Transmit Antenna Selection and Beamformer Design for Secure Spatial Modulation with Rough CSI of Eve

Guiyang Xia, Yan Lin, *Member, IEEE*, Tingting Liu, *Member, IEEE*, Feng Shu, *Member, IEEE*, and Lajos Hanzo, *Fellow, IEEE*

Abstract—The security of spatial modulation (SM) aided networks can always be improved by reducing the desired link's power at the cost of degrading its bit error ratio performance and assuming the power consumed to artificial noise (AN) projection (ANP). We formulate the joint optimization problem of maximizing the secrecy rate (Max-SR) over the transmit antenna selection and ANP in the context of secure SM-aided networks. In order to solve this problem, we provide a pair of solutions, namely joint and separate solutions. Specifically, an accurate approximation of the SR is used for reducing the computational complexity, and the optimal AN covariance matrix (ANCM) is found by convex optimization for any given active antenna group (AAG). Then, given a large set of AAGs, simulated annealing mechanism is invoked for optimizing the choice of AAG, where the corresponding ANCM is recomputed by this optimization method as well when the AAG changes. To further reduce the complexity of the above-mentioned joint optimization, a low-complexity two-stage separate optimization method is also proposed. Moreover, when the number of transmit antennas tends to infinity, the Max-SR problem becomes equivalent to that of maximizing the ratio of the desired user's signal-to-interference-plus-noise ratio to the eavesdropper's. Thus, our original problem reduces to a fractional programming problem and a significant computational complexity reduction can be achieved. Finally, our simulation results verify the efficiency of the proposed methods in terms of the SR performance attained.

Index Terms—Spatial modulation, active antenna group selection, artificial noise, secure transmission, finite-alphabet input.

I. INTRODUCTION

Guiyang Xia, Yan Lin, and Feng Shu are with the School of Electronic and Optical Engineering, Nanjing University of Science and Technology, 210094, China. Email: xiaguiyang@njust.edu.cn; yanlin@njust.edu.cn; shufeng@njust.edu.cn.

Tingting Liu is with the School of Information and Communication Engineering, Nanjing Institute of Technology, Nanjing 211167, China. Email:liutt@njit.edu.cn.

Lajos Hanzo is with the school of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. Email:lh@ecs.soton.ac.uk.

AS a promising technique, spatial modulation (SM) [1] invokes the more general index modulation (IM) concept for the transmit antennas (TAs) to convey extra information [2] [3], which has attracted tremendous attention over the past decade. Recently, SM has shown advantages in terms of its spectral efficiency versus energy efficiency in various communication networks, including cooperative, full-duplex, single/multi-user and cognitive radio systems. Therefore, SM has become a popular candidate for next-generation systems [4]–[9]. Due to the broadcast nature of radio propagation, its security problem has to be considered, since various types of wireless access devices may overhear the private messages. Traditionally, the security of a communication system has been ensured through cryptography and authentication in the network layer, which often imposes additional computational complexity for key generation and complex decryption algorithms [10] [11]. However, the key distribution and management is challenging for large-scale wireless networks. Nevertheless, physical layer security (PLS) [12]–[14] does not require a key for ensuring security, where the fundamental philosophy is to exploit the randomness of communication channels. Then, the transmitter (Alice) aims for conveying private information securely to the desired receiver and to keep the illegitimate receiver as ignorant of the private information as possible [15]–[17].

To improve the security against an eavesdropper, several SM-based PLS schemes have been proposed. In [18], a full-duplex receiver was employed at the desired receiver (Bob), where Bob receives confidential messages and simultaneously transmits artificial noise (AN) to corrupt the illegal receiver (Eve). In [19] and [20], AN was transmitted along the null-space of the legitimate channel for enhancing its security without any prior knowledge of Eve's location. A precoding-aided SM scheme was proposed in [21] to exploit the index of receive antennas to convey spatial information, where the precoding matrix was designed for maximizing the signal-to-noise ratio (SNR) at the desired receiver, whilst minimizing the eavesdropper's SNR. Another interesting proposal of Wu *et al.* [22] was that of injecting AN in the null-space of the legitimate channel to combat passive eavesdroppers. Additionally, several time-varying mapping schemes were proposed in [23] [24] to enhance the security of SM systems, where the time-varying characteristic was the unique channel fading, which was only known to transmitter and to the legitimate receiver. Nevertheless, these researches have not considered the effect

of a varying number of antennas on security.

When the number of TAs is not a power of two, selecting an active antenna group (AAG) can be adopted to further improve the security of SM systems [25]–[27]. Although the AAG selection based on the minimum Euclidean distance criterion and minimizing the bit-error-rate has drawn considerable attention in conventional SM systems [28]–[31], none of these preceding contributions have considered the presence of eavesdroppers. Recently, the AAG selection scheme was also been highlighted as an efficient way of enhancing the security and creating a secure SM (SSM) network. To be specific, with the aid of AN, a leakage-based null-space projection (NSP) method (LNSP) was proposed in [25] operating at an extremely low-complexity, which achieved an acceptable secrecy rate (SR) performance. As a powerful measure of mitigating the computational complexity, the cut-off rate was invoked in [32] for optimizing the mutual information (MI) or the channel capacity of SM-based systems. Subsequently, the authors proposed several low-complexity precoding schemes for maximizing the cut-off rate and showed impressive performance gains. Moreover, the authors in [26] proposed a set of AAG selection schemes for maximizing the SR, in which the cut-off rate based scheme shown almost the same SR profit as that of exhaustive search (ES) over the original SR expression. However, these AAG selection schemes only considered the SR performance gain gleaned from channel diversity, which causes a serious SR performance degradation in the high-SNR region.

As mentioned, AN was used to improve the SR performance by directly projecting into the null-space of the legitimate channel, which was achieved by exploiting its closed-form expression. However, this approach has its limitations, because it only allows the AN avoid affecting the detection of the desired receiver, but dispenses with more holistic considerations. In addition, for circumventing that the size of the AAG combinations grows exponentially upon increasing the number of TAs, simulated annealing (SA) [33] [34] was invoked for reducing the search complexity of this combinatorial optimization problem. The key benefit of SA is that it avoids convergence to local minima. In this context, we assume that the rough channel state information (CSI) of the illegal channel can be obtained at the transmitter, and then exploit the PLS of an SSM system, where both TA selection and AN design are invoked for enhancing the security. The main contributions of this paper can be summarized as follows:

1) To enhance the security of SM systems, a joint AAG selection and AN projection design problem is formulated. Furthermore, for reducing the complexity of the original SR expression involving multiple integrals, a lower bound of the approximate MI (AMI) between Alice and Bob as well as an upper bound of the AMI between Alice and Eve are derived to formulate a concave maximization problem over ANCM for any given AAG.
2) In order to solve the intractable 0-1 combinatorial optimization problem of AAG selection, the SA mechanism is invoked for approaching the global optimum. A joint SA-based maximization of the approximate SR (ASR) (SA-Max-ASR) is proposed by repeatedly optimizing the Max-ASR criterion for finding the optimal ANCM once a new AAG is generated. Additionally, a low-complexity two-stage separate SA-Max-ASR algorithm is also proposed.
3) As the number of TAs is increased, the double summation operation in the ASR expression over the legitimate combination of TAs and modulated symbols is eliminated and a simplified expression is derived. Correspondingly, a low-complexity method of maximizing the ratio between the signal-to-interference-plus-noise ratio (SINR) at the desired receiver and that at the eavesdropper is conceived for efficiently designing the AAG and the ANCM.

The remainder of this treatise is organized as follows. In Section II, an SSM system is described and the definition of its average SR is given. Subsequently, a gradient descend (GD) based method of optimizing the AN projection matrix (ANPM) is proposed. In Section III, a closed-form ASR expression is derived and a concave maximization problem is formulated. Then an SA-based AAG selection scheme is presented. In Section IV, a simple optimization objective function (OF) is derived to replace the original function as the number of TAs tends to a large value, and a low complexity scheme is presented. In Section V, both the convergence and the complexity of the proposed methods is analysed, followed by our numerical results in Section VI. Finally, Section VII offers our conclusions.

Notations: Vectors and matrices are represented in boldface. $\lfloor \cdot \rfloor$, $\| \cdot \|^2$, C_m^n denote the floor function, the Frobenius norm and the binomial coefficient, respectively. The superscripts $(\cdot)^T$ and $(\cdot)^H$ represent the transpose and the conjugate transpose operations, respectively. Besides, $\text{tr}(\cdot)$, $\text{diag}(\cdot)$ and $\text{det}(\cdot)$ denote the trace, the determinant and the diagonal of a matrix, respectively. $\mathbb{E}(\cdot)$ means the expectation operation. Matrix $\mathbf{I}_N$ refers to the N-by-N identity matrix. $\mathcal{CN}(\mu, \sigma^2)$ implies a complex Gaussian distribution with μ mean and σ^2 variance.

II. SYSTEM MODEL AND PROBLEM FORMULATION

Consider the SSM system illustrated in Fig. 1, where N_t transmit antennas (TAs) are employed at Alice, while both Bob and Eve are equipped with a single antenna. Here, Eve intends to intercept the confidential messages from Alice to Bob.

Considering the fact that N_t is not always a power of two, it becomes necessary to select an AAG choosing $N_s = 2^{\lfloor \log_2 N_t \rfloor}$ elements. In accordance with the concept of SM, Alice activates one out of N_s antenna indices to convey $\lfloor \log_2 N_t \rfloor$ bits of information. Then, $\lfloor \log_2 N_t \rfloor + \log_2 M$ bits can be transmitted per channel use in total, where M is the size of the $\mathcal{M}$-ary classic modulation constellation pointer. Additionally, it can be observed that there are $L = C_{N_t}^{N_s}$ possible AAG combinations. Let us denote the set of all possible AAGs as $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, \cdots, \mathcal{C}_L\}$, where $\mathcal{C}_l \, (l = 1, \cdots, L)$ denotes the l-th AAG and each AAG contains N_s TAs.
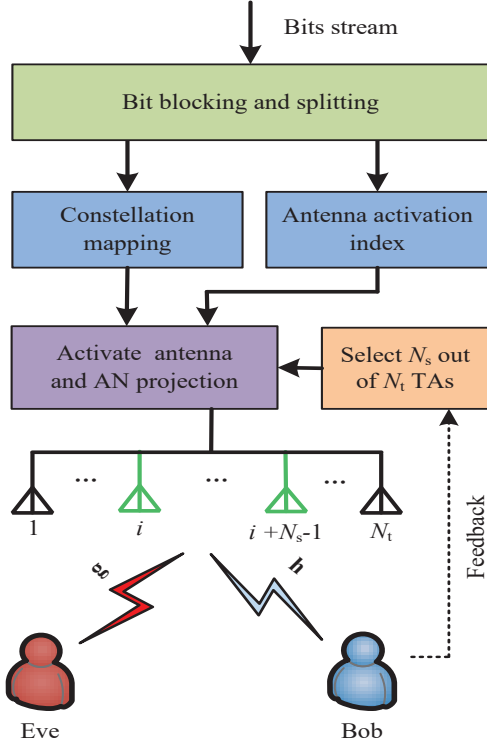
Fig. 1. A secure SM system model.

For a given AAG $\mathcal{C}_l$, an SM symbol $\boldsymbol{x}$ associated with AN is given by

$$\begin{aligned}
\boldsymbol{x} &= \sqrt{P_1}\mathbf{s}_i^j + \sqrt{P_2}\mathbf{Tn} \\
&= \sqrt{P_1}\mathbf{e}_i b_j + \sqrt{P_2}\mathbf{Tn}, \quad (1)
\end{aligned}$$

where P_1 and P_2 represent the power associated with $P_1 + P_2 = P_s$, and P_s denotes the total transmit power. Vector $\mathbf{e}_i$ represents the i-th column of $\mathbf{I}_{N_s}$, which implies that the i-th ($i = 1, \cdots, N_s$) TA of $\mathcal{C}_l$ is activated. Moreover, $b_j \in \mathcal{M} = \{b_1, \cdots, b_M\}$ is the j-th amplitude phase modulation (APM) symbol in one $\mathcal{M}$-ary constellation. Additionally, $\mathbf{n} \in \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_s})$ is the AN vector, and $\mathbf{T}$ is the ANPM with $\mathbb{E}\left(\mathbf{TT}^H\right) = \mathbf{Q}$, where $\mathbf{Q}$ is the ANCM and satisfies $\mathsf{tr}(\mathbf{Q}) = 1$. The corresponding signals received at Bob and at Eve are respectively represented as

$$\begin{aligned}
y_B &= \mathbf{hS}_l\boldsymbol{x} + n_B \\
&= \sqrt{P_1}\mathbf{h}_l\mathbf{e}_i b_j + \sqrt{P_2}\mathbf{h}_l\mathbf{Tn} + n_B, \quad (2) \\
y_E &= \mathbf{gS}_l\boldsymbol{x} + n_E \\
&= \sqrt{P_1}\mathbf{g}_l\mathbf{e}_i b_j + \sqrt{P_2}\mathbf{g}_l\mathbf{Tn} + n_E, \quad (3)
\end{aligned}$$

where $\mathbf{S}_l$ is the AAG selection matrix, while $\mathbf{h}_l = \mathbf{hS}_l \in \mathbb{C}^{1 \times N_s}$ and $\mathbf{g}_l = \mathbf{gS}_l \in \mathbb{C}^{1 \times N_s}$ are the sub-channels of $\mathbf{h}$ and $\mathbf{g}$ that depends on $\mathcal{C}_l$. Furthermore, $\mathbf{h} \in \mathbb{C}^{1 \times N_t}$ and $\mathbf{g} \in \mathbb{C}^{1 \times N_t}$ are the complex channel gain vectors spanning from Alice to Bob and from Alice to Eve, respectively. Additionally, $n_B \in \mathcal{CN}(0, \sigma_B^2)$ and $n_E \in \mathcal{CN}(0, \sigma_E^2)$ are the independent complex Gaussian noises at Bob and at Eve. Based upon the received signal in (2) for a given $\mathcal{C}_l$, the maximum likelihood (ML)

detector may be utilized by Bob to jointly detect the spatial symbol and the conventional APM symbol, formulated as:

$$(\hat{i}, \hat{j}) = \arg\min_{\mathcal{S}} |y_B - \sqrt{P_1}\mathbf{h}_l\mathbf{e}_i b_j|^2, \quad (4)$$

where $\mathcal{S} = \mathcal{C}_l \times \mathcal{M}$ is the super-alphabet set that contains all possible combinations of the active antennas and the conventional symbols. As a matter of fact, the joint ML principle of (4) is the optimal detection since each component of $\mathcal{S}$ has equal probability to be selected [35].

