Efficient Memory-Bounded Optimal Detection for GSM-MIMO Systems

Ke He[®], Le He[®], Lisheng Fan[®], Xianfu Lei[®], *Member, IEEE*, Yansha Deng[®], *Member, IEEE*, and George K. Karagiannidis[®], *Fellow, IEEE*

Abstract—We investigate the optimal signal detection problem in large-scale multiple-input multiple-output (MIMO) system with the generalized spatial modulation (GSM) scheme, which can be formulated as a closest lattice point search (CLPS). To identify invalid signals, an efficient pruning strategy is needed while searching on the GSM decision tree. However, the existing algorithms have exponential complexity, whereas they are infeasible in large-scale GSM-MIMO systems. In order to tackle this problem, we propose a memory-efficient pruning strategy by leveraging the combinatorial nature of the GSM signal structure. Thus, the required memory size is squared to the number of transmit antennas. We further propose an efficient memory-bounded maximum likelihood (ML) search (EM-MLS) algorithm by jointly employing the proposed pruning strategy and the memory-bounded best-first algorithm. Theoretical and simulation results show that our proposed algorithm can achieve the optimal bit error rate (BER) performance, while its memory size can be bounded. Moreover, the expected time complexity decreases exponentially with increasing the signal-to-noise ratio (SNR) as well as the system's excess degree of freedom, and it often converges to squared time under practical scenarios.

Index Terms—Signal detection, MIMO, maximum likelihood detection, generalized spatial modulation, tree search algorithm, sphere decoding.

Manuscript received 9 November 2021; revised 28 March 2022; accepted 9 May 2022. Date of publication 20 May 2022; date of current version 15 July 2022. This work was supported in part by the NSFC (Nos. 61871139/62101145), by the Natural Science Foundation of Guang-dong Province (No. 2021A1515011392), and by the research program of Guangzhou University (No. YJ2021003). The work of Xianfu Lei was supported by the National Key R&D Program of China under Grant 2019YFB1803400, the National Natural Science Foundation of China under Grant 61971360, the Fundamental Research Funds for the Central Universities under Grant XJ2021KJZK007, and the open research fund of National Mobile Communications Research Laboratory, Southeast University(No. 2021D05). The associate editor coordinating the review of this article and approving it for publication was H. Zhang. (*Corresponding author: Lisheng Fan.*)

Ke He is with the School of Computer Science and Cyber Engineering, Guangzhou University, Guangzhou 510006, China, and also with the Interdisciplinary Centre for Security, Reliability and Trust (SnT), Signal Processing and Satellite Communications Research Group (SIGCOM), University of Luxembourg, 1855 Esch-sur-Alzette, Luxembourg (e-mail: heke2018@e.gzhu.edu.cn).

Le He and Lisheng Fan are with the School of Computer Science and Cyber Engineering, Guangzhou University, Guangzhou 510006, China (e-mail: hele20141841@163.com; lsfan@gzhu.edu.cn).

Xianfu Lei is with the School of Information Science and Technology, Institute of Mobile Communications, Southwest Jiaotong University, Chengdu 610031, China (e-mail: xflei@home.swjtu.edu.cn).

Yansha Deng is with the Department of Engineering, King's College London, London WC2R 2LS, U.K. (e-mail: yansha.deng@kcl.ac.uk).

George K. Karagiannidis is with the Department of Electrical Computer Engineering, Aristotle University of Thessaloniki, 54636 Thessaloniki, Greece (e-mail: geokarag@auth.gr).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCOMM.2022.3176649.

Digital Object Identifier 10.1109/TCOMM.2022.3176649

I. INTRODUCTION

ASSIVE multiple-input multiple-output (MIMO) has became one of the basic component of future wireless communication networks [1], [2]. The use of massive antennas at both transmitter and receiver offers multiplexing and diversity, and thereby a massive MIMO system is established to support ultra-reliable and low-latency services. However, a major concern of the conventional MIMO schema is the rapid increase of complexity and deployment cost [3]-[5]. In order to tackle this problem, spatial modulation (SM) has been extensively investigated as an innovative and promising technology [6], [7]. Despite the fact that multiple antennas are equipped at the transmitter, SM activates only one antenna to convey a certain symbol at any instant. In SM, the transmitted information depends on both the activated antenna index and the transmitted symbol. Therefore, an overall increase in spectral and energy efficiency is achieved, and the inter-channel interference (ICI) can be reduced as well. However, a main drawback of SM is that the overall increase of performance is limited for massive MIMO systems. Therefore, generalized spatial modulation (GSM) was proposed as an extension of SM [8]. Instead of activating only one antenna in SM, GSM activates multiple antennas during a symbol. Thus, it has the potential to provide a comprehensive trade-off among spectral and energy efficiency by adaptively setting the number of activated antennas [7].

In MIMO systems, one important approach to improve the quality of service (QoS) is to reduce the detection errors. However, due to ICI and varying activated transmit antenna combinations (TACs), it becomes much more challenging to recover the transmitted information in GSM. Specifically, the maximum likelihood detection (MLD) in GSM involves a brute-force enumeration over the set of all valid TACs as well as the associated symbols drawn from a specific constellation. Thus, the computational complexity of MLD increases exponentially with the number of activated transmit antennas and the constellation size, which is prohibitively high in large-scale MIMO systems [9], [10]. Although several low-complexity detection algorithms, including ordered block minimum mean square error (OB-MMSE) [11], Gaussian approximation [12], Bayesian cooperative detection [13] and sparsity recovery algorithms [14]–[19], have been proposed to reduce the complexity of GSM signal detection, they actually underperform the optimal MLD with considerably inevitable performance loss. To tackle this problem, researchers have proposed several algorithms to search the shortest valid path

0090-6778 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information. on the GSM decision tree [20]–[27], with the purpose of achieving exact maximum likelihood (ML) performance and meanwhile reducing the average complexity. In general, such existing search algorithms can be categorized into one stage and two stage search algorithms, which are described in details in the following.

A. Literature Overview

The basic idea of one-stage search algorithms is to identify the inactivated antennas by extending the constellation with "0" or NULL symbols [20]-[24]. Under this setting, one is able to jointly determine the selected TAC as well as the transmitted symbols. It is straightforward to find the optimal ML estimation by applying tree search algorithms (e.g. sphere decoding (SD) [20] or K-best SD [22]) to find the shortest valid path. However, only a few number of paths representing valid GSM signals should be considered in the search space of large-scale GSM-MIMO systems. Therefore, pruning the unpromising branches becomes an essential strategy for one-stage search algorithms. From this perspective, researchers proposed several methods to address this problem. In [20], the authors proposed to employ the structure of GSM to restrain the visiting branches leading to valid candidates, which was accomplished by introducing a boolean matrix as a mask or filter at each tree level. In addition, a layer ordering and TAC partitioning aided SD algorithm were proposed in [23], which employed a similar pruning strategy as [20]. Moreover, the authors in [21], [22], [24] proposed to employ a GSM look-up table or codebook to help prune unpromising branches. Such methods are classical implementation of the idea of sacrificing space to store a look-up table or codebook to improve efficiency.

In order to avoid pruning many invalid branches, researchers proposed to search according to the combinatorial nature of GSM signals. Those algorithms are called as two-stage search algorithms, as they typically involves two stages to find the shortest valid path [25], [26]. Two-stage search algorithms will first determine the transmit antenna configuration, and then estimate the transmitted symbols by solving the least-squares problem for all effectively smaller sub-systems. For instance, the authors in [25] proposed a sorting-aided successive SD algorithm (SA-SSDA), which first orders the effective sub-systems, and then successively performs SD on the sub-systems to achieve the MLD performance with a lower computational complexity. Moreover, a similar successive search algorithm was proposed in [26], which employed box-optimization and initial radius broadcasting to further reduce the complexity of SA-SSDA.

Although many one-stage and two-stage search algorithms have been proposed in the past years, there still exists some issues regarding large-scale GSM-MIMO systems. Specifically, for the one-stage search algorithms introduced in [20]–[24], the boolean matrix or look-up table can be prohibitively large in large-scale systems. Storing a huge size of boolean matrix or look-up table can be very challenging for many storage-sensitive internet of things (IoT) devices, and querying can become much more inefficient as well. For the two-stage search algorithms introduced in [25] and [26], ordering subproblems require to exhaustively compute the costs of all possible TACs, which is however again an ineffective strategy for large-scale systems. Generally, the existing one-stage or two-stage algorithms are mainly based on depth-first search [20], [21], [23]–[26] or breadth-first strategy [22]. Despite the fact that depth-first and breath-first search require less memory space than the best-first search, they usually visit much more nodes and run very slowly in large-scale MIMO systems [28]–[30]. Therefore, finding a feasible search algorithm with near-optimal detection performance still remains a critical challenge for large-scale GSM-MIMO systems.

B. Contribution

In this paper, we are interested in achieving the exact optimal detection performance in large-scale GSM-MIMO systems, while trying to significantly improve the search efficiency and memory efficiency. In order to accomplish this goal, we propose an efficient pruning strategy based on combinatorial codes. Moreover, we modify the memorybounded best-first search algorithm with the proposed pruning strategy, and thereby an efficient memory-bounded ML search (EM-MLS) algorithm is proposed to achieve the exact ML performance and considerable complexity reduction. To summarize our work, the main contributions of this paper are as follows:

- By leveraging the combinatorial nature of the GSM signal structure, we propose an efficient pruning strategy whose time and space complexities grow linearly and squared with the number of transmitted antennas, respectively. Thus, the proposed pruning strategy is much more efficient with comparison to the existing strategies, which involve exponentially high complexities.
- With the help of memory-bounded pruning and memorybounded best-first searching, we show that the proposed EM-MLS is able to perform the best-first style search on the GSM decision tree, reach the optimal BER performance with limited memory, and can run 30 times faster than the existing optimal algorithms.
- We further show that the average computational complexity of the proposed search algorithm decreases exponentially with increasing the SNR and system's excess degree of freedom. In fact, it often converges to squared time under practical scenarios. This finding suggests that the proposed algorithm is feasible for large-scale GSM-MIMO systems, especially when the system is highly overdetermined.

