $\hat{s}_l = \Re(\hat{s}_l)_{ML} + j\Im(\hat{s}_l)_{ML}$ . Thus, from (8) and (9) we obtain  $N_t$  complex-valued symbols associated with the  $N_t$  hypotheses, since  $1 \leq l \leq N_t$ . These symbols are represented by the set  $S_{min} = \{\hat{s}_1, \hat{s}_2, \dots, \hat{s}_{N_t}\}$ . Thus, we have

$$(\hat{l}, \hat{s})_{ML} = \arg \min_{l \in L, s \in S} \|\mathbf{y} - \mathbf{h}_l s\|_2^2, \quad (10)$$

$$= \arg \min_{l \in L} \|\mathbf{y} - \mathbf{h}_l \hat{s}_l\|_2^2. \quad (11)$$

Furthermore, by considering that

$$\|\mathbf{y} - \mathbf{h}_l \hat{s}_l\|_2^2 = \|\mathbf{y}\|^2 + \|\mathbf{h}_l\|^2 |\hat{s}_l|^2 - 2\Re(\mathbf{h}_l^H \mathbf{y} \hat{s}_l^*), \quad (12)$$

$$= \|\mathbf{y}\|^2 + \|\mathbf{h}_l\|^2 |\hat{s}_l|^2 - 2\Re(\hat{y}_l \hat{s}_l^*) \|\mathbf{h}_l\|^2, \quad (13)$$

$$= \|\mathbf{y}\|^2 + \{|\hat{s}_l|^2 - 2\Re(\hat{y}_l \hat{s}_l^*)\} \|\mathbf{h}_l\|^2, \quad (14)$$

$$= \|\mathbf{y}\|^2 + \{|\hat{y}_l - \hat{s}_l|^2 - |\hat{y}_l|^2\} \|\mathbf{h}_l\|^2, \quad (15)$$

TABLE I  
COMPARISON OF SOME OF THE THE KNOWN EXISTING DETECTORS CONCEIVED FOR THE SM SYSTEM.

Detectors	ML Optimality	Order of complexity	Computational complexity	Complexity grows with $M$ ?
ML detector of (3)	optimal	$O_{ML} = N_t M$	$C_{ML} = 6N_t N_r M$	YES
Matched-Filtering (MF) based detector [6]	sub-optimal	$O_{MF} = N_t M/4$ (refer to (22) in [6])	$C_{MF} = 12N_r N_t + (N_t + 4)M$ (refer to (35) in [6])	YES
Signal Vector based Detector (SVD) [7]	sub-optimal	$O_{SVD} = \max(M, N_t)$	$C_{SVD} = (6N_r + 4)N_t + 2N_r + 6N_r M$ (refer to (5) in [7])	YES
Developed MF (DMF) based detector [8]	sub-optimal	$O_{DMF} \geq O_{MF}$	$C_{DMF} \geq C_{MF}$ (refer to Section II-B [8])	YES
Distance Based ordered Detection (DBD) [9]	Conditionally optimal ( $p = N_t$ )	$O_{DBD} = \max(M, p)$ $1 \leq p \leq N_t$	$C_{DBD} = 8N_r N_t + 6N_r p + 3N_t + 2N_t M$ (refer to Section II-F [9])	YES
<b>Proposed detector</b>	<b>optimal</b>	<b><math>N_t</math></b>	<b><math>C_{HL-ML} = (6N_r + 11)N_t</math></b>	<b>NO</b>

from (15) and (11) we arrive at:

$$(\hat{l}, \hat{s})_{ML} = \arg \min_{l \in L} (|\hat{y}_l - \hat{s}_l|^2 - |\hat{y}_l|^2) \|\mathbf{h}_l\|^2. \quad (16)$$

#### A. Optimality of the HL-ML detector under imperfect CSIR

Assuming that the true channel matrix is  $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$  and the channel estimated by the receiver is  $\check{\mathbf{H}} \in \mathbb{C}^{N_r \times N_t}$ , the ML detector employed by the SM system is given by

$$(\hat{l}, \hat{s})_{ML} = \arg \min_{l \in L, s \in S} \|\mathbf{y} - \check{\mathbf{h}}_l s\|_2^2, \quad (17)$$

$$\triangleq \arg \min_{l \in L, s \in S} |\check{y}_l - s|_2^2, \quad (18)$$

where  $\check{y}_l = \frac{\check{\mathbf{h}}_l^H \mathbf{y}}{\|\check{\mathbf{h}}_l\|^2}$ . Proceeding along similar lines to those of Proposition 1, it is straightforward to show that the proposed HL-ML detector remains optimal under imperfect CSIR. In the remaining part of the paper, we focus our attention on achieving a performance, which is same as that of the SM system operating with perfect CSIR, when the system is actually operating in an imperfect CSIR scenario.

#### B. Sphere Detection in SM systems

Sphere detection is a well-known method of approaching the ML performance in MIMO systems at a significantly reduced computational complexity [10]-[13]. Recently, specially tailored SDs were proposed in [14], [15], for SM systems having  $N_t \leq N_r$ . In [16], SDs were proposed for SM systems having  $N_t > N_r$  by regularizing the underdetermined channel matrix. Furthermore, in [17] the SD of [15] was extended for SM systems having  $N_r < N_t \leq 2N_r - 1$ , and a generalized ML-optimal decoder was proposed for SM systems having arbitrary  $N_t$ . All these detectors are optimal, when the search radius is sufficiently large so that at least one point is found within the radius of search. These detectors essentially involve the computation of the QR decomposition [27] of the real-valued channel matrix (for example, refer to (15) in [15]) which imposes a complexity of  $C_{QR} = \sum_{k=1}^N [2f(k) + f(k)^2 + 2f(k)^3 + 1] - N_r^3$ , where  $N = \min(N_r - 1, N_t)$  (refer to (22) in [15]). The additional computational complexity includes that of  $C_{\mathbf{z}} = 4N_t N_r$  and  $C_{\|\mathbf{Q}_2^T \bar{\mathbf{y}}\|^2} = 2N_r(2N_r - 2N_t + 1)$  (refer to Section IV-A in [15] for more details). Therefore, regardless of the number of

points found inside the search radius, the SD incurs a fixed complexity of  $C_{fix} = C_{QR} + C_{\mathbf{z}} + C_{\|\mathbf{Q}_2^T \bar{\mathbf{y}}\|^2}$ <sup>3</sup>. In Section V, we compare the complexity imposed by the proposed detector to that of the SD and show that  $C_{fix}$  of the SD exceeds the overall computational complexity of the proposed HL-ML detector by a large margin.

#### IV. CAPACITY OPTIMIZED TRAINING AND POWER ALLOCATION

In Section III we presented a low-complexity detector designed for SM systems, which was shown to be ML-optimal under both perfect and imperfect CSIR conditions. In a practical scenario the channel estimated at the receiver with the aid of training/pilot symbols is always prone to estimation errors due to the inherent imperfection of the channel estimation process. A straightforward approach to reduce the estimation error is to increase the number of training symbols, but this reduces the useful data rate, hence it is undesirable. In [18]-[20], the performance of the SM system relying on realistic imperfect CSIR was studied. In [18], it was shown that by employing a sufficiently large number of training symbols the performance loss due to imperfect CSIR can be significantly reduced. In [19], an iterative channel-estimation/data-detection algorithm was proposed for reducing the performance loss in the SM system due to imperfect CSIR, while using a minimum number of training symbols. However, this solution incurs a significant computational complexity at the receiver. Our objective in this section is to minimize the performance loss due to imperfect CSIR without increasing the number of pilot symbols or increasing the receiver complexity. This is achieved by sharing the available transmission power optimally between the training and data symbols, as described in the following steps:

- 1) A non-trivial lower bound on the capacity of the SM system relying on the estimated channel is obtained

<sup>3</sup>Analogous to QR decomposition used in [14], [15], the SD based detector in [16] uses Cholesky decomposition. It can be easily verified the complexity imposed by pre-computations involved in SD based detector of [16] itself exceeds the overall computational complexity of the the proposed HL-ML detector. Note that the computational complexity of various SD based detectors presented in [14]-[16] are normalized by that of the ML detector. In this paper, we have considered the absolute computational complexity of various detectors for the sake of explicit clarity. The insightful comments of the anonymous Reviewer are much appreciated, which eliminated any ambiguity concerning the system's complexity in comparison to [16].

and the reduction in the SM system's capacity due to imperfect CSIR is analyzed.

- 2) The power sharing between the training and data symbols is optimized so that the lower bound on the SM system's capacity relying on the estimated CSIR is maximized. Thus, better estimates of the channel are obtained without any increase in the number of training symbols or increase in the receiver complexity.

Note that the above approach is commonly used in the existing literature [28]-[30] for optimizing the transmit power, training period, training sequence and so on, in both single and multiuser scenarios.