In this context, we assume that Alice has perfect CSI on the main channel. An accurate CSI estimate of the channel from Alice to Bob may be derived by using training sequences, which is then sent back to the transmitter through dedicated feedback links [17]. Additionally, it is also assumed that Alice can obtain a rough CSI estimate of the eavesdropper's channel, which corresponds to the scenario that Eve is an active user in wireless networks. According to [36] [37], the additive uncertainty model for the CSI of Eve at Alice is given by

$$\mathbf{g} = \tilde{\mathbf{g}} + \Delta\mathbf{g}, \quad (5)$$

where $\tilde{\mathbf{g}}$ is the estimated channel of $\mathbf{g} \sim \mathcal{CN}(0, \mathbf{I})$, and $\Delta\mathbf{g}$ is the corresponding estimation error, which is assumed to be a zero-mean Gaussian random variable associated with the covariance of $\sigma_e^2 \in (0, 1)$, i.e., we have $\Delta\mathbf{g} \sim \mathcal{CN}(0, \sigma_e^2\mathbf{I})$. It is worth mentioning that, we consider the worst case that $\mathbf{g}$ is perfectly estimated by Eve. In the remainder of this treatise, we will develop schemes of enhancing the security of such SSM networks.

In general, the transmit symbol b_j is equiprobably drawn from a discrete $\mathcal{M}$-ary constellation. For a specific channel realization and fixed $\mathcal{C}_l$, the MI between Alice and Bob can be expressed as

$$I(\boldsymbol{x}; y_B) = \log_2 N_s M -$$
$$\frac{1}{N_s M}\sum_{i=1}^{N_s M} \mathbb{E}_{n_B}\left\{\log_2 \sum_{j=1}^{N_s M} \exp\left(\frac{-f_{b,i,j}}{P_2\mathbf{h}_l\mathbf{Q}\mathbf{h}_l^H + \sigma_B^2}\right)\right\}, \quad (6)$$

where

$$f_{b,i,j} = |\sqrt{P_1}\mathbf{h}_l(\boldsymbol{x}_i - \boldsymbol{x}_j) + n_B|^2 - |n_B|^2, \quad (7)$$

and $\boldsymbol{x}_{\{\cdot\}}$ represents a legitimate transmit symbol in the set $\mathcal{S}$. Similarly, we have the MI between Alice and Eve:

$$I(\boldsymbol{x}; y_E) = \log_2 N_s M -$$
$$\frac{1}{N_s M}\sum_{m=1}^{N_s M} \mathbb{E}_{n_E}\left\{\log_2 \sum_{k=1}^{N_s M} \exp\left(\frac{-f_{e,m,k}}{P_2\mathbf{g}_l\mathbf{Q}\mathbf{g}_l^H + \sigma_E^2}\right)\right\}, \quad (8)$$

where

$$f_{e,m,k} = |\sqrt{P_1}\mathbf{g}_l(\boldsymbol{x}_m - \boldsymbol{x}_k) + n_E|^2 - |n_E|^2. \quad (9)$$

Notably, this work considers the worst case scenario that the eavesdropper knows the mapping rule of both the modulated symbols and TA indices, thus the maximal MI can be attained by Eve. Combining (6) and (8) yields the ergodic SR defined as

$$\bar{R}_s = \mathbb{E}_{\mathbf{h},\mathbf{g}}\left[I(\boldsymbol{x}; y_B) - I(\boldsymbol{x}; y_E), 0\right]^+, \quad (10)$$

where $[a]^+ = \max(a, 0)$ and $\hat{R}_s = I(\boldsymbol{x}; y_B) - I(\boldsymbol{x}; y_E)$ is the instantaneous SR for a specific channel realization. Similar to [38], we adopt a lower bound on $\hat{R}_s$ to formulate a corresponding optimization problem in the face of having a random CSI error of $\boldsymbol{\Delta g}$, where the lower bound is $R_s = I(\boldsymbol{x}; y_B) - \mathbb{E}_{\boldsymbol{\Delta g}} I(\boldsymbol{x}; y_E)$. Since the number of TAs is not a power of two, it is necessary to select an AAG and to design AN for enhancing the security. The optimization problem of maximizing the SR (Max-SR) can be cast as

$$\max_{\mathbf{s}_l, \mathbf{Q}} \quad R_s \tag{11a}$$

$$s.t. \quad \mathsf{tr}(\mathbf{Q}) = 1, \tag{11b}$$

$$\mathbf{Q} \succeq 0, \tag{11c}$$

$$\sum_{i=1}^{N_t} s_i = N_s, \tag{11d}$$

$$s_i \in \{0, 1\}, i = 1, \cdots, N_t, \tag{11e}$$

where s_i is the i-th element of the AAG vector $\mathbf{s} \in R^{N_t \times 1}$. Mathematically, N_s '1' elements are assigned to the diagonal locations of a $N_t \times N_t$ matrix of zeros. For instance, we have to select 4 TAs to form an AAG when $N_t = 6$. If the 5-th AAG of say $\mathbf{s}_5 = \{1, 2, 4, 6\}$, has the optimal SR performance, the corresponding AAG selection matrix becomes:

$$\mathbf{S}_5 = \mathsf{diag}(\mathbf{s}_5) \triangleq \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}^T. \tag{12}$$

For a given $\mathbf{S}_l$, the actual effective sub-channel spanning from Alice to Bob is $\mathbf{h}_l = \mathbf{h}\mathbf{S}_l$. Similarly, the sub-channel from Alice to Eve is $\mathbf{g}_l = \mathbf{g}\mathbf{S}_l$. Observe from (11) that the optimization problem is a mixed integer programming problem, where (11b) and (11c) are continuous constraints whereas (11d) and (11e) are 0-1 integer constraints. More particularly, due to the constraints (11d) and (11e), the optimization problem in (11) is an NP-hard problem, hence it requires an excessive search complexity to find the optimal subset of maximizing R_s. Additionally, the OF is non-convex and it is not in closed-form, thus it is a challenging task to solve problem (11).

III. AAG SELECTION AND AN DESIGN FOR MAXIMIZING THE SR PERFORMANCE

In this section, a GD-based method is first presented to design the ANPM, where ES is adopted to find the optimal AAG. This scheme is used as our benchmark. Then, a joint SA-Max-ASR scheme of AAG and ANCM is proposed to achieve a high SR performance. However, to reduce the complexity, a low-complexity separate SA-Max-ASR scheme is also proposed at the cost of a slight performance loss.

A. GD-based design of AN projection matrix

Due to the non-convexity of the OF R_s, obtaining a closed-form solution for (11) becomes intractable. However, it is natural to adopt numerical algorithms to search for local maxima of the OF such as GD. Given a fixed $\mathbf{S}_l$, $\mathbf{T}$ can be optimized first, and then averaged over all possible AAGs

---

Algorithm 1 Numerical search for maximizing SR (ES plus GD)

1: List all possible $\mathbf{S}_l$, $l = (1, \cdots, L)$.
2: **For** $l = 1 : L$
3: Initial $\mathbf{T}_1$ with constraint $\mathsf{tr}(\mathbf{T}_1 \mathbf{T}_1^H) = 1$. Set step size μ and minimum tolerance μ_{min}.
4: Set $k = 1$, calculate $R_s(k) = R_s(\mathbf{S}_l, \mathbf{T}_1)$.
5: If $\mu \geq \mu_{min}$ goto step 6, otherwise stop algorithm and return $\mathbf{T}_k$.
6: Calculate $\mathbf{T}_k' = \mathbf{T}_k + \mu \nabla_{\mathbf{T}_k} R_s(\mathbf{S}_l)$, and normalize $\mathbf{T}_k'$.
7: Calculate $R_s' = R_s(\mathbf{T}_k')$.
8: If $R_s' \geq R_s(\mathbf{T}_k)$, update $R_s(k+1) = R_s'$ and $\mathbf{T}_{k+1} = \mathbf{T}_k'$, then goto step 9; Otherwise, $\mu = \mu/2$ and goto step 5.
9: $k = k + 1$ goto step 6.
10: Storing $\mathbf{R}(l) = [R_{s,k+1}, \mathbf{T}_{k+1}]$.
11: **End**
12: Output $[\mathbf{S}^*, \mathbf{T}^*] = \arg\max_l \mathbf{R}(l)$.

---

to find the best $\mathbf{S}_l$ and $\mathbf{T}$. The optimization problem can be rewritten as

$$\max_{\mathbf{T}} \quad R_s(\mathbf{S}_l) \tag{13a}$$

$$s.t. \quad \mathsf{tr}(\mathbf{T}\mathbf{T}^H) = 1. \tag{13b}$$

Accordingly, the Lagrangian function can be directly written as:

$$L(\mathbf{T}, \lambda) = -R_s(\mathbf{S}_l) + \lambda \left[\mathsf{tr}(\mathbf{T}\mathbf{T}^H) - 1 \right], \tag{14}$$

where λ is the Lagrange multiplier. Then, the GD-based method can be applied to solve this optimization problem and its detailed process is illustrated in Algorithm 1. By taking the gradient of $R_s(\mathbf{S}_l)$ with respect to $\mathbf{T}$ and setting it to zero, we have

$$-\nabla_{\mathbf{T}} R_s(\mathbf{S}_l) + \lambda \mathbf{T} = 0, \tag{15}$$

where $\nabla_{\mathbf{T}} R_s(\mathbf{S}_l)$ is shown in (16), and

$$\kappa_E = \sum_{k=1}^{N_s M} \exp\left(\frac{-f_{e,m,k}}{P_2(\tilde{\mathbf{g}}_l + \boldsymbol{\Delta g}_l)\mathbf{T}\mathbf{T}^H(\tilde{\mathbf{g}}_l + \boldsymbol{\Delta g}_l)^H + \sigma_E^2} \right), \tag{17}$$

$$\kappa_B = \sum_{j=1}^{N_s M} \exp\left(\frac{-f_{b,i,j}}{P_2 \mathbf{h}_l \mathbf{T}\mathbf{T}^H \mathbf{h}_l^H + \sigma_B^2} \right). \tag{18}$$

Next, by substituting (17) and (18) into (16), the ES plus GD method shown in Algorithm 1 can be exploited to obtain $\mathbf{S}_l$ and $\mathbf{T}$ numerically. Hence, Algorithm 1 can be guaranteed to converge to a local optimum. Finally, a near-optimal $\mathbf{S}^*$ and $\mathbf{T}^*$ may be obtained by repeating Algorithm 1 using a number of distinct initializations.

B. Proposed joint SA-Max-ASR optimization of AAG and ANCM

In the preceding subsection, the variables $\mathbf{S}$ and $\mathbf{T}$ were obtained by numerical search algorithm, which involves a large

$$\nabla_{\mathbf{T}} R_s(\mathbf{S}_l) = \frac{1}{\ln 2 \cdot N_s M} \times$$

$$\left\{ \sum_{m=1}^{N_s M} \mathbb{E}_{\boldsymbol{\Delta}\mathbf{g}_l, n_E} \left(\frac{1}{\kappa_E} \sum_{k=1}^{N_s M} \left(\frac{P_2 f_{e,m,k} (\tilde{\mathbf{g}}_l + \boldsymbol{\Delta}\mathbf{g}_l)^H (\tilde{\mathbf{g}}_l + \boldsymbol{\Delta}\mathbf{g}_l) \mathbf{T}}{\left(P_2 (\tilde{\mathbf{g}}_l + \boldsymbol{\Delta}\mathbf{g}_l) \mathbf{T}\mathbf{T}^H (\tilde{\mathbf{g}}_l + \boldsymbol{\Delta}\mathbf{g}_l)^H + \sigma_E^2 \right)^2} \right) \exp \left(\frac{-f_{e,m,k}}{P_2 (\tilde{\mathbf{g}}_l + \boldsymbol{\Delta}\mathbf{g}_l) \mathbf{T}\mathbf{T}^H (\tilde{\mathbf{g}}_l + \boldsymbol{\Delta}\mathbf{g}_l)^H + \sigma_E^2} \right) \right)$$

$$- \sum_{i=1}^{N_s M} \mathbb{E}_{n_B} \left(\frac{1}{\kappa_B} \sum_{j=1}^{N_s M} \left(\frac{P_2 f_{b,i,j} \mathbf{h}_l^H \mathbf{h}_l \mathbf{T}}{\left(P_2 \mathbf{h}_l \mathbf{T}\mathbf{T}^H \mathbf{h}_l^H + \sigma_B^2 \right)^2} \right) \exp \left(\frac{-f_{b,i,j}}{P_2 \mathbf{h}_l \mathbf{T}\mathbf{T}^H \mathbf{h}_l^H + \sigma_B^2} \right) \right) \right\} \quad (16)$$