II. GSM-MIMO SYSTEM AND ML DETECTION

In this section, we will introduce the considered GSM-MIMO system, as well as several traditional approaches to address ML detection in the system.

A. GSM-MIMO System Model

In this paper, we consider a large-scale GSM-MIMO system with N_t and N_r antennas at the transmitter and receiver,



Fig. 1. Transmitter and receiver diagrams for the GSM-MIMO system.

TABLE I MAPPING TABLE FOR TAC SELECTION IN GSM-MIMO SYSTEMS, WHERE $N_t=5$ and $N_a=2$

| Index | Bits for TAC selection | TAC |
|-------|------------------------|--------|
| 0 | 000 | (1, 2) |
| 1 | 001 | (1, 3) |
| 2 | 010 | (2, 3) |
| 3 | 011 | (1, 4) |
| 4 | 100 | (2, 4) |
| 5 | 101 | (3, 4) |
| 6 | 110 | (1, 5) |
| 7 | 111 | (2, 5) |
| 8 | Invalid | (3, 5) |
| 9 | Invalid | (4, 5) |

where only N_a $(N_r \ge N_t \ge N_a \gg 1)$ transmit antennas are activated. The excess degree of freedom is denoted by $N_e = N_r - N_t$, and the total number of TACs is given by $\binom{N_t}{N_a}$, where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ is the binomial coefficient. It should be noted that not all TACs are available in the system, since only $K = 2^{\lfloor \log_2 \binom{N_t}{N_a} \rfloor}$ TACs can be considered for transmission, where $\lfloor \cdot \rfloor$ denotes the floor operation. Besides, the symbols transmitted by the activated transmit antennas are modulated by *M*-ary quadrature amplitude modulation (*M*-QAM) or phase shift keying (*M*-PSK) modulation with alphabet \mathcal{X} . Hence, one is able to convey $L = \log_2 K + N_a \log_2 M$ bits in total at a symbol duration, and only N_a out of N_t radio frequency (RF) chains are required with comparison to the conventional MIMO systems.

As shown in Fig. 1, the transmission process in the GSM-MIMO system include two stages. Given a block of L bits, in the first stage, the first $\log_2 K$ bits will be used to select the transmit antennas according to the underlying GSM mapping, and an example of GSM mapping table for TAC selection is shown in Table. I. After the transmit antennas are determined, in the second stage, the remaining $N_a \log_2 M$ bits are modulated by M-QAM or M-PSK modulation, and the resulting signal $x \in \mathcal{X}^{N_a}$ will be transmitted via a Rayleigh flat fading wireless channel. Formally, the received signal at the receiver is given by

$$y = Hs + w, \tag{1}$$

where $\boldsymbol{H} \in \mathbb{C}^{N_r \times N_t}$ denotes the channel state information (CSI) matrix with independent and identically distributed (i.i.d.) components following $\mathcal{CN}(0,\rho)$, $s \in S^{N_t} \subset \mathbb{C}^{N_t \times 1}$ denotes the transmitted signal, and $w \sim \mathcal{CN}(0, I)$ represents the additive white Gaussian noise (AWGN). Note that $S = \mathcal{X} \cup \{0\}$ denotes the alphabet of the effectively extended constellation with "0" identifying inactivated antennas (e.g. $\mathcal{X} = \{-1, +1\}$ and $S = \{-1, +1, 0\}$ for BPSK modulation). Alternatively, the above system model can be expressed as an effective sub-system, which is given by

$$\boldsymbol{y} = \sum_{j=1}^{N_t} \boldsymbol{h}_j \boldsymbol{s}_j + \boldsymbol{w} = \boldsymbol{H}_k \boldsymbol{x} + \boldsymbol{w}, \qquad (2)$$

where $H_k = [h_{k_{N_a}}, \dots, h_{k_2}, h_{k_1}]$ is the effective CSI matrix with columns corresponding to the nonzero elements of s. It should be noted that there are K effective sub-systems in total.

B. ML Detection

Based on the system model of (1), the mathematically optimal detector is the MLD given by

$$\boldsymbol{s}^* = \operatorname*{arg\,min}_{\hat{\boldsymbol{s}} \in \mathcal{S}_v} \left\| \boldsymbol{y} - \boldsymbol{H} \hat{\boldsymbol{s}} \right\|^2, \tag{3}$$

whose principle is to find the closest valid candidate in terms of Euclidean distance. Note that the search space $S_v \subseteq S^{N_t}$ is the subset containing all valid GSM signal vectors. In addition, based on the effective sub-system model of (2), the ML solution can be found by solving a sequence of sub-problems

$$\langle k^*, \boldsymbol{x}^* \rangle = \operatorname*{arg\,min}_{k, \hat{\boldsymbol{x}}} \left\| \boldsymbol{y} - \boldsymbol{H}_k \hat{\boldsymbol{x}} \right\|^2,$$
 (4)

which requires to estimate both the TAC index and transmitted symbols successively. Based on k^* and x^* , one can thereby recover the whole signal vector s^* . However, directly solving (3) or (4) both involve an exponentially increasing computational complexity [30]. Hence, we need to construct a decision tree by applying QR factorization as

$$H = \underbrace{\left[Q_1 \quad Q_2\right]}_{Q} \begin{bmatrix} R\\ 0 \end{bmatrix} = Q \begin{bmatrix} R\\ 0 \end{bmatrix}, \qquad (5)$$

where $Q_1 \in \mathbb{C}^{N_r \times N_t}$ and $Q_2 \in \mathbb{C}^{N_r \times N_e}$ both have orthogonal columns, and $R \in \mathbb{C}^{N_t \times N_t}$ is the upper triangular matrix.

 $egin{aligned} m{s}^* &= rgmin_{\hat{m{s}}\in\mathcal{S}_v} \left\|m{Q}^Hm{y} - egin{bmatrix}m{R}\m{0} \end{bmatrix} \hat{m{s}}
ight\|^2 \end{aligned}$ lar matrix. Then, (4) can be written as (6)

Fig. 2. Structure of a GSM-MIMO decision tree with $N_t = 3$, $N_a = 2$ and BPSK modulation. The solid black nodes lead to invalid goal nodes, while the

$$= \underset{\hat{s}\in\mathcal{S}_{v}}{\operatorname{arg\,min}} \sum_{\substack{j=1\\N_{t}}}^{N_{t}} \left| z_{k} - \sum_{j=0}^{k} r_{k,j} s_{j} \right|^{2}$$
(7)

$$= \underset{\hat{s} \in \mathcal{S}_{v}}{\operatorname{arg\,min}} \sum_{j=1}^{N_{v}} B(\hat{s}^{j}), \tag{8}$$

where $\boldsymbol{z} \triangleq \boldsymbol{Q}_1^H \boldsymbol{y}, r_{i,j}$ denotes the (i,j)-th element of \boldsymbol{R} counting from the right-bottom to the left-top, and the *j*-th cost increment is given by $B(\hat{s}^k) = \left| z_k - \sum_{j=0}^k r_{k,j} s_j \right|^2$, which only depends on the partially determined vector $\hat{s}^k = [\hat{s}_k, \hat{s}_{k-1}, \cdots, \hat{s}_1]$. In addition, the cumulative cost is $G(\hat{s}^k) = \sum_{i=1}^k B(\hat{s}^i)$, and the successive identity $G(\hat{s}^{k+1}) = G(\hat{s}^k) + B(\hat{s}^{k+1})$ holds.

solid red nodes lead to valid goal nodes.

Then, (3) can be written as

Thus, a perfect |S|-way and N_t -level decision tree is constructed [30], [31]. In the resulting tree, we denote a node on the k-th level as \hat{s}^k , which represents the path leading from the root to that node as well. In particular, the cumulative cost of a goal node (i.e. a node $\hat{s}^{\hat{k}}$ at the deepest level with $k = N_t$), represents the Euclidean distance of the associated signal vector. It should be noted that not every goal node is associated with a valid GSM signal. Therefore, one has to find the ML solution of (3) by finding the shortest path leading from the root to a valid goal node. To illustrate this, an example is provided in Fig. 2 with $N_t = 3$, $N_a = 2$ and BPSK modulation.

Analogously, one can also construct a forest containing a sequence of $|\mathcal{X}|$ -way N_a -level decision trees by factorizing H_k for every effective sub-system as

$$\boldsymbol{H}_{k} = \begin{bmatrix} \bar{\boldsymbol{Q}}_{k} & \tilde{\boldsymbol{Q}}_{k} \end{bmatrix} \begin{bmatrix} \boldsymbol{R}_{k} \\ \boldsymbol{0} \end{bmatrix}, \qquad (9)$$

where $\bar{\boldsymbol{Q}}_k \in \mathbb{C}^{N_r \times N_a}$ and $\tilde{\boldsymbol{Q}}_k \in \mathbb{C}^{N_r \times (N_r - N_a)}$ both have orthogonal columns, and $\boldsymbol{R}_k \in \mathbb{C}^{N_a \times N_a}$ is the upper triangu-

$$\langle k^*, \boldsymbol{x}^* \rangle = \arg\min_{k} \left(\| \tilde{\boldsymbol{Q}}_k^H \boldsymbol{y} \|^2 + \min_{\hat{\boldsymbol{x}} \in \mathcal{X}^{N_a}} \| \bar{\boldsymbol{Q}}_k^H \boldsymbol{y} - \boldsymbol{R}_k \hat{\boldsymbol{x}} \|^2 \right).$$
(10)

Obviously, the above equation defines a decision forest of K small decision trees, and the cost of the k-th root is $\|\tilde{\boldsymbol{Q}}_k^H \boldsymbol{y}\|^2$. Based on the resulting decision forest, one is able to find the ML solution of (4) in two stages. The first stage is to order all the trees in an ascending order according to the costs of their roots. The second stage is to find the shortest path of each tree, and select the shortest path among the shortest paths of all trees as the optimal solution.