#### A. Bounds on the Capacity of the SM System

Note that (1) can also be written as  $\mathbf{y} = \sqrt{\rho}\mathbf{h}_l s + \mathbf{n}$ , where  $\rho$  represents the average received signal-to-noise ratio at each receive antenna and elements of  $\mathbf{n}$  are from  $\mathcal{CN}(0, 1)$ . Since, the transmitted symbol  $s$  as well as the antenna index  $l$  convey information in an SM system, the MI is given by

$$I(\mathbf{y}; s, l) = I(\mathbf{y}; s|l) + I(\mathbf{y}; l). \quad (19)$$

Note that the SM system's capacity was studied in [31], where neither  $I(\mathbf{y}; l)$  was evaluated nor any bounds were derived (refer to (10) in [31]). In this paper, we bound  $I(\mathbf{y}; l)$  using Jensen's inequality and obtain a closed form expression, which in turn provides a better capacity estimate for the SM system than that given by the SIMO capacity. Let  $f_s(s)$  and  $f_l(l)$  represent the p.d.f. and p.m.f. of the transmitted symbol and of the antenna index, respectively. Note that  $s \in \mathbb{C}$  and  $l$  can only assume values from the set  $\{i\}_{i=1}^{N_t}$ . Then the capacity of the SM system is given by

$$\begin{aligned} \max_{f_s(s), f_l(l)} I(\mathbf{y}; s, l) &= \max_{f_s(s), f_l(l)} \{I(\mathbf{y}; s|l) + I(\mathbf{y}; l)\}, \\ &\leq \max_{f_s(s), f_l(l)} I(\mathbf{y}; s|l) + \max_{f_s(s), f_l(l)} I(\mathbf{y}; l). \end{aligned}$$

Since,  $I(\mathbf{y}; l) = H(l) - H(l|\mathbf{y})$ , we have

$$\begin{aligned} \max_{f_s(s), f_l(l)} I(\mathbf{y}; l) &= \max_{f_s(s), f_l(l)} \{H(l) - H(l|\mathbf{y})\}, \\ &\leq \max_{f_s(s), f_l(l)} H(l) = \log N_t, \end{aligned}$$

where  $l$  is uniformly distributed over the set  $\{i\}_{i=1}^{N_t}$ . Thus, we have

$$I(\mathbf{y}; s|l) = \mathbb{E}_l [I(\mathbf{y}; s)|l] = \frac{1}{N_t} \sum_{l=1}^{N_t} [I(\mathbf{y}; s)|l].$$

For a given antenna index  $l$ , it is straightforward to show that  $I(\mathbf{y}; s)$  is maximized, when  $s$  is from the zero-mean complex-valued Gaussian distribution [32], therefore we have  $\max_{f_s(s)} I(\mathbf{y}; s|l) = \frac{1}{N_t} \sum_{l=1}^{N_t} \log(1 + \rho \|\mathbf{h}_l\|^2)$ . Thus,

$$C_{SM} = \max_{f_s(s), f_l(l)} I(\mathbf{y}; s, l) \leq \frac{1}{N_t} \sum_{l=1}^{N_t} \log(1 + \rho \|\mathbf{h}_l\|^2) + \log N_t. \quad (20)$$

By exploiting the fact that  $I(\mathbf{y}; s, l) \geq I(\mathbf{y}; s|l)$  [32], the following trivial bound can be obtained

$$C_{SM} \geq C_{SIMO} = \frac{1}{N_t} \sum_{l=1}^{N_t} \log(1 + \rho \|\mathbf{h}_l\|^2). \quad (21)$$

Thus, from (20) and (21) we have

$$\frac{1}{N_t} \sum_{l=1}^{N_t} \log(1 + \rho \|\mathbf{h}_l\|^2) \leq C_{SM} \leq \frac{1}{N_t} \sum_{l=1}^{N_t} \log(1 + \rho \|\mathbf{h}_l\|^2) + \log N_t. \quad (22)$$

The following proposition gives a non-trivial lower bound for  $C_{SM}$ .

*Proposition 2:* The capacity of an SM system relying on  $N_t$  transmit antennas can be lower bounded as

$$\begin{aligned} C_{SM} &\geq \frac{1}{N_t} \sum_{l=1}^{N_t} \log(1 + \rho \|\mathbf{h}_l\|^2) + \log(N_t) \\ &\quad - \frac{1}{N_t} \sum_{l=1}^{N_t} \log \left[ \sum_{l'=1}^{N_t} \frac{e(\|\mathbf{h}_l\|^2 \rho + 1)}{(\|\mathbf{h}_l\|^2 + \|\mathbf{h}_{l'}\|^2) \rho + 2} \right]. \end{aligned} \quad (23)$$

*Proof:* Proof is given in Appendix B.

At high SNRs the lower bound given in (23) may be approximated as

$$\begin{aligned} C_{SM} &\geq \frac{1}{N_t} \sum_{l=1}^{N_t} \log(\rho \|\mathbf{h}_l\|^2) + \log(N_t) \\ &\quad - \frac{1}{N_t} \sum_{l=1}^{N_t} \log \left[ \sum_{l'=1}^{N_t} \frac{e(\|\mathbf{h}_l\|^2)}{(\|\mathbf{h}_l\|^2 + \|\mathbf{h}_{l'}\|^2)} \right]. \end{aligned} \quad (24)$$

Thus, at high SNR the bound on the ergodic MI of the SM system is given by (25). Note that in (25), we have assumed that the SM system is operating with perfect CSIR. In the next subsection, we consider the SM system operating with minimum mean-square error (MMSE) CE and obtain the capacity bound equivalent to that given in (25).

#### B. Capacity-optimized Training in SM Systems

Let  $n$  represent the number of channel uses per transmission frame such that  $n \leq T_c$ , where  $T_c$  is the coherence period in terms of the number of channel uses. Let  $n_t$  and  $n_d$  represent the number of training and data symbols per frame, respectively, so that  $n = (n_t + n_d)$  and  $n_d \gg n_t$ . Let  $\rho_t$  and  $\rho_d$  represent the transmitted signal power during the training and data period, respectively, so that  $n\rho = (n_t\rho_t + n_d\rho_d)$ , where  $n\rho$  is the total available energy per frame. Furthermore, let  $n_t^{(i)}$  represent the training period associated with each transmit antenna, so that  $\sum_{i=1}^{N_t} n_t^{(i)} = n_t$ , while  $\mathbf{X}_t = [\mathbf{x}_t(1) \ \mathbf{x}_t(2) \ \dots \ \mathbf{x}_t(n_t)] \in \mathbb{C}^{N_t \times n_t}$  and  $\mathbf{X}_d = [\mathbf{x}_d(1) \ \mathbf{x}_d(2) \ \dots \ \mathbf{x}_d(n_d)] \in \mathbb{C}^{N_t \times n_d}$  represent the training and data sequences, respectively. The sequence received during the training phase is given by

$$\mathbf{Y}_t = \sqrt{\rho_t} \mathbf{H} \mathbf{X}_t + \mathbf{N}_t,$$

while that during the data transmission phase is given by

$$\mathbf{Y}_d = \sqrt{\rho_d} \mathbf{H} \mathbf{X}_d + \mathbf{N}_d,$$

$$C_{SM} \geq \mathbb{E} \left\{ \frac{1}{N_t} \sum_{l=1}^{N_t} \log(\rho \|\mathbf{h}_l\|^2) + \log(N_t) - \frac{1}{N_t} \sum_{l=1}^{N_t} \log \left[ \sum_{l'=1}^{N_t} \frac{e(\|\mathbf{h}_l\|^2)}{(\|\mathbf{h}_l\|^2 + \|\mathbf{h}_{l'}\|^2)} \right] \right\}. \quad (25)$$

where the elements of  $\mathbf{N}_t \in \mathbb{C}^{N_r \times n_t}$  and  $\mathbf{N}_d \in \mathbb{C}^{N_r \times n_d}$  are from  $\mathcal{CN}(0, 1)$ . The MMSE estimate of the channel matrix is given by

$$\check{\mathbf{H}} = \frac{1}{\sqrt{\rho_t}} \mathbf{Y}_t \mathbf{X}_t^H \left[ \frac{1}{\rho_t} \mathbf{I}_{N_t} + \mathbf{X}_t \mathbf{X}_t^H \right]^{-1}. \quad (26)$$

It is shown in [33], [34] that the optimal training sequence that minimizes the estimation error must satisfy the condition  $\mathbf{X}_t \mathbf{X}_t^H \propto \mathbf{I}_{N_t}$ . This may be easily satisfied in the SM system by choosing the training sequence to be a sequence of scaled diagonal matrices. Hence, we have an identical training overhead across the antennas, yielding  $n_t^{(i)} = \frac{n_t}{N_t}$  for  $1 \leq i \leq N_t$ . Note that when  $\mathbf{R}_H = \mathbb{E}[\mathbf{H}^H \mathbf{H}] \neq \mathbf{I}_{N_t}$ , that is when the channels of the transmit antennas are correlated, it was shown in [35] that the columns of the optimal training matrix must be proportional to the columns of  $\mathbf{Q}$ , where the columns of  $\mathbf{Q}$  are the Eigenvectors of  $\mathbf{R}_H$ . Thus, when the channels of transmit antennas are correlated, optimal training is not possible in the SM system due to the limitation imposed on the number of active transmit antennas per channel use. In this paper, we consider the uncorrelated fading scenario, where  $\mathbf{R}_H = \mathbf{I}_{N_t}$ .

The capacity bound in (25) was obtained by considering perfect channel knowledge at the receiver. In what follows, we determine the worst-case lower bound on the capacity of the SM system considering the estimated channel matrix instead of the actual channel matrix and obtain the optimal values of  $\rho_t$  and  $\rho_d$  that maximize the worst-case lower bound on the SM system's capacity.