number of SR evaluations and combinations, when N_t is large. Hence, it only can be used as a performance benchmark for small-scale scenarios. In what follows, we will present a lower complexity algorithm for optimizing $\mathbf{S}$ and $\mathbf{Q} = \mathbb{E}(\mathbf{T}\mathbf{T}^H)$, which has the capability to approach the SR performance of Algorithm 1. Once the corresponding optimal solution is obtained, the ergodic SR is evaluated by (10). We circumvent the difficulty by avoiding the calculation of multiple integrals in the SR expression, where the corresponding approximate MI (AMI) is given by

$$I_B^a(\mathbf{S}_l, \mathbf{Q}) = \zeta - \log_2 \sum_{i=1}^{N_s M} \sum_{j=1}^{N_s M} \exp \left(\frac{-P_1 \mathbf{d}_{ij}^H \mathbf{h}_l^H \mathbf{h}_l \mathbf{d}_{ij}}{4 \left(P_2 \mathbf{h}_l \mathbf{Q} \mathbf{h}_l^H + \sigma_B^2 \right)} \right), \quad (19)$$

where $\zeta = 2\log_2 N_s M$ and $\mathbf{d}_{ij} = \boldsymbol{x}_i - \boldsymbol{x}_j$. Similarly, the AMI for Eve is given by

$$I_E^a(\mathbf{S}_l, \mathbf{Q}) = \zeta - \mathbb{E}_{\boldsymbol{\Delta}\mathbf{g}_l} \log_2 \sum_{m=1}^{N_s M} \sum_{k=1}^{N_s M}$$
$$\exp \left(\frac{-P_1 \mathbf{d}_{mk}^H (\tilde{\mathbf{g}}_l + \boldsymbol{\Delta}\mathbf{g}_l)^H (\tilde{\mathbf{g}}_l + \boldsymbol{\Delta}\mathbf{g}_l) \mathbf{d}_{mk}}{4 \left(P_2 (\tilde{\mathbf{g}}_l + \boldsymbol{\Delta}\mathbf{g}_l) \mathbf{Q} (\tilde{\mathbf{g}}_l + \boldsymbol{\Delta}\mathbf{g}_l)^H + \sigma_E^2 \right)} \right), \quad (20)$$

where $\mathbf{d}_{mk} = \boldsymbol{x}_m - \boldsymbol{x}_k$. For a similar derivation process please refer to the Appendix A of [39]. Then, upon replacing R_s by (19) and (20), the ASR with a given AAG becomes

$$R_s^a(\mathbf{S}_l, \mathbf{Q}) = E(\mathbf{S}_l, \mathbf{Q}) - B(\mathbf{S}_l, \mathbf{Q}), \quad (21)$$

where

$$E(\mathbf{S}_l, \mathbf{Q}) = \mathbb{E}_{\boldsymbol{\Delta}\mathbf{g}_l} \log_2 \sum_{m=1}^{N_s M} \sum_{k=1}^{N_s M}$$
$$\exp \left(\frac{-P_1 \mathbf{d}_{mk}^H (\tilde{\mathbf{g}}_l + \boldsymbol{\Delta}\mathbf{g}_l)^H (\tilde{\mathbf{g}}_l + \boldsymbol{\Delta}\mathbf{g}_l) \mathbf{d}_{mk}}{4 \left(P_2 (\tilde{\mathbf{g}}_l + \boldsymbol{\Delta}\mathbf{g}_l) \mathbf{Q} (\tilde{\mathbf{g}}_l + \boldsymbol{\Delta}\mathbf{g}_l)^H + \sigma_E^2 \right)} \right), \quad (22)$$

$$B(\mathbf{S}_l, \mathbf{Q}) = \log_2 \sum_{i=1}^{N_s M} \sum_{j=1}^{N_s M} \exp \left(\frac{-P_1 \mathbf{d}_{ij}^H \mathbf{h}_l^H \mathbf{h}_l \mathbf{d}_{ij}}{4 \left(P_2 \mathbf{h}_l \mathbf{Q} \mathbf{h}_l^H + \sigma_B^2 \right)} \right). \quad (23)$$

It is noteworthy that replacing (6) by the AMI (19) is an efficient way of reducing the computational complexity [39].

Via applying Jensen's inequality, (22) can be lower bounded as

$$\mathbb{E}_{\boldsymbol{\Delta}\mathbf{g}} \log_2 \sum_{m=1}^{N_s M} \sum_{k=1}^{N_s M} \exp$$
$$\left(\frac{-P_1 \mathbf{d}_{mk}^H (\tilde{\mathbf{g}}_l + \boldsymbol{\Delta}\mathbf{g}_l)^H (\tilde{\mathbf{g}}_l + \boldsymbol{\Delta}\mathbf{g}_l) \mathbf{d}_{mk}}{4 \left(P_2 (\tilde{\mathbf{g}}_l + \boldsymbol{\Delta}\mathbf{g}_l) \mathbf{Q} (\tilde{\mathbf{g}}_l + \boldsymbol{\Delta}\mathbf{g}_l)^H + \sigma_E^2 \right)} \right) \geq$$
$$\log_2 \sum_{m=1}^{N_s M} \sum_{k=1}^{N_s M} \exp$$
$$\mathbb{E}_{\boldsymbol{\Delta}\mathbf{g}} \left(\frac{-P_1 \mathbf{d}_{mk}^H (\tilde{\mathbf{g}}_l + \boldsymbol{\Delta}\mathbf{g}_l)^H (\tilde{\mathbf{g}}_l + \boldsymbol{\Delta}\mathbf{g}_l) \mathbf{d}_{mk}}{4 \left(P_2 (\tilde{\mathbf{g}}_l + \boldsymbol{\Delta}\mathbf{g}_l) \mathbf{Q} (\tilde{\mathbf{g}}_l + \boldsymbol{\Delta}\mathbf{g}_l)^H + \sigma_E^2 \right)} \right). \quad (24)$$

According to [40] and [41], the following effective approximation can be adopted:

$$\mathbb{E}_{\boldsymbol{\Delta}\mathbf{g}} \frac{-P_1 \mathbf{d}_{mk}^H (\tilde{\mathbf{g}}_l + \boldsymbol{\Delta}\mathbf{g}_l)^H (\tilde{\mathbf{g}}_l + \boldsymbol{\Delta}\mathbf{g}_l) \mathbf{d}_{mk}}{4 \left(P_2 (\tilde{\mathbf{g}}_l + \boldsymbol{\Delta}\mathbf{g}_l) \mathbf{Q} (\tilde{\mathbf{g}}_l + \boldsymbol{\Delta}\mathbf{g}_l)^H + \sigma_E^2 \right)} \approx$$
$$\frac{-P_1 \mathbb{E}_{\boldsymbol{\Delta}\mathbf{g}} \left(\mathbf{d}_{mk}^H (\tilde{\mathbf{g}}_l + \boldsymbol{\Delta}\mathbf{g}_l)^H (\tilde{\mathbf{g}}_l + \boldsymbol{\Delta}\mathbf{g}_l) \mathbf{d}_{mk} \right)}{4 \left(P_2 \mathbb{E}_{\boldsymbol{\Delta}\mathbf{g}} \left((\tilde{\mathbf{g}}_l + \boldsymbol{\Delta}\mathbf{g}_l) \mathbf{Q} (\tilde{\mathbf{g}}_l + \boldsymbol{\Delta}\mathbf{g}_l)^H \right) + \sigma_E^2 \right)}. \quad (25)$$

Then, $E(\mathbf{S}_l, \mathbf{Q})$ can be rewritten as

$$\tilde{E}(\mathbf{S}_l, \mathbf{Q}) =$$
$$\log_2 \sum_{m=1}^{N_s M} \sum_{k=1}^{N_s M} \exp \left\{ \frac{-P_1 \left(\mathbf{d}_{mk}^H (\tilde{\mathbf{g}}_l^H \tilde{\mathbf{g}}_l + \sigma_e^2 \mathbf{I}) \mathbf{d}_{mk} \right)}{4 \left(P_2 \tilde{\mathbf{g}}_l \mathbf{Q} \tilde{\mathbf{g}}_l^H + \sigma_e^2 P_2 + \sigma_E^2 \right)} \right\}. \quad (26)$$

Next, let us define a closed-form expression for the ASR as follows

$$R_A(\mathbf{S}_l, \mathbf{Q}) = \tilde{E}(\mathbf{S}_l, \mathbf{Q}) - B(\mathbf{S}_l, \mathbf{Q}). \quad (27)$$

Therefore, the optimization problem in (11) can be converted into

$$\max_{\mathbf{S}_l, \mathbf{Q}} \quad R_A(\mathbf{S}_l, \mathbf{Q}) \quad (28a)$$
$$s.t. \quad \text{tr}(\mathbf{Q}) = 1, \mathbf{Q} \succeq 0, \quad (28b)$$
$$\sum_{i=1}^{N_t} s_i = N_s, \quad (28c)$$
$$s_i \in \{0, 1\}, i = 1, \cdots, N_t. \quad (28d)$$

However, (28a) is still a non-concave function of the continuous optimization variable $\mathbf{Q} \succeq 0$ with a fixed $\mathbf{S}_l$. Next we

convert $R_A(\mathbf{S}_l, \mathbf{Q})$ into a concave function, so that a unique solution $\mathbf{Q}$ can be obtained.

To elaborate, we first derive a convex function as an upper bound of $B(\mathbf{S}_l, \mathbf{Q})$. Due to the fact that the concave function of

$$B_{ij}(\mathbf{S}_l, \mathbf{Q}) = \frac{-A_{ij}}{\mathbf{h}_l \mathbf{Q} \mathbf{h}_l^H + b} \tag{29}$$

can be upper bounded by its first-order approximation $B_{ij}^{(1)}(\mathbf{S}_l, \mathbf{Q})$, i.e., its tangent at point $\mathbf{Q}_0$, thus we have

$$B_{ij}(\mathbf{S}_l, \mathbf{Q}) \leq B_{ij}^{(1)}(\mathbf{S}_l, \mathbf{Q})$$
$$\triangleq \frac{-A_{ij}}{\mathbf{h}_l \mathbf{Q}_0 \mathbf{h}_l^H + b} + \text{tr}\left\{ \frac{A_{ij}\mathbf{h}_l^H \mathbf{h}_l}{\left(\mathbf{h}_l \mathbf{Q}_0 \mathbf{h}_l^H + b\right)^2} (\mathbf{Q} - \mathbf{Q}_0) \right\}, \tag{30}$$

where $A_{ij} = P_1 \mathbf{d}_{ij}^H \mathbf{h}_l^H \mathbf{h}_l \mathbf{d}_{ij}/4P_2$ and $b = \sigma_B^2/P_2$. The inequality (30) holds due to $A_{ij} \geq 0$ and $b > 0$. Substituting the above inequality into (23), we have the following inequality

$$B(\mathbf{S}_l, \mathbf{Q}) \leq \tilde{B}(\mathbf{S}_l, \mathbf{Q}) = \log_2 \sum_{i=1}^{N_s M} \sum_{j=1}^{N_s M} \exp\left(B_{ij}^{(1)}(\mathbf{S}_l, \mathbf{Q}) \right), \tag{31}$$

where $B_{ij}^{(1)}(\mathbf{S}_l, \mathbf{Q})$ is a linear function of $\mathbf{Q}$. It is plausible that $\tilde{B}(\mathbf{S}_l, \mathbf{Q})$ is convex and it is also an upper bound of $B(\mathbf{S}_l, \mathbf{Q})$. At the same time, considering that $B(\mathbf{S}_l, \mathbf{Q})$ is always larger than or equal to 0, $B(\mathbf{S}_l, \mathbf{Q})$ can be further upper bounded as

$$B(\mathbf{S}_l, \mathbf{Q}) \leq B'(\mathbf{S}_l, \mathbf{Q}) = \max\left\{ \tilde{B}(\mathbf{S}_l, \mathbf{Q}),\ 0 \right\}. \tag{32}$$

In the following, a concave function related to the lower bound of $E(\mathbf{Q})$ is also derived. Firstly, we reformulate the exponential function at a feasible point $\mathbf{Q}_0$, given by

$$\exp\left(\frac{-C_{mk}}{\tilde{\mathbf{g}}_l \mathbf{Q} \tilde{\mathbf{g}}_l^H + c} \right) \geq E_{mk}^{(1)}(\mathbf{S}_l, \mathbf{Q})$$
$$\triangleq \exp\left(\frac{-C_{mk}}{\tilde{\mathbf{g}}_l \mathbf{Q}_0 \tilde{\mathbf{g}}_l^H + c} \right) \left(1 + \frac{C_{mk}}{\tilde{\mathbf{g}}_l \mathbf{Q}_0 \tilde{\mathbf{g}}_l^H + c} - \frac{C_{mk}}{\tilde{\mathbf{g}}_l \mathbf{Q} \tilde{\mathbf{g}}_l^H + c} \right), \tag{33}$$

where

$$C_{mk} = \frac{P_1 \mathbf{d}_{mk}^H \left(\tilde{\mathbf{g}}_l^H \tilde{\mathbf{g}}_l + \sigma_e^2 \mathbf{I} \right) \mathbf{d}_{mk}}{4P_2}, \tag{34}$$

and $c = \sigma_e^2 + \sigma_E^2/P_2$. Explicitly, $E_{mk}^{(1)}(\mathbf{S}_l, \mathbf{Q})$ is a concave function of $\mathbf{Q}$. Note that $E_{mk}^{(1)}(\mathbf{S}_l, \mathbf{Q})$ can be negative, hence we extend the domain of $\log_2(x)$ to the field of real numbers, given by

$$\dot{\log}_2(x) = \begin{cases} \log_2(x), & x > 0 \\ -\infty, & x \leq 0 \end{cases}. \tag{35}$$