Basically, one-stage search algorithms are developed based on (6) [20]–[24], while two-stage search algorithms are developed based on (10) [25], [26]. For the two-stage search algorithms introduced in [25], [26], they have to factorize all the sub-matrices and compute the costs of all roots based on (10). Hence, it is hard to scale those algorithms to largescale problems, especially when the system is highly overdetermined. As to the one-stage search algorithms introduced in [20]–[24], they require exponentially increased memory space to store mapping tables. Unfortunately, the number of valid signal vectors ($|S_v| = K|\mathcal{X}|^{N_a}$ in total) is often significantly less than the size of the whole search space ($|\mathcal{S}|^{N_t}$ in total) in large-scale systems. For example, only about 0.01%signal vectors are valid when $N_t = 32$, $N_a = 16$ and QPSK modulation is adopted. In conclusion, the main drawback of the existing search methods is that they either require a huge size of memory space or a lot of computing time that increases exponentially with the problem scale, which motivates us to find an efficient memory-bounded search strategy for largescale GSM-MIMO systems.



TABLE II BIJECTIVE MAPPING TABLE FOR GSM-MIMO SYSTEMS WITH $N_t = 5$ and $N_a = 2$

| Index | Bits for TAC selection | TAC | BASS |
|-------|------------------------|--------|-----------------|
| 0 | 000 | (1, 2) | [1, 1, 0, 0, 0] |
| 1 | 001 | (1, 3) | [1, 0, 1, 0, 0] |
| 2 | 010 | (2, 3) | [0, 1, 1, 0, 0] |
| 3 | 011 | (1, 4) | [1, 0, 0, 1, 0] |
| 4 | 100 | (2, 4) | [0, 1, 0, 1, 0] |
| 5 | 101 | (3, 4) | [0, 0, 1, 1, 0] |
| 6 | 110 | (1, 5) | [1, 0, 0, 0, 1] |
| 7 | 111 | (2, 5) | [0, 1, 0, 0, 1] |
| 8 | Invalid | (3, 5) | [0, 0, 1, 0, 1] |
| 9 | Invalid | (4, 5) | [0, 0, 0, 1, 1] |

III. PROPOSED EFFICIENT MEMORY-BOUNDED ML SEARCH STRATEGY

In this section, we will first introduce our proposed pruning strategy. After that, we will show the proposed efficient memory-bounded ML detection algorithm, and analyze its expected computational complexity afterwards.

A. Efficient Pruning With Combinatorial Coding Schema

According to the GSM signal structure, a key observation is that once we rank TACs in lexicographical order, a bijective mapping can be established between the indices and antenna combinations. This can be done by interpreting the TAC as a binary antenna selection sequence (BASS), and computing the corresponding lexicographical ordered index number. As long as we have an arrangement on only using the first K TACs for space modulation, we can check if a signal vector is valid by verifying the index number (starts from zero) of the associated BASS. Specifically, a signal vector is valid if it has the correct number of activated antennas, and the index is less than K. Following this, we first define the BASS of each TAC as

$$\boldsymbol{c} = [c_0, c_1, \dots, c_i, \dots, c_{N_t-1}], \quad c_i \in \{0, 1\},$$
(11)

where $c_i = 1$ indicates that the (i + 1)-th antenna is activated, and the resulting BASS of a valid GSM signal must satisfy $\sum_{i=0}^{N_t-1} c_i = N_a$. Note that this is also called as N_a -hot vector in the literature, since there are at most N_a elements which are activated. In addition, we will use $\mathbf{c} = C(\mathbf{s}^k)$ to represent the associated BASS of \mathbf{s}^k , and we have $\sum_{i=0}^{N_t-1} c_i \leq N_a$ if $k < N_t$. For illustration, an example of bijective mapping is presented in Table. II, where $N_t = 5$ and $N_a = 2$.

In order to compute the corresponding index $i_c(c)$ of a BASS c, we employ an algorithm introduced in [32] to calculate the index as

$$i_c(\boldsymbol{c}) = \sum_{n=0}^{N_t-1} c_n \binom{n}{k_n},$$
(12)

where $k_n = \sum_{i=0}^{n} c_i$ and $\binom{n}{k_n} = 0$ for all $n < k_n$. Now one is able to construct a combinatorial coding trellis [33] based on (12). The corresponding encoding and decoding procedures are presented in Algorithm 1. Based on this algorithm, one can encode a BASS to the corresponding index number, and decode an index number back to the corresponding BASS.

| Algorithm 1 | Encoding | and Decod | ling Proced | lures for | Combi- |
|---------------|-------------|-----------|-------------|-----------|--------|
| natorial Codi | ing Trellis | | | | |

| 8 |
|--|
| 1: procedure Encode(c) |
| 2: Compute the index $i_c(c)$ according to (12) |
| 3: return $i_c(c)$ |
| 4: end procedure |
| 5: procedure $Decode(i_c)$ |
| 6: Initialize $\boldsymbol{c} = [c_0, c_1, \dots, c_n, \dots, c_{N_t}]$ with zeros |
| 7: $r = i_c$ |
| 8: for $k = N_a$ to 1 do |
| 9: for $n = N_e + k - 1$ to $k - 1$ do |
| 10: if $r \ge \binom{n}{k}$ then |
| 11: $c_n = 1$ |
| 12: $r = r - \binom{n}{k}$ |
| 13: break |
| 14: end if |
| 15: end for |
| 16: end for |
| 17: return <i>c</i> |
| 18: end procedure |
| |

To further check whether an arbitrary intermediate node \hat{s}^k is a valid node, we obtain the following theorem.

Theorem 1: Let \hat{s}^k be an arbitrary node on the GSM decision tree of (6), $\mathbf{c} = C(\hat{s}^k)$ be the corresponding BASS of \hat{s}^k , and $n_a = \sum_{n=N_t-k}^{N_t-1} c_n$ be the number of currently activated antennas. Then, \hat{s}^k is a valid node if

$$N_a + k - N_t \le n_a \le N_a,\tag{13}$$

and

$$\sum_{k=N_t-k}^{N_t-1} c_n \binom{n}{k_n + N_a - n_a} < K, \tag{14}$$

where $k_n = \sum_{i=N_t-k}^n c_i$ denotes the counter of activated antennas.

Proof: From the GSM decision tree, it is clear that an intermediate node \hat{s}^k is a valid node if and only if there is a valid goal node on the sub-tree of \hat{s}^k . In other words, \hat{s}^k is a valid node if the minimum index number among the goal nodes on the sub-tree of \hat{s}^k is less than K. Following this, we expand the associated BASS of \hat{s}^k as

$$c = [0, \dots, 0, c_{N_t-k}, \dots, c_{N_t-1}],$$
 (15)
(N_t - k) terms

where only the last k elements are determined and the remaining $N_t - k$ undetermined elements are all zeros. In addition, the number of currently activated antennas should satisfy (13). Since the associated TACs are lexicographically ordered for all valid goal nodes, the minimum BASS among the valid goal nodes on the sub-tree can be expressed as

$$\boldsymbol{c}^{\dagger} = \begin{bmatrix} \underbrace{1, 1, \dots, 1}_{(N_a - n_a) \text{ ones}} & 0, \dots, 0, \underbrace{c_{N_t - k}, \dots, c_{N_t - 1}}_{k \text{ terms}} \end{bmatrix}, \quad (16)$$

where the first $N_a - n_a$ terms are all ones, and the remaining terms are the same as c. From (12), the corresponding

combinatorial index of c^{\dagger} can be computed as

$$i_{c}(\boldsymbol{c}^{\dagger}) = \sum_{n=0}^{N_{a}-n_{a}-1} \binom{n}{n+1} + \sum_{n=N_{t}-k}^{N_{t}-1} c_{n} \binom{n}{k_{n}+N_{a}-n_{a}}$$
(17)

$$=\sum_{n=N_t-k}^{N_t-1} c_n \binom{n}{k_n+N_a-n_a},$$
(18)

and it should satisfy $i_c(c^{\dagger}) < K$. Thus, we have finished the proof of Theorem 1.

By utilizing Theorem 1, one can easily check if an interme diate node is valid during search on the decision tree. In orde to analyze the complexity of the pruning strategy, we shoul focus on the complexity of the decoding algorithm, which can be viewed as a traverse along the trellis. As shown i Fig. 3, the decoding algorithm starts with the combinatoria terms at the upper-left corner, and then compares the inde of a signal candidate with the combinatorial term on th diagonally downward transition. If the index is greater that the combinatorial term, then the combinatorial term will b selected for transition, and meanwhile the index is update by subtracting with the combinatorial term. Otherwise, th algorithm continues to find the next node on the right sid transition. Obviously, we can find from the trellis that only $N_a(N_t-N_a+1)$ combinatorial terms need to be stored. There fore, the space complexity of the proposed pruning strategy is $\mathcal{O}(N_t^2)$ with $N_a = \lfloor \frac{N_t+1}{2} \rfloor$. Moreover, since the combinatorial terms are pre-computed, the worst-case time complexity is $\mathcal{O}(N_t N_a)$, which grows linearly with the fixed N_a or N_t . For comparison, at least $K|\mathcal{X}|^{N_t}$ terms should be stored for the mapping table based pruning strategies. Therefore, we are able to prune the tree efficiently by leveraging the combinatorial nature of the GSM signal structure.