The received vector in the  $i^{\text{th}}$  channel use during the data transmission phase is given by

$$\mathbf{y}_d(i) = \sqrt{\rho_d} \mathbf{H} \mathbf{x}(i) + \mathbf{n}(i), \quad (27)$$

$$= \sqrt{\rho_d} \mathbf{h}_{l_i} s_i + \mathbf{n}(i), \quad (28)$$

where  $l_i$  and  $s_i$  represent the antenna index and the transmitted symbol at the  $i^{\text{th}}$  channel use. If  $\check{\mathbf{H}} = [\check{\mathbf{h}}_1 \check{\mathbf{h}}_2 \dots \check{\mathbf{h}}_{N_t}]$  represents the MMSE estimate of the channel gain matrix, then (28) may be written in terms of the estimated channel as

$$\mathbf{y}_d(i) = \sqrt{\rho_d} \check{\mathbf{h}}_{l_i} s_i + \underbrace{\sqrt{\rho_d} \tilde{\mathbf{h}}_{l_i} s_i + \mathbf{n}(i)}_{\mathbf{n}'(i)},$$

where  $\tilde{\mathbf{h}}_{l_i} = \mathbf{h}_{l_i} - \check{\mathbf{h}}_{l_i}$  is the estimation error. Note that  $\mathbf{n}'(i)$  is not independent, but it is uncorrelated with the data, since

$$\mathbb{E} \left[ s_i \left( \sqrt{\rho_d} \tilde{\mathbf{h}}_{l_i} s_i + \mathbf{n}(i) \right)^* \right] = \sqrt{\rho_d} \mathbb{E} \left[ \tilde{\mathbf{h}}_{l_i}^* |s_i|^2 \right] + \mathbb{E} [s_i \mathbf{n}(i)^*], \\ = \mathbf{0},$$

where we have exploited the fact that the MMSE estimate is unbiased [26], i.e.,  $\mathbb{E}[\check{\mathbf{h}}_{l_i}] = \mathbf{0}$ . From *Theorem 1* of [28], it is known that the specific  $\mathbf{n}'(i)$  that gives the worst-case

capacity is a zero-mean Gaussian vector with the covariance matrix  $\sigma_{\mathbf{n}'}^2 \mathbf{I}_{N_r}$ , where

$$\sigma_{\mathbf{n}'}^2 = \frac{1}{N_r} \text{tr} \left\{ \mathbb{E} \left[ \left( \sqrt{\rho_d} \tilde{\mathbf{h}}_{l_i} s_i + \mathbf{n}(i) \right) \left( \sqrt{\rho_d} \tilde{\mathbf{h}}_{l_i} s_i + \mathbf{n}(i) \right)^H \right] \right\} \\ = \frac{1}{N_r} \text{tr} \left\{ \rho_d \mathbb{E} \left[ \tilde{\mathbf{h}}_{l_i} \tilde{\mathbf{h}}_{l_i}^H \right] + \mathbf{I}_{N_r} \right\}.$$

Note that  $\mathbb{E}[\tilde{\mathbf{h}}_{l_i} \tilde{\mathbf{h}}_{l_i}^H]$  is the covariance matrix of the MMSE, which can be shown to be  $\frac{1}{(1+\rho_t n_t^{(l_i)})} \mathbf{I}_{N_r}$ . Details of this derivation can be found in Appendix C. Thus, we have  $\sigma_{\mathbf{n}'}^2 = \frac{1}{N_r} \text{tr} \left\{ \rho_d \mathbb{E} \left[ \tilde{\mathbf{h}}_{l_i} \tilde{\mathbf{h}}_{l_i}^H \right] + \mathbf{I}_{N_r} \right\} = \frac{\rho_d}{(1+\rho_t n_t^{(l_i)})} + 1$ . Now, replacing the true channel in (23) by the estimated channel  $\check{\mathbf{h}}_i$ , the SNR  $\rho$  by  $\frac{\rho_d}{\sigma_{\mathbf{n}'}^2}$  and taking into account that only  $(n - n_t)$  out of  $n$  channel uses are dedicated to data transmission, we have the worst case capacity bound given by (29) (in the next page), where  $C_{SM}^{imp}$  represents the bound on the capacity of the SM system operating with *imperfect CSIR*.

At high SNR,  $C_{SM}^{imp}$  can be approximated (30). Exploiting that  $\check{\mathbf{h}}_l = \frac{1}{\sigma_{\check{\mathbf{h}}_l}} \check{\mathbf{h}}_l \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_r})$ , where  $\sigma_{\check{\mathbf{h}}_l}^2 = \frac{\rho_t n_t^{(l)}}{1+\rho_t n_t^{(l)}}$ , (30) can be written as (31). Since  $\frac{\sigma_{\check{\mathbf{h}}_l}^2}{\sigma_{\mathbf{n}'}^2} = \frac{\rho_t n_t^{(l)}}{1+\rho_d+\rho_t n_t^{(l)}}$  and  $\frac{\sigma_{\check{\mathbf{h}}_l}^2}{\sigma_{\check{\mathbf{h}}_{l'}}^2} = \frac{n_t^{(l)}(1+\rho_t n_t^{(l')})}{n_t^{(l')(1+\rho_t n_t^{(l)})} \approx 1$  for a large  $\rho_t$  and any  $(l, l')$  pair, (31) can be written as (32).

Note that the capacity bounds given in (25) and (32) are identical, except for  $\rho_{eff} = \frac{\rho_d \rho_t n_t^{(l)}}{1+\rho_d+\rho_t n_t^{(l)}}$  that captures the SNR reduction due to imperfect CSIR and for the factor  $\frac{n-n_t}{n}$ , which captures the reduction in the number of channel uses available for data transmission owing to the training overhead. Taking  $n_t^{(i)} = n_t/N_t$  for  $1 \leq i \leq N_t$ , corresponding to equal training overheads across all transmit antennas, we arrive at  $\rho_{eff} = \frac{\rho_d \rho_t n_t}{N_t(1+\rho_d)+\rho_t n_t}$ . Note that under equal training overheads across all transmit antennas results in a  $\rho_{eff}$ , which is the same as that of the conventional MIMO system, where all the  $N_t$  transmit antennas are activated in any channel use [28]. Let  $\beta$  define the fraction of total energy allocated for data transmission and  $(1 - \beta)$  represent the fraction of total energy allocated for training, which results in  $\rho_d n_d = \beta \rho n$  and  $\rho_t n_t = (1 - \beta) \rho n$  for  $0 < \beta < 1$ . The optimal  $\beta$  that maximizes (32) is obtained by substituting both  $\rho_d = \frac{\beta \rho n}{n_d}$  as well as  $\rho_t = \frac{(1-\beta)\rho n}{n_t}$  in  $\rho_{eff}$  and then maximizing it over  $\beta$ , which results in  $\beta_{opt} = \delta - \sqrt{\delta(\delta - 1)}$ , where  $\delta = \frac{N_t + \rho n}{\rho n (1 - \frac{N_t}{n_d})}$ .

Using optimal power splitting between the data and training sequences, it may be shown that the optimal length of the training sequence for the SM system is  $n_{t,opt} = N_t$ . Since the proof of this result follows similar lines to that of *Theorem 3* in [28], we omit the details of this part.

When the coherence intervals of the links of the various transmit antennas differ significantly or when the SM system is

$$C_{SM}^{imp} \geq \mathbb{E} \left\{ \frac{n - n_t}{n} \left( \frac{1}{N_t} \sum_{l=1}^{N_t} \log(1 + \frac{\rho_d}{\sigma_{\mathbf{n}'}^2} \|\check{\mathbf{h}}_l\|^2) + \log(N_t) - \frac{1}{N_t} \sum_{l=1}^{N_t} \log \left[ \sum_{l'=1}^{N_t} \frac{e(\|\check{\mathbf{h}}_l\|^2 \frac{\rho_d}{\sigma_{\mathbf{n}'}^2} + 1)}{(\|\check{\mathbf{h}}_l\|^2 + \|\check{\mathbf{h}}_{l'}\|^2) \frac{\rho_d}{\sigma_{\mathbf{n}'}^2} + 2} \right] \right) \right\} \quad (29)$$

$$C_{SM}^{imp} \geq \mathbb{E} \left\{ \frac{n - n_t}{n} \left( \frac{1}{N_t} \sum_{l=1}^{N_t} \log(1 + \frac{\rho_d}{\sigma_{\mathbf{n}'}^2} \|\check{\mathbf{h}}_l\|^2) + \log(N_t) - \frac{1}{N_t} \sum_{l=1}^{N_t} \log \left[ \sum_{l'=1}^{N_t} \frac{e(\|\check{\mathbf{h}}_l\|^2)}{(\|\check{\mathbf{h}}_l\|^2 + \|\check{\mathbf{h}}_{l'}\|^2)} \right] \right) \right\} \quad (30)$$

$$C_{SM}^{imp} \geq \mathbb{E} \left\{ \frac{n - n_t}{n} \left( \frac{1}{N_t} \sum_{l=1}^{N_t} \log(1 + \frac{\rho_d \sigma_{\mathbf{h}_l}^2}{\sigma_{\mathbf{n}'}^2} \|\bar{\mathbf{h}}_l\|^2) + \log(N_t) - \frac{1}{N_t} \sum_{l=1}^{N_t} \log \left[ \sum_{l'=1}^{N_t} \frac{e}{1 + \frac{\|\bar{\mathbf{h}}_{l'}\|^2 \sigma_{\mathbf{h}_l}^2}{\|\bar{\mathbf{h}}_l\|^2 \sigma_{\mathbf{h}_{l'}}^2}} \right] \right) \right\} \quad (31)$$

$$C_{SM}^{imp} \geq \mathbb{E} \left\{ \frac{n - n_t}{n} \left( \frac{1}{N_t} \sum_{l=1}^{N_t} \log \left( 1 + \underbrace{\frac{\rho_d \rho_t n_t^{(l)}}{1 + \rho_d + \rho_t n_t^{(l)}}}_{\rho_{eff}} \|\bar{\mathbf{h}}_l\|^2 \right) + \log(N_t) - \frac{1}{N_t} \sum_{l=1}^{N_t} \log \left[ \sum_{l'=1}^{N_t} \frac{e \|\bar{\mathbf{h}}_l\|^2}{\|\bar{\mathbf{h}}_l\|^2 + \|\bar{\mathbf{h}}_{l'}\|^2} \right] \right) \right\} \quad (32)$$

equipped with redundant transmit antennas and it is employing a selection algorithm, we may have to have different  $n_t^{(i)}$  for  $1 \leq i \leq N_t$ . The detailed study of this is postponed for future research.