Using the inequality (33), we rewrite $\tilde{E}(\mathbf{S}_l, \mathbf{Q})$ with the aid of a lower bound as follows

$$\tilde{E}(\mathbf{S}_l, \mathbf{Q}) \geq \bar{E}(\mathbf{S}_l, \mathbf{Q}) = \dot{\log}_2 \sum_{m=1}^{N_s M} \sum_{k=1}^{N_s M} E_{mk}^{(1)}(\mathbf{Q}). \tag{36}$$

Meanwhile, upon considering that $\bar{E}(\mathbf{Q})$ is less than $2\log_2 N_s M$ and taking into account (36), we have

$$\tilde{E}(\mathbf{S}_l, \mathbf{Q}) \geq E'(\mathbf{S}_l, \mathbf{Q}) = \min\left\{ \bar{E}(\mathbf{S}_l, \mathbf{Q}),\ 2\log_2 N_s M \right\}, \tag{37}$$

where $E'(\mathbf{S}_l, \mathbf{Q})$ is a point-wise maximum of a concave function and a constant, and thus it is also a concave function of $\mathbf{Q}$ for a given $\mathbf{S}_l$. By replacing $E'(\mathbf{S}_l, \mathbf{Q})$ and $B'(\mathbf{S}_l, \mathbf{Q})$ with $\tilde{E}(\mathbf{S}_l, \mathbf{Q})$ and $B(\mathbf{S}_l, \mathbf{Q})$ respectively, a concave maximization problem can be formulated for $\mathbf{Q}$ as follows:

$$\max_{\mathbf{S}_l, \mathbf{Q}} \quad R_s^c(\mathbf{S}_l, \mathbf{Q}) = E'(\mathbf{S}_l, \mathbf{Q}) - B'(\mathbf{S}_l, \mathbf{Q}) \tag{38a}$$
$$s.t. \quad \text{tr}(\mathbf{Q}) = 1, \mathbf{Q} \succeq 0. \tag{38b}$$

Then, the optimal $\mathbf{Q}$ for (38) can be obtained iteratively with a random given feasible point.

Remark 1: Based upon (19)-(37), the OF in (11) has been transformed into a concave function with respect to $\mathbf{Q}$. Generally, obtaining the optimal solution of (11) is infeasible, since the SR cannot be performed in a closed form, when finite-alphabet inputs are considered. To this end, we invoked two approximations: firstly, we used the ASR for eliminating the expectations over n_B and n_E, and secondly, we adopted a loose approximation in (25) to circumvent the calculation of the expectation over $\Delta \mathbf{g}$. Importantly, the efficiency of using our ASR expression to substitute the SR term has been demonstrated in [26] [39]. Moreover, the efficiency of the loose approximation adopted has also been verified in [40] [41]. Due to the approximation error, the bound derived in (38a) is not tight, hence the solution obtained is suboptimal for problem (11).

In the above section, we converted the continuous optimization problem into a concave one, thus a unique solution $\mathbf{Q}$ can be obtained for any given $\mathbf{S}_l$ $(l = 1, \cdots, L)$. Next we focus our attention on solving the 0-1 programming problem for AAG selection. As it is widely known, in contrast to GD, SA explores the entire search space in a random guided fashion by sometimes degrading the OF value in an attempt to avoid getting trapped in local minima. In other words, the SA method iterates by perturbing the current configuration and measuring the change in cost. When the change in cost is positive, the new AAG is automatically accepted, otherwise the probability of accepting the OF reduction is calculated by evaluating the so-called Boltzmann factor C_k. If this probability is higher than a random number in the interval $[0, 1)$, the new AAG is accepted, otherwise, it is rejected. This acceptance rule, namely the 'Metropolitan', can be expressed as [33]

$$\min\left\{ 1, \exp\left(\frac{R_s^c(\mathbf{S}_l, \mathbf{Q}) - R_s^c(\mathbf{S}_l^o, \mathbf{Q}^o)}{C_0} \right) \right\} > \eta, \tag{39}$$

where $\mathbf{S}_l^o$ is a random neighbour AAG of the current $\mathbf{S}_l$ and $\mathbf{Q}^o$ is the corresponding ANCM computed by (38), while C_0 is the initial control parameter associated with $C_0 > 0$ and $\eta \in [0, 1)$.

The proposed SA-based AAG selection scheme includes the following procedures: 1) generate neighbour AAG to impose a perturbation, 2) stochastic motion to avoid getting trapped in local maxima, 3) reduce the mutation parameter C_k to

Algorithm 2 Joint AAG selection and AN design for maximizing ASR (Joint SA-Max-ASR)

1: Given an initial AAG vector $\mathbf{s} \in \mathcal{C}$;
   Initialization of simulation mutation parameters $C_0 > C_f > 0$;
   Set an iterative counter $k = 0$ and give a sampling number S.

2: Mutation process:
   1). Generating a random solution $\mathbf{s}' \in N(\mathbf{s})$, and evaluating $\triangle R = R_s^c(\mathbf{s}') - R_s^c(\mathbf{s})$.
   2). If 'Metropolis criterion' is satisfied, i.e., min $\{1, \exp(\triangle R/C_k)\} > \eta \in [0,1)$, then $\mathbf{s} = \mathbf{s}'$.
   3). If 'Metropolis equilibrium' under C_k is realized, then go to 3; Otherwise, go to Step 2.1.

3: Integer sampling process:
   1). Evaluate $R_s^c(\mathbf{s})$.
   2). Give a temporary set $V = \mathbf{s}$, and set $R_{pre}^c = R_s^c(\mathbf{s})$.
   3). Select a solution $\mathbf{s}' \in (N(\mathbf{s}) - V)$ randomly, and $V = V \cup \{\mathbf{s}'\}$; Evaluating $\triangle R = R_s^c(\mathbf{s}') - R_s^c(\mathbf{s})$.
   4). If $\triangle R > 0$, then $\mathbf{s} = \mathbf{s}'$.
   5). If $|V| = S$, then go to Step 3.6; else go to Step 3.3.
   6). If $R_s^c(\mathbf{s}) > R_{pre}^c$, then go to Step 3.2; else go to Step 4.

4: Annealing process: reducing simulation mutation parameter $C_{k+1} = C_k - \triangle C_k$, $\triangle C_k > 0$.

5: If 'stop criterion' is not satisfied, i.e., $C_k > C_f$, then setting $k = k+1$, go to Step 2; otherwise, output: $\mathbf{s}_{opt} = \mathbf{s}$.

6: **End**

Procedure 1 Generating neighborhood AAG vector

1: Given $\mathbf{s} = (s_1, \cdots, s_{N_t})$ with $\sum_{i=1}^{N_t} s_i = N_s$.

2: Find the indexes with 1 elements and 0 elements of $\mathbf{s}$, and put them in $\mathbf{I}_1 = (e_1, \cdots, e_{N_s})$ and $\mathbf{I}_0 = (j_1, \cdots, j_{N_t-N_s})$, respectively.

3: Setting $r_1 = \text{Randint}[1, N_s]$, $r_0 = \text{Randint}[1, N_t - N_s]$.

4: Setting $s_{I_1(r_1)} = 0$ and $s_{I_0(r_0)} = 1$.

5: $\mathbf{s}' = (s_1, \cdots, s_{I_1(r_1)}, \cdots, s_{I_0(r_0)}, \cdots, s_{N_t})$.

6: **return**

As shown in Algorithm 2, the implementation of the proposed SA-based AAG selection scheme requires the design of the following two distinct processes. The first is the process of generating a new solution by Procedure 1, which exploits a specific generation mechanism and compares the two solutions in term of their cost. Then a decision is made as to whether or not the new AAG could be accepted. The other process is the evolution control strategy, which requires an initial value of the mutation parameter and a decrement function of the mutation parameter. Remarkably, the annealing process was found to constantly decrease the mutation parameter C_k until it reaches the 'stop criterion', namely the optimum level C_f.

C. Separate optimization of AAG by SA and of the ANCM by Max-ASR (Separate SA-Max-ASR)

Similar to [25], the following low complexity method is proposed for separately optimizing $\mathbf{S}_l$ and $\mathbf{Q}$. Considering that MI is originated from the active TA indices and APM symbols of the SM system, thus the AAG may be optimized before designing $\mathbf{Q}$. Once the AAG has been determined, the corresponding $\mathbf{Q}$ is optimized by (38). According to [26], the ASR in terms of the AAG is

$$R_s^s = I_e^s - I_b^s, \tag{41}$$

where I_e^s is the approximate rate in terms of the AAG for Eve on the face of a realistic channel estimation error, given by

$$I_e^s = \log_2 \sum_{m=1}^{N_s M} \sum_{k=1}^{N_s M} \exp \left(\frac{-P_1 \mathbf{d}_{mk}^H (\tilde{\mathbf{g}}_l^H \tilde{\mathbf{g}}_l + \sigma_e^2 \mathbf{I}) \mathbf{d}_{mk}}{4\sigma_E^2} \right). \tag{42}$$

Similarly, the approximate rate of Bob can be expressed as

$$I_b^s = \log_2 \sum_{i=1}^{N_s M} \sum_{j=1}^{N_s M} \exp \left(\frac{-P_1 \mathbf{d}_{ij}^H \mathbf{h}_l^H \mathbf{h}_l \mathbf{d}_{ij}}{4\sigma_B^2} \right). \tag{43}$$

In order to reduce the computational complexity, the repeated calculations can be avoided by defining a pair of upper triangular matrices, $\mathbf{U}_B \in \mathbb{R}^{N_t \times N_t}$ and $\mathbf{U}_E \in \mathbb{R}^{N_t \times N_t}$, whose (u,v)-th entry is respectively given by

$$b_{u,v} = \begin{cases} \sum_{m,n \in \mathbb{M}} \exp \left(\frac{-P_1 |h_u s_m - h_v s_n|^2}{4\sigma_B^2} \right), u < v \\ \sum_{m \neq n \in \mathbb{M}} \exp \left(\frac{-P_1 |h_u|^2 |s_m - s_n|^2}{4\sigma_B^2} \right), u = v, m > n, \end{cases} \tag{44}$$

increase the search precision. To elaborate a little further, procedure 2) invokes the 'Metropolis equilibrium' criterion to terminate the stochastic motion, when a predetermined number of perturbations is completed. Let us denote the set of optimal points as

$$S_{\text{opt}} = \{\mathbf{S}_l = \text{diag}(\mathbf{s}_{\text{opt}}) : R_s^c(\mathbf{s}_{\text{opt}}) \geq R_s^c(\mathbf{s}), \mathbf{s} \in \mathcal{C}\}, \tag{40}$$

and $N(\mathbf{s}) \in \mathcal{C}$ is the neighborhood of the solution $\mathbf{s}$. Together with (38), the proposed joint SA-Max-ASR scheme using SA is listed in Algorithm 2, where $R_s^c(\mathbf{s})$ denotes $R_s^c(\mathbf{S}_l, \mathbf{Q})$ for simplicity.

More explicitly, Algorithm 2 combines the mutation process and self-reproduction strategy into an evolutionary process for approaching the optimum AAG. This search process is performed repeatedly upon the self-reproduction processes and annealing strategy, where the self-reproduction processes search within the immediate neighborhood for an improved solution. Algorithm 2 always starts with a random AAG in reach for a local minimum and then escapes from the suboptimal local 'traps'. To generate a random neighborhood AAG vector $\mathbf{s}' \in N(\mathbf{s})$ in Step 2.1 of Algorithm 2, the neighborhood sampling technique of Procedure 1 is proposed. The initial AAG vector $\mathbf{s}$ contains N_s '1' elements and $N_t - N_s$ '0' elements, where the function $\text{Randint}[n_1, n_2]$ means that a uniformly distributed integer is randomly generated from the interval $[n_1, n_2]$. Procedure 1 ensures that the number of elements '1' of $\mathbf{s}'$ equals to N_s.

$$e_{u,v} =$$

$$\begin{cases} \sum_{m,n \in \mathbb{M}} \exp\left(\frac{-P_1|(|g_u|^2+\sigma_e^2)s_m-(|g_v|^2+\sigma_e^2)s_n|^2}{4\sigma_B^2}\right), u < v \\ \sum_{m \neq n \in \mathbb{M}} \exp\left(\frac{-P_1(|g_u|^2+\sigma_e^2)|s_m-s_n|^2}{4\sigma_B^2}\right), u = v, m > n. \end{cases} \tag{45}$$

Upon $\mathbf{U}_B$ and $\mathbf{U}_E$, the AAG associated with the highest SR can be promptly found by utilizing the SA strategy, upon replacing the OF $R_s^c(\mathbf{s})$ by

$$R_s^s = \log_2 D_E - \log_2 D_B \tag{46}$$

in Algorithm 2, where D_E and D_B are the summations of sub-triangular entries corresponding to the selected AAG, respectively. Once the AAG is obtained, the corresponding $\mathbf{Q}$ is designed by (38). The separate optimization scheme dramatically reduces the complexity, because the procedure of designing $\mathbf{Q}$ is avoided as the AAG changes. Our simulation results will show that this decoupled design strategy strikes a compelling performance versus complexity tradeoff.

IV. ASYMPTOTIC SIMPLE EQUIVALENCE OF SR IN LARGE-SCALE SCENARIO AND ASSOCIATED OPTIMIZATION

The proposed joint and separate SA-Max-ASR optimization schemes of Subsections III-B and III-C possess low complexities compared to the direct of Max-SR optimization of Algorithm 1. However, it may still be a complex task as N_t tends to large values, because the computational complexity of the iterative algorithm grows exponentially upon increasing the number of TAs. With this motivation, we provide a new method that removes the two-layer sum over the legitimate transmit vectors of $\mathcal{S}$, hence dramatically reducing the computational complexity.