B. Efficient Memory-Bounded Best-First Search

In order to reduce the number of visited nodes, and limit the memory usage at the same time, we introduce an efficient memory-bounded search algorithm, which jointly uses the aforementioned pruning strategy, and the memory-bounded best-first search strategy [34] to achieve the exact ML performance. Before showing the details, it is necessary to introduce the following important concepts:

- ACTIVE is an ordered list with possibly limited size, whose elements are ordered according to their costs. In practice, it can be a priority queue or self-balancing binary search tree (a.k.a. AVLTree) to meet the need of fast querying.
- A node is said to be *generated* if it has been visited during the expansion of its parent.
- A node is said to be *in memory* if it occupies space of ACTIVE.
- A node is said to be *expanded* if all of its valid successors are in memory.

The pseudo code of the proposed algorithm is presented in Algorithm 2, and the pseudo code of utility functions is presented in Algorithm 3. In the decision tree, every goal node

Algorithm 2 Efficient Memory-Bounded ML Search (EM-MLS) Algorithm

| Inp | put: z and R |
|--------------|---|
| Ou | tput: ŝ |
| 1: I | nsert the root node s^0 into ACTIVE |
| 2: l | oop |
| 3: | if ACTIVE is empty then |
| 4: | return \emptyset with failure |
| 5: | end if |
| 6: | $\hat{s}^{\kappa} \leftarrow deepest \ least-cost \ node \ in \ ACTIVE$ |
| 7: | if s^k is a goal node $(k = N_t)$ then |
| 8: | return \hat{s}^{κ} with success |
| 9: | end if |
| 10: | $\hat{s}^{k+1} \leftarrow$ next not generated valid successor according |
| to | Theorem 1 or best forgotten successor of \hat{s}^{κ} |
| 11: | Insert \hat{s}^{k+1} into its parent's generated successor list |
| 12: | HANDLE $(\hat{s}_{k}^{\kappa}, \hat{s}^{\kappa+1})$ |
| 13: | $\operatorname{Adjust}(\hat{s}^{\kappa})$ |
| 14: | MAKESPACE() |
| 15: | Insert \hat{s}^{k+1} in ACTIVE |
| 16: | if \hat{s}^{κ} is expanded then |
| 17: | remove \hat{s}^{k} from ACTIVE |
| 18: | end if |
| 19: e | nd loop |

is associated with a specific signal candidate, every branch represents the cost of the associated symbol decision, and the shortest path which has the least cumulative costs represents the optimal ML estimate [30], [34]. As shown in the algorithm 2, the algorithm maintains a partially expanded decision tree of the whole tree, generates only one valid successor during each expansion, and forgets the most unpromising leaf node. The search procedure always starts from a dummy root node s^0 , chooses the deepest least-cost leaf node¹ s^k from the memory, and expands it at each iteration. In particular, as shown in Line 10 of Algorithm 2, the algorithm finds the next not-generated valid successor of s^k according to Theorem 1 during the expansion. Hence, all the invalid intermediate nodes and goal nodes can be pruned very efficiently, which significantly reduces the complexity of the algorithm.

Besides, the proposed algorithm will keep expanding the best leaf node until ACTIVE is full. Since the memory size is limited, EM-MLS will forget the most unpromising node from ACTIVE, and remember the forgotten node's key information in its parent. After safely deleting the most unpromising node, the proposed algorithm is able to move forward, and recover forgotten nodes if there is not any other path better than the forgotten path. Thus, the memory size can be bounded, and EM-MLS guarantees to find the shortest path, which represents the exact ML estimate. By using the above search strategy, the memory usage of the proposed algorithm can be limited. Meanwhile, the algorithm can take full advantage of all the available memory during the search.

¹The leaf node is with respect to the partially expanded decision tree maintained by the algorithm. That means the forgotten node is not necessary to be the goal node of the whole decision tree of (6).



Fig. 3. Trellis diagram for decoding index $(N_a = 3)$.

Algorithm 3 Utility Functions for the EM-MLS Algorithm 1: function HANDLE(\hat{s}^k, \hat{s}^{k+1}) if \hat{s}^{k+1} is not a forgotten node then $G(\hat{s}^{k+1}) \leftarrow \max\left(G(\hat{s}^k), G(\hat{s}^{k+1})\right)$ 2: 3: else 4: Recover \hat{s}^{k+1} 's cost from its parent 5: end if 6: 7: end function function ADJUST(\hat{s}^k) 8: if all of \hat{s}^k 's successors are generated then 9: $\hat{s}^{k+1} \leftarrow$ least-cost successor among all generated 10: successors and forgotten successors of \hat{s}^{k+1} if \hat{s}^{k+1} 's cost is finite and not equal to \hat{s}^k 's cost then 11: Update \hat{s}^{k} 's cost to \hat{s}^{k+1} 's cost 12: \hat{A} DJUST(\hat{s}^k 's parent) 13: end if 14· end if 15: 16: end function 17: function MAKESPACE() if ACTIVE is not full then 18: return 19: end if 20: Remove shallowest highest-cost leaf node \hat{s}_{i}^{k} from 21: ACTIVE Remove \hat{s}_{j}^{k} from its parent's generated successor list 22: Remember \hat{s}_j^k 's cost in its parent's forgotten successor 23: list 24: if the parent is not in ACTIVE then Insert the parent into ACTIVE MAKESPACE 25: end if 26:

27: end function

It should be noted that the BER performance is not related to the memory size, as long as the least required memory space² is satisfied. In principle, with different sizes of memory space, the proposed algorithm still generates the same set of nodes, but may visit some nodes more than once depending on the memory size [34]. This means that the proposed algorithm

 2 To find a complete path from the root node to any goal node, the memory size must be at least larger than the length of the path.

will run faster with a larger memory size, as it has no need to frequently forget unpromising nodes. In particular, if the memory size is unlimited, EM-MLS behaves exactly the same as the other best-first search algorithms, except for the ability to prune the nodes which will break the structure of the GSM signal. In the simulations, we will show that the proposed algorithm can reach nearly the optimal performance with memory sizes which is linear to the problem scale. Although EM-MLS can perform very well with small memory space, we should provide as much memory as possible in practice, as EM-MLS can perform better with larger memory space. Consequently, with the help of the proposed efficient pruning strategy and memory-bounded best-first search, the proposed search algorithm guarantees to obtain the exact ML estimate, while its space complexity can be bounded.

C. Expected Computational Complexity

To analyze the computational complexity of the proposed algorithm, we first provide the following theorem.

Theorem 2 (Conditions for Node Expansion): Let \hat{s}^k be any valid node satisfying Theorem 1 on the tree of (6), then a necessary condition for expanding \hat{s}^k is

$$G(\hat{\boldsymbol{s}}^k) \le G(\boldsymbol{s}^*),\tag{19}$$

and the sufficient condition is

$$G(\hat{\boldsymbol{s}}^k) < G(\boldsymbol{s}^*),\tag{20}$$

where s^* denotes the ML estimate.

Proof: We first prove the necessary condition. Since s^* is the ML estimate, it will be the first goal node expanded by the algorithm. For any valid node \hat{s}^k satisfying Theorem 1, the algorithm expands s^* before \hat{s}^k if $G(\hat{s}^k) > G(s^*)$. As the algorithm terminates at s^* , \hat{s}^k will not be expanded by the algorithm. Hence, the necessary condition holds. For the sufficient condition, the algorithm expands \hat{s}^k before s^* if $G(\hat{s}^k) < G(s^*)$. When $G(\hat{s}^k) = G(s^*)$, s^* may be expanded before \hat{s}^k due to the algorithm's visitation priority. At this time, the algorithm terminates, and \hat{s}^k will not be expanded. Hence, the sufficient condition holds as well.

Following Theorem 2, it is clear that only one goal node will be eventually expanded by the algorithm. As a contrast, the depth-first or breadth-first algorithms like SD or K-best SD may expand more than one goal nodes [28]. As the computational complexity of search algorithms mainly depend on the number of expanded nodes, the proposed algorithm often runs much faster than SD in practice. Generally, the complexity of SD can be regarded as a loose upper bound of the complexity of the proposed algorithm, and both algorithms will require exponential complexity at the worst case [30]. However, since the decision tree is constructed based on the random system model in (1), we can further characterize the expected complexity of the proposed algorithm based on the following theorem.

Theorem 3: For the GSM-MIMO system in (1) with fixed N_t and finite lattice points, the expected computational complexity of the proposed EM-MLS algorithm decreases exponentially with both the increasing SNR which is proportional to ρ , and the increasing excess degree of freedom N_e . In particular, the expected computational complexity converges to $\mathcal{O}(N_t^2)$ with sufficiently large ρ and N_e .

Proof: See Appendix.

Following Theorem 3, we can observe that EM-MLS not only benefits from increasing SNR, but also benefits from increasing the excess degree of freedom of the system. The latter reveals that one is able to significantly improve the efficiency of the proposed algorithm without additional energy consumption. As to the other GSM-MIMO signal detection algorithms like Gaussian approximation and OB-MMSE, they do not utilize the excess degree of freedom to reduce their computational complexities. In particular, when N_e grows large enough, the proposed algorithm can reach the lowest average complexity that is squared of N_t . Therefore, we believe that the proposed algorithm is feasible for the large-scale GSM-MIMO system, especially when the system is highly overdetermined.