## V. SIMULATION RESULTS AND DISCUSSIONS

In the first part of this section, we present the results of our numerical simulations for characterizing the performance of the proposed HL-ML detector with respect to that of the exhaustive search based ML detector. Furthermore, we study the computational complexity of the proposed detector in comparison to various detectors proposed in the existing literature. In the second part, we study the capacity of the SM system operating with imperfect CSIR and the resultant capacity improvement due to the proposed optimal power splitting over that of equal power splitting between the training and data symbols. We also compare the performance of the SM system employing the iterative detection/estimation algorithm of [19] to that of the SM system employing optimal power sharing, when both are operating with imperfect CSIR.

### A. Proposed HL-ML detector - Performance and Complexity analysis

**Simulation Scenario 1:** In all our simulations we have assumed a frequency-flat block Rayleigh fading channel and used at least  $10^{t+1}$  symbols at a Symbol Error Rate (SER) of  $10^{-t}$ . The receiver is assumed to have perfect CSIR, when employing both the proposed and the exhaustive search based ML detector.

Considering an SM system having  $N_t = 4$ ,  $N_r = 2$  and employing 16-, 32-, 64- and 128-QAM signal sets, Fig. 1 gives the SER comparison of the proposed HL-ML detector to that of the exhaustive search based ML detector. It is clear

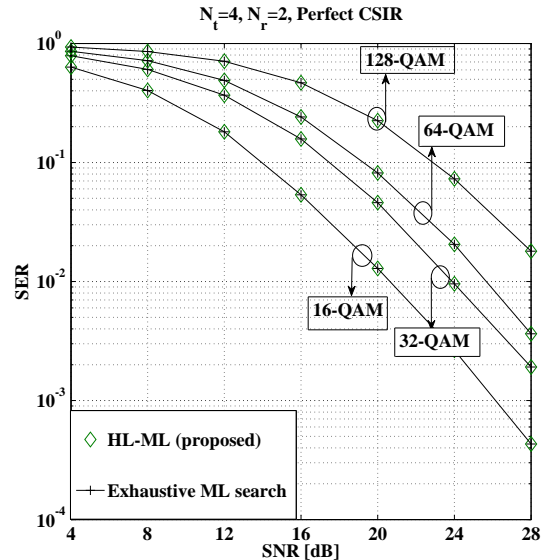


Fig. 1. SER comparison of the proposed HL-ML detector with that of the exhaustive search based ML detector in an SM system having  $N_t = 4$ ,  $N_r = 2$  and employing 16-, 32-, 64- and 128-QAM signal sets in the presence of Rayleigh fading and perfect CSIR.

from Fig. 1 that the proposed detector attains an identical performance compared to that of the exhaustive search based ML detector for all SNR points and data rates, which validates our *Proposition 1*. Note that the existing detectors listed in Table I are sub-optimal and hence their SER performance is expected to be poorer than that of the exhaustive search based ML detector.

Fig. 2 gives the comparison of the computational complexity per symbol detection of the proposed HL-ML detector to that of the various existing detectors as a function of the signal

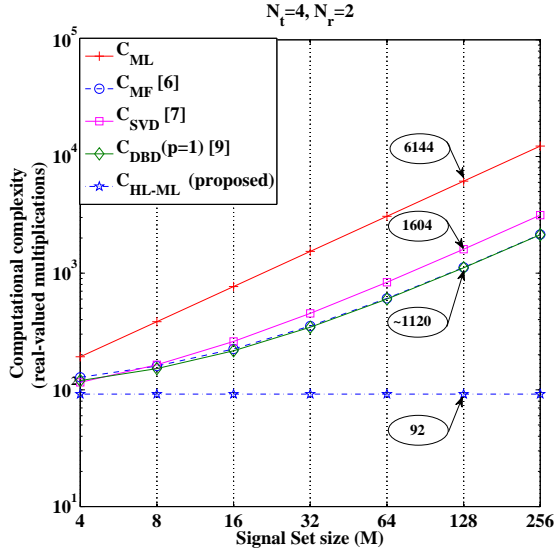


Fig. 2. Comparison of the computational complexity per symbol detection of the proposed detector with various existing detectors as a function of the signal set size  $M$ .

set size  $M$ . It is clear from Fig. 2 that the complexity of the proposed detector does not grow with  $M$ . This property is very desirable in closed-loop operations such as adaptive modulation or adaptive modulation and antenna selection, where the modulation order is changed suitably based on the limited feedback information sent by the receiver. Furthermore, it is easy to see from Fig. 2 that the proposed HL-ML detector imposes a significantly reduced complexity compared to that of the existing detectors. More specifically, when using the 128-QAM signal set, the computational complexity of the HL-ML detector is about 98.5% lower than that of the exhaustive search based ML detector, about 94% lower than that of the MF based detector, and about 91% lesser compared to that of SVD [7] and DBD ( $p = 1$ ) [9].

Fig 3 gives a comparison of the proposed detector's complexity to various existing detectors as a function of  $N_t$  for fixed values of  $M = 16$  and  $N_r = 2, 8$ . It is clear from both the plots of Fig. 3 that the proposed detector imposes the lowest complexity among the existing detectors. More specifically, when considering  $N_r = 2$  and  $N_t = 8$  (refer to Plot (a) in Fig. 3), the computational complexity of the proposed detector is about 88% lower compared to that of the exhaustive search based ML detector, about 56% lower than that of the MF and DBD, and about 43% lower compared to that of the SVD. When considering  $N_r = 8$  and  $N_t = 4$  (refer to Plot (b) in Fig. 3), the computational complexity of the proposed detector is about 92% lower than that of the exhaustive search based ML detector, about 90% lower compared to that of the SD, about 76% lower compared to that of the SVD, about 54% lower compared to that of the MF based detector and about 46% lower compared to that of the DBD. Note that the true computational complexity of the SD [15] is higher than that considered ( $C_{fix}$ ). Thus, we emphasize that using a SD is not essential in the SM system for achieving an ML performance

at a reduced complexity.

### B. Capacity optimized training in the SM system

**Simulation Scenario 2:** In all our simulations the imperfect CSIR scenario relies on the MMSE CE obtained using an  $n_t$ -length training block, where the training block used is a set of scaled identity matrices. In evaluating the ergodic MI, we have considered at least 5000 channel realizations and a coherence interval of  $n = 100$  channel uses. As before, we have used at least  $10^{t+1}$  symbols at a SER of  $10^{-t}$ .

Fig. 4 gives a comparison of the ergodic MI as a function of SNR in an SM system operating with both perfect and estimated channel information. The trivial upper and lower bounds on the capacity of the SM system given in (22) are also plotted for comparison. The SM system is assumed to have  $N_t = 4$ ,  $n = 100$ ,  $n_t = 4$  and  $N_r = 1, 2$ . Plot (a) and Plot (b) in Fig. 4 correspond to  $N_r = 1$  and  $N_r = 2$ , respectively. It is readily seen from both the plots of Fig. 4 that

- the SM system suffers from a significant MI loss when operating with imperfect CSIR. Specifically, at an SNR of 20 dB, with equal power allocation between data and training symbols, there is a reduction in MI of about 16% for  $N_r = 1$  and 17% for  $N_r = 2$ . and with optimal power sharing, the MI reduction due to imperfect CSIR is observed to be about 7% for  $N_r = 1$  and 8% for  $N_r = 2$ .

In Fig. 4 we have considered the frame length (or equivalently the coherence interval) to be 100 channel uses. Now, considering a fixed SNR of 20 dB, we study the MI of the SM system operating with imperfect CSIR as a function of the frame length for both the equal and optimal power sharing scenarios. Fig. 5 gives a comparison of the ergodic MI in an SM system for  $N_t = 4$ ,  $n_t = 4$ ,  $N_r = 1, 2$ , and employing both equal and optimal power splitting between the data and training symbols. Plot (a) and Plot (b) in Fig. 5 correspond to  $N_r = 1$  and  $N_r = 2$ , respectively.

- It is clear from both the plots of Fig. 5 that equal power splitting suffers from a higher MI reduction compared to that of the optimal splitting. Specifically, at a frame length of 200 channel uses, equal power splitting suffers from about 16% reduction for  $N_r = 1$  and about 14% reduction for  $N_r = 2$ , whereas the optimal power splitting results in about 7% reduction for both  $N_r = 1$  and  $N_r = 2$ . Thus, optimal power splitting mitigates the MI reduction due to imperfect CSIR by more than 50% compared to equal power splitting.
- Also, it is clear from Fig. 5 that although a higher frame length (or coherence interval) yields a higher MI, the gain in MI diminishes as the frame length increases, which is more prominent in case of equal power sharing than for optimal power allocation.