Theorem 1: As the number of TAs tends to a large scale, the optimization problem of maximizing the SR of (27) can be reduced to maximizing the ratio of the SINR at the desired receiver to that at eavesdropper (Max-R-SINR) as follows

$$\max_{\mathbf{Q}} \ R_L'(\mathbf{S}_l, \mathbf{Q}) \tag{47a}$$

$$s.t. \ \ \mathrm{tr}(\mathbf{Q}) = 1, \mathbf{Q} \succeq 0, \tag{47b}$$

$$\sum_{i=1}^{N_t} s_i = N_s, \tag{47c}$$

$$s_i \in \{0,1\}, i = 1, \cdots, N_t, \tag{47d}$$

where we have:

$$R_L'(\mathbf{S}_l, \mathbf{Q}) = \frac{\|\mathbf{h}_l\|^2 \left(P_2\tilde{\mathbf{g}}_l\mathbf{Q}\tilde{\mathbf{g}}_l^H + \sigma_E'^2\right)}{(\|\tilde{\mathbf{g}}_l\|^2 + N_s\sigma_e^2)\left(P_2\mathbf{h}_l\mathbf{Q}\mathbf{h}_l^H + \sigma_B^2\right)}. \tag{48}$$

Proof: See Appendix A. ∎

Remark 2: Upon comparing $R_L'(\mathbf{S}_l, \mathbf{Q})$ to (27), the key benefit of using $R_L'(\mathbf{S}_l, \mathbf{Q})$ to substitute (27) is that the double summation of legitimate transmit symbols is removed from the OF for large-scale SSM systems. This can be explained as follows: Having a large number of independently fading communication channels provides a high diversity gain, hence the transmit symbols have a rather negligible effect on the received energy of the desired receiver and of the eavesdropper.

As a result, the optimization problem can be translated into an energy maximization problem that is only related to the communication channels.

Noting that the OF in (47a) is a linear fractional function and always non-convex, in accordance with the technique in [42], the problem can be rewritten for a given $\mathcal{C}_l$ as

$$\max_{\mathbf{Q}} \ S_E(\mathbf{Q}) - \lambda S_B(\mathbf{Q}) \tag{49a}$$

$$s.t. \ \ \mathbf{Q} \succeq 0, \tag{49b}$$

$$\mathrm{tr}(\mathbf{Q}) = 1, \tag{49c}$$

where

$$S_E(\mathbf{Q}) = P_2\tilde{\mathbf{g}}_l\mathbf{Q}\tilde{\mathbf{g}}_l^H + \sigma_E'^2, \tag{50}$$

$$S_B(\mathbf{Q}) = P_2\mathbf{h}_l\mathbf{Q}\mathbf{h}_l^H + \sigma_B^2, \tag{51}$$

and λ is an auxiliary variable, which is iteratively updated by

$$\lambda[t+1] = \frac{S_E(\mathbf{Q}[t])}{S_B(\mathbf{Q}[t])}, \tag{52}$$

where t is the iteration index. It has been shown in [42] that the convergence is guaranteed by alternatively updating λ using (52) and solving it for $\mathbf{Q}$ with the aid of (49), because λ is nondecreasing after each iteration.

Upon replacing $R_s^c(\mathbf{S}_l, \mathbf{Q})$ by $R_L'(\mathbf{S}_l, \mathbf{Q})$, the SA-based AAG selection can solve the mixed integer optimization problem using Algorithm 2. Nevertheless, its complexity may still be excessive, because $\mathbf{Q}$ is determined by a series of iterations once the AAG changes, and the number of changes is always high. In view of this, it is necessary to design a method that can promptly find a potential AAG to improve the security. Upon assuming $\mathbf{Q}_l^*$ is the optimal ANCM for the l-th AAG, we have

$$\max_{\mathbf{S}_l} \frac{\|\mathbf{h}_l\|^2 \left(P_2\tilde{\mathbf{g}}_l\mathbf{Q}_l^*\tilde{\mathbf{g}}_l^H + \sigma_E'^2\right)}{(\|\tilde{\mathbf{g}}_l\|^2 + \sigma_e^2 N_s)\left(P_2\mathbf{h}_l\mathbf{Q}_l^*\mathbf{h}_l^H + \sigma_B^2\right)} =$$
$$\max_{\mathbf{S}_l} \frac{\|\mathbf{h}_l\|^2}{(\|\tilde{\mathbf{g}}_l\|^2 + \sigma_e^2 N_s)} \frac{\mathrm{tr}\left((\tilde{\mathbf{g}}_l^H\tilde{\mathbf{g}}_l + \psi_E\mathbf{I})\mathbf{Q}_l^*\right)}{\mathrm{tr}\left((\mathbf{h}_l^H\mathbf{h}_l + \psi_B\mathbf{I})\mathbf{Q}_l^*\right)}, \tag{53}$$

where $\psi_E = \sigma_E'^2/P_2$ and $\psi_B = \sigma_B^2/P_2$. According to [42] [43], (53) is non-convex and hence no closed-form solution exists. However, such a trace ratio problem can be transformed into a simpler ratio tracing problem by sacrificing some of the accuracy. Then, it becomes equivalent to the determinant ratio problem of [44],

$$\frac{\|\mathbf{h}_l\|^2}{(\|\tilde{\mathbf{g}}_l\|^2 + \sigma_e^2 N_s)}\left(\frac{\det\left((\tilde{\mathbf{g}}_l^H\tilde{\mathbf{g}}_l + \psi_E\mathbf{I})\mathbf{Q}_l^*\right)}{\det\left((\mathbf{h}_l^H\mathbf{h}_l + \psi_B\mathbf{I})\mathbf{Q}_l^*\right)}\right)$$
$$= \frac{\|\mathbf{h}_l\|^2}{(\|\tilde{\mathbf{g}}_l\|^2 + \sigma_e^2 N_s)}\left(\frac{\det\left((\tilde{\mathbf{g}}_l^H\tilde{\mathbf{g}}_l + \psi_E\mathbf{I})\right)}{\det\left((\mathbf{h}_l^H\mathbf{h}_l + \psi_B\mathbf{I})\right)}\right) \tag{54}$$
$$= \frac{\|\mathbf{h}_l\|^2\left(\psi_E^{N_s} + \psi_E^{N_s-1}\|\tilde{\mathbf{g}}_l\|^2\right)}{(\|\tilde{\mathbf{g}}_l\|^2 + \sigma_e^2 N_s)\left(\psi_B^{N_s} + \psi_B^{N_s-1}\|\mathbf{h}_l\|^2\right)}. \tag{55}$$

Equation (54) holds as a result of $\det(\mathbf{AB}) = \det(\mathbf{A}) \cdot \det(\mathbf{B})$ when the square matrices $\mathbf{A}$ and $\mathbf{B}$ have the same size. Thus the optimization problem can be further reduced to

$$\max_{\mathbf{S}_l} \frac{\|\mathbf{h}_l\|^2 \left(\psi_E^{N_s} + \psi_E^{N_s-1}\|\tilde{\mathbf{g}}_l\|^2\right)}{\left(\|\tilde{\mathbf{g}}_l\|^2 + \sigma_e^2 N_s\right)\left(\psi_B^{N_s} + \psi_B^{N_s-1}\|\mathbf{h}_l\|^2\right)} \quad (56a)$$

$$\sum_{i=1}^{N_t} s_i = N_s, \quad (56b)$$

$$s_i \in \{0,1\}, i = 1, \cdots, N_t. \quad (56c)$$

In this way, the AAG can be pre-determined before designing $\mathbf{Q}$, and the complexity will be dramatically reduced, hence the algorithm can be applied in real-time situations. Once the AAG is determined, the corresponding $\mathbf{Q}^*$ is obtained by (49).

V. CONVERGENCE AND COMPLEXITY ANALYSIS

In this section, we investigate the convergence of the proposed SA-based AAG selection scheme. It is noted that the continuous optimization problem of $\mathbf{Q}$ is converted into a concave form for any given AAG, thus $\mathbf{Q}$ is optimal during the evolution process of AAG.

A. Proof of Convergence

As shown in Algorithm 2, the proposed SA-based AAG selection algorithm can be viewed as a stochastic process, where the outcome of each iteration strictly depends on the outcome of the previous iteration. Hence, we model and analyze the SA-based AAG selection method using the theory of finite Markov chains.

In [45], the authors have demonstrated that the Markov chain associated with our SA-based 0-1 programming problem exhibits a strong Markov-like properties, and the condition of asymptotical convergence in the homogeneous case of

$$\forall \mathbf{s}, \mathbf{s}' \in \mathcal{C}, \exists\, p > 1, \exists\, \mathbf{s}_0, \mathbf{s}_1, \cdots, \mathbf{s}_p \in \mathcal{C}, \text{with}$$
$$\mathbf{s}_0 = \mathbf{s}, \mathbf{s}_p = \mathbf{s}', \text{and } G_{\mathbf{s}_k,\mathbf{s}_{k+1}}^c > 0, k = 0, \cdots, p-1, \quad (57)$$

where $G_{\mathbf{ss}'}^c$ stands for the generation probability corresponding to the integer sampling procedure of Algorithm 2.

Lemma 1. The generation probability of $G_{\mathbf{s},\mathbf{s}'}^c$ is equal to

$$G_{\mathbf{s},\mathbf{s}'}^c = \begin{cases} \frac{1}{N_s N_t - N_s^2} & \mathbf{s}' \in \mathcal{C} \\ 0 & \mathbf{s}' \notin \mathcal{C} \end{cases} \quad (58)$$

with $\mathbf{s}, \mathbf{s}' \in \mathcal{C}$.

Proof: A neighborhood AAG vector $\mathbf{s}' \in N(\mathbf{s})$ of a given AAG vector $\mathbf{s} \in \mathcal{C}$ is close to $\mathbf{s}$ with

$$N(\mathbf{s}) = \{\mathbf{s}' \in \mathcal{C} : \mathbf{s}' \text{ is constructed by randomly removing one}$$
$$\text{TA in } \mathbf{I}_1 \text{ and meanwhile randomly activating}$$
$$\text{one silent TA in } \mathbf{I}_0\}, \quad (59)$$

where $\mathbf{I}_1$ and $\mathbf{I}_0$ are the active TA index sets of the current AAG and of the silent TAs, respectively. The integer neighborhood sampling procedure randomly deactivates an active antenna in $\mathbf{I}_1$, and randomly activates a silent antenna in $\mathbf{I}_0$ at the same time. Therefore the size of neighborhoods is

$$|N(\mathbf{s})| = N_s(N_t - N_s), \quad (60)$$

which completes the proof of Lemma 1. ∎

Theorem 2: The proposed SA-based AAG selection method of maximizing $R_s^c(\mathbf{s})$ converges asymptotically to the globally optimal $\mathbf{s}_{\text{opt}}$, where the generation probabilities $G_{\mathbf{s},\mathbf{s}'}^c$ given in (58) asymptotically satisfy the condition in (57) from a global viewpoint.

Proof: See Appendix B. ∎

For any $\mathbf{S}_l$, the optimal solution $\mathbf{Q}$ of R_s^c can be obtained as a benefit of its concavity. Upon combining it with the proposed SA-based AAG selection method, we can conclude that our proposed joint Max-ASR scheme can approach the globally optimal $\mathbf{S}_l$ and $\mathbf{Q}$.

B. Complexity Analysis

In this subsection, the complexities of the different algorithms are calculated in terms of the number of floating-operations (FLOPs). For the direct solution in Algorithm 1, the computational and search complexity is excessive because a large number of sample points ($N_{\text{samp}} \geq 500$) for $\Delta\mathbf{g}_l$, n_B and n_E are required to evaluate the accurate SR. The total number of FLOPs of the ES plus GD method of Algorithm 1 can be expressed as

$$\mathcal{O}_{\text{ES plus GD}} = 3LN_{\text{samp}}D_1 N_s^2 M^2 \left(4N_s^3 + 7N_s^2 + N_s + 6\right), \quad (61)$$

where D_1 is the number of iterations for Algorithm 1. It is remarkable that L will become significantly large as N_t increases.

For our proposed joint SA-Max-ASR in Algorithm 2, the complexity is imposed by three parts: the annealing process, solving the concave maximization problem and the sampling procedure. The complexity of solving problem (38) each time is [40]

$$\mathcal{C}_{\mathbf{Q}} = 2M^2 N_s^2(3N_s^3 + 4N_s^2) + \mathcal{O}\left[N_s^{4.5}\ln(1/\epsilon)\right], \quad (62)$$

where the first term denotes the complexity of calculating (38a). Therefore, the total complexity of the joint SA-Max-ASR can be expressed as

$$\mathcal{O}_{\text{Joint SA-Max-ASR}} = KSD_2 \left[\mathcal{C}_{\mathbf{Q}} + 2M^2 N_s^2(2N_s^3 + 3N_s^2)\right], \quad (63)$$

where D_2 denotes the number of iterations required for solving (38), K is the number of mutations necessitated for reaching the termination threshold C_f and S is the size of the samples in Algorithm 2.