IV. SIMULATION RESULTS AND DISCUSSION

In this section, we present simulation results to verify the effectiveness of our proposed EM-MLS algorithm. Specifically, we perform simulations on several GSM-MIMO systems where the transmitted signals experience Rayleigh flat fading channels [35], [36]. If not specified, the presented simulations are performed with $N_r = N_t = 16$, $N_a = 10$ and QPSK modulation. Under this setting, there are 2^{32} signal candidates in total, which is prohibitively large such that the brute-force search is totally impractical. In simulations, we compare our proposed EM-MLS algorithm with the following competitive GSM-MIMO detection algorithms.

- OMP: Orthogonal matching pursuit algorithm for fast sparse signal recovery introduced in [17].
- GA: Maximum likelihood detector with Gaussian approximation introduced in [12].
- OB-MMSE: Ordered block minimum mean squared error detector introduced in [11].
- SASSD: Sorting assisted successive sphere decoding algorithm introduced in [25].

In order to verify the memory efficiency of the proposed algorithm, we will show the performance of our proposed



Fig. 4. BER comparison versus SNR with $N_r = N_t = 16$, $N_a = 10$ and QPSK modulation.

algorithm in terms of memory space. Specifically, we use EM-MLS(L) to indicate that the capacity of ACTIVE is set to L. In particular, EM-MLS(∞) denotes that the memory size is unlimited, and EM-MLS will visit each node no more than once in this situation.

Fig. 4 plots the BER performances of the aforementioned detection algorithms, where SNR varies from 5 dB to 15 dB. We can find from the figure that the proposed EM-MLS algorithm reaches the exact ML performance, while the performances of OMP, GA, and OB-MMSE are far from reaching the optimal BER performance. This is because that the proposed algorithm guarantees to find the shortest path of the GSM decision tree as long as the minimum size of memory space is established. Note that we neglect the memory size of EM-MLS in BER performance comparisons, since the memory size does not affect the BER performance. In addition, we can find from the figure that the performance gap between the proposed algorithm and the other sub-optimal algorithms enlarges with increasing SNR. Specifically, when SNR =15 dB, EM-MLS can significantly reduce the errors of OMP, GA and OB-MMSE to about 0.001%, 0.01% and 0.76%, respectively. In addition, the SNR gains of EM-MLS over GA and OB-MMSE are about 5 dB and 2 dB at the BER level of 10^{-2} . These results show that the performance gap between the sub-optimal algorithms and our proposed EM-MLS rapidly enlarges in high SNR regime, which indicates that pursuing the optimal ML performance is of importance for large-scale systems.

Fig. 5 compares the computing times of the aforementioned algorithms, where the SNR varies from 5 dB to 20 dB. In particular, we present the computing time of EM-MLS with different sizes of memory space in order to verify the memory efficiency of EM-MLS. We can find from the figure that the computing time of both SASSD and EM-MLS decreases with increasing SNR, and converges in the high SNR regime. However, the computing time of EM-MLS is much lower than that of SASSD, especially when SNR is high. Specifically, EM-MLS is about 37 times faster than SASSD, and reduces



Fig. 5. Computing time comparison versus SNR with $N_r = N_t = 16$, $N_a = 10$ and QPSK modulation.

the computing time of SASSD to about 97.34% when SNR = 20 dB. More importantly, EM-MLS can be even faster than the sub-optimal algorithms in the high SNR regime. Specifically, EM-MLS runs about 4.6 times faster than OMP when SNR = 20 dB. This indicates that our proposed EM-MLS can prune the unpromising nodes efficiently and visit significantly fewer nodes to find the shortest valid path in large-scale search space. Furthermore, we can find from Fig. 5 that the computing time of EM-MLS is not sensitive to the size of memory space, since EM-MLS(64) consumes nearly the same computing time as EM-MLS(4096) and EM-MLS(∞) in a wide range of SNR. These results sufficiently verify the effectiveness of the proposed EM-MLS algorithm.

Generally, the computing time of search algorithms is proportional to the total number of floating point operations (FLOPs) and the total number of visited nodes. Hence, Figs. 6-7 are presented to further compare the two components of the aforementioned algorithms. Specifically, Fig. 6 depicts the number of visited nodes for the aforementioned algorithms under the same environment setting of Fig. 5. One can conclude from Fig. 6 that the number of nodes visited by EM-MLS decreases very rapidly with the increasing SNR. The number of nodes visited by SASSD decreases as well, but converges to a much high level with comparison to EM-MLS. This is because SASSD finds the shortest valid path in a depth-first manner, which will visit more nodes than the best-first style search. In addition, the size of memory space has impact on the number of nodes visited by EM-MLS in the low SNR regime, but it quickly vanishes with the increasing SNR. Actually, with different memory space, EM-MLS still generates the same set of nodes, but may visit some nodes more than once depending on the size of memory space. This phenomenon further indicates that the complexity of EM-MLS is not sensitive to the size of memory space, especially for high SNR regime.

Fig. 7 demonstrates the number of FLOPs for the competing algorithms in the same environment setting of Fig. 5. Since EM-MLS generates the same set of nodes with different



Fig. 6. Number of visited nodes versus SNR with $N_r = N_t = 16$, $N_a = 10$ and QPSK modulation.



Fig. 7. FLOPs comparison versus SNR with $N_r = N_t = 16$, $N_a = 10$ and QPSK modulation.

memory sizes, it performs the same number of FLOPs as well. Hence, we plot only the number of FLOPs performed by EM-MLS(64) in the figure for brevity. From Fig. 7, it can be found that the number of FLOPs of EM-MLS decreases rapidly in the high SNR regime, while the number of FLOPs of SASSD varies insignificantly at the same time. By combining the simulation results in Fig. 6, we can conclude that this is because even the number of nodes visited by SASSD decreases, it still has to do preprocessing for all subsystems (see (4)) as we have pointed out in Sec II-B, which is a bad strategy in large-scale systems. Hence, the number of FLOPs at the preprocessing stage dominates in the total number of FLOPs by SASSD, which indicates that SASSD is hard to scale to large-scale systems. As a contrast, EM-MLS only need to perform preprocessing once, thus the number of FLOPs decreases linearly with the number of generated nodes. Therefore, EM-MLS becomes much faster and more efficient than SASSD in large-scale systems.



Fig. 8. Computing time versus SNR with $N_t = 32$, $N_a = 16$ and QPSK modulation.



Fig. 9. Computing time versus the excess degree of freedom with $N_t = 32$, $N_a = 16$ and QPSK modulation.

To further validate Theorem 3, we present simulation results regarding complexity comparisons in Figs. 8-9. Specifically, Fig. 8 illustrates the computing time comparison versus different SNRs, where $N_t = 32$, $N_a = 16$ and SNR varies from 15 dB to 25 dB. Fig. 9 illustrates the computing time comparison versus different excess degree of freedom, where $N_t = 32, N_a = 16$ and N_e varies from 0 to 32. The size of search space is 2^{61} in the simulations, and the memory size of EM-MLS is set to 1024. We can find from Fig. 8 that the computing time decreases exponentially with the increasing SNR for all the values of N_e . Similarly, we can also find from Fig. 9 that the computing time decreases exponentially with the increasing excess degree of freedom for all SNR settings. In addition, the computing time converges to a certain value with a sufficiently large N_e and a sufficiently high SNR. Besides, the computing time of EM-MLS converges faster with higher SNR and larger N_e . These observations obviously support our theoretical complexity analysis. In further, we can



Fig. 10. Computing time versus SNR with $N_t = 32$, $N_a = 16$ and 16-QAM modulation.

see from the figures that the expected complexity usually reaches the lowest complexity under practical conditions, and the proposed algorithm performs much better under highly overdetermined systems. In conclusion, these results suggest that the proposed EM-MLS algorithm is feasible for largescale GSM-MIMO systems, especially when the system is highly overdetermined.

In order to show the effectiveness of the proposed algorithm with high-order modulation schemes, Fig. 10 is used to present the computing time results versus SNR, where 16-QAM is adopted, and the numbers of transmit antennas, receive antennas and RF chains are 16, 32, and 10, respectively. We can find from this figure that the computing time of the proposed algorithm still outperforms the aforementioned algorithms. Specifically, EM-MLS(64) runs 20 and 2.5 times faster than SA-SSD and OMP, respectively. Moreover, the computing time of EM-MLS(64) is very close to that of EM-MLS(∞), which indicates that the proposed algorithm can still run very efficiently with low memory requirement in 16-QAM 16 × 32 GSM-MIMO systems. These results further verify the effectiveness of the proposed algorithm.

V. CONCLUSION

In this paper, we have studied the classical optimal signal estimation problem in large-scale GSM-MIMO systems. We designed an efficient memory-bounded optimal signal detection algorithm, which not only limits the memory usage, but also can achieve the exact optimal maximum likelihood detection performance with an affordable computational complexity. To do so, we have proposed a memory-efficient pruning strategy for search algorithms by leveraging the combinatorial nature of the GSM signal structure. Based on this proposed pruning strategy, we have further proposed an efficient memory-bounded tree search algorithm, namely EM-MLS. We have shown that the proposed algorithm guarantees to find the optimal solution with limited memory space. Meanwhile, the computational complexity of the proposed algorithm decreases exponentially with the increasing SNR and the excess degree of freedom, and usually converges to squared time under practical conditions. Therefore, we confirm that the proposed algorithm is feasible for large-scale GSM-MIMO systems, especially when the system is highly overdetermined.

Nevertheless, we only consider a fixed number of activated antennas at the transmitter in this work. In practice, the number of activated transmit antennas can be dynamic, which is one of the interesting topics for future works. Besides, we can further consider improving the search speed and robustness of the proposed algorithm with data-driven methods, which should be the major objective of our future research.

APPENDIX Proof of Theorem 3

To prove this theorem, we first introduce the following useful results.