Plot (a) in Fig. 6 quantifies the power allocated for training and data blocks as a function of the SNR, while Plot (b) in Fig. 6 gives the mean channel estimation error as a function of SNR, when considering both equal and optimal power sharing between data and training blocks. It is clear from



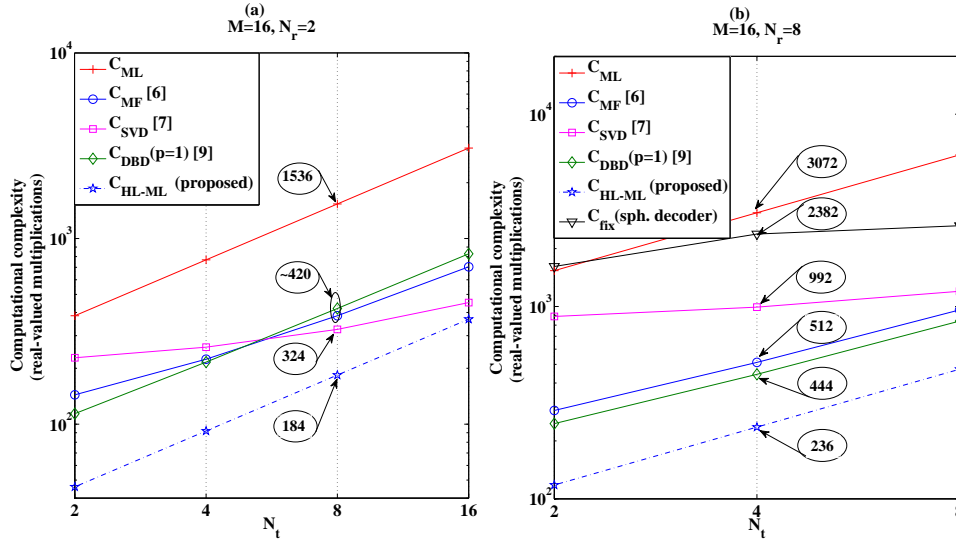


Fig. 3. Comparison of the computational complexity per symbol detection of the proposed detector with the various existing detectors as a function of number of transmit antennas  $N_t$  when the signal set size is fixed to be  $M = 16$ . Plot (a) corresponds to  $N_r = 2$  case and Plot (b) corresponds to that of  $N_r = 8$ .

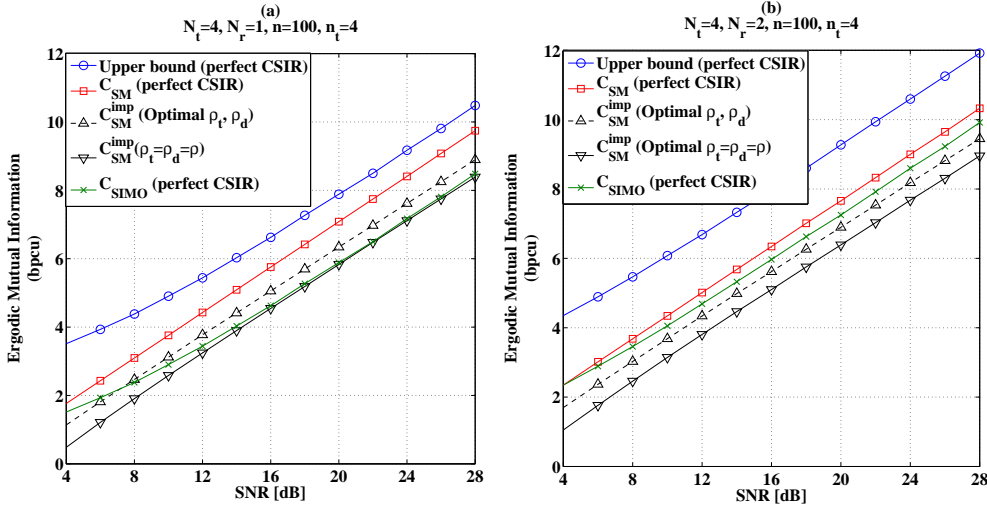


Fig. 4. Comparison of the ergodic MI as a function of SNR in the SM system operating with perfect CSIR with that of the SM system operating with imperfect CSIR and employing equal and optimal power splitting between training and data symbols. The SM system is assumed to have  $N_t = 4, n_t = 4, n = 100$  and  $N_r = 1, 2$ . Plot (a) corresponds to  $N_r = 1$  case and Plot (b) corresponds to that of  $N_r = 2$ .

Plot (a) in Fig. 6 that the power allocated for the training block is significantly higher than that allocated for data blocks. Plot (b) in Fig. 6 gives the mean channel estimation error for various training periods, when considering equal power sharing between the data and training blocks.

- It is clear from this plot that the optimal allocation using a minimal training length of  $n_t = 4$  gives a better CE than that of the equal power allocation associated with  $n_t = 12$ . Therefore, optimal power sharing reduces the loss in the data throughput due to lower training overhead and with a better CE, can be expected to give a better SER performance.

Let us now consider an SM system operating with imperfect CSIR and employing a conventional QAM signal set, and then

compare the SER performance offered by the iterative data-detection/channel-estimation algorithm [19] to that offered by the SM system employing optimal power splitting. In order to benchmark the performance of the SM system having imperfect CSIR against that of the conventional MIMO system, the SER performance of the Alamouti code [36] is provided, considering both cases of equal and optimal power sharing [28].

Note that, although it is essential to consider a coded system in order to compare the system performance to the fundamental capacity/achievable MI limits, we restrict our study in this paper to an uncoded system, since our focus is on achieving same performance as that of the system operating with perfect CSIR, without aiming for approaching the system

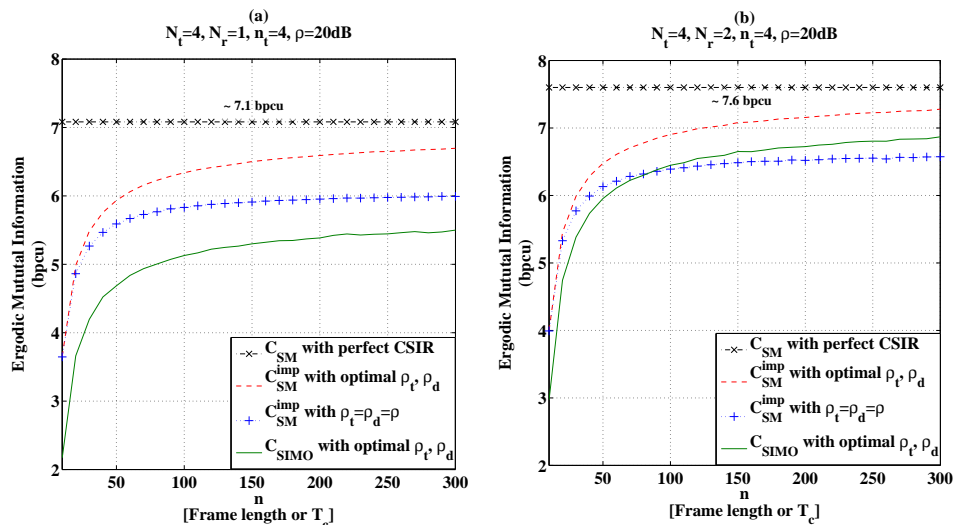


Fig. 5. Comparison of the ergodic MI as a function of frame length  $n$  (or coherence interval) in the SM system operating with perfect CSIR with that of the SM system operating with imperfect CSIR and employing equal and optimal power splitting between training and data symbols.

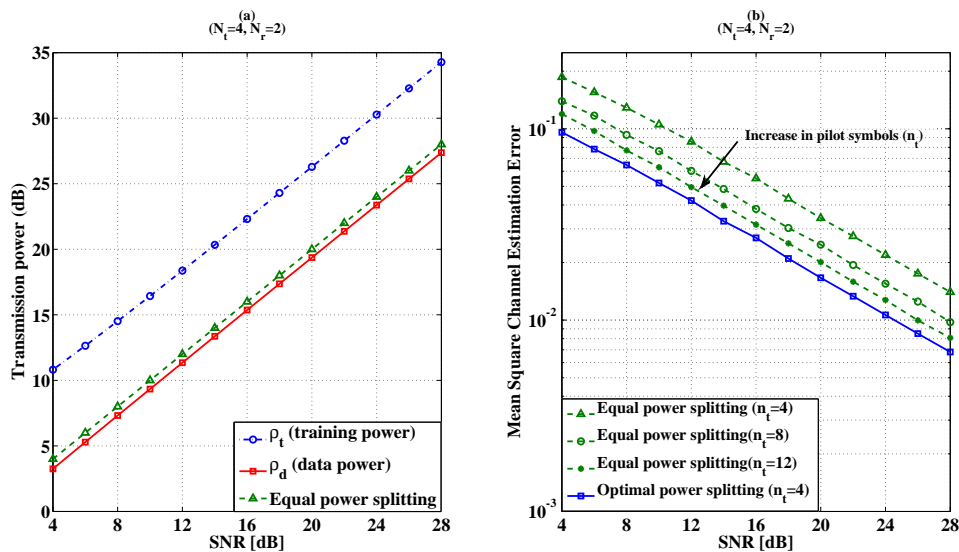


Fig. 6. Plot (a) gives the capacity optimized power splitting in an SM system having  $n = 100$ ,  $n_t = N_t = 4$  and  $N_r = 2$ . Plot (b) compares the mean squared error of the estimated channel in optimal power sharing scenario with that of the equal power sharing scenario for various training lengths.

capacity. Considering various throughputs, Fig. 7 gives the SER curves of the

- 1) SM system operating with perfect CSIR;
- 2) SM system operating with imperfect CSIR and employing the proposed capacity-optimized power splitting;
- 3) SM system operating with imperfect CSIR and employing equal power splitting;
- 4) SM system operating with imperfect CSIR and employing the iterative data-detection/channel-estimation algorithm of [19];
- 5) Alamouti code with perfect CSIR;
- 6) Alamouti code with imperfect CSIR, and
- 7) Alamouti code with imperfect CSIR and employing optimal power splitting between data and training blocks [28].

It is clear from Fig. 7 that both the SM system as well as the Conventional MIMO (CMIMO) system employing the Alamouti code suffer from a significant performance loss compared to their perfect CSIR counterparts. Specifically, at an SER of about  $10^{-2}$  the SM system suffers about 2 dB, 2.5 dB and 2.5 dB SNR loss, when operating at 6, 7 and 8 bpcu, respectively, and the system using Alamouti code suffers from about 5 dB SNR loss, when operating at the same throughputs as that of the SM system.

- The SM system is more robust to channel imperfections compared to that of the conventional MIMO system.
- It is clear from Fig. 7 that employing optimal power splitting between data and training blocks significantly reduces the performance loss owing to imperfect CSIR in both the SM and CMIMO systems.

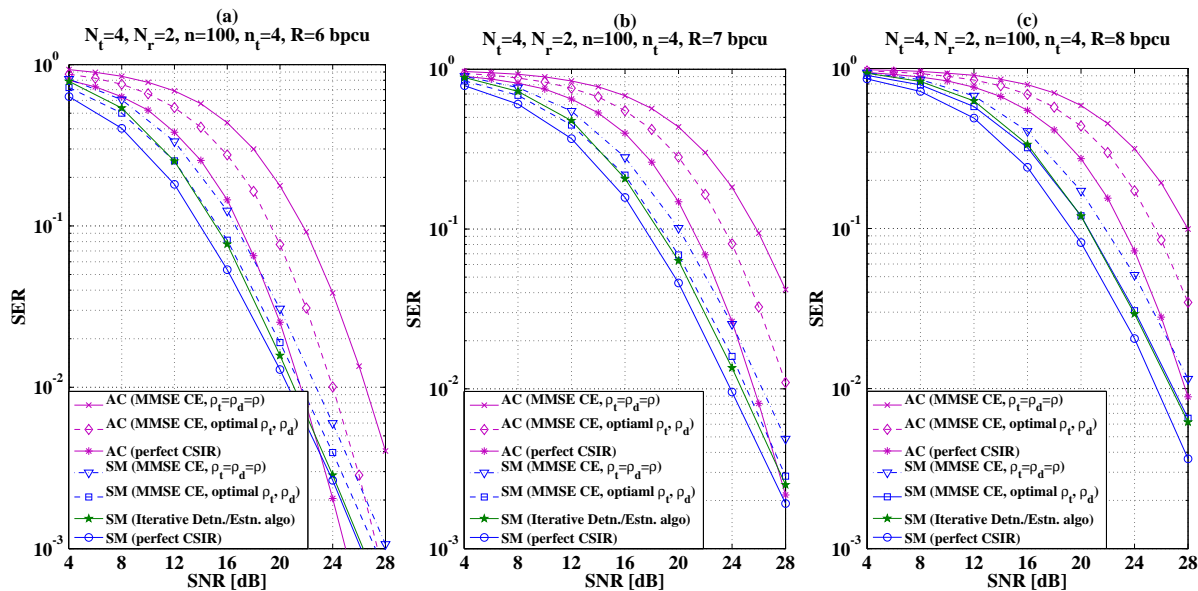


Fig. 7. SER performance comparison of the SM and the Alamouti code (AC) under imperfect CSIR condition. The performance of the iterative data-detection/channel-estimation algorithm [19] in SM system is also provided. The SM system is assumed to have  $N_t = 4$ ,  $N_r = 2$ ,  $n = 100$  and  $n_t = 4$ .

- As the throughput increases, the SER performance of the SM system using imperfect CSIR and employing optimal power splitting coincides with that of the iterative detection/estimation algorithm.

Therefore, the proposed capacity-optimized power splitting enables us to achieve almost the same performance as that of the algorithm in [19] without any additional computational burden at the receiver.

It is evident from Fig. 7 that the SM system operating with imperfect CSIR and employing optimal power splitting gives a better SER performance compared to its CMIMO counterpart in the low and medium SNR region. Specifically, at an SER of about  $10^{-2}$ , it is observed that the SM system gives an SNR gain of about 2.5 dB, 3 dB and 3.5 dB at 6, 7 and 8 bpcu, respectively, with respect to the CMIMO system employing Alamouti's code. However, at high SNR region the Alamouti's code gives better SER performance owing to a higher diversity order.

## VI. CONCLUSIONS

In this paper, we have proposed a hard-limiting based ML-optimal detector for the SM system whose computational complexity does not grow with the size of the signal set, provided that the employed signal set is either square- or rectangular-QAM. The proposed detector is shown to have the lowest computational complexity among the various existing detectors proposed for SM systems, including that of SD. We have considered the SM system operating with imperfect CSIR and obtained a non-trivial lower bound on its capacity. The power allocation between data and training blocks is optimized by maximizing this lower bound, and the resultant capacity improvement and the quality of the estimated channel at the receiver is studied. The SER performance of the SM system using both the equal and the proposed optimal power

allocation is compared to that of the SM system operating with the aid of the iterative data-detection/channel-estimation algorithm [19] and of the conventional MIMO system employing Alamouti's code. The SM system employing the proposed optimal power splitting is observed to give a SER performance which is the same as that of the SM system using the iterative data-detection/channel-estimation algorithm of [19] and significantly outperforms the conventional MIMO system employing Alamouti's code.

## VII. APPENDIX A

### COMPUTATIONAL COMPLEXITY OF VARIOUS DETECTORS CONCEIVED FOR THE SM SYSTEM

We provide a brief description of the computational complexity of various existing detectors given in Table 1.

#### A. MF Based Detector [6]

Consider (35) of [6], which gives the computational complexity of the MF based detector, formulated as:

$$C = \frac{(4MNTQ + 4NTQ + QL')/\tau + 4NTQ + 3QL' + 4L}{\log_2(QL)} \quad (33)$$

Neglecting the normalization term  $\log_2(QL)$  and considering  $Q = N_t$ ,  $M = 1$ ,  $N = N_r$ ,  $T = 1$ ,  $\tau = 1$ ,  $L' = M/4 = |S|/4$ , and  $L = M = |S|$  which correspond to the SM system having  $N_t$  transmit as well as  $N_r$  receive antennas and employing a signal set  $S$  having  $|S| = M$ , we arrive at  $C_{MF} = 12N_rN_t + (N_t + 4)M$ .

### B. SVD [7]

Consider (5) of [7], which is given by

$$k = \arg \min_j \theta(j), \text{ where } \theta_j = \arccos \frac{|\mathbf{h}_j^H \mathbf{y}|}{\|\mathbf{h}_j\| \|\mathbf{y}\|} \quad (34)$$

and

$$\hat{s} = \arg \min_{s \in S} \|\mathbf{y} - \mathbf{h}_k s\|^2. \quad (35)$$

The computational complexity imposed by the above equations is summarized as follows:

- 1)  $\mathbf{h}_j^H \mathbf{y}$  takes  $4N_r$  real-valued multiplications.
- 2)  $|\mathbf{h}_j^H \mathbf{y}| = \sqrt{[\Re(\mathbf{h}_j^H \mathbf{y})]^2 + [\Im(\mathbf{h}_j^H \mathbf{y})]^2}$  takes 2 real-valued multiplications.
- 3)  $\|\mathbf{h}_j\| = \sqrt{\sum_{k=1}^{N_r} \Re(h_{j,k})^2 + \Im(h_{j,k})^2}$  takes  $2N_r$  real-valued multiplications.
- 4)  $\|\mathbf{y}\| = \sqrt{\sum_{k=1}^{N_r} \Re(y_k)^2 + \Im(y_k)^2}$  takes  $2N_r$  real-valued multiplications.
- 5) The product  $\|\mathbf{y}\| \|\mathbf{h}_j\|$  takes 1 real-valued multiplication and the division  $\frac{|\mathbf{h}_j^H \mathbf{y}|}{\|\mathbf{h}_j\| \|\mathbf{y}\|}$  is considered as 1 multiplication.

Thus, the number of real-valued multiplications involved in obtaining  $k$  is  $(4N_r + 2N_r + 2 + 1)N_t + 2N_r = (6N_r + 4)N_t + 2N_r$ . Note that  $\|\mathbf{y}\|$  needs to be computed only once. Furthermore, for a given  $s$ ,  $\mathbf{h}_k s$  takes  $4N_r$  real-valued multiplications and  $\|\mathbf{y} - \mathbf{h}_k s\|^2 = \|\mathbf{b}\|^2 = \sum_{i=1}^{N_r} [\Re(b_i)]^2 + [\Im(b_i)]^2$  takes  $2N_r$  real-valued multiplications. Thus, the overall complexity is given by  $C_{SVD} = (6N_r + 4)N_t + 2N_r + 6N_r M$ , since  $|S| = M$ . Note that the complexity involved in computing the square root operations in 2) to 4) (above) and the arccos operation, which has to be computed  $N_t$  number of times for obtaining  $k$  are not considered in the complexity analysis.

### C. DMF Based Detector [8]

By referring to Section II-B of [8] it is clear that the computational complexity of DMF based detector is lower bounded by that of the MF based detector. Thus, we omit the details of the complexity analysis of this detector.

### D. DBD [9]

Consider (28) of [9], which is given by

$$C_{DBD} = \frac{(T/\tau + T + 3N_t + 6N_r p)}{L}, \quad (36)$$

where  $T = 4N_r N_t$ ,  $1 \leq p \leq N_t$ ,  $L = \log_2(N_t M)$  and  $\tau$  is the coherence interval. Note that (36) does not include the computational complexity involved in the classic symbol demodulation, which is given by  $2N_t M$  real-valued multiplications (refer to (5) of [9]). Thus, considering the coherence interval  $\tau = 1$  and neglecting the normalization term  $L$ , the overall complexity of the DBD is  $C_{DBD} = 8N_r N_t + 3N_t + 6N_r p + 2N_t M$ .

### E. Proposed HL-ML detector

For a given  $l$ , the computation of  $\hat{y}_l = \frac{\mathbf{h}_l^H \mathbf{y}}{\|\mathbf{h}_l\|^2}$  takes  $6N_r + 2$  real-valued multiplications. Since

- 1)  $\mathbf{h}^H \mathbf{y}$  takes  $4N_r$  real-valued multiplications,
- 2)  $\|\mathbf{h}_l\|^2 = \sum_{k=1}^{N_r} \Re(h_{l,k})^2 + \Im(h_{l,k})^2$  takes  $2N_r$  real-valued multiplications, and
- 3) the divisions  $\frac{\Re(\mathbf{h}^H \mathbf{y})}{\|\mathbf{h}_l\|^2}$  and  $\frac{\Im(\mathbf{h}^H \mathbf{y})}{\|\mathbf{h}_l\|^2}$  are considered as two multiplications, which results in a total complexity of  $6N_r + 2$  operations.

Since  $l$  varies from 1 to  $N_t$ , the computation of the set  $\{\hat{y}_l\}_{l=1}^{N_t}$  takes  $(6N_r + 2)N_t$  real-valued multiplications.

Now consider (8), which is of the form

$$\Re(\hat{s}_l) = \min(a, N_1 - 1),$$

$$\text{where } a = \max(b, -N_1 + 1) \text{ and } b = 2 \left\lfloor \frac{u_1 + 1}{2} \right\rfloor - 1, \quad (37)$$

$u_1 = \Re(\hat{y}_l)$  for a given  $l$ . It is clear from (37) that the computation of  $b$  takes 2 real-valued multiplications and the computation of  $\max(b, -N_1 + 1)$  and  $\min(a, N_1 - 1)$  involve no multiplications. Furthermore, it is straightforward to see from (37) that the complexity involved in obtaining  $\Re(\hat{s}_l)$  and  $a$  does not depend on the particular value that  $N_1$  takes. Similarly, the computation of  $\Im(\hat{s}_l)$  takes 2 real-valued multiplications. Thus,  $\hat{s}_l = \Re(\hat{s}_l) + j\Im(\hat{s}_l)$  takes 4 real-valued multiplications. Therefore, the computation of the set  $S_{min} = \{\hat{s}_i\}_{i=1}^{N_t}$  takes  $4N_t$  real-valued multiplications.

Considering (16), for a given  $l$  the computation of

- 1)  $|\hat{y}_l|^2 = \Re(\hat{y}_l)^2 + \Im(\hat{y}_l)^2$  takes 2 real-valued multiplications,
- 2)  $|\hat{y}_l - \hat{s}_l|^2 = \Re(\hat{y}_l - \hat{s}_l)^2 + \Im(\hat{y}_l - \hat{s}_l)^2$  takes 2 real-valued multiplications, and
- 3) the product of  $|\hat{y}_l - \hat{s}_l|^2 - |\hat{y}_l|^2$  and  $\|\mathbf{h}\|^2$  takes 1 real-valued multiplication.

Note that  $\|\mathbf{h}_l\|^2$  has already been computed while obtaining  $\hat{y}_l$ . Thus, for a given  $l$  the computation of (16) takes 5 real-valued multiplications. Again, since  $l$  ranges from 1 to  $N_t$  the overall complexity in computing (16) is  $5N_t$ , the computational complexity of the proposed detector is  $C_{HL-ML} = (6N_r + 2)N_t + 4N_t + 5N_t = (6N_r + 11)N_t$ .

## VIII. APPENDIX B PROOF OF PROPOSITION 2

A non-trivial lower bound for  $C_{SM}$  is derived as follows. From (19), we have

$$I(\mathbf{y}; l) = H(l) - H(l|\mathbf{y}) = \log(N_t) - H(l|\mathbf{y}), \text{ and}$$

$$H(l|\mathbf{y}) = \sum_{l=1}^{N_t} \int_{y'} f_{l,y'}(l, y') \log \frac{1}{f_{l,y'}(l|y')} dy', \quad (38)$$

where for a given  $l$  we have  $y' = \sqrt{\rho}\|\mathbf{h}_l\|s + n'$  and  $n' \sim \mathcal{CN}(0, 1)$ . Using Bayes' rule, (38) can be written as

$$\begin{aligned} H(l|\mathbf{y}) &= \sum_{l=1}^{N_t} \int_{y'} f_{y'|l}(y'|l) f_l(l) \log \frac{\sum_{l'=1}^{N_t} f_{y'|l'}(y'|l')}{f_{y'|l}(y'|l)} dy', \\ &= \frac{1}{N_t} \sum_{l=1}^{N_t} \int_{y'} f_{y'|l}(y'|l) \log \frac{1}{f_{y'|l}(y'|l)} dy' + \\ &\quad \frac{1}{N_t} \sum_{l=1}^{N_t} \int_{y'} f_{y'|l}(y'|l) \log \left( \sum_{l'=1}^{N_t} f_{y'|l'}(y'|l') \right) dy', \\ &= \frac{1}{N_t} \sum_{l=1}^{N_t} \log [\pi e (1 + \|\mathbf{h}_l\|^2 \rho)] + \\ &\quad \frac{1}{N_t} \sum_{l=1}^{N_t} \int_{y'} \frac{1}{\pi \sigma_l^2} e^{-\frac{|y'|^2}{\sigma_l^2}} \log \left( \sum_{l'=1}^{N_t} \frac{1}{\pi \sigma_{l'}^2} e^{-\frac{|y'|^2}{\sigma_{l'}^2}} \right) dy', \end{aligned}$$

where  $\sigma_l^2 = \rho\|\mathbf{h}_l\|^2 + 1$ . Furthermore, changing the variable  $y'$  to  $y'_l$  for ease of presentation, we can write

$$\begin{aligned} &\frac{1}{N_t} \sum_{l=1}^{N_t} \int_{y'} \frac{1}{\pi \sigma_l^2} e^{-\frac{|y'|^2}{\sigma_l^2}} \log \left( \sum_{l'=1}^{N_t} \frac{1}{\pi \sigma_{l'}^2} e^{-\frac{|y'|^2}{\sigma_{l'}^2}} \right) dy' \\ &= \frac{1}{N_t} \sum_{l=1}^{N_t} \mathbb{E}_{y'_l} \left[ \log \left( \sum_{l'=1}^{N_t} \frac{1}{\pi \sigma_{l'}^2} e^{-\frac{|y'_l|^2}{\sigma_{l'}^2}} \right) \right]. \quad (39) \end{aligned}$$

From Jensen's inequality we have  $\mathbb{E}[\log(X)] \leq \log(\mathbb{E}[X])$ , thus, (39) can be bounded as

$$\begin{aligned} &\frac{1}{N_t} \sum_{l=1}^{N_t} \mathbb{E}_{y'_l} \left[ \log \left( \sum_{l'=1}^{N_t} \frac{1}{\pi \sigma_{l'}^2} e^{-\frac{|y'_l|^2}{\sigma_{l'}^2}} \right) \right] \\ &\leq \frac{1}{N_t} \sum_{l=1}^{N_t} \log \left[ \mathbb{E}_{y'_l} \left( \sum_{l'=1}^{N_t} \frac{1}{\pi \sigma_{l'}^2} e^{-\frac{|y'_l|^2}{\sigma_{l'}^2}} \right) \right], \\ &= \frac{1}{N_t} \sum_{l=1}^{N_t} \log \left[ \sum_{l'=1}^{N_t} \frac{1}{\pi (\|\mathbf{h}_l\|^2 + \|\mathbf{h}_{l'}\|^2) \rho + 2\pi} \right]. \end{aligned}$$

Thus,

$$\begin{aligned} H(l|\mathbf{y}) &\leq \frac{1}{N_t} \sum_{l=1}^{N_t} \log [\pi e (1 + \|\mathbf{h}_l\|^2 \rho)] + \\ &\quad \frac{1}{N_t} \sum_{l=1}^{N_t} \log \left[ \sum_{l'=1}^{N_t} \frac{1}{\pi (\|\mathbf{h}_l\|^2 + \|\mathbf{h}_{l'}\|^2) \rho + 2\pi} \right], \\ &= \frac{1}{N_t} \sum_{l=1}^{N_t} \log \left[ \sum_{l'=1}^{N_t} \frac{e(\|\mathbf{h}_l\|^2 \rho + 1)}{(\|\mathbf{h}_l\|^2 + \|\mathbf{h}_{l'}\|^2) \rho + 2} \right]. \end{aligned}$$

Therefore, we have

$$I(\mathbf{y}; l) \geq \log(N_t) - \frac{1}{N_t} \sum_{l=1}^{N_t} \log \left[ \sum_{l'=1}^{N_t} \frac{e(\|\mathbf{h}_l\|^2 \rho + 1)}{(\|\mathbf{h}_l\|^2 + \|\mathbf{h}_{l'}\|^2) \rho + 2} \right],$$

and

$$\begin{aligned} C_{SM} &\geq \frac{1}{N_t} \sum_{l=1}^{N_t} \log(1 + \rho\|\mathbf{h}_l\|^2) + \log(N_t) \\ &\quad - \frac{1}{N_t} \sum_{l=1}^{N_t} \log \left[ \sum_{l'=1}^{N_t} \frac{e(\|\mathbf{h}_l\|^2 \rho + 1)}{(\|\mathbf{h}_l\|^2 + \|\mathbf{h}_{l'}\|^2) \rho + 2} \right]. \end{aligned}$$

## IX. APPENDIX C

### EVALUATION OF MMSE COVARIANCE MATRIX

Since  $\check{\mathbf{h}}_i = \mathbf{h}_i - \hat{\mathbf{h}}_i$ , we have

$$\mathbb{E}[\check{\mathbf{h}}_i \check{\mathbf{h}}_i^H] = \mathbb{E}[\mathbf{h}_i \mathbf{h}_i^H] + \mathbb{E}[\check{\mathbf{h}}_i \check{\mathbf{h}}_i^H] - \mathbb{E}[\mathbf{h}_i \check{\mathbf{h}}_i^H] - \mathbb{E}[\check{\mathbf{h}}_i \mathbf{h}_i^H]. \quad (40)$$

It is straightforward to show that  $\mathbb{E}[\mathbf{h}_i \mathbf{h}_i^H] = \mathbf{I}_{N_r}$ . Consider the received matrix corresponding to the training phase of the  $l_i^{\text{th}}$  transmit antenna as  $\mathbf{Y}_t^{(l_i)} = \sqrt{\rho_t} \mathbf{h}_i \mathbf{x}_t + \mathbf{N}_t^{(l_i)} \in \mathbb{C}^{N_r \times n_t^{(l_i)}}$ , where  $\mathbf{x}_t = [1, 1, \dots, 1] \in \mathbb{C}^{1 \times n_t^{(l_i)}}$ . From (26), we have  $\check{\mathbf{h}}_i = \left( \frac{\rho_t n_t^{(l_i)}}{1 + \rho_t n_t^{(l_i)}} \right) \left( \mathbf{h}_i + \frac{1}{\sqrt{\rho_t}} \mathbf{n}' \right)$ , where  $\mathbf{n}' = \frac{1}{n_t^{(l_i)}} \mathbf{N}_t^{(l_i)} \mathbf{x}_t^H \sim \mathcal{CN}(\mathbf{0}, \frac{1}{n_t^{(l_i)}} \mathbf{I}_{N_r})$ . Therefore, we have

$$\begin{aligned} \mathbb{E}[\check{\mathbf{h}}_i \check{\mathbf{h}}_i^H] &= \frac{\rho_t n_t^{(l_i)}}{1 + \rho_t n_t^{(l_i)}} \mathbf{I}_{N_r}, \\ \mathbb{E}[\check{\mathbf{h}}_i \mathbf{h}_i^H] &= \mathbb{E}[\mathbf{h}_i \check{\mathbf{h}}_i^H] = \frac{\rho_t n_t^{(l_i)}}{1 + \rho_t n_t^{(l_i)}} \mathbf{I}_{N_r}. \end{aligned}$$

Thus, from (40) we have  $\mathbb{E}[\check{\mathbf{h}}_i \check{\mathbf{h}}_i^H] = \frac{1}{1 + \rho_t n_t^{(l_i)}} \mathbf{I}_{N_r}$ .

## REFERENCES

- [1] R. Mesleh, H. Haas, C. Ahn, and S. Yun, "Spatial modulation - a new low complexity spectral efficiency enhancing technique," in *Proc. First International Conf. Commun. Netw., Beijing, China.*, pp. 1-5, Oct. 2006.
- [2] R. Mesleh, H. Haas, S. Sinanovic, C. Ahn, and S. Yun, "Spatial modulation," *IEEE Trans. Veh. Technol.*, vol. 57, no. 4, pp. 2228-2242, July 2008.
- [3] M. Di Renzo, H. Haas, and P. M. Grant, "Spatial Modulation for Multiple-Antenna Wireless Systems - A Survey", *IEEE Communications Magazine*, vol. 49, no. 12, pp. 182-191, Dec. 2011.
- [4] M. Di Renzo and H. Haas, "Bit Error Probability of Spatial Modulation (SM-) MIMO over Generalized Fading Channels," *IEEE Trans. Veh. Technol.*, vol. 61, no. 3, pp. 1124-1144, Mar. 2012.
- [5] 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.
- [6] S. Sugiura, C. Xu, S. X. Ng, and L. Hanzo, "Reduced-complexity coherent versus non-coherent QAM-aided space-time shift keying," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 3090-3101, Nov. 2011.
- [7] J. Wang, S. Jia and J. Song, "Signal Vector Based Detection Scheme for Spatial Modulation," *IEEE Commun. Lett.*, vol. 16, no. 1, pp. 19-21, Jan. 2012.
- [8] P. Yang, Y. Xiao, L. Li, Q. Tang and S. Li, "An Improved Matched-Filter Based Detection Algorithm for Space-Time Shift Keying Systems," *IEEE Signal Process. Lett.*, vol. 19, no. 5, pp. 271-274, May 2012.
- [9] Q. Tang, Y. Xiao, P. Yang, Q. Yu, and S. Li, "A new low-complexity near-ML detection algorithm for spatial modulation," *IEEE Wireless Comm. Lett.*, Dec. 2012 (available under Early Access Articles).
- [10] U. Fincke and M. Pohst, "Improved methods for calculating vectors of short length in a lattice, including a complexity analysis," *Mathematics of Computation*, vol. 44, no. 170, pp. 463-471, Apr. 1985.
- [11] J. Jalden and B. Ottersten, "On the complexity of sphere decoding in digital communications," *IEEE Trans. Signal Proc.*, vol. 53, no. 4, pp. 1474-1484, Apr. 2005.

- [12] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1639-1642, July 1999.
- [13] B. Hassibi and H. Vikalo, "On the expected complexity of integer least-squares problems," in *Proc. IEEE ICASSP*, vol. 2, pp. 1497-1500, May 2002.
- [14] A. Younis, R. Mesleh, H. Haas, and P. Grant, "Reduced Complexity Sphere Decoder for Spatial Modulation Detection Receivers," in *Proc. IEEE GLOBECOM*, Dec. 2010.
- [15] A. Younis, M. Renzo, R. Mesleh, and H. Haas, "Sphere decoding for spatial modulation," in *Proc. IEEE ICC 2011*, June 2011.
- [16] A. Younis, S. Sinanovic, M. Di Renzo, R. Mesleh, and H. Haas, "Generalised sphere decoding for spatial modulation," *IEEE Transactions on Communications*, vol.61, no.7, pp. 2805-2815, July 2013.
- [17] R. Rajashekar and K. V. S. Hari, "Low complexity maximum likelihood detection in spatial modulation systems." Available: arXiv:1206.6190v1, June 2012.
- [18] E. Basar, U. Aygolu, E. Panayirci, and H. V. Poor, "Performance of spatial modulation in the presence of channel estimation errors," *IEEE Commun. Lett.*, vol. 16, no. 2, pp. 176-179, Feb. 2012.
- [19] S. Sugiura and L. Hanzo, "Effects of channel estimation on spatial modulation," *IEEE Signal Proc. Lett.*, vol. 19, no. 12, pp. 805-808, Dec. 2012.
- [20] M. Di Renzo, D. De Leonardi, F. Graziosi, and H. Haas, "Space Shift Keying (SSK-) MIMO with practical channel estimates," *IEEE Trans. Commun.*, vol. 60, no. 4, pp. 998-1112, April 2012.
- [21] S. Ikki and R. Mesleh, "A general framework for performance analysis of space shift keying (SSK) modulation in the presence of gaussian imperfect estimations," *IEEE Commun. Lett.*, vol. 16, no. 2, pp. 228-230, Feb. 2012.
- [22] R. Y. Chang, S.-J. Lin, and W.-H. Chung, "Energy efficient transmission over space shift keying modulated MIMO channels," *IEEE Trans. Commun.*, vol. 60, no. 10, pp. 2950-2959, Oct. 2012.
- [23] M. Di Renzo and H. Haas, "Improving the performance of space shift keying (SSK) modulation via opportunistic power allocation," *IEEE Commun. Lett.*, vol. 14, no. 6, pp. 500-502, June 2010.
- [24] S. Sugiura, S. Chen and L. Hanzo, "Coherent and differential space-time shift keying: A dispersion matrix approach," *IEEE Trans. Commun.*, vol. 58, no. 11, pp. 3219-3230, Nov. 2010.
- [25] L. P. Natarajan and B. S. Rajan, "Generalized distributive law for ML decoding of space-time block codes," *IEEE Trans. Inform. Theory*, vol. 59, no. 5, pp. 2914-2935, May 2013.
- [26] D. Tse and P. Viswanath, "Fundamentals of Wireless Communication." *Cambridge University Press*, 2005.
- [27] Gene H. Golub, Charles F. Van Loan, "Matrix Computations" (3rd ed.), *The Johns Hopkins University Press*, 1996.
- [28] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE. Trans. Inform. Theory*, vol. 49, no. 4, pp. 951-963, Apr. 2003.
- [29] T. Marzetta, "How much training is required for multiuser MIMO ?," in *Proc. 40th Asilomar Conf. Signals Syst. Comput.*, pp. 359-363, 2006.
- [30] M. Kobayashi, G. Caire, and N. Jindal, "How much training and feedback are needed in MIMO broadcast channels?," in *Proc. IEEE Int. Symp. Inf. Theory*, pp. 2663-2667, July 2008.
- [31] Y. Yang and B. Jiao, "Information-guided channel-hopping for high data rate wireless communication," *IEEE Commun. Lett.*, vol. 12, no. 4, pp. 225-227, Apr. 2008.
- [32] T. M. Cover and J. A. Thomas, "Elements of information theory" (2nd ed.), *John Wiley & Sons*, 2006.
- [33] T. L. Marzetta, "BLAST training: Estimating channel characteristics for high-capacity space-time wireless," in *Proc. 37th Annu. Allerton Conf. Communications, Control, and Computing*, Sept. 22-24, 1999.
- [34] J.-C. Guey, M. Fitz, M. Bell, and W.-Y. Kuo, "Signal design for transmitter diversity wireless communication systems over Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 4, pp. 527-537, Apr. 1999.
- [35] M. Biguesh and A. B. Gershman, "MIMO channel estimation: optimal training and tradeoffs between estimation techniques," in *Proc. ICC, Paris, France*, vol. 5, pp. 2658-2662, June 2004.
- [36] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451-1458, Oct. 1998.