For the separate SA-Max-ASR scheme, the pair of matrices $\mathbf{D}_B$ and $\mathbf{D}_E$ have to be calculated firstly, which requires about $M^2\left(N_t^2 + N_t\right)$ FLOPs. Adding the above complexities and including that of the SA algorithm for finding the AAG, the total complexity of the separate SA-Max-ASR optimization is

$$\mathcal{O}_{\text{Separate SA-Max-ASR}} = D_2 \mathcal{C}_{\mathbf{Q}} + KSM^2\left(N_t^2 + N_t\right). \quad (64)$$

Once $\mathbf{D}_B$ and $\mathbf{D}_E$ have been calculated, the SA-based algorithm can be used for rapidly finding the optimal AAGs because only some summation operations are required for the sub-matrices. Moreover, the problem in (38) only has to be solved once in the separate SA-Max-ASR scheme, thus the

computational cost will be dramatically reduced. Based upon the above complexity analysis, it can be observed that the complexity of the proposed joint and separate SA-Max-ASR optimization schemes are much lower than that of Algorithm 1 due to having $N_{\text{samp}} \gg N_s$ and $L \gg KS$.

For the large-scale SSM system, the optimization OF is converted into the ratio of SINR_B to SINR_E. The complexity of the Max-R-SINR optimization can be approximated as [46]

$$\mathcal{O}_{\text{Max-R-SINR}} \approx nD_3\mathcal{O}(N_s^{3.5}) + 4KSN_s, \qquad (65)$$

where D_3 is the number of iterations required for solving (49). Compared to the joint SA-Max-ASR, the complexity of the separate SA-Max-ASR optimization reduced to the order of $\mathcal{O}(N_s^{4.5})$. In large-scale scenarios, our proposed Max-R-SINR optimization method only requires about $\mathcal{O}(N_s^{3.5})$ FLOPs to optimize the variable $\mathbf{Q}$, which dramatically reduces the computational complexity.

VI. Simulation Results and Discussions

In this section, numerical simulation results are presented for evaluating the SR performance of the proposed methods, where the LNSP method of [25] and a generalized method of [47], which aims for maximizing the ratio of the received AN's power of Eve to that of Bob, namely GMax-R-PAN method, are used as our performance benchmarks. Specifically, the system parameters are set as follows: $P_s = N_s$ and $C_f = 0.001$, the termination condition for all algorithms is set to $\epsilon = 0.001$ and the initial mutation parameter C_0 is computed according to [48]. The noise levels at the desired receiver and at the eavesdropping receiver are assumed to be identical, i.e., $\sigma_B^2 = \sigma_E^2$. Additionally, the ergodic SR is averaged over 500 channel realizations according to (10).
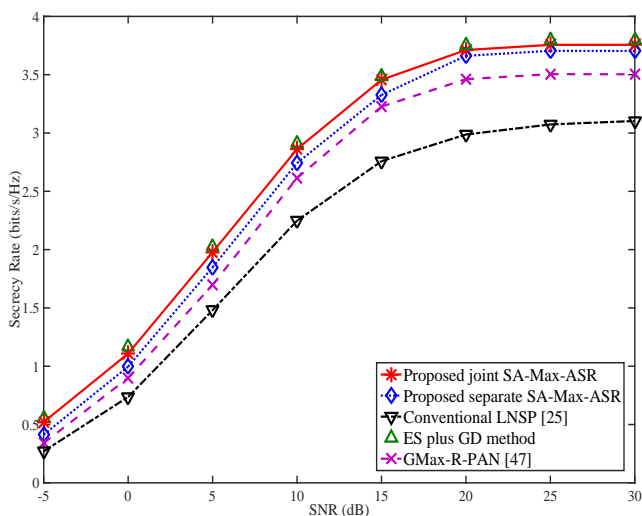

Fig. 2. Achievable SR versus SNR with $N_t = 7$, $N_s = 4$, and $\sigma_e^2 = 0.25$, and the modulation scheme is QPSK.

Fig. 2 plots the achievable SR of the proposed joint and separate SA-Max-ASR optimization methods versus the SNR with $\sigma_e^2 = 0.25$. It can be seen from this figure that the proposed methods achieve higher SR performance gains than both the LNSP method of [25] and GMax-R-PAN method of

[47]. To be specific, the commonly-used NSP method imposes a serious SR performance loss, because it only considers the interference signal without giving cognizance to the entire secure SM network as a whole. Moreover, the GMax-R-PAN method ignores the effect of the phases of the legitimate symbols on security, which thus results in a SR performance degradation. Additionally, the SR performance of the joint SA-Max-ASR optimization is close to that of ES plus GD scheme, in which 5 random initializations of $\mathbf{T}$ are repeated for Algorithm 1. Remarkably, the separate SA-Max-ASR performs slightly worse than the joint method, while it has a much lower complexity. Hence, it strikes a beneficial performance versus complexity tradeoff.
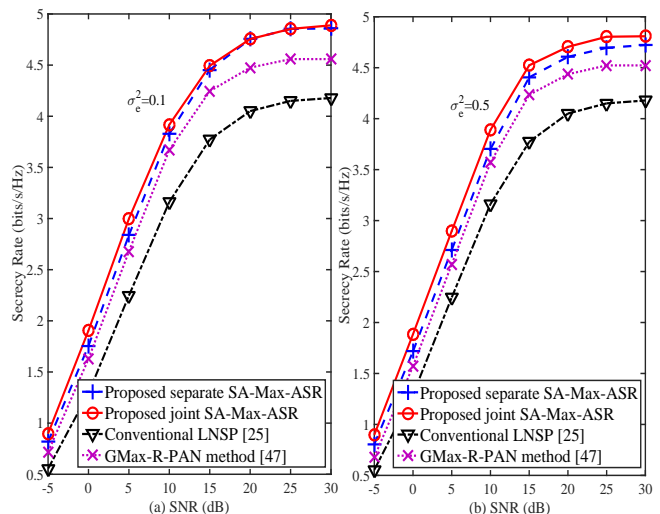

Fig. 3. Comparison of achievable SR versus SNR in the case of $N_t = 15$ and $N_s = 8$, where (a) $\sigma_e^2 = 0.1$ and (b) $\sigma_e^2 = 0.5$.

Fig. 3 shows the achievable SR of the proposed joint and separate SA-Max-ASR optimization methods for $N_t = 15$ and $N_s = 8$ with $\sigma_e^2 = 0.1$ and 0.5, respectively. Due to the prohibitive complexity ($L = 6435$), we do not consider the performance curve of Algorithm 1 in Fig. 3 for comparison. It becomes evident from Fig. 3 (a) that the SR of our proposed separate SA-Max-ASR optimization is close to that of the joint SA-Max-ASR scheme. Additionally, Fig. 3 (b) shows that the SR performance of our proposed methods is better than that of the LNSP method regardless of the estimation error of the illegitimate channel. Specifically, Fig. 4 shows the evolution process of the achievable SR versus the number of iterations for our proposed joint SA-Max-ASR optimization both at SNR=0dB and 10dB, respectively. Observing the two sub-figures, it follows that the SR performance of the joint SA-Max-ASR optimization eventually converges to a stable level exhibiting slight regional oscillations. In other words, the convergence of the proposed joint SA-Max-ASR scheme can be ensured as a benefit of having a probability of accepting a new generated AAG upon decreasing C_k.

Let us now consider the SR performance of the proposed Max-R-SINR optimization scheme. Fig. 5 (a) shows the achievable SR versus SNR for $N_t = 100$ and $N_s = 64$ with $\sigma_e^2 = 0.25$. In this case, we let the improvements along the
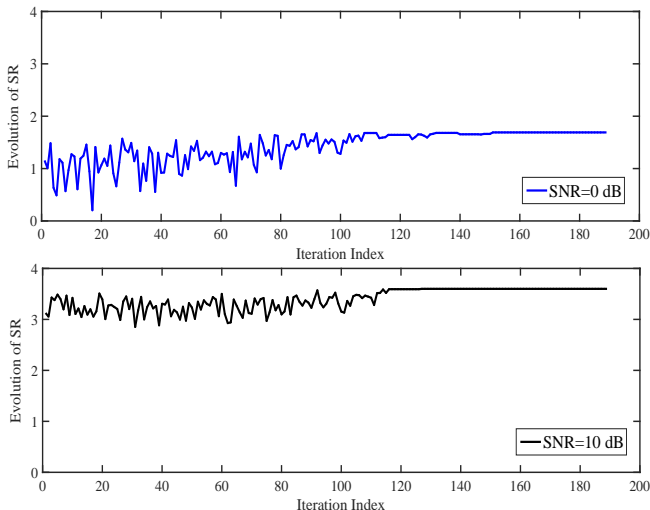
Fig. 4. Evolution of the achievable SR upon increasing the number of iterations at SNR=0dB and SNR=10dB, using the same configurations as in Fig. 3 (a).
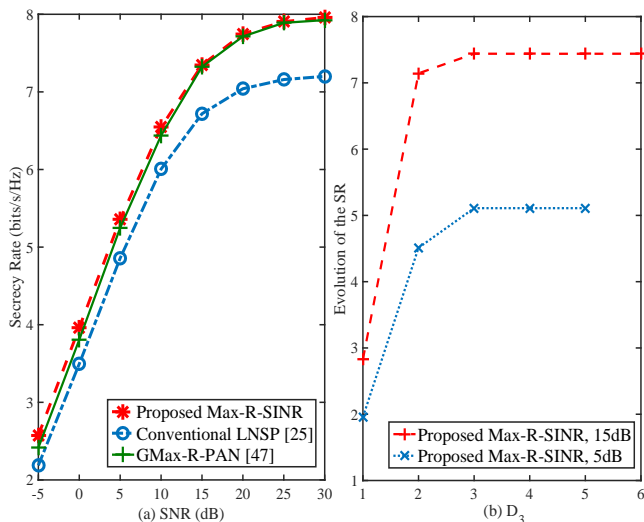

Fig. 5. (a) Achievable SR versus SNR and (b) evolution of the SR with respect to D_3, where $N_t = 100$, $N_s = 64$, $\sigma_e^2 = 0.25$ and the modulation scheme is QPSK.

gradients to reach the steepest descent (i.e., $|V| = |N(\mathbf{s})|$) in SA. Observe that the proposed Max-R-SINR scheme is capable of providing a significant SR performance benefit over the LNSP scheme. In addition, the GMax-R-PAN method achieves a similar SR performance to our proposed Max-R-SINR scheme in the high-SNR region. As a matter of fact, our proposed Max-R-SINR scheme can be approximated by the GMax-R-PAN method, as $\sigma_B^2 \to 0$ and $\sigma_E^2 \to 0$. However, the GMax-R-PAN is a heuristic scheme, which lacks theoretical foundation. In particular, the SR performance of our proposed Max-R-SINR scheme approaches $\log_2 N_s M = 8$ in the high-SNR region. Correspondingly, Fig. 5 (b) illustrates the evolution of the SR with respect to D_3 for our proposed Max-R-SINR method at the predefined SNRs of 5dB and 15dB. It is worth mentioning that a few iterations are sufficient for a near-optimal ANCM. The results in Fig. 5 verify the efficiency
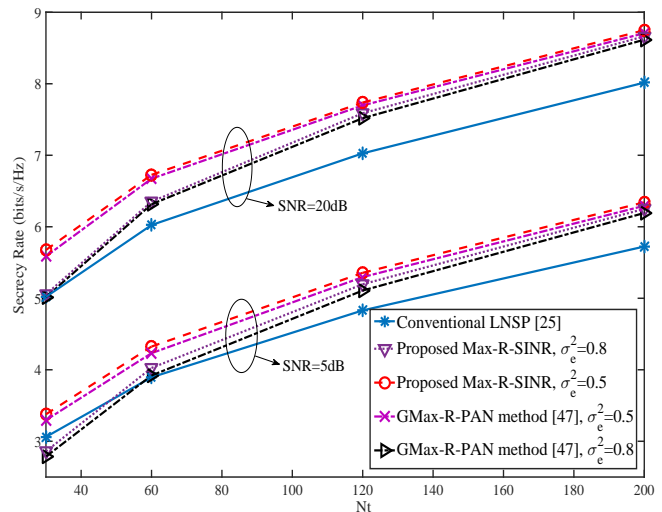
of the proposed Max-R-SINR scheme.

Fig. 6. Achievable SR versus the number of TAs for SNR=5dB and 20dB with $\sigma_e^2 = 0.5$ and $\sigma_e^2 = 0.8$.

Finally, in Fig. 6, we evaluate the accuracy of the proposed Max-R-SINR method for different number of TAs, where $N_t = 30, 60, 120, 200$ are considered and the corresponding number of active TAs is $16, 32, 64$ and 128, respectively. We observe from Fig. 6 that the SR performance of the proposed Max-R-SINR is low for the LNSP scheme, when $N_t = 30$ and $\sigma_e^2 = 0.8$ at SNR=5dB. This is because N_t has to reach a certain size to fully exploit the advantage of our proposed Max-R-SINR scheme in terms of its SR. Additionally, it can be seen that the SR performance of the Max-R-SINR becomes better than that of the LNSP for $N_t \geq 60$, and its benefit becomes more apparent upon increasing N_t. Specially, Fig. 6 underlines the merit of Max-R-SINR optimization in terms of its SR when $N_t = 200$, albeit σ_e^2 increases from 0.5 to 0.8. On the other hand, our proposed Max-R-SINR scheme has a consistently higher SR than the GMax-R-PAN method, regardless of the specific situation. In summary, we can conclude that the Max-R-SINR scheme efficiently reduces the computational cost in large-scale SSM scenarios.

VII. CONCLUSIONS

In this paper, the joint AAG selection and ANCM design were studied, when only rough partial CSI of Eve is obtained at transmitter. Due to the high-complexity of the ES plus GD method, both a joint and a separate SA-Max-ASR optimization scheme was proposed for optimizing the AAG and ANCM. Compared to the latter, the former achieves a better SR performance which is close to that of the ES plus GD method, while the latter has a lower complexity at the cost of a slight SR performance loss. To mitigate the complexity of the Max-ASR method when N_t tends to be large, we conceived the Max-R-SINR scheme. Our simulation results have quantified the SR performance gains of our proposed schemes compared to both the existing LNSP method and GMax-R-PAN method, and shown the tradeoff between the SR performance and the complexity. Our future work will focus on optimizing the AAG

$$\lim_{N_s \to \infty} R_A(\mathbf{S}_l, \mathbf{Q})$$

$$= \lim_{N_s \to \infty} \left\{ \log_2 \sum_{m=1}^{N_s M} \sum_{k=1}^{N_s M} \exp\left(-\frac{1}{4} \mathbf{d}_{mk}^H \mathbf{M}_E \mathbf{d}_{mk}\right) - \log_2 \sum_{i=1}^{N_s M} \sum_{j=1}^{N_s M} \exp\left(-\frac{1}{4} \mathbf{d}_{ij}^H \mathbf{M}_B \mathbf{d}_{ij}\right) \right\}$$

$$= \lim_{N_s \to \infty} \log_2 \left\{ \frac{N_s M + \sum_{m=1}^{N_s M} \sum_{k=1, k \neq m}^{N_s M} \exp\left(-\frac{1}{4} \mathbf{d}_{mk}^H \mathbf{M}_E \mathbf{d}_{mk}\right)}{N_s M + \sum_{i=1}^{N_s M} \sum_{j=1, j \neq i}^{N_s M} \exp\left(-\frac{1}{4} \mathbf{d}_{ij}^H \mathbf{M}_B \mathbf{d}_{ij}\right)} \right\} \tag{73}$$

$$= \lim_{N_s \to \infty} \log_2 \left\{ \frac{N_s M + N_s \cdot \sum_{\iota_E=1}^{M(N_s M - 1)} \left(\frac{1}{N_s} \sum_{i=1}^{N_s} \exp\left(-\frac{1}{4} \frac{P_1 \mathsf{tr}(\mathbf{u}\mathbf{u}^H)}{P_2 \tilde{\mathbf{g}}_l \mathbf{Q} \tilde{\mathbf{g}}_l^H + \sigma_E'^2} \lambda_{\iota_E} u_i^H u_i\right)\right)}{N_s M + N_s \cdot \sum_{\iota_B=1}^{(N_s M - 1)M} \left(\frac{1}{N_s} \sum_{j=1}^{N_s} \exp\left(-\frac{1}{4} \frac{P_1 \mathsf{tr}(\mathbf{h}_l \mathbf{h}_l^H)}{P_2 \mathbf{h}_l \mathbf{Q} \mathbf{h}_l^H + \sigma_B^2} \lambda_{\iota_B} u_j^H u_j\right)\right)} \right\} \tag{74}$$

$$= \log_2 \left\{ \frac{M + \sum_{\iota_E=1}^{(N_s M - 1)M} \left(\int_{u_i} \exp\left(-\left(\frac{P_1 \mathsf{tr}(\mathbf{u}\mathbf{u}^H)}{4\left(P_2 \tilde{\mathbf{g}}_l \mathbf{Q} \tilde{\mathbf{g}}_l^H + \sigma_E'^2\right)}\right) \lambda_{\iota_E} u_i^H u_i\right) du_i\right)}{M + \sum_{\iota_B=1}^{(N_s M - 1)M} \left(\int_{u_j} \exp\left(-\left(\frac{P_1 \mathsf{tr}(\mathbf{h}_l \mathbf{h}_l^H)}{4\left(P_2 \mathbf{h}_l \mathbf{Q} \mathbf{h}_l^H + \sigma_B^2\right)}\right) \lambda_{\iota_B} u_j^H u_j\right) du_j\right)} \right\} \tag{75}$$

$$= \log_2 \left\{ \frac{M + \sum_{\iota_E=1}^{(N_s M - 1)M} \frac{4\pi\left(P_2 \tilde{\mathbf{g}}_l \mathbf{Q} \tilde{\mathbf{g}}_l^H + \sigma_E'^2\right)}{\lambda_{\iota_E} P_1 \mathsf{tr}(\mathbf{u}\mathbf{u}^H)}}{M + \sum_{\iota_B=1}^{(N_s M - 1)M} \frac{4\pi\left(P_2 \mathbf{h}_l \mathbf{Q} \mathbf{h}_l^H + \sigma_B^2\right)}{\lambda_{\iota_B} P_1 \mathsf{tr}(\mathbf{h}_l \mathbf{h}_l^H)}} \right\} \approx \log_2 \left\{ \frac{\frac{\left(P_2 \tilde{\mathbf{g}}_l \mathbf{Q} \tilde{\mathbf{g}}_l^H + \sigma_E'^2\right)}{P_1 \mathsf{tr}(\mathbf{u}\mathbf{u}^H)} \left(\frac{1}{\lambda_1} + \cdots + \frac{1}{\lambda_{(N_s M - 1)M}}\right)}{\frac{\left(P_2 \mathbf{h}_l \mathbf{Q} \mathbf{h}_l^H + \sigma_B^2\right)}{P_1 \mathsf{tr}(\mathbf{h}_l \mathbf{h}_l^H)} \left(\frac{1}{\lambda_1} + \cdots + \frac{1}{\lambda_{(N_s M - 1)M}}\right)} \right\} \tag{76}$$

$$= \log_2 \left\{ \frac{\frac{\left(P_2 \tilde{\mathbf{g}}_l \mathbf{Q} \tilde{\mathbf{g}}_l^H + \sigma_E'^2\right)}{P_1 \mathsf{tr}(\mathbf{u}\mathbf{u}^H)}}{\frac{\left(P_2 \mathbf{h}_l \mathbf{Q} \mathbf{h}_l^H + \sigma_B^2\right)}{P_1 \mathsf{tr}(\mathbf{h}_l \mathbf{h}_l^H)}} \right\} = \log_2 \left\{ \frac{P_1 \|\mathbf{h}_l\|^2 \left(P_2 \tilde{\mathbf{g}}_l \mathbf{Q} \tilde{\mathbf{g}}_l^H + \sigma_E'^2\right)}{P_1 \left(\|\tilde{\mathbf{g}}_l\|^2 + N_s \sigma_e^2\right) \left(P_2 \mathbf{h}_l \mathbf{Q} \mathbf{h}_l^H + \sigma_B^2\right)} \right\}. \tag{77}$$

---

selection matrix and ANCM in the face of both desired channel and wiretap channel estimation errors.

APPENDIX A
PROOF OF THEOREM 1

Here, we investigate the characteristics of SSM system when N_t tends to large values. For convenience, let us define the matrices

$$\mathbf{M}_E = \frac{P_1(\tilde{\mathbf{g}}_l^H \tilde{\mathbf{g}}_l + \sigma_e^2 \mathbf{I})}{P_2 \tilde{\mathbf{g}}_l \mathbf{Q} \tilde{\mathbf{g}}_l^H + \sigma_E'^2}, \tag{66}$$

$$\mathbf{M}_B = \frac{P_1 \mathbf{h}_l^H \mathbf{h}_l}{P_2 \mathbf{h}_l \mathbf{Q} \mathbf{h}_l^H + \sigma_B^2}. \tag{67}$$

As $N_t \to \infty$, the components of the channel vector can be considered to obey the Gaussian distribution. Upon exploiting that

$$\mathbf{d}_{mk}^H \mathbf{M}_E \mathbf{d}_{mk} = \left(\frac{P_1}{P_2 \tilde{\mathbf{g}}_l \mathbf{Q} \tilde{\mathbf{g}}_l^H + \sigma_E'^2}\right) \mathbf{u}^H \mathbf{D}_{mk} \mathbf{u}, \tag{68}$$

where $\mathbf{D}_{mk} = \mathbf{d}_{mk} \mathbf{d}_{mk}^H$, and $\tilde{\mathbf{g}}_l^H \tilde{\mathbf{g}}_l + \sigma_e^2 \mathbf{I} = \mathbf{u}\mathbf{u}^H$ since $\tilde{\mathbf{g}}_l^H \tilde{\mathbf{g}}_l + \sigma_e^2 \mathbf{I}$ is symmetric and normalizing $\mathbf{u}$ as a standard Gaussian

distribution $\bar{\mathbf{u}}$, we arrive at

$$\left(\frac{P_1}{P_2 \tilde{\mathbf{g}}_l \mathbf{Q} \tilde{\mathbf{g}}_l^H + \sigma_E'^2}\right) \mathbf{u}^H \mathbf{D}_{mk} \mathbf{u} = \frac{P_1 \mathsf{tr}(\mathbf{u}\mathbf{u}^H)}{P_2 \tilde{\mathbf{g}}_l \mathbf{Q} \tilde{\mathbf{g}}_l^H + \sigma_E'^2} \bar{\mathbf{u}}^H \mathbf{D}_{mk} \bar{\mathbf{u}}. \tag{69}$$

When $m \neq k$, the rank of $\mathbf{D}_{mk} = (\boldsymbol{x}_m - \boldsymbol{x}_k)(\boldsymbol{x}_m - \boldsymbol{x}_k)^H$ equals to 1 and then $\mathbf{D}_{mk}$ can be rewritten as $\mathbf{D}_{mk} = \mathbf{U}_{mk}^H \mathsf{diag}(\lambda_{mk}, 0, \cdots, 0) \mathbf{U}_{mk}$, where λ_{mk} represents the unique nonzero eigenvalue of $\mathbf{D}_{mk}$ and $\mathbf{U}_{mk}$ is the unitary matrix whose columns are the eigenvectors of $\mathbf{D}_{mk}$.

As a further step, the expression (69) may be shown to be equivalent to

$$\mathbf{d}_{mk}^H \mathbf{M}_E \mathbf{d}_{mk} = \frac{P_1 \mathsf{tr}(\mathbf{u}\mathbf{u}^H)}{P_2 \tilde{\mathbf{g}}_l \mathbf{Q} \tilde{\mathbf{g}}_l^H + \sigma_E'^2} \lambda_{mk} u_i^H u_i, \quad (m \neq k), \tag{70}$$

in which u_i is the first element of the vector $\mathbf{U}_{mk} \bar{\mathbf{u}}$. The components of $\mathbf{U}_{mk} \bar{\mathbf{u}}$ still follow the Gaussian distribution due to the fact that the vectors $\bar{\mathbf{u}}$ and $\mathbf{U}_{mk} \bar{\mathbf{u}}$ have the same

statistics. When $\lambda_{mk} \neq 0$, as $N_t(N_s) \to \infty$, we have

$$\lim_{N_s \to \infty} \frac{1}{N_s} \sum_{i=1}^{N_s} \exp\left(-\frac{1}{4}\mathbf{d}_{mk}^H \mathbf{M}_E \mathbf{d}_{mk}\right) =$$

$$\int_{u_i} \exp\left(-\left(\frac{P_1 \text{tr}(\mathbf{u}\mathbf{u}^H)}{4\left(P_2\tilde{\mathbf{g}}_l \mathbf{Q}\tilde{\mathbf{g}}_l^H + \sigma_E'^2\right)}\right)\lambda_{mk} u_i^H u_i\right) du_i \tag{71}$$

$$= \frac{4\pi\left(P_2\tilde{\mathbf{g}}_l \mathbf{Q}\tilde{\mathbf{g}}_l^H + \sigma_E'^2\right)}{\lambda_{mk} P_1 \text{tr}(\mathbf{u}\mathbf{u}^H)}. \tag{72}$$

Upon exploiting that there are $N_s M$ possibilities of $\mathbf{d}_{mk}^H \mathbf{M}_E \mathbf{d}_{mk}$ $(m, k \in (1, \cdots, N_s M))$ equal to 0, we can derive expression (77), shown at the top of the previous page, where λ_{ij} is the eigenvalue of $\mathbf{D}_{ij} = (\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)^H$, and u_j is the first component of the vector $\mathbf{U}_{ij}\bar{\mathbf{u}}$.

It can be inferred from (77) that maximizing $R_A(\mathbf{S}_l, \mathbf{Q})$ in (27) for our large-scale SSM system can be further reduced to maximizing the ratio of SINR_B at the desired receiver to SINR_E at the eavesdropper, where we have:

$$\text{SINR}_B = \frac{P_1 \text{tr}(\mathbf{h}_l \mathbf{h}_l^H)}{P_2 \mathbf{h}_l \mathbf{Q}\mathbf{h}_l^H + \sigma_B^2}, \tag{78}$$

$$\text{SINR}_E = \frac{P_1 \text{tr}(\tilde{\mathbf{g}}_l^H \tilde{\mathbf{g}}_l + \sigma_e^2 \mathbf{I})}{P_2 \tilde{\mathbf{g}}_l \mathbf{Q}\tilde{\mathbf{g}}_l^H + \sigma_E'^2}. \tag{79}$$

This completes the proof of Theorem 1.

APPENDIX B
PROOF OF THEOREM 2

It can be seen from Procedure 1 that two components, e_i $(i = 1, \cdots, N_s)$ in $\mathbf{I}_1$ and j_c $(c = 1, \cdots, N_t - N_s)$ in $\mathbf{I}_0$ have to be chosen randomly to swap positions, i.e., $e_i \to 0$ and $j_c \to 1$. The probability for a neighborhood AAG to be selected equals to $1/(N_s N_t - N_s^2)$, thus we have

$$\forall \mathbf{s}, \mathbf{s}' \in \mathcal{C}, \exists\, p > 1, \exists\, \mathbf{s}_0, \mathbf{s}_1, \cdots, \mathbf{s}_p \in \mathcal{C}, \text{with}$$
$$\mathbf{s}_0 = \mathbf{s}, \mathbf{s}_p = \mathbf{s}', \text{and } G_{\mathbf{s},\mathbf{s}'}^c > 0, k = 0, \cdots, p-1. \tag{80}$$

Consequently, the components of the stationary distribution $\mathbf{q}(C)$ of the Markov chain $\xi(C)$ satisfy

$$q_{\mathbf{s}}(C) = \lim_{k \to \infty} P\left\{\xi_C(k) = \mathbf{s}|\xi_C(0) = \mathbf{s}'\right\}$$

$$= \frac{|N(\mathbf{s})|\exp\left(\frac{-R_s^c(\mathbf{s})}{C}\right)}{\sum_{\mathbf{s}' \in \mathcal{C}} |N(\mathbf{s}')|\exp\left(\frac{-R_s^c(\mathbf{s}')}{C}\right)}. \tag{81}$$

Then,

$$\lim_{C \to 0} q_{\mathbf{s}}(C) = q_{\mathbf{s}}^* = \begin{cases} \frac{1}{|S_{\text{opt}}|} & \mathbf{s} \in S_{\text{opt}} \\ 0 & \mathbf{s} \notin S_{\text{opt}}. \end{cases} \tag{82}$$

Finally, we have

$$\lim_{C \to 0} \lim_{k \to \infty} P\left\{\xi_C(k) = \mathbf{s}\right\} = \lim_{C \to 0} q_{\mathbf{s}}(C) = q_{\mathbf{s}}^*, \tag{83}$$

or

$$\lim_{C \to 0} \lim_{k \to \infty} P\left\{\xi_C(k) \in S_{\text{opt}}\right\} = \sum_{s \in S_{\text{opt}}} q_{\mathbf{s}}^* = 1, \tag{84}$$

which completes the proof of Theorem 2.

REFERENCES

[1] R. Y. Mesleh, H. Haas, S. Sinanovic, C. W. Ahn, and S. Yun, "Spatial modulation," *IEEE Trans. on Veh. Technol.*, vol. 57, no. 4, pp. 2228–2241, Jul. 2008.

[2] E. Basar, M. Wen, R. Mesleh, M. D. Renzo, Y. Xiao, and H. Haas, "Index modulation techniques for next-generation wireless networks," *IEEE Access*, vol. 5, pp. 16693–16746, 2017.

[3] P. Yang, B. Zhang, Y. Xiao, B. Dong, S. Li, M. El-Hajjar, and L. Hanzo, "Detect-and-forward relaying aided cooperative spatial modulation for wireless networks," *IEEE Trans. on Commun.*, vol. 61, no. 11, pp. 4500–4511, Nov. 2013.

[4] E. Basar, "On multiple-input multiple-output OFDM with index modulation for next generation wireless networks," *IEEE Trans. on Signal Process.*, vol. 64, no. 15, pp. 3868–3878, Aug. 2016.

[5] L. He, J. Wang, J. Song, and L. Hanzo, "On the multi-user multi-cell massive spatial modulation uplink: How many antennas for each user?" *IEEE Trans. on Wireless Commun.*, vol. 16, no. 3, pp. 1437–1451, Mar. 2017.

[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] P. Yang, Y. L. Guan, Y. Xiao, M. D. Renzo, S. Li, and L. Hanzo, "Transmit precoded spatial modulation: Maximizing the minimum euclidean distance versus minimizing the bit error ratio," *IEEE Trans. on Wireless Commun.*, vol. 15, no. 3, pp. 2054–2068, Mar. 2016.

[8] T. Lakshmi Narasimhan, P. Raviteja, and A. Chockalingam, "Generalized spatial modulation in large-scale multiuser MIMO systems," *IEEE Trans. on Wireless Commun.*, vol. 14, no. 7, pp. 3764–3779, Jul. 2015.

[9] Q. Wu, G. Y. Li, W. Chen, D. W. K. Ng, and R. Schober, "An overview of sustainable green 5G networks," *IEEE Wireless Commun.*, vol. 24, no. 4, pp. 72–80, Aug. 2017.

[10] J. Jin, C. Xiao, M. Tao, and W. Chen, "Linear precoding for fading cognitive multiple-access wiretap channel with finite-alphabet inputs," *IEEE Trans. on Veh. Technol.*, vol. 66, no. 4, pp. 3059–3070, Apr. 2017.

[11] X. Chen, D. W. K. Ng, W. H. Gerstacker, and H. H. Chen, "A survey on multiple-antenna techniques for physical layer security," *IEEE Commun. Surv. Tut.*, vol. 19, no. 2, pp. 1027–1053, 2017.

[12] Y. Zou, J. Zhu, X. Wang, and L. Hanzo, "A survey on wireless security: Technical challenges, recent advances, and future trends," *Proc. IEEE*, vol. 104, no. 9, pp. 1727–1765, Sep. 2016.

[13] F. Wang, C. Liu, Q. Wang, J. Zhang, R. Zhang, L. Yang, and L. Hanzo, "Optical jamming enhances the secrecy performance of the generalized space-shift-keying-aided visible-light downlink," *IEEE Trans. on Commun.*, vol. 66, no. 9, pp. 4087–4102, Sep. 2018.

[14] S. Yan, N. Yang, I. Land, R. Malaney, and J. Yuan, "Three artificial-noise-aided secure transmission schemes in wiretap channels," *IEEE Trans. on Veh. Technol.*, vol. 67, no. 4, pp. 3669–3673, Apr. 2018.

[15] H. M. Wang, Q. Yin, and X. G. Xia, "Distributed beamforming for physical-layer security of two-way relay networks," *IEEE Trans. on Signal Process.*, vol. 60, no. 7, pp. 3532–3545, Jul. 2012.

[16] Y. Wu, C. Xiao, Z. Ding, X. Gao, and S. Jin, "Linear precoding for finite-alphabet signaling over MIMOME wiretap channels," *IEEE Trans. on Veh. Technol.*, vol. 61, no. 6, pp. 2599–2612, Jul. 2012.

[17] Y. Wu, J. B. Wang, J. Wang, R. Schober, and C. Xiao, "Secure transmission with large numbers of antennas and finite alphabet inputs," *IEEE Trans. on Commun.*, vol. 65, no. 8, pp. 3614–3628, Aug. 2017.

[18] C. Liu, L. L. Yang, and W. Wang, "Secure spatial modulation with a full-duplex receiver," *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 838–841, Dec. 2017.

[19] L. Wang, S. Bashar, Y. Wei, and R. Li, "Secrecy enhancement analysis against unknown eavesdropping in spatial modulation," *IEEE Commun. Lett.*, vol. 19, no. 8, pp. 1351–1354, Aug. 2015.

[20] Y. Chen, L. Wang, Z. Zhao, M. Ma, and B. Jiao, "Secure multiuser MIMO downlink transmission via precoding-aided spatial modulation," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1116–1119, Jun. 2016.

[21] F. Wu, R. Zhang, L. Yang, and W. Wang, "Transmitter precoding-aided spatial modulation for secrecy communications," *IEEE Trans. on Veh. Technol.*, vol. 65, no. 1, pp. 467–471, Jan. 2016.

[22] F. Wu, L. L. Yang, W. Wang, and Z. Kong, "Secret precoding-aided spatial modulation," *IEEE Commun. Lett.*, vol. 19, no. 9, pp. 1544–1547, Sep. 2015.

[23] X. Q. Jiang, M. Wen, H. Hai, J. Li, and S. Kim, "Secrecy-enhancing scheme for spatial modulation," *IEEE Commun. Lett.*, vol. 22, no. 3, pp. 550–553, Mar. 2018.

[24] Y. Yang and M. Guizani, "Mapping-varied spatial modulation for physical layer security: Transmission strategy and secrecy rate," *IEEE J. Sel. Area. Commun.*, vol. 36, no. 4, pp. 877–889, Apr. 2018.

[25] F. Shu, Z. Wang, R. Chen, Y. Wu, and J. Wang, "Two high-performance schemes of transmit antenna selection for secure spatial modulation," *IEEE Trans. on Veh. Technol.*, vol. 67, no. 9, pp. 8969–8973, Sep. 2018.

[26] G. Xia, F. Shu, Y. Zhang, J. Wang, S. ten Brink, and J. Speidel, "Antenna selection method of maximizing secrecy rate for green secure spatial modulation," *IEEE Trans. on Green Commun. and Netw.*, vol. 3, no. 2, pp. 288–301, Jun. 2019.

[27] S. Yan, N. Yang, R. Malaney, and J. Yuan, "Transmit antenna selection with alamouti coding and power allocation in MIMO wiretap channels," *IEEE Trans. on Wireless Commun.*, vol. 13, no. 3, pp. 1656–1667, Mar. 2014.

[28] R. Rajashekar, K. V. S. Hari, and L. Hanzo, "Antenna selection in spatial modulation systems," *IEEE Commun. Lett.*, vol. 17, no. 3, pp. 521–524, Mar. 2013.

[29] Z. Zhou, N. Ge, and X. Lin, "Reduced-complexity antenna selection schemes in spatial modulation," *IEEE Commun. Lett.*, vol. 18, no. 1, pp. 14–17, Jan. 2014.

[30] Z. Sun, Y. Xiao, P. Yang, S. Li, and W. Xiang, "Transmit antenna selection schemes for spatial modulation systems: Search complexity reduction and large-scale MIMO applications," *IEEE Trans. on Veh. Technol.*, vol. 66, no. 9, pp. 8010–8021, Sep. 2017.

[31] P. Yang, Y. Xiao, Y. L. Guan, S. Li, and L. Hanzo, "Transmit antenna selection for multiple-input multiple-output spatial modulation systems," *IEEE Trans. on Commun.*, vol. 64, no. 5, pp. 2035–2048, May 2016.

[32] N. S. Perovic, P. Liu, J. Blumenstein, M. Di Renzo, and A. Springer, "Optimization of the cut-off rate of generalized spatial modulation with transmit precoding," *IEEE Trans. on Commun.*, vol. 66, no. 10, pp. 4578–4595, Oct. 2018.

[33] B. Wang, H. Xie, X. Xia, and X. Zhang, "A NSGA-II algorithm hybridizing local simulated-annealing operators for a bi-criteria robust job-shop scheduling problem under scenarios," *IEEE Trans. on Fuzzy Syst.*, vol. 27, no. 5, pp. 1075–1084, May 2019.

[34] S. Zhan, Z. Zhang, L. Wang, and Y. Zhong, "List-based simulated annealing algorithm with hybrid greedy repair and optimization operator for 0-1 knapsack problem," *IEEE Access*, vol. 6, pp. 54 447–54 458, 2018.

[35] R. Zhang, L. L. Yang, and L. Hanzo, "Error probability and capacity analysis of generalised pre-coding aided spatial modulation," *IEEE Trans. on Wireless Commun.*, vol. 14, no. 1, pp. 364–375, Jan. 2015.

[36] Y. Tang, J. Xiong, D. Ma, and X. Zhang, "Robust artificial noise aided transmit design for MISO wiretap channels with channel uncertainty," *IEEE Commun. Lett.*, vol. 17, no. 11, pp. 2096–2099, Nov. 2013.

[37] X. Yu, Y. Hu, Q. Pan, X. Dang, N. Li, and M. H. Shan, "Secrecy performance analysis of artificial-noise-aided spatial modulation in the presence of imperfect CSI," *IEEE Access*, vol. 6, pp. 41 060–41 067, 2018.

[38] S. Bashar, Z. Ding, and C. Xiao, "On the secrecy rate of multi-antenna wiretap channel under finite-alphabet input," *IEEE Commun. Lett.*, vol. 15, no. 5, pp. 527–529, May 2011.

[39] S. R. Aghdam and T. M. Duman, "Joint precoder and artificial noise design for MIMO wiretap channels with finite-alphabet inputs based on the cut-off rate," *IEEE Trans. on Wireless Commun.*, vol. 16, no. 6, pp. 3913–3923, Jun. 2017.

[40] Z. Zhu, Z. Chu, N. Wang, S. Huang, Z. Wang, and I. Lee, "Beamforming and power splitting designs for AN-aided secure multi-user MIMO SWIPT systems," *IEEE Trans. on Inf. Forensics Security*, vol. 12, no. 12, pp. 2861–2874, Dec. 2017.

[41] P. Lin, S. Lai, S. Lin, and H. Su, "On secrecy rate of the generalized artificial-noise assisted secure beamforming for wiretap channels," *IEEE J. Sel. Area. Commun.*, vol. 31, no. 9, pp. 1728–1740, Sep. 2013.

[42] F. Nie, S. Xiang, Y. Jia, C. Zhang, and S. Yan, "Trace ratio criterion for feature selection," in *Proc. 23rd AAAl Conf. Artif. Intell.*, 2008, pp. 671–676.

[43] H. Wang, S. Yan, D. Xu, X. Tang, and T. Huang, "Trace ratio vs. ratio trace for dimensionality reduction," in *Proc. IEEE Conf. on Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[44] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. second ed. Academic Press, 1991.

[45] P. Tian, J. Ma, and D.-M. Zhang, "Non-linear integer programming by darwin and boltzmann mixed strategy," *Eur. J. Oper. Res.*, vol. 105, no. 1, pp. 224–235, 1998.

[46] Z. Chu, Z. Zhu, M. Johnston, and S. Y. L. Goff, "Simultaneous wireless information power transfer for MISO secrecy channel," *IEEE Trans. on Veh. Technol.*, vol. 65, no. 9, pp. 6913–6925, Sep. 2016.

[47] X. Zhou, J. Li, F. Shu, Q. Wu, Y. Wu, W. Chen, and L. Hanzo, "Secure SWIPT for directional modulation-aided AF relaying networks," *IEEE J. Sel. Area. Commun.*, vol. 37, no. 2, pp. 253–268, Feb. 2019.

[48] W. Ben-Ameur, "Computing the initial temperature of simulated annealing," *Comput. Optim. Appl.*, vol. 29, no. 3, pp. 369–385, 2004.