A. Upper Bound on the Gaussian Hypergeometric Function

Lemma 1: For any positive numbers $a \ge b > 0$ and c > 1, we have the following inequality

$$_{2}F_{1}(1,a;b+1;\frac{1}{c}) \le \left(\frac{c}{c-1}\right)^{a},$$
 (A.1)

where ${}_{2}F_{1}(\cdot, \cdot; \cdot; \cdot)$ denotes the Gaussian hypergeometric function.

Proof: According to the identities of the Gaussian hypergeometric function [37], [38], we have

$${}_{2}F_{1}\left(1,a;b+1;\frac{1}{c}\right) = \left(\frac{c}{c-1}\right)^{a} {}_{2}F_{1}\left(b,a;b+1;\frac{1}{1-c}\right)$$
(A.2)

$$= \left(\frac{c}{c-1}\right) {}_{2}F_{1}\left(a,b;b+1;\frac{1}{1-c}\right)$$
(A.3)
$$\left(\frac{c}{c}\right)^{a} \int_{-1}^{1} t^{b-1}$$

$$= \left(\frac{c}{c-1}\right) b \int_0^{\infty} \frac{t^{c-1}}{\left(1 + \frac{t}{c-1}\right)^a} dt.$$
(A.4)

Under the given conditions, it is clear that the following inequality holds

$$\int_{0}^{1} \frac{t^{b-1}}{\left(1 + \frac{t}{c-1}\right)^{a}} dt \le \int_{0}^{1} t^{b-1} dt = \frac{1}{b}.$$
 (A.5)

Thus, we have

$$_{2}F_{1}\left(1,a;b+1;\frac{1}{c}\right) \leq \left(\frac{c}{c-1}\right)^{a}.$$
 (A.6)

B. Inequalities for the Gamma Function

Lemma 2 (See Theorem 5 in [39]): Let a > 0 and b > 0, then

$$\Gamma\left(\frac{a+b}{2}\right) \le [\Gamma(a)]^{\frac{1}{2}}[\Gamma(b)]^{\frac{1}{2}},\tag{A.7}$$

where $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ is the Gamma function.

Lemma 3: Let a > 0 be a positive real number, then we have the following inequality

$$\Gamma(2a) \le \frac{2^{4a-1}}{\pi a \left(a + \frac{1}{2}\right)^2} \left[\Gamma(a+1)\right]^2.$$
 (A.8)

Proof: For a > 0, the *Legendre Duplication Formula* [40, p. 58] states that

$$\Gamma(2a) = \frac{2^{2a-1}}{\sqrt{\pi}} \Gamma(a) \Gamma\left(a + \frac{1}{2}\right).$$
 (A.9)

In addition, we have the following inequality [39, Corollary 5]

$$\Gamma(a) \le \frac{2^{2a-1}}{\sqrt{\pi}a^2} \Gamma\left(a + \frac{1}{2}\right). \tag{A.10}$$

Then, evaluating (A.10) successively gives

$$\Gamma\left(a+\frac{1}{2}\right) \le \frac{2^{2a}}{\sqrt{\pi}\left(a+\frac{1}{2}\right)^2} \Gamma\left(a+1\right).$$
(A.11)

Thus, Lemma 3 can be proved by using the fact that $\Gamma(a+1) = a\Gamma(a)$ and substituting (A.11) into (A.9).

C. Proof

Now we are ready to complete the proof. Let S_e be the set of valid nodes to be expanded. From Theorem 2, we can bound the probability of expanding a valid node \hat{s}^k $(1 \le k \le N_t)$ as

$$\Pr\left(\hat{\boldsymbol{s}}^{k} \in \mathcal{S}_{e}\right) \leq \Pr\left(G(\hat{\boldsymbol{s}}^{k}) \leq G(\boldsymbol{s}^{*})\right) \qquad (A.12)$$

$$\leq \Pr\left(G(\hat{\boldsymbol{s}}^k) \leq G(\boldsymbol{s})\right),$$
 (A.13)

where *s* is the transmitted signal vector, and s^* denotes the ML estimate of *s*. Let us partition the upper triangular matrix R and the vector $v \triangleq Q^H w$ as

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{R}_{N_t-k,N_t-k} & \boldsymbol{R}_{N_t-k,k} \\ \boldsymbol{0} & \boldsymbol{R}_{k,k} \end{bmatrix} \text{ and } \boldsymbol{v} = \begin{bmatrix} \boldsymbol{v}^{N_t-k} \\ \boldsymbol{v}^k \\ \boldsymbol{v}^{N_e} \end{bmatrix},$$
(A.14)

where \mathbf{R}_{N_t-k,N_t-k} , $\mathbf{R}_{N_t-k,k}$ and $\mathbf{R}_{k,k}$ are the corresponding $(N_t - k) \times (N_t - k)$, $(N_t - k) \times k$ and $k \times k$ parts of \mathbf{R} , respectively. Analogously, \mathbf{v}^{N_t-k} , \mathbf{v}^k and \mathbf{v}^{N_e} are the corresponding $(N_t - k) \times 1$, $k \times 1$ and $N_e \times 1$ parts of \mathbf{v} , respectively. Then, we have

$$\Pr\left(G(\hat{\boldsymbol{s}}^{k}) \leq G(\boldsymbol{s})\right)$$

$$= \Pr\left(\left\| \begin{bmatrix} \boldsymbol{R}_{k,k} \\ \boldsymbol{0} \end{bmatrix} \tilde{\boldsymbol{s}}^{k} + \begin{bmatrix} \boldsymbol{v}^{k} \\ \boldsymbol{v}^{N_{e}} \end{bmatrix} \right\|^{2} \leq \left\| \begin{bmatrix} \boldsymbol{v}^{N_{t}-k} \\ \boldsymbol{v}^{k} \\ \boldsymbol{v}^{N_{e}} \end{bmatrix} \right\|^{2}\right),$$
(A.15)

where $\tilde{s}^k \triangleq s^k - \hat{s}^k$ denotes the difference vector. It is important to emphasize that we only consider the probability of expanding the valid nodes lying on the wrong paths in the proof, i.e. $\tilde{s}^k \neq 0$, since the nodes lying on the correct path must be expanded. According to [31] and [41], one can always find a $(N_e + k) \times (N_e + k)$ unitary matrix Θ whose distribution is independent of $\mathbf{R}_{k,k}$, such that

$$\bar{H} \triangleq \Theta \begin{bmatrix} R_{k,k} \\ 0 \end{bmatrix},$$
 (A.16)

whose distribution is the same as the lower $(N_e + k) \times k$ part of H. Accordingly, we also have

$$\bar{\boldsymbol{w}} \triangleq \boldsymbol{\Theta} \begin{bmatrix} \boldsymbol{v}^k \\ \boldsymbol{v}^{N_e} \end{bmatrix}, \qquad (A.17)$$

whose distribution is the same as the lower $(N_e + k) \times 1$ part of w. Then, (A.15) can be written as

$$\Pr\left(G(\hat{\boldsymbol{s}}^{k}) \leq G(\boldsymbol{s})\right)$$
(A.18)
=
$$\Pr\left(\left\|\bar{\boldsymbol{H}}\tilde{\boldsymbol{s}}^{k}\right\|^{2} + 2\bar{\boldsymbol{w}}^{H}\bar{\boldsymbol{H}}\tilde{\boldsymbol{s}}^{k} + \left\|\begin{bmatrix}\boldsymbol{v}^{k}\\\boldsymbol{v}^{N_{e}}\end{bmatrix}\right\|^{2} \leq \|\boldsymbol{v}\|^{2}\right)$$
(A.19)

$$= \Pr\left(\underbrace{\left\|\bar{\boldsymbol{H}}\tilde{\boldsymbol{s}}^{k}\right\|^{2} + 2\bar{\boldsymbol{w}}^{H}\bar{\boldsymbol{H}}\tilde{\boldsymbol{s}}^{k}}_{\delta_{k}} \leq \underbrace{\|\boldsymbol{v}^{N_{t}-k}\|^{2}}_{d_{k}}\right) \qquad (A.20)$$

$$= \Pr\left(\delta_k \le d_k\right),\tag{A.21}$$

where $\delta_k \triangleq \left\| \bar{\boldsymbol{H}} \tilde{\boldsymbol{s}}^k \right\|^2 + 2 \bar{\boldsymbol{w}}^H \bar{\boldsymbol{H}} \tilde{\boldsymbol{s}}^k$ and $d_k \triangleq \| \boldsymbol{v}^{N_t - k} \|^2 \ge 0$. Given $\bar{\boldsymbol{H}} \tilde{\boldsymbol{s}}^k$, it is obvious that δ_k is normally distributed with

$$\mathbb{E}\{\delta_k|\eta_k\} = \eta_k \text{ and } \mathsf{Var}\{\delta_k|\eta_k\} = 4\eta_k, \quad (A.22)$$

where $\eta_k \triangleq \left\| \bar{\boldsymbol{H}} \tilde{\boldsymbol{s}}^k \right\|^2 \ge 0$. Besides, $\frac{\eta_k}{\rho \| \tilde{\boldsymbol{s}}^k \|^2}$ is a chi-square random variable with $(N_e + k)$ degree of freedom, while d_k is a chi-square random variable with $(N_t - k)$ degree of freedom. Note that $\| \tilde{\boldsymbol{s}}^k \|^2 \ge 1$ holds for any $\hat{\boldsymbol{s}}^k \neq \boldsymbol{s}^k$. Using the Chernoff bound, we can bound the probability as

$$\Pr(\delta_k \le d_k | \eta_k, d_k) = Q(\frac{\eta_k - d_k}{2\sqrt{\eta_k}})$$
$$\le \begin{cases} e^{-\frac{(\eta_k - d_k)^2}{8\eta_k}}, & \eta_k > d_k\\ 1, & \eta_k < d_k \end{cases}$$
(A.23)

where $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{t^2}{2}} dt$ is the Gaussian *Q*-function. Since the three random variables $\bar{H}\tilde{s}^k$, \bar{w} and v^{N_t-k} are mutually independent, averaging over η_k and d_k gives the upper bound on the expected probability as

$$\frac{\Pr\left(\delta_{k} \leq d_{k}\right)}{= \mathbb{E}_{\left\{\eta_{k}, d_{k}\right\}} \left\{ \Pr\left(\delta_{k} \leq d_{k} \middle| \eta_{k}, d_{k}\right) \right\}}$$
(A.24)

$$\leq \mathbb{E}_{d_k} \left\{ \int_0^{d_k} f_{\eta_k}(t) \mathrm{d}t + \int_{d_k}^\infty e^{-\frac{(t-d_k)^2}{8t}} f_{\eta_k}(t) \mathrm{d}t \right\}$$
(A.25)

$$\leq \mathbb{E}_{d_k} \left\{ \frac{\gamma\left(\frac{N_e+k}{2}, \frac{d_k}{2\rho \|\tilde{s}^k\|^2}\right)}{\Gamma\left(\frac{N_e+k}{2}\right)} + e^{\frac{d_k}{4}} \int_0^\infty e^{-\frac{t}{8}} f_{\eta_k}(t) \mathrm{d}t \right\}$$
(A.26)

$$\leq \int_0^\infty \frac{\gamma\left(\frac{N_e+k}{2}, \frac{t}{2\rho}\right)}{\Gamma\left(\frac{N_e+k}{2}\right)} f_{d_k}(t) \mathrm{d}t + \int_0^\infty \frac{e^{\frac{d_k}{4}}}{(1+\frac{\rho}{4})^{\frac{N_e+k}{2}}} f_{d_k}(t) \mathrm{d}t$$
(A.27)

For the first integral in (A.27), it can be bounded as

$$\int_{0}^{\infty} \frac{\gamma\left(\frac{N_{e}+k}{2}, \frac{t}{2\rho}\right)}{\Gamma\left(\frac{N_{e}+k}{2}\right)} f_{d_{k}}(t) \mathrm{d}t \qquad (A.28)$$

$$\leq \frac{\Gamma\left(\frac{N_{e}+N_{t}}{2}\right) \rho^{-\frac{N_{e}+k}{2}} {}_{2}F_{1}\left(1, \frac{N_{e}+N_{t}}{2}; \frac{N_{e}+k}{2}+1; \frac{1}{1+\rho}\right)}{\Gamma\left(\frac{N_{t}-k}{2}\right) \Gamma\left(\frac{N_{e}+k}{2}+1\right) \left(\frac{1+\rho}{\rho}\right)^{\frac{N_{e}+N_{t}}{2}} \qquad (A.29)$$

$$\leq \frac{\Gamma\left(\frac{N_e+N_t}{2}\right)\rho^{-\frac{N_e+k}{2}}}{\Gamma\left(\frac{N_t-k}{2}\right)\Gamma\left(\frac{N_e+k}{2}+1\right)} \tag{A.30}$$

$$\leq \frac{[\Gamma(N_t - k)]^{\frac{1}{2}} [\Gamma(N_e + k)]^{\frac{1}{2}} \rho^{-\frac{N_e + k}{2}}}{\Gamma(\frac{N_t - k}{2}) \Gamma(\frac{N_e - k}{2} + 1)}$$
(A.31)

$$<\frac{\left[\Gamma\left(N_t-k\right)\right]^{\frac{1}{2}}}{\Gamma\left(\frac{N_t-k}{2}\right)}\left(\frac{\rho}{4}\right)^{-\frac{N_e+k}{2}},\tag{A.32}$$

where $\gamma(\cdot, \cdot)$ denotes the incomplete gamma function. Note that the inequality in (A.30) holds from Lemma 1, the inequality in (A.31) holds from Lemma 2, and the inequality in (A.32) holds according to Lemma 3. For the second integral in (A.27), it can be bounded as

$$\int_{0}^{\infty} \frac{e^{\frac{d_{k}}{4}}}{(1+\frac{\rho}{4})^{\frac{N_{e}+k}{2}}} f_{d_{k}}(t) dt = \frac{2^{\frac{N_{t}-k}{2}}}{(1+\frac{\rho}{4})^{\frac{N_{e}+k}{2}}}$$
(A.33)
$$< 2^{\frac{N_{t}-k}{2}} \left(\frac{\rho}{4}\right)^{-\frac{N_{e}+k}{2}}.$$
(A.34)

By substituting (A.32) and (A.34) into (A.27), we have

$$\overline{\Pr}\left(\hat{\boldsymbol{s}}^{k} \in \mathcal{S}_{e}\right) < \left(\frac{\left[\Gamma\left(N_{t}-k\right)\right]^{\frac{1}{2}}}{\Gamma\left(\frac{N_{t}-k}{2}\right)} + 2^{\frac{N_{t}-k}{2}}\right) \left(\frac{\rho}{4}\right)^{-\frac{N_{e}+k}{2}}.$$
(A.35)

Now it is clear to find from (A.35) that with fixed N_t , the expected probability of expanding a k-level valid node lying on the wrong path, will decrease exponentially with increasing ρ and N_e . Since the total number of valid nodes are fixed, the expected complexity of the proposed algorithm will decrease exponentially as well. In particular, when ρ and N_e are sufficiently large, the expected probability in (A.35) vanishes for each wrong valid node. In this case, only the correct nodes will be eventually expanded by the algorithm. Since the computational complexity of the algorithm converges to $\mathcal{O}(N_t^2)$, which is the time complexity of computing the cumulative cost of the correct path.

REFERENCES

M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," *IEEE Commun. Mag.*, vol. 58, no. 3, pp. 55–61, Mar. 2020.

- [2] T.-H. Vu, T.-V. Nguyen, D. B. da Costa, and S. Kim, "Performance analysis and deep learning design of underlay cognitive NOMA-based CDRT networks with imperfect SIC and co-channel interference," *IEEE Trans. Commun.*, vol. 69, no. 12, pp. 8159–8174, Dec. 2021.
- [3] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Jun. 2014.
- [4] A. S. de Sena, "Massive MIMO-NOMA networks with imperfect SIC: Design and fairness enhancement," *IEEE Trans. Wireless Commun.*, vol. 19, no. 9, pp. 6100–6115, Jun. 2020.
- [5] S. Zheng, C. Shen, and X. Chen, "Design and analysis of uplink and downlink communications for federated learning," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 7, pp. 2150–2167, Jul. 2021.
- [6] M. D. Renzo, H. Haas, A. Ghrayeb, S. Sugiura, and L. Hanzo, "Spatial modulation for generalized MIMO: Challenges, opportunities, and implementation," *Proc. IEEE*, vol. 102, no. 1, pp. 56–103, Jan. 2014.
- [7] M. Wen et al., "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.
- [8] A. Younis, R. Mesleh, M. D. Renzo, and H. Haas, "Generalised spatial modulation for large-scale MIMO," in *[Proc. 22nd Eur. Signal Process. Conf., (EUSIPCO)*, Lisbon, Portugal, Sep. 2014, pp. 346–350.
- [9] J. Liao, J. Zhao, F. Gao, and G. Y. Li, "A model-driven deep learning method for massive MIMO detection," *IEEE Commun. Lett.*, vol. 24, no. 8, pp. 1724–1728, Aug. 2020.
- [10] F. Liang, C. Shen, and F. Wu, "An iterative BP-CNN architecture for channel decoding," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 144–159, Feb. 2018.
- [11] Y. Xiao et al., "Low-complexity signal detection for generalized spatial modulation," *IEEE Commun. Lett.*, vol. 18, no. 3, pp. 403–406, Mar. 2014.
- [12] C.-T. Lin, W.-R. Wu, and C.-Y. Liu, "Low-complexity ML detectors for generalized spatial modulation systems," *IEEE Trans. Commun.*, vol. 63, no. 11, pp. 4214–4230, Nov. 2015.
- [13] Z. Ma, F. Gao, J. Jiang, and Y.-C. Liang, "Cooperative detection for ambient backscatter assisted generalized spatial modulation," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2019, pp. 1–6.
- [14] L. Liu, S. Huang, and B. M. Kurkoski, "Memory approximate message passing," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2021, pp. 1379–1384.
- [15] L. Liu, C. Yuen, Y. L. Guan, Y. Li, and C. Huang, "Gaussian message passing for overloaded massive MIMO-NOMA," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 210–226, Jan. 2019.
- [16] C. Wang, P. Cheng, Z. Chen, J. A. Zhang, Y. Xiao, and L. Gui, "Near-ML low-complexity detection for generalized spatial modulation," *IEEE Commun. Lett.*, vol. 20, no. 3, pp. 618–621, Mar. 2016.
- [17] L. Xiao *et al.*, "Efficient compressive sensing detectors for generalized spatial modulation systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1284–1298, Feb. 2017.
- [18] Y. Wu, H. Ying, X.-Q. Jiang, and H. Hai, "A joint data mapping and detection for high performance generalized spatial modulation," *IEEE Commun. Lett.*, vol. 23, no. 11, pp. 2008–2011, Nov. 2019.
- [19] H. Albinsaid, K. Singh, S. Biswas, C.-P. Li, and M.-S. Alouini, "Block deep neural network-based signal detector for generalized spatial modulation," *IEEE Commun. Lett.*, vol. 24, no. 12, pp. 2775–2779, Dec. 2020.
- [20] J. A. Cal-Braz and R. Sampaio-Neto, "Low-complexity sphere decoding detector for generalized spatial modulation systems," *IEEE Commun. Lett.*, vol. 18, no. 6, pp. 949–952, Jun. 2014.
- [21] B. Zheng, X. Wang, M. Wen, and F. Chen, "Soft demodulation algorithms for generalized spatial modulation using deterministic sequential Monte Carlo," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 3953–3967, Jun. 2017.
- [22] B. Zheng, M. Wen, F. Chen, N. Huang, F. Ji, and H. Yu, "The K-best sphere decoding for soft detection of generalized spatial modulation," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4803–4816, Nov. 2017.
- [23] T. Q. Tran, S. Sugiura, and K. Lee, "Ordering- and partitioning-aided sphere decoding for generalized spatial modulation," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 10087–10091, Oct. 2018.
- [24] Y.-M. Chen, W.-C. Cheng, C.-P. Li, and Z. J. Haas, "Low-complexity generalized spatial modulation schemes using codebook-assisted MIMO detectors," *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 12358–12362, Dec. 2018.
- [25] T.-H. Liu, C.-E. Chen, and C.-H. Liu, "Fast maximum likelihood detection of the generalized spatially modulated signals using successive sphere decoding algorithms," *IEEE Commun. Lett.*, vol. 23, no. 4, pp. 656–659, Apr. 2019.

- [26] V. M. Garcia-Molla, F. J. Martínez-Zaldívar, M. Angeles Simarro, and A. Gonzalez, "Maximum likelihood low-complexity GSM detection for large MIMO systems," *Signal Process.*, vol. 175, Oct. 2020, Art. no. 107661.
- [27] J. Sun, Y. Zhang, J. Xue, and Z. Xu, "Learning to search for MIMO detection," *IEEE Trans. Wireless Commun.*, vol. 19, no. 11, pp. 7571–7584, Nov. 2020.
- [28] Y. Dai and Z. Yan, "Memory-constrained tree search detection and new ordering schemes," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 6, pp. 1026–1037, Dec. 2009.
- [29] R. Y. Chang and W.-H. Chung, "Best-first tree search with probabilistic node ordering for MIMO detection: Generalization and performancecomplexity tradeoff," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 780–789, Feb. 2012.
- [30] L. He, K. He, L. Fan, X. Lei, A. Nallanathan, and G. K. Karagiannidis, "Towards optimally efficient search with deep learning for large-scale MIMO systems," *IEEE Trans. Commun.*, vol. 70, no. 2, pp. 101–116, May 2022.
- [31] A. D. Murugan, H. El Gamal, M. O. Damen, and G. Caire, "A unified framework for tree search decoding: Rediscovering the sequential decoder," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 933–953, Mar. 2006.
- [32] J. Schalkwijk, "An algorithm for source coding," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 3, pp. 395–399, May 1972.
- [33] P. Kabal. (2018). Combinatorial Coding and Lexicographic Ordering. [Online]. Available: http://www-mmsp.ece.mcgill.ca/Documents/ Reports/2018/KabalR2018.pdf
- [34] S. J. Russell, "Efficient memory-bounded search methods," in *Proc. Eur. Conf. Artif. Intell. (ECAI)*, 1992, pp. 1–5.
- [35] X. Lai, "Outdated access point selection for mobile edge computing with cochannel interference," *IEEE Trans. Vehic. Tech.*, early access, Apr. 14, 2022, doi: 10.1109/TVT.2022.3167405.
- [36] S. Tang, "Dilated convolution based CSI feedback compression for massive MIMO systems," *IEEE Trans. Vehic. Tech.*, vol. 71, no. 5, pp. 211–216, Jun. 2022.
- [37] F. Beukers, "Gauss' hypergeometric function," in Arithmetic Geometry Around Hypergeometric Functions. Cham, Switzerland: Springer, 2007, pp. 23–42.
- [38] G. Nemes and A. B. Olde Daalhuis, "Large-parameter asymptotic expansions for the legendre and allied functions," *SIAM J. Math. Anal.*, vol. 52, no. 1, pp. 437–470, Jan. 2020.
- [39] S. Dragomir, R. Agarwal, and N. Barnett, "Inequalities for beta and gamma functions via some classical and new integral inequalities," *J. Inequal. Appl.*, vol. 2000, no. 2, 2000, Art. no. 504054.
- [40] L. Andrews, "Special functions for engineers and applied mathematicians," Appl. Opt., vol. 25, no. 18, p. 3096, 1986.
- [41] B. Hassibi and H. Vikalo, "On the sphere-decoding algorithm I. Expected complexity," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2806–2818, Aug. 2005.



Ke He received the bachelor's degree from the Wuhan University of Technology in 2015 and the master's degree in computer engineering from Guangzhou University in 2021. He is currently pursuing the Ph.D. degree with the Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg. His current research interests mainly focus on signal processing and machine learning.



Le He received the bachelor's degree from Northeastern University in 2018. He is currently pursuing the master's degree with the Network and Intelligent Computing Technology (NICT) Laboratory, School of Computer Science, Guangzhou University. His current research interests include machine learning, signal processing, and their real world applications.



Lisheng Fan received the bachelor's degree from the Department of Electronic Engineering, Fudan University, China, in 2002, the master's degree from the Department of Electronic Engineering, Tsinghua University, China, in 2005, and the Ph.D. degree from the Department of Communications and Integrated Systems, Tokyo Institute of Technology, Japan, in 2008. He is currently a Professor with Guangzhou University. He has published many articles in international journals, such as IEEE TRANS-ACTIONS ON WIRELESS COMMUNICATIONS, IEEE

TRANSACTIONS ON COMMUNICATIONS, and IEEE TRANSACTIONS ON INFORMATION THEORY, as well as papers in conferences, such as IEEE ICC, IEEE GLOBECOM, and IEEE WCNC. His research interests span in the areas of wireless cooperative communications, physical-layer secure communications, interference modeling, and system performance evaluation. He has served as a member of technical program committees for IEEE conferences, such as GLOBECOM, ICC, WCNC, and VTC. He has served as the Chair for Wireless Communications and Networking Symposium for Chinacom 2014. He is a Guest Editor of *EURASIP Journal on Wireless Communications and Networking*.



Yansha Deng (Member, IEEE) received the Ph.D. degree in electrical engineering from the Queen Mary University of London, U.K., in 2015. From 2015 to 2017, she was a Post-Doctoral Research Fellow with King's College London, U.K., where she is currently a Senior Lecturer (an Associate Professor) with the Department of Engineering. Her research interests include molecular communication and machine learning for 5G/6G wireless networks. She has served as a TPC Member for many IEEE conferences, such as IEEE GLOBECOM and

ICC. She was a recipient of the Best Paper Awards from ICC 2016 and GLOBECOM 2017 as the first author. She was an Exemplary Reviewer of the IEEE TRANSACTIONS ON COMMUNICATIONS in 2016 and 2017 and IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS in 2018. She is an Associate Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS and IEEE TRANSACTIONS ON MOLECULAR, BIOLOGICAL AND MULTI-SCALE COMMUNICATIONS, a Senior Editor of the IEEE COMMUNICATIONS LETTERS, and the Vertical Area Editor of *IEEE Internet of Things Magazine*.



Xianfu Lei (Member, IEEE) received the Ph.D. degree from Southwest Jiaotong University in 2012. From 2012 to 2014, he has worked as a Research Fellow with the Department of Electrical and Computer Engineering, Utah State University. He has been an Associate Professor with the School of Information Science and Technology, Southwest Jiaotong University, since 2015. His research interests are 5G/6G networks, cooperative and energy harvesting networks, and physical-layer security. He has received the Best Paper Award in IEEE/CIC

ICCC2020, the Best Paper Award in WCSP2018, the WCSP 10-Year Anniversary Excellent Paper Award, the IEEE COMMUNICATIONS LETTERS Exemplary Editor 2019, and the Natural Science Award of China Institute of Communications (2019). He has served as a Senior/Associate Editor for IEEE COMMUNICATIONS LETTERS from 2014 to 2019. He has been serving as an Area Editor for IEEE COMMUNICATIONS LETTERS and an Associate Editor for IEEE WIRELESS COMMUNICATIONS LETTERS and IEEE TRANSACTIONS ON COMMUNICATIONS.



George K. Karagiannidis (Fellow, IEEE) was born in Pithagorion, Samos Island, Greece. He received the University Diploma (five years) and Ph.D. degrees in electrical and computer engineering from the University of Patras in 1987 and 1999, respectively.

From 2000 to 2004, he was a Senior Researcher at the Institute for Space Applications and Remote Sensing, National Observatory of Athens, Greece. In June 2004, he joined the Faculty of the Aristotle University of Thessaloniki, Greece, where he is cur-

rently a Professor with the Electrical and Computer Engineering Department and the Head of the Wireless Communications and Information Processing (WCIP) Group. He is also an Honorary Professor at Southwest Jiaotong University, Chengdu, China. His research interests are in the broad area of digital communications systems and signal processing, with emphasis on wireless communications, optical wireless communications, wireless power transfer, and applications and communications and signal processing for biomedical engineering.

Dr. Karagiannidis has been involved as the general chair, the technical program chair, and a member of technical program committees in several IEEE and non-IEEE conferences. He is one of the highly-cited authors across all areas of electrical engineering, recognized from Clarivate Analytics as the Web of Science Highly-Cited Researcher in the six consecutive years 2015–2020. In the past, he was an editor in several IEEE journals. From 2012 to 2015, he was the Editor-in Chief of IEEE COMMUNICATIONS LETTERS. He serves as the Associate Editor-in Chief for IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY.