Received 1 July 2022; accepted 18 July 2022. Date of publication 22 July 2022; date of current version 5 August 2022.

Digital Object Identifier 10.1109/OJCOMS.2022.3193068

Mutual Information Analysis for Generalized Spatial Modulation Systems With Multilevel Coding

VOLKER KÜHN [1] (Member, IEEE), DANIEL NICOLAS BAILON [2] (Graduate Student Member, IEEE), AND JÜRGEN FREUDENBERGER [2] (Member, IEEE)

[1] Faculty of Computer Science and Electrical Engineering, University of Rostock, 18051 Rostock, Germany

[2] Institute for System Dynamics, HTWG Konstanz, University of Applied Sciences, 78462 Konstanz, Germany

CORRESPONDING AUTHOR: J. FREUDENBERGER (e-mail: juergen.freudenberger@htwg-konstanz.de)

This work was supported by the German Federal Ministry of Research and Education (BMBF, 16ES1045) through the PENTA Project 17013 XSR-FMC.

ABSTRACT Generalized Spatial Modulation (GSM) enables a trade-off between very high spectral efficiencies and low hardware costs for massive MIMO systems. This is achieved by transmitting information via the selection of active antennas from a set of available antennas besides the transmission of conventional data symbols. GSM systems have been investigated concerning various aspects like suitable signal constellations, efficient detection algorithms, hardware implementations, spatial precoding, and error control coding. On the other hand, determining the capacity of GSM is challenging because no closed-form expressions have been found so far. This paper investigates the mutual information for different GSM variants. We consider a multilevel coding approach, where the antenna selection and IQ modulation are encoded independently. Combined with multistage decoding, such an approach enables low-complexity capacity-achieving coded modulation. The influence of the data symbols on the mutual information is illuminated. We analyze the portions of mutual information related to antenna selection and the IQ modulation processes which depend on the GSM variant and the signal constellation. Moreover, the potential of spatial modulation for massive MIMO systems with many transmit antennas is investigated. Especially in systems with many transmit antennas much information can be conveyed by antenna selection.

INDEX TERMS Spatial modulation, information theory, capacity.

I. INTRODUCTION

MULTIPLE antenna systems have become popular since the seminal work of [1] and are an integral part of many state-of-the-art communication standards. With the trend to higher carrier frequencies like mm-wave or even beyond, antennas grew extremely small and many elements can be incorporated into an antenna array of moderate size. Massive Multiple-Input Multiple-Output (MIMO) systems are currently discussed where hundreds of antennas at the base station lead to the channel hardening effect [2], [3] enabling high spectral efficiencies with very simple signal processing techniques. However, with a growing antenna number, the number of AD/DA converters, mixers, and amplifiers grows as well.

A completely different approach is pursued with GSM where information is conveyed by antenna selection. In GSM systems, only a small subset of transmit antennas is activated simultaneously, and, consequently, all remaining antennas are inactive. Originally, Spatial Modulation (SM) uses only a single active transmit antenna in each time slot, whereas GSM activates multiple transmit antennas at a time [4]–[6]. Furthermore, GSM and Generalized Multistream Spatial Modulation GMSM can be distinguished. While the former system transmits the same data symbol over all active antennas, the latter allows transmitting independent data streams over active antennas. If antenna selection can be implemented by fast switching devices, only a few radio frequency (RF) chains are needed, lowering the implementation costs. In massive MIMO systems, extremely high spectral efficiencies can be achieved with SM. However, an appropriate number of receive antennas is required to enable the receiver to detect the transmit signal. Furthermore, the computational complexity of the optimal Maximum Likelihood (ML) detector for SM increases

rapidly as more antennas or larger modulation orders are used.

Inspired by the favorable properties, a lot of research has been spent on SM. Detailed information about the various facets of SM and a large number of applications can be found in the overview papers and surveys [7]–[10] and the references therein.

Plenty of work has been spent to optimize signal constellations and detection algorithms in uncoded SM systems. For instance, signal constellations have been investigated to improve the performance of SM [5], [6], [11]–[17]. For these uncoded approaches, suboptimal antenna detection algorithms are analyzed in [11], [18]–[23]. Furthermore, precoding for SM systems has been investigated in order to mitigate spatial correlations [24], [25]. Fewer publications consider coded spatial modulation schemes. Some focus on combinations of SM with space-time coding [26]–[28]. Others generalize traditional coding techniques like trellis-coded modulation to SM transmission [25], [29], [30].

While the analysis of coded and uncoded SM in terms of error rate performance provides valuable insights about structural properties of SM systems, the ultimate performance limits have to be tackled by looking at the system capacity. The GSM capacity is difficult to determine because no closed-form solution for the differential entropy of a Gaussian Mixture Model (GMM) exists. Therefore, the mutual information is usually approximated by lower and upper bounds [31]–[34] or computed numerically. The latter approach is limited to only small system dimensions due to the curse of dimensionality [35]. In [36], the capacity of an SM system with a single active transmit and a single receive antenna is analyzed. The mutual information is approximated by exploiting an ML detection resulting in a discrete memoryless channel with pairwise error probabilities. A significant gap between the numerically computed mutual information and this approximation is observed. A lower bound on the mutual information for the non-orthogonal downlink transmission with finite transmit alphabets and spatial modulation is derived in [37]. The authors in [38] considered a GSM system with linear precoding using $N_a = N_t$ active antennas. Two different precoders are considered and their performance in terms of overall mutual information as well as antenna detection mutual information is compared to a conventional GSM system. The mutual information is approximated in the same way as in [36] using the pairwise error probability of a maximum likelihood detector.

Many of these publications consider rather small SM systems or they concentrate on the overall mutual information and do not analyze the specific mutual information obtained from the antenna selection process. Recently, the authors in [39] proposed to use the cut-off rate instead of the mutual information for analyzing generalized spatial modulation systems with transmit precoding. Another interesting approach is Huffman coding-based spatial modulation presented in [40]. Moreover, the authors provide upper and lower bounds for the mutual information of an SM system with one active transmit antenna.

In this work, we investigate the capacity of different variants of SM systems. Similar to coded modulation [41], we partition the SM system into two levels, i.e., the antenna selection level and the IQ modulation level. In combination with multistage decoding, such an approach enables low-complexity capacity-achieving coded modulation. To achieve the overall channel capacity, the code rates have to be chosen according to the capacity rule [41] for each of the levels. We investigate the portions of mutual information related to antenna selection and the IQ modulation processes which depend on the GSM transmission system and the signal constellation. The presented analysis provides the following contributions.

1) We derive upper and lower bounds on the mutual information for different versions of SM systems. The tightness of the bounds is analyzed for small system sizes by comparing them with results from numerical integration.
2) The mutual information of different SM variants is investigated. In particular, the rates for the antenna selection level and the IQ modulation level are considered.
3) The influence of the data symbols on the antenna detection is investigated. We demonstrate that the signal constellations have little impact on the overall mutual information until the information of the IQ modulation level saturates due to the finite alphabet size. However, the partitioning of the rates for the antenna selection level and the IQ modulation level depends on the signal constellation.
4) The bounds can be used to investigate the mutual information of massive MIMO systems with many transmit but only a few active antennas. We demonstrate that much information can be conveyed by antenna selection for large systems, whereas the portion of the mutual information provided by the IQ modulation level becomes comparably small.

The paper is organized as follows: Section II introduces the considered system model and various variants of SM. The mutual information achievable with SM is derived in Section III. Upper and lower bounds are proposed for each SM variation in Section IV. Numerical results are presented in Section V before Section VI concludes this paper.

II. SPATIAL MODULATION SYSTEM MODEL

In this section, we introduce the notation and describe the system model. Random variables are denoted by calligraphic letters $\mathcal{X}$, $\mathcal{Y}$, and the corresponding signals by small letters x and y. They can take values from the sets $\mathbb{X}$ and $\mathbb{Y}$ with cardinalities $|\mathbb{X}|$ and $|\mathbb{Y}|$, respectively. The probability of discrete random variables is denoted by $\Pr\{\mathcal{X} = x\}$, and the terms $p(y|x)$, $p(x, y)$ represent the conditional and joint probability density functions. The mutual information of two random variables $\mathcal{X}$ and $\mathcal{Y}$ is termed $I(\mathcal{X}; \mathcal{Y})$. Moreover,

vectors and multivariate random variables are denoted in bold letters $\mathbf{y}$ and $\boldsymbol{\mathcal{Y}}$, respectively.

A. GENERAL SYSTEM DESCRIPTION

GSM is a MIMO transmission strategy conveying information by selecting N_a active antennas from a set of N_t transmit antennas. Hence, the transmit vector $\mathbf{x} = \begin{bmatrix} x_1 & \cdots & x_{N_t} \end{bmatrix}^{\mathrm{T}}$ is sparse and contains $N_t - N_a$ zeros. The assignment of data symbols onto the active antennas can be expressed by the $N_t \times N_a$ mapping matrix $\mathbf{F}_l$ having a single one in each row and N_a ones in total. For instance, the matrix

$$\mathbf{F}_l = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

maps symbol s_1 onto antenna 1 and symbol s_2 onto antenna 2. The index l denotes the set of active antennas termed antenna pattern. It can be chosen from $l \in \{1, 2, \ldots, N_p\}$ with $N_p = \binom{N_t}{N_a}$. Therefore, the selection of N_a active antennas and the mapping of data symbols onto these antennas can be described by a bijective function

$$\mathbf{x} = f(\mathbf{s}, l) = \mathbf{F}_l \cdot \mathbf{s}. \tag{1}$$

The $N_a \times 1$ vector $\mathbf{s}$ comprises the data symbols $s_\mu \in \mathbb{S}$ to be transmitted. They are assumed to have zero mean and an average power $\mathrm{E}\{|\mathcal{S}|^2\} = 1$ per symbol.

The receiver is equipped with N_r antennas and has the tasks to detect the antenna selection index l as well as the data symbols $\mathbf{s}$ transmitted over all active antennas. The received vector can be described by

$$\mathbf{y} = \rho \cdot \mathbf{H} \cdot \mathbf{x} + \mathbf{n} = \rho \cdot \mathbf{H}\mathbf{F}_l \cdot \mathbf{s} + \mathbf{n} = \rho \cdot \mathbf{H}_l \cdot \mathbf{s} + \mathbf{n} \tag{2}$$

with the $N_r \times N_t$ channel matrix $\mathbf{H}$ and the noise vector $\mathbf{n}$. The channel matrix $\mathbf{H}$ is assumed to consist of i.i.d. complex Gaussian random variables with zero mean and variance $\sigma^2_{\mathcal{H}} = 1$. The matrix $\mathbf{H}_l = \mathbf{H} \cdot \mathbf{F}_l$ contains only those columns of $\mathbf{H}$ associated to active antenna elements. The factor ρ scales the overall transmit power. Assuming Additive White Gaussian Noise (AWGN) with average power $\sigma^2_{\mathcal{N}} = 1$ per receive antenna, the likelihood function of the SM system has the form

$$\begin{aligned} p(\mathbf{y} \mid \mathbf{H}, \mathbf{x}) &= p(\mathbf{y} \mid \mathbf{H}, \mathbf{s}, l) \\ &= \frac{1}{\pi^{N_r}} e^{-\|\mathbf{y} - \rho \mathbf{H}\mathbf{x}\|^2} = \frac{1}{\pi^{N_r}} e^{-\|\mathbf{y} - \rho \mathbf{H}_l \mathbf{s}\|^2} . \end{aligned} \tag{3}$$

At each receive antenna, we obtain the average Signal-to-Noise Ratio (SNR)

$$\tilde{\gamma} = \rho^2 \cdot \frac{N_a \sigma^2_{\mathcal{H}} \mathrm{E}\{|\mathcal{S}|^2\}}{\sigma^2_{\mathcal{N}}} = \rho^2 N_a. \tag{4}$$

Due to $\sigma^2_{\mathcal{H}} = 1$, the overall received signal power depends on the number of receive antennas. In order to ensure a fair comparison of systems with different N_r, we also incorporate the number of receive antennas into the SNR measure and obtain

$$\gamma = \tilde{\gamma} \cdot N_r = \rho^2 N_a N_r \ . \tag{5}$$

The SNR defined in (5) is used in subsequent sections when presenting numerical results.

B. VARIANTS OF SPATIAL MODULATION

Different variants of SM exist in the literature. In this paper, we distinguish the following approaches assuming equally likely antenna patterns, i.e., $\Pr\{\mathcal{L} = l\} = \frac{1}{N_p}$ holds.

1) Generalized Space Shift Keying

With Generalized Space Shift Keying (GSSK), information is only conveyed via the selection of N_a active antennas and no data symbols are transmitted. For this strategy, the data symbol vector $\mathbf{s} = \mathbf{1}$ becomes the all-one vector of length N_a. For a fixed channel matrix $\mathbf{H}$ and a particular antenna pattern l, the received vector

$$\mathbf{y} = \rho \cdot \mathbf{H}_l \mathbf{1} + \mathbf{n} = \rho \cdot \tilde{\mathbf{h}}_l + \mathbf{n} \tag{6}$$

is Gaussian distributed with conditional mean $\mu_l = \mathrm{E}\{\boldsymbol{\mathcal{Y}} | \mathbf{H}_l\} = \rho \tilde{\mathbf{h}}_l$ and covariance matrix $\Sigma_l = \mathbf{I}$. For $N_a = 1$, the vector $\tilde{\mathbf{h}}_l$ is simply the l-th column of $\mathbf{H}$, whereas $\tilde{\mathbf{h}}_l$ equals the sum of those columns of $\mathbf{H}$ corresponding to active antennas for $N_a > 1$. Without conditioning on $\mathbf{H}_l$, the distribution of $\mathbf{y}$ becomes a Gaussian Mixture Model (GMM)

$$p(\mathbf{y}|\mathbf{H}) = \frac{1}{N_p} \cdot \sum_{l=1}^{N_p} p(\mathbf{y}|\mathbf{H}, l) \tag{7}$$

with mean

$$\mu_{\boldsymbol{\mathcal{Y}}|\mathbf{H}} = \frac{1}{N_p} \cdot \sum_{l=1}^{N_p} \mu_l = \frac{\rho}{N_p} \cdot \sum_{l=1}^{N_p} \tilde{\mathbf{h}}_l$$

and covariance matrix

$$\Sigma_{\boldsymbol{\mathcal{Y}}|\mathbf{H}} = \mathbf{I} + \frac{1}{N_{\mathrm{P}}} \sum_{l=1}^{N_{\mathrm{P}}} \mu_l \cdot \left(\mu_l - \mu_{\boldsymbol{\mathcal{Y}}|\mathbf{H}}\right)^{\mathrm{H}}. \tag{8}$$

Space Shift Keying (SSK) with $N_a = 1$ is included as a special case.

2) Generalized Spatial Modulation

In contrast to GSSK, GSM transmits the same data symbol s over N_a active antennas. In this case, $\mathbf{s} = s \cdot \mathbf{1}$ holds leading to the received vector

$$\mathbf{y} = \rho \cdot \mathbf{H}_l \mathbf{1} \cdot s + \mathbf{n} = \rho \cdot \tilde{\mathbf{h}}_l s + \mathbf{n} \ . \tag{9}$$

Conditioned on $\mathbf{H}$ and l, $\mathbf{y}$ is Gaussian distributed with zero mean and covariance matrix

$$\Sigma_l = \Sigma_{\boldsymbol{\mathcal{Y}}|\mathbf{H}_l} = \rho^2 \tilde{\mathbf{h}}_l \tilde{\mathbf{h}}_l^{\mathrm{H}} + \mathbf{I} \ . \tag{10}$$

The covariance matrix of the resulting GMM becomes

$$\Sigma_{\boldsymbol{\mathcal{Y}}|\mathbf{H}} = \frac{1}{N_p} \sum_{l=1}^{N_p} \Sigma_l = \mathbf{I} + \frac{\rho^2}{N_p} \sum_{l=1}^{N_p} \tilde{\mathbf{h}}_l \tilde{\mathbf{h}}_l^{\mathrm{H}} \ . \tag{11}$$

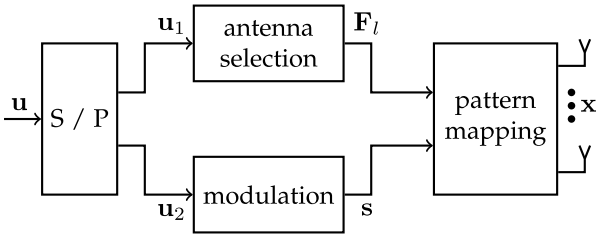

FIGURE 1. Transmitter model for multilevel coded SM.

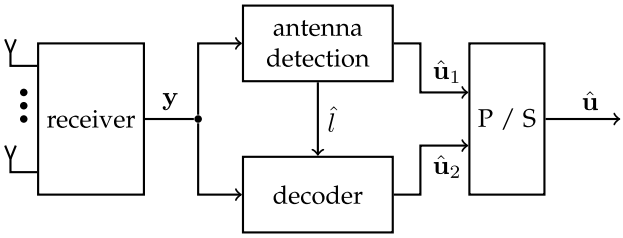

FIGURE 2. Receiver model for multilevel coded SM.

3) Generalized Multistream Spatial Modulation

For GMSM, the vector $\mathbf{s}$ contains N_a independent data symbols, each transmitted over one of the active antennas. The received vector $\mathbf{y}$ has the general form as described in (2) and is conditionally Gaussian distributed with zero mean and covariance matrix

$$\Sigma_l = \Sigma_{\mathcal{Y}|\mathbf{H}_l} = \rho^2 \mathbf{H}_l \mathbf{H}_l^{\mathrm{H}} + \mathbf{I}. \tag{12}$$

The covariance matrix of the resulting GMM is

$$\Sigma_{\mathcal{Y}|\mathbf{H}} = \mathbf{I} + \frac{\rho^2}{N_p} \sum_{l=1}^{N_p} \mathbf{H}_l \mathbf{H}_l^{\mathrm{H}} \ . \tag{13}$$

III. MUTUAL INFORMATION OF SM SYSTEMS

As described in the introduction, the capacity of SM systems representing the maximum mutual information $I(\mathcal{X}; \mathcal{Y}|\mathbf{H}) = h(\mathcal{Y}|\mathbf{H}) - h(\mathcal{Y}|\mathcal{X}, \mathbf{H})$ has been analyzed under various assumptions. A major difficulty is the computation of the differential entropy $h(\mathcal{Y}|\mathbf{H})$ because no closed-form solutions exist for Gaussian mixture models. Furthermore, the computation via numerical integration is restricted to small system dimensions only, because the sampling space increases exponentially with N_r and N_t.

In this section, we first exploit the system descriptions in Section II and provide expressions for the mutual information of GMSM, GSM, and GSSK systems. Subsequently, results obtained by numerical integration are presented for small N_r. This includes the investigation of the influence of the modulation scheme onto the mutual information.

A. MULTI-LEVEL CODING FOR SM SYSTEMS

The bijective mapping $\mathbf{x} = f(\mathbf{s}, l)$ introduced in Section II suggests to apply a multi-level coding approach as depicted in Figs. 1 and 2. Similar to the coded modulation scheme proposed in [41], [42], the SM system is split into two levels by multiplexing the stream of information bits $\mathbf{u}$ onto two branches. In the upper branch, indices l are generated for the selection of active antennas. The lower branch determines the data symbols $\mathbf{s}$ to be transmitted. These data symbols are then assigned to the active antennas in the block *pattern mapping* delivering the transmit vector $\mathbf{x}$.

At the receiver, multi-stage decoding is performed. First, the antenna pattern is identified. Based on the decision $\hat{l}$, the data symbols can be detected. The corresponding estimated information bits $\hat{\mathbf{u}}_1$ and $\hat{\mathbf{u}}_2$ are demultiplexed for obtaining the final bit vector $\hat{\mathbf{u}}$.

Applying the chain rule [43], the channel dependent mutual information $I(\mathcal{X}; \mathcal{Y} \mid \mathbf{H})$ of a general GMSM system can be described by

$$\begin{aligned} I(\mathcal{X}; \mathcal{Y} \mid \mathbf{H}) &= I(\mathcal{S}, \mathcal{L}; \mathcal{Y} \mid \mathbf{H}) \\ &= I(\mathcal{L}; \mathcal{Y} \mid \mathbf{H}) + I(\mathcal{S}; \mathcal{Y} \mid \mathcal{L}, \mathbf{H}). \end{aligned} \tag{14}$$

The term $I(\mathcal{L}; \mathcal{Y}|\mathbf{H})$ in (14) represents the mutual information related to the antenna selection and detection process and corresponds to the upper branches in Figs. 1 and 2. The second term $I(\mathcal{S}; \mathcal{Y}|\mathcal{L}, \mathbf{H})$ denotes the mutual information for a known antenna pattern l and represents a conventional $N_a \times N_r$ MIMO system with the channel matrices $\mathbf{H}_l$. It corresponds to the lower levels in Figs. 1 and 2.

1) MUTUAL INFORMATION OF IQ MODULATION LEVEL

For the general GMSM system, the mutual information $I(\mathcal{S}; \mathcal{Y}|\mathcal{L}, \mathbf{H})$ of the IQ modulation level is maximized for Gaussian distributed transmit symbols yielding the classical MIMO capacity

$$\begin{aligned} I(\mathcal{S}; \mathcal{Y} \mid \mathbf{H}_l) &= C(\mathbf{H}_l) \\ &= \log_2 \det\left(\mathbf{I} + \rho^2 \mathbf{H}_l \mathbf{H}_l^{\mathrm{H}}\right). \end{aligned} \tag{15}$$

For GSM transmitting the same data symbol s over all active antennas, maximum ratio combining $\tilde{\mathbf{h}}_l^{\mathrm{H}} \mathbf{y} = \rho |\tilde{\mathbf{h}}_l|^2 s + \tilde{\mathbf{h}}_l^{\mathrm{H}} \mathbf{n}$ can be performed once the antenna pattern has been detected leading to

$$I(\mathcal{S}; \mathcal{Y}|\mathbf{H}_l) = \log_2\left(1 + \rho^2 \|\tilde{\mathbf{h}}_l\|^2\right). \tag{16}$$

For GSSK, no data symbols are transmitted and $I(\mathcal{S}; \mathcal{Y}|\mathbf{H}_l) = 0$ holds. Averaging over all antenna patterns and channel realizations yields

$$I(\mathcal{S}; \mathcal{Y} \mid \mathcal{L}) = \mathrm{E}_{\mathbf{H}} \left\{ \frac{1}{N_p} \sum_{l=1}^{N_p} I(\mathcal{S}; \mathcal{Y} \mid \mathbf{H}_l) \right\}. \tag{17}$$

2) MUTUAL INFORMATION OF ANTENNA SELECTION LEVEL

The computation of the mutual information $I(\mathcal{L}; \mathcal{Y}|\mathbf{H})$ in (14) is more involved. It has to be mentioned that even though the chain rule allows the separation of antenna selection and data symbol transmission, the mutual information $I(\mathcal{L}; \mathcal{Y}|\mathbf{H})$ is affected by the type of modulation because

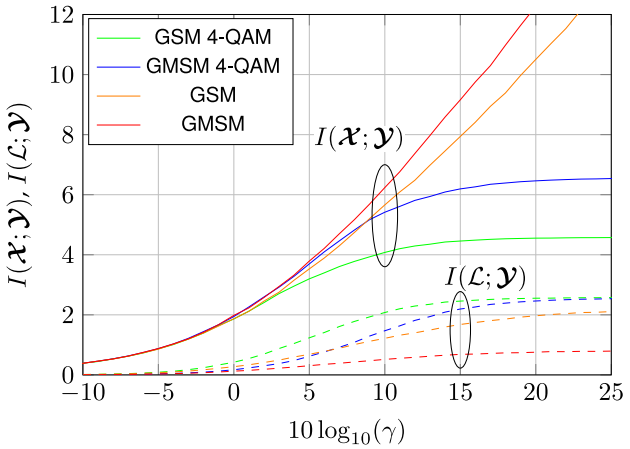

FIGURE 3. Capacities of antenna selection level with $N_t = 4$, $N_a = 2$, $N_r = 2$ and QAM modulation schemes.

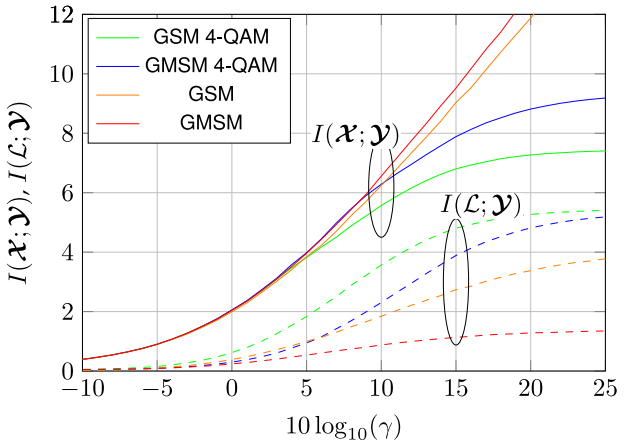

FIGURE 4. Capacities of antenna selection level with $N_t = 10$, $N_a = 2$, $N_r = 2$ and QAM modulation schemes.

this influences the distribution of $\mathcal{Y}$. For a particular channel matrix $\mathbf{H}$, it is defined by

$$I(\mathcal{L};\mathcal{Y} \mid \mathbf{H}) = h(\mathcal{Y} \mid \mathbf{H}) - h(\mathcal{Y} \mid \mathcal{L}, \mathbf{H}).$$

According to Section II-B, $\mathcal{Y}$ conditioned on l and $\mathbf{H}$ is again Gaussian distributed. For GMSM and GSM, $\mathcal{Y}$ has zero mean and covariance matrix Σ_l. Its differential entropy can be computed as

$$h(\mathcal{Y}|l,\mathbf{H}) = \log_2 \det(\pi e \Sigma_l). \tag{18}$$

For the GSSK variant, the conditional mean is nonzero and the covariance matrix equals the identity matrix leading to $h(\mathcal{Y}|l,\mathbf{H}) = \log_2 \det(\pi e \mathbf{I}) = N_r \log_2(\pi e)$. The computation of the differential entropy $h(\mathcal{Y} \mid \mathbf{H})$ requires the density

$$p(\mathbf{y} \mid \mathbf{H}) = \tfrac{1}{N_p} \cdot \textstyle\sum_{l=1}^{N_p} p(\mathbf{y} \mid \mathbf{H}_l) \tag{19}$$

of a GMM. A simple analytical treatment via covariance matrices is no longer possible, as $p(\mathbf{y}|\mathbf{H})$ is not Gaussian [32] and no closed-form solution for the differential entropy exists [44]. A straightforward way is to compute $p(\mathbf{y}|\mathbf{H})$ and subsequently $h(\mathcal{Y}|\mathbf{H})$ numerically. This requires an appropriate sampling of a $2N_r$-dimensional space which becomes quickly infeasible due to the curse of dimensionality and necessitates the development of efficient bounds. Next, we discuss some numerical results for small system dimensions.

B. NUMERICAL RESULTS

For small system dimensions, the level-specific mutual information can be numerically computed. Fig. 3 shows results for GMSM as well as for GSM and allows a comparison of the level capacities. The considered SM system employs $N_t = 4$ transmit, $N_r = 2$ receive antennas, and $N_a = 2$ transmit antennas are active. The total mutual information $I(\mathcal{X};\mathcal{Y})$ is represented by solid lines, whereas dashed lines represent the information $I(\mathcal{L};\mathcal{Y})$ transmitted via the antenna selection. The MIMO part $I(\mathcal{S};\mathcal{Y}|\mathcal{L})$ is not shown for clarity. The distribution of the overall mutual information $I(\mathcal{X};\mathcal{Y})$ into the two levels differs for GMSM and GSM. Moreover, it also depends on the modulation alphabet. To show the influence of the modulation alphabet, results with Gaussian random variables as well as with 4-QAM symbols are provided.

Regarding the modulation level and the high SNR regime, the curve for GMSM with Gaussian symbols has a slope of 2 bit/s/Hz per 3 dB increase in SNR, i.e., GMSM benefits from a Degree of Freedom (DoF) of 2. The slope of GSM is smaller and will asymptotically become 1 bit/s/Hz per 3 dB increase in SNR. The curves with 4-QAM saturate at $\log_2(M^{N_a}\binom{N_t}{N_a}) = \log_2(4 \cdot 6) \approx 4.6$ bit/s/Hz for GSM and at $\log_2(4^2 \cdot 6) \approx 6.6$ bit/s/Hz for GMSM, respectively. On the other hand, for low SNR values, all systems obtain a similar performance. For GSM with 4-QAM, a significant part of the overall mutual information results from the antenna selection level which partly compensates the loss in the modulation level compared to Gaussian data symbols.

Fig. 4 shows similar results for $N_t = 10$ transmit antennas. However, the antenna selection level contributes a larger share to the overall information $I(\mathcal{X};\mathcal{Y})$. Compared with the case for $N_t = 4$, $I(\mathcal{L};\mathcal{Y})$ increases significantly, particularly for the digital 4-QAM modulation. Up to an SNR of 10 dB, GSM with 4-QAM achieves a performance close to the optimal Gaussian distribution of data symbols. Moreover, the differences between GMSM and GSM become smaller.

Next, we consider the influence of the signal constellation on the mutual information. In [5], [6], [11]–[13], [45], various two-dimensional signal constellations have been investigated for SM with a single active antenna and compared to conventional Phase Shift Keying (PSK) and square Quadrature Amplitude Modulation (QAM) constellations. Particularly, [12], [45] proposed the star-QAM constellations. The star-QAM has two amplitude levels, $M/2$ symbols per level, which were optimized according to the constellation figure of merit.

Fig. 5 presents the mutual information for different signal constellations with $M = 16$ using GMSM. At an SNR of approx. 16 dB, $I(\mathcal{S};\mathcal{Y}|\mathcal{L})$ begins to stagnate due to the finite alphabet size $|\mathbb{S}|$. Up to this point,

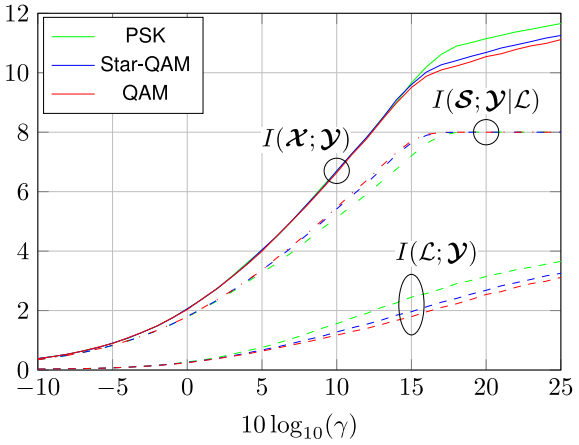

FIGURE 5. Comparison of mutual information of different signal constellations with $M = 16$ using GMSM with $N_t = 10$, $N_a = 2$, $N_r = 2$.

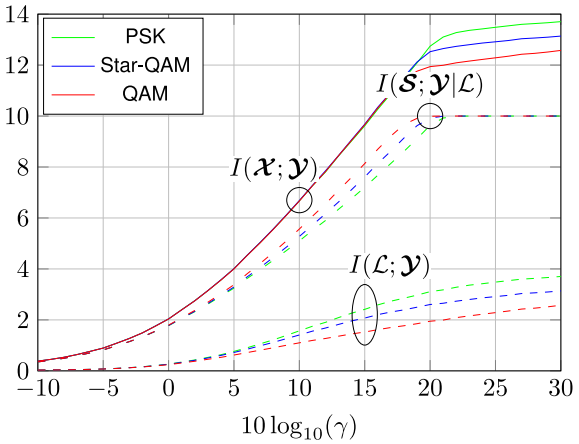

FIGURE 6. Comparison of mutual information of different signal constellations with $M = 32$ using GMSM with $N_t = 10$, $N_a = 2$, $N_r = 2$.

all constellations show the same total mutual information $I(\mathcal{X};\mathcal{Y})$. Similarly, the multidimensional signal constellations from [14], [17] obtained no performance gain compared to the two-dimensional signal constellations. Beyond 16 dB, the PSK constellation leads to a slightly larger mutual information than the star-QAM constellations with two different magnitudes or square QAM with three magnitudes. It seems that a smaller number of symbol magnitudes is beneficial for the antenna detection and increases $I(\mathcal{L};\mathcal{Y})$. Note that the rates $I(\mathcal{L};\mathcal{Y})$ and $I(\mathcal{S};\mathcal{Y}|\mathcal{L})$ both depend on the signal constellation.

The results for 32-valued alphabets in Fig. 6 show greater gains of PSK and star-QAM over square QAM. In general, the Gaussian symbol alphabet leads to the largest overall mutual information. However, finite alphabets imply only very small losses for low SNR values. Interestingly, the antenna selection level benefits from finite alphabets and becomes more important.

IV. BOUNDS ON MUTUAL INFORMATION

As mentioned in the previous section, the entropy of GMMs cannot be easily computed. Closed-form expressions do not exist and numerical integration becomes cumbersome when the number of dimensions increases. Therefore, upper and lower bounds have been derived [32], [44], [46]. We adopt these results to the SM variants discussed in Sections II-B and III-A. Essentially, bounds on the critical term $h(\mathcal{Y}|\mathbf{H})$ will be derived from which approximations of the mutual information terms can be concluded. Throughout the following derivation, only Gaussian distributed data symbols except for Generalized Space Shift Keying (GSSK) are assumed. We start the analysis for a particular channel realization $\mathbf{H}$ and determine the expectation w.r.t. the channel statistics afterwards.

A. LOWER BOUND ON MUTUAL INFORMATION

A lower bound on $h(\mathcal{Y}|\mathbf{H})$ can be obtained by applying Jensen's inequality. As the logarithm is a concave function,

$$\mathrm{E}\{\log_2 \mathcal{Y}\} \leq \log_2(\mathrm{E}\{\mathcal{Y}\})$$

holds. Inserting this inequality into the expression for the differential entropy $h(\mathcal{Y}|\mathbf{H})$, we obtain

$$\begin{aligned} h(\mathcal{Y}|\mathbf{H}) &= -\int_{\mathbb{C}^{N_r}} \frac{1}{N_p}\sum_{l=1}^{N_p} p(\mathbf{y}|\mathbf{H}_l)\log_2 \frac{1}{N_p}\sum_{i=1}^{N_p} p(\mathbf{y}|\mathbf{H}_i)d\mathbf{y} \\ &\geq -\frac{1}{N_p}\sum_{l=1}^{N_p}\log_2\left(\frac{1}{N_p}\sum_{i=1}^{N_p}\int_{\mathbb{C}^{N_r}} p(\mathbf{y}|\mathbf{H}_l)p(\mathbf{y}|\mathbf{H}_i)d\mathbf{y}\right). \end{aligned}$$

The integral over the product of two Gaussian distributions with means μ_l, μ_i, and covariance matrices Σ_l, Σ_i, yields a Gaussian distribution [44]

$$\begin{aligned} z_{l,i} &= \int_{\mathbb{C}^{N_r}} p(\mathbf{y}|\mathbf{H}_l)\cdot p(\mathbf{y}|\mathbf{H}_i)\ d\mathbf{y} \\ &= \frac{e^{-(\mu_l-\mu_i)^{\mathrm{H}}(\Sigma_l+\Sigma_i)^{-1}(\mu_l-\mu_i)}}{\det(\pi(\Sigma_l+\Sigma_i))}. \end{aligned} \tag{20}$$

With this intermediate result, the differential entropy is lower bounded by

$$h(\mathcal{Y}|\mathbf{H}) \geq \log_2 N_p - \frac{1}{N_p}\cdot\sum_{l=1}^{N_p}\log_2\left(\sum_{i=1}^{N_p} z_{l,i}\right).$$

1) LOWER BOUND FOR GSSK

For GSSK, the received vector given in (6) leads to identical covariance matrices $\Sigma_l = \mathbf{I}$ but different means μ_l. With $\Sigma_l + \Sigma_i = 2\mathbf{I}$ and (20), the variables $z_{l,i}$ become

$$z_{l,i} = (2\pi)^{-N_r}\cdot e^{-\frac{1}{2}\rho^2\left\|\tilde{\mathbf{h}}_l-\tilde{\mathbf{h}}_i\right\|^2}$$

leading to lower bounds on the differential entropy

$$\begin{aligned} h(\mathcal{Y}|\mathbf{H}) &\geq \log_2 N_p - \frac{1}{N_p}\sum_{l=1}^{N_p}\log_2\left(\sum_{i=1}^{N_p}\frac{e^{-\frac{1}{2}\rho^2\left\|\tilde{\mathbf{h}}_l-\tilde{\mathbf{h}}_i\right\|^2}}{(2\pi)^{N_r}}\right) \\ &= \log_2(N_p) + N_r\cdot\log_2(2\pi) \\ &\quad - \frac{1}{N_p}\cdot\sum_{l=1}^{N_p}\log_2\left(\sum_{i=1}^{N_p} e^{-\frac{1}{2}\rho^2\left\|\tilde{\mathbf{h}}_l-\tilde{\mathbf{h}}_i\right\|^2}\right) \end{aligned}$$

and the mutual information

$$I(\mathcal{L};\boldsymbol{\mathcal{Y}}|\mathbf{H}) = h(\boldsymbol{\mathcal{Y}}|\mathbf{H}) - h(\boldsymbol{\mathcal{Y}}|\mathcal{L},\mathbf{H}) = h(\boldsymbol{\mathcal{Y}}|\mathbf{H}) - h(\boldsymbol{\mathcal{N}}) \geq \log_2\big(N_p\big) + N_r\big(1-\log_2(e)\big) - \frac{1}{N_p}\cdot\sum_{l=1}^{N_p}\log_2\left(\sum_{i=1}^{N_p} e^{-\frac{1}{2}\rho^2\left\|\tilde{\mathbf{h}}_l-\tilde{\mathbf{h}}_i\right\|^2}\right). \quad (21)$$

2) LOWER BOUND FOR GSM AND GMSM

Compared to GSSK, the received vector $\mathbf{y}$ follows a GMM with zero mean, i.e., $\mu_l = \mathbf{0}\ \forall l$, but varying covariance matrices. Hence, the variables $z_{l,i}$ in (20) become

$$z_{l,i} = \frac{1}{\det(\pi(\Sigma_l+\Sigma_i))}$$

leading to the lower bounds on the differential entropy

$$h(\boldsymbol{\mathcal{Y}}|\mathbf{H}) \geq \log_2\big(N_p\big) + N_r\log_2(\pi) - \frac{1}{N_p}\cdot\sum_{l=1}^{N_p}\log_2\left(\sum_{i=1}^{N_p}\frac{1}{\det(\Sigma_l+\Sigma_i)}\right)$$

and the mutual information

$$I(\mathcal{L};\boldsymbol{\mathcal{Y}}|\mathbf{H}) \geq \log_2\big(N_p\big) - N_r\log_2(e) - \frac{1}{N_p}\cdot\sum_{l=1}^{N_p}\log_2\left(\sum_{i=1}^{N_p}\frac{\det\Sigma_l}{\det(\Sigma_l+\Sigma_i)}\right). \quad (22)$$

Lower bound will be denoted by $I_l(\mathcal{L};\boldsymbol{\mathcal{Y}}|\mathbf{H})$ and $I_l(\boldsymbol{\mathcal{X}};\boldsymbol{\mathcal{Y}}|\mathbf{H})$, respectively.

B. UPPER BOUNDS ON MUTUAL INFORMATION

1) TRIVIAL UPPER BOUND ON MUTUAL INFORMATION

An upper bound on the differential entropy of a Gaussian mixture model is obtained by rewriting the sum in the logarithm and neglecting one non-negative term [44].

$$\begin{aligned} h(\boldsymbol{\mathcal{Y}}|\mathbf{H}) &= -\mathrm{E}_{\boldsymbol{\mathcal{Y}}|\mathbf{H}}\left\{\log_2\left(\frac{1}{N_p}\cdot\sum_{i=1}^{N_p}p(\mathbf{y}|\mathbf{H}_l)\right)\right\} \\ &= -\mathrm{E}_{\boldsymbol{\mathcal{Y}}|\mathbf{H}}\left\{\log_2\left(\frac{1}{N_p}\cdot p(\mathbf{y}|\mathbf{H}_l)\cdot(1+\epsilon_l)\right)\right\} \\ &\quad \text{with}\quad \epsilon_l = \frac{\sum_{i\neq l}^{N_p}p(\mathbf{y}|\mathbf{H}_i)}{p(\mathbf{y}|\mathbf{H}_l)} \\ &= \log_2 N_p - \mathrm{E}_{\boldsymbol{\mathcal{Y}}|\mathbf{H}}\big\{\log_2(p(\mathbf{y}|\mathbf{H}_l)) + \log_2(1+\epsilon_l)\big\} \\ &\leq \log_2 N_p - \mathrm{E}_{\boldsymbol{\mathcal{Y}}|\mathbf{H}}\big\{\log_2(p(\mathbf{y}|\mathbf{H}_l))\big\} \\ &= \log_2 N_p + h(\boldsymbol{\mathcal{Y}}|\mathcal{L},\mathbf{H}) \end{aligned} \quad (23)$$

The inequality in (23) is the consequence from $\log_2(1+\epsilon_l) \geq 0$ holds for $\epsilon_l \geq 0$. Inserting the upper bound on $h(\boldsymbol{\mathcal{Y}}|\mathbf{H})$ into the mutual information yields the trivial upper bound

$$\begin{aligned} I(\mathcal{L};\boldsymbol{\mathcal{Y}}|\mathbf{H}) &\leq \log_2 N_p + h(\boldsymbol{\mathcal{Y}}|\mathcal{L},\mathbf{H}) - h(\boldsymbol{\mathcal{Y}}|\mathcal{L},\mathbf{H}) \\ &= \log_2 N_p. \end{aligned} \quad (24)$$

Obviously, this is a rather loose upper bound as it just considers the entropy of the antenna selection process without the information loss during transmission.

2) REFINED UPPER BOUND ON MUTUAL INFORMATION

The upper bound can be tightened by merging a subset of L mixture components $p(\mathbf{y}|\mathbf{H}_l)$ of the GMM to a new Gaussian distribution $\tilde{f}_L(\mathbf{y})$ with the first two moments being the mean of those of the merged components. The optimization of this approximation requires to check all $\binom{N_p}{L}$ possible subsets for all $1 \leq L \leq N_p$ and to determine the one with the smallest differential entropy because approximating a GMM by a Gaussian distribution increases entropy. In order to keep the complexity moderate, we simplify this optimization and consider only N_p subsets $\{1,\ldots,L\}$ for $1 \leq L \leq N_p$. $L = N_p$ means that the GMM is completely replaced by a single Gaussian distribution. For the merged Gaussian components, the expressions

- $f_L(\mathbf{y}|\mathbf{H}) = \sum_{l=1}^{L} p(\mathbf{y}|\mathbf{H}_l)$ $\rightarrow\ \tilde{f}_L(\mathbf{y}|\mathbf{H}_l) \sim L\cdot\mathcal{CN}(\mathbf{y};\mu_{f_L},\Sigma_{f_L})$
- $\mu_{f_L} = \frac{1}{L}\sum_{l=1}^{L}\mu_l$
- $\Sigma_{f_L} = \frac{1}{L}\sum_{l=1}^{L}\Sigma_l + \mu_l\cdot(\mu_l - \frac{1}{L}\sum_{i=1}^{L}\mu_i)^{\mathrm{H}}$

hold. We obtain a refined upper bound on the differential entropy

$$\begin{aligned} h(\boldsymbol{\mathcal{Y}}|\mathbf{H}) &= -\mathrm{E}_{\boldsymbol{\mathcal{Y}}|\mathbf{H}}\left\{\log_2\left(\frac{1}{N_p}\left[f_L(\mathbf{y}) + \sum_{i=L+1}^{N_p}p(\mathbf{y}|\mathbf{H}_i)\right]\right)\right\} \\ &= -\frac{1}{N_\mathrm{P}}\int_{\mathbb{C}^{N_r}} f_L(\mathbf{y}) \log_2\left(\frac{1}{N_\mathrm{P}}f_L(\mathbf{y})\left[1+\frac{\sum_{i=L+1}^{N_\mathrm{P}}p(\mathbf{y}|\mathbf{H}_i)}{f_L(\mathbf{y})}\right]\right) \\ &\quad + \sum_{l=L+1}^{N_\mathrm{P}} p(\mathbf{y}|\mathbf{H}_l)\log_2\left(\frac{1}{N_\mathrm{P}}p(\mathbf{y}|\mathbf{H}_l)[1+\epsilon_l]\right)d\mathbf{y} \\ &\leq -\frac{1}{N_\mathrm{P}}\int_{\mathbb{C}^{N_r}} f_L(\mathbf{y})\log_2\left(\frac{1}{N_\mathrm{P}}f_L(\mathbf{y})\right)d\mathbf{y} \\ &\quad - \frac{1}{N_\mathrm{P}}\sum_{l=L+1}^{N_\mathrm{P}}\int_{\mathbb{C}^{N_r}} p(\mathbf{y}|\mathbf{H}_l)\log_2\left(\frac{1}{N_\mathrm{P}}p(\mathbf{y}|\mathbf{H}_l)\right)d\mathbf{y} \\ &\leq -\frac{1}{N_\mathrm{P}}\int_{\mathbb{C}^{N_r}} \tilde{f}_L(\mathbf{y})\log_2\left(\frac{1}{N_\mathrm{P}}\tilde{f}_L(\mathbf{y})\right)d\mathbf{y} \\ &\quad -\frac{1}{N_\mathrm{P}}\sum_{l=L+1}^{N_\mathrm{P}}\int_{\mathbb{C}^{N_r}} p(\mathbf{y}|\mathbf{H}_l)\log_2\left(\frac{1}{N_\mathrm{P}}p(\mathbf{y}|\mathbf{H}_l)\right)d\mathbf{y}. \end{aligned}$$

The last inequality holds because the GMM distribution $f_L(\mathbf{y})$ is replaced by a Gaussian distribution $\tilde{f}_L(\mathbf{y})$ with identical covariance matrix. With this approach, a refined upper bound

$$\begin{aligned} I(\mathcal{L};\boldsymbol{\mathcal{Y}}|\mathbf{H}) &\leq I_u(\mathcal{L};\boldsymbol{\mathcal{Y}}|\mathbf{H},L) \\ &= \log_2 N_p - \frac{L}{N_p}\log_2(L) + \frac{L}{N_p}\cdot\log_2\det\big(\Sigma_{f_L}\big) \\ &\quad - \frac{1}{N_p}\cdot\sum_{l=1}^{L}\log_2\det(\Sigma_l) \end{aligned} \quad (25)$$

on the mutual information is obtained. The tightest upper bound is $I_u(\mathcal{L};\boldsymbol{\mathcal{Y}}|\mathbf{H}) = \min_L I_u(\mathcal{L};\boldsymbol{\mathcal{Y}}|\mathbf{H},L)$. Numerical investigations reveal that in almost all cases $L = N_p$ is the

best choice, i.e., the distributions of all antenna patterns are merged into a single Gaussian distribution. As the mutual information $I_u(\mathcal{L};\mathcal{Y}|\mathbf{H})$ is always non-negative and furthermore upper bounded by $\log_2 N_p$, we can further tighten the bound by applying

$$I_u(\mathcal{L};\mathcal{Y}|\mathbf{H}) \leftarrow \min\{\max[0, I_u(\mathcal{L};\mathcal{Y}|\mathbf{H})], \log_2 N_p\} \quad (26)$$

before averaging w.r.t. the channel statistics.

3) VARIATIONAL UPPER BOUND ON MUTUAL INFORMATION

Based on the derivations in [46], we derive now a variational upper bound. We introduce variational parameters $\phi_{l'|l}$ which have to fulfill the condition $\sum_{l'=1}^{N_\mathrm{P}} \phi_{l'|l} = 1 \ \forall l$. The differential entropy can be rewritten to

$$\begin{aligned} h(\mathcal{Y} \mid \mathbf{H}) &= -\frac{1}{N_p}\sum_{l=1}^{N_\mathrm{P}} \int p(\mathbf{y}|\mathbf{H}_l) \\ &\quad \times \log_2 \sum_{l'=1}^{N_\mathrm{P}} \phi_{l'|l} \cdot \frac{p(\mathbf{y}|\mathbf{H}_{l'})}{N_p \phi_{l'|l}} d\mathbf{y} \\ &\leq -\frac{1}{N_p}\sum_{l=1}^{N_\mathrm{P}}\sum_{l'=1}^{N_\mathrm{P}} \phi_{l'|l} \\ &\quad \times \int p(\mathbf{y}|\mathbf{H}_l) \cdot \log_2\left(\frac{p(\mathbf{y}|\mathbf{H}_{l'})}{N_p \phi_{l'|l}}\right) d\mathbf{y} \\ &= \log_2 N_\mathrm{P} + \frac{1}{N_\mathrm{P}}\sum_{l=1}^{N_\mathrm{P}} \\ &\quad \left[\sum_{l'=1}^{N_\mathrm{P}} \phi_{l'|l} \cdot \left(h_{l,l'}(\mathcal{Y}) + \log_2 \phi_{l'|l}\right)\right]. \end{aligned} \quad (27)$$

The inequality holds due to Jensen's inequality because the sum over l' with coefficients $\phi_{l'|l}$ can be interpreted as an expectation being moved outside of the logarithm. The function

$$h_{l,l'}(\mathcal{Y}) = -\int p(\mathbf{y}|\mathbf{H}_l) \cdot \log_2 p(\mathbf{y}|\mathbf{H}_{l'}) d\mathbf{y}$$

in (27) denotes the cross entropy between the conditional distributions $p(\mathbf{y}|\mathbf{H}_l)$ and $p(\mathbf{y}|\mathbf{H}_{l'})$. For two Gaussian distributions, it has the general form

$$\begin{aligned} h_{l,l'}(\mathcal{Y}) &= \log_2 \det(\pi \Sigma_{l'}) + \log_2 e^{\mathrm{tr}\left\{\Sigma_{l'}^{-1}\Sigma_l\right\}} \\ &\quad + \log_2 e^{(\mu_l - \mu_{l'})^\mathrm{H} \Sigma_{l'}^{-1} (\mu_l - \mu_{l'})} \end{aligned} \quad (28)$$

and will be specified for particular SM versions below.

The upper bound can be tightened by choosing the variational parameters $\phi_{l'|l}$ such that they minimize the term in the square brackets of (27) for each outer index l. Equating the partial derivative w.r.t. $\phi_{l'|l}$ to zero and normalizing the result to $\sum_{l'} \phi_{l'|l} = 1$ yields

$$\phi_{l'|l} = \frac{2^{-h_{l,l'}(\mathcal{Y})}}{\sum_{l''} 2^{-h_{l,l''}(\mathcal{Y})}}. \quad (29)$$

Inserting the variational coefficients in (29) into (27) leads to the solution

$$h(\mathcal{Y} \mid \mathbf{H}) \leq \log_2 N_\mathrm{P} - \frac{1}{N_\mathrm{P}}\sum_{l=1}^{N_\mathrm{P}} \log_2 \sum_{l'} 2^{-h_{l,l'}(\mathcal{Y})}. \quad (30)$$

Variational Upper Bound for GSSK: For GSSK, $\Sigma_l = \mathbf{I}$ holds for all l and the cross-entropy in (28) becomes

$$h_{l,l'}(\mathcal{Y}) = N_\mathrm{R} \log_2(e\pi) + \log_2 e^{\rho^2 \|\tilde{\mathbf{h}}_l - \tilde{\mathbf{h}}_{l'}\|^2}. \quad (31)$$

Inserting (31) into (30) leads to the upper bounds

$$\begin{aligned} h(\mathcal{Y} \mid \mathbf{H}) &\leq \log_2 N_\mathrm{P} + N_\mathrm{R} \log_2(e\pi) \\ &\quad - \frac{1}{N_\mathrm{P}}\sum_{l=1}^{N_\mathrm{P}} \log_2\left(\sum_{l'} e^{-\rho^2 \|\tilde{\mathbf{h}}_l - \tilde{\mathbf{h}}_{l'}\|^2}\right) \end{aligned} \quad (32)$$

on the differential entropy and

$$I_v(\mathcal{L};\mathcal{Y} \mid \mathbf{H}) = \log_2 N_\mathrm{P} - \frac{1}{N_\mathrm{P}}\sum_{l=1}^{N_\mathrm{P}} \log_2\left(\sum_{l'} e^{-\rho^2 \|\tilde{\mathbf{h}}_l - \tilde{\mathbf{h}}_{l'}\|^2}\right) \quad (33)$$

on the mutual information of the antenna selection branch.

Variational Upper Bound for GSM and GMSM: For GSM and GMSM, the received signals have zero mean and the covariance matrices for a given antenna pattern l become $\Sigma_l = \Sigma_{\mathcal{Y}|\mathbf{H}_l} = \rho^2 \tilde{\mathbf{h}}_l \tilde{\mathbf{h}}_l^\mathrm{H} + \mathbf{I}$ and $\Sigma_l = \rho^2 \mathbf{H}_l \mathbf{H}_l^\mathrm{H} + \mathbf{I}$, respectively. As a function of Σ_l, the cross-entropy in (28) and the variational upper bounds have identical forms. The cross-entropy becomes

$$h_{l,l'}(\mathcal{Y}) = \log_2 \det(\pi \Sigma_{l'}) + \log_2 e^{\mathrm{tr}\left\{\Sigma_{l'}^{-1}\Sigma_l\right\}}. \quad (34)$$

Inserting (34) into (30) leads to the upper bound on the differential entropy

$$\begin{aligned} h(\mathcal{Y} \mid \mathbf{H}) &\leq \log_2 N_\mathrm{P} + N_\mathrm{R} \log_2(\pi) \\ &\quad - \frac{1}{N_\mathrm{P}}\sum_{l=1}^{N_\mathrm{P}} \log_2\left(\sum_{l'} \frac{e^{-\mathrm{tr}\left\{\Sigma_{l'}^{-1}\Sigma_l\right\}}}{\det \Sigma_{l'}}\right). \end{aligned} \quad (35)$$

With this intermediate result, the variational upper bound on the mutual information of the antenna selection branch becomes

$$\begin{aligned} I_v(\mathcal{L};\mathcal{Y} \mid \mathbf{H}) &= \log_2 N_\mathrm{P} - N_\mathrm{R} \log_2(e) \\ &\quad - \frac{1}{N_\mathrm{P}}\sum_{l=1}^{N_\mathrm{P}} \log_2\left(\sum_{l'} \frac{\det \Sigma_l}{\det \Sigma_{l'}} e^{-\mathrm{tr}\left\{\Sigma_{l'}^{-1}\Sigma_l\right\}}\right). \end{aligned} \quad (36)$$

C. ERGODIC UPPER BOUND ON MUTUAL INFORMATION

In [47], the capacity of SM systems was derived. The authors argue that the antenna selection process essentially selects code words defined by the channel matrix $\mathbf{H}$. Therefore, the channel coefficients can be interpreted as data symbols and the expectation w.r.t. the channel statistics becomes an expectation w.r.t. the source statistics. In the case of Rayleigh

fading, the channel coefficients are complex Gaussian distributed and achieve the channel capacity. For general channel statistics, the approach in [47] requires a shaping of the input distribution and knowledge of the channel statistics at the transmitter in order to achieve capacity.

Mathematically, this interpretation moves the expectation w.r.t. the channel statistics in (17) into the logarithm. Due to Jensen's inequality, the capacity formula derived in [47]

$$C_{\text{AWGN}} = N_r \cdot \log_2(1 + \text{SNR}) \tag{37}$$

is an upper bound on the achievable mutual information. It represents the capacity of N_r parallel AWGN channels having identical SNRs. Particularly, it seems to be independent of N_t and N_a.

We follow a different interpretation to obtain the result in (37) and consider a general MIMO system $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$ with Gaussian distributed transmit vector $\mathbf{x}$. The MIMO system has the instantaneous channel dependent capacity

$$C(\mathbf{H}) = \log_2 \det(\rho^2 \mathbf{H}\mathbf{H}^{\text{H}} + \mathbf{I}) \tag{38}$$

and the ergodic capacity

$$\bar{C} = \text{E}_{\mathbf{H}}\{C(\mathbf{H})\} \tag{39}$$

For fixed N_r, N_a and $\sigma_{\mathcal{H}}^2 = \frac{1}{N_t}$ being the variance of a single channel coefficient, the capacity becomes in the asymptotic limit $N_t \to \infty$

$$\begin{aligned} C(\mathbf{H}) &\xrightarrow[N_t \to \infty]{} \log_2 \det(\rho^2 \mathbf{I} + \mathbf{I}) = \log_2 \prod_{i=1}^{N_r} \left(1 + \rho^2\right) \\ &= N_r \cdot \log_2\left(1 + \rho^2\right) = C_{\text{AWGN}}. \end{aligned} \tag{40}$$

Hence, the capacity asymptotically approaches the result in (37). In order to obtain the same average SNR as defined in (4), the total transmit power $\rho^2 N_a$ has to be distributed onto the N_r parallel AWGN channels, i.e., $\text{SNR} = \rho^2 N_a / N_r$ holds in (37).

V. NUMERICAL RESULTS

While the results presented in Section III-B focused on small systems with at most $N_t = 10$ transmit and up to $N_r = 2$ receive antennas, this section now considers larger systems with much more antennas. This is accomplished by using upper and lower bounds on the mutual information derived in the previous section. The channel matrices have been generated using i.i.d. complex Gaussian coefficients with $\sigma_{\mathcal{H}}^2 = 1$. For small system dimensions, 1000 channels have been sampled to obtain accurate average mutual information results. For the system with $N_t = 100$ and $N_a = 2$, only 10 channels have been used due to the high computational costs. For GMSM and GSM, the data symbols are complex Gaussian distributed with variance $\sigma_{\mathcal{S}}^2 = 1$.

First, we analyze the tightness of the bounds by comparing them with results from numerical integration for systems with a small number of receive antennas. Second, we analyze the performance of different SM variants and consider the influence of data symbols on the mutual information of the antenna selection. Finally, we compare the SM performance with the ergodic MIMO capacity in (39) and the asymptotic capacity in (40), particularly for large systems.

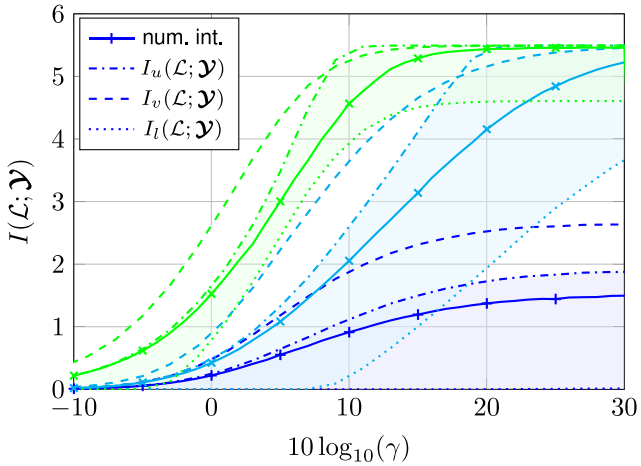

FIGURE 7. Level-specific mutual information $I(\mathcal{L};\mathcal{Y})$ of SM systems with $N_t = 10$, $N_a = 2$, $N_r = 2$ and Gaussian transmit symbols (blue: GMSM, cyan: GSM, green: GSSK).

A. ACCURACY OF BOUNDS

We start the discussion by investigating the accuracy of the bounds on $I(\mathcal{L};\mathcal{Y})$ derived in Section IV. As the tightness of bounds is evaluated by comparison with results from numerical integration, we look at system dimensions for which numerical integration is still feasible.

Fig. 7 illustrates results for a system with $N_t = 10$ transmit, $N_r = 2$ receive and $N_a = 2$ active antennas. For the GMSM system presented by blue curves, the lower bound is close to zero regardless of the SNR and not useful. It is generally very loose and provides results significantly larger than zero only for $N_r > N_a$. The refined upper bound $I_u(\mathcal{L};\mathcal{Y})$ is rather close to the true mutual information $I(\mathcal{L};\mathcal{Y})$ but exhibits a constant gap even for large SNR. The variational bound $I_v(\mathcal{L};\mathcal{Y})$ is less tight than the refined upper bound. Also for GSM depicted by cyan curves, upper and lower bounds are not accurately predicting the true mutual information. An intersection of refined and variational upper bounds can be observed at an SNR of 18 dB. Beyond this point, the variational upper bound becomes tighter and it might be useful to take the minimum of the two upper bounds for each SNR. In the high SNR regime, the original version of the refined upper bound (not shown in Fig. 7) defined in (25) becomes very loose and it grows without bound. It has been tightened in (26) by limiting it to the theoretical maximum of $\log_2 N_p$ before averaging over the channel statistics. For GSSK illustrated by green curves, upper and lower bound become tighter but still exhibit a gap of several dB. The intersection of refined and variational upper bounds can also be observed, but the advantage of $I_v(\mathcal{L};\mathcal{Y})$ at high SNR is rather small. For the considered scenario with $N_r = 2$ receive antennas, the maximum mutual information $\log_2 N_p$ =

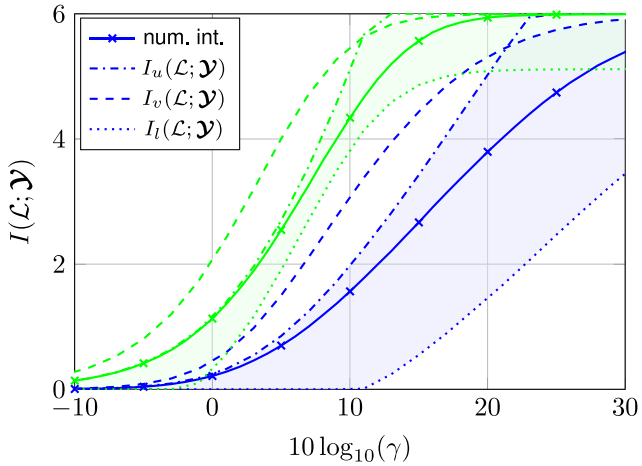

FIGURE 8. Accuracy of upper and lower bounds for GSM system with $N_t = 64$, $N_a = 1$, $N_r = 2$ (blue: GSM, green: GSSK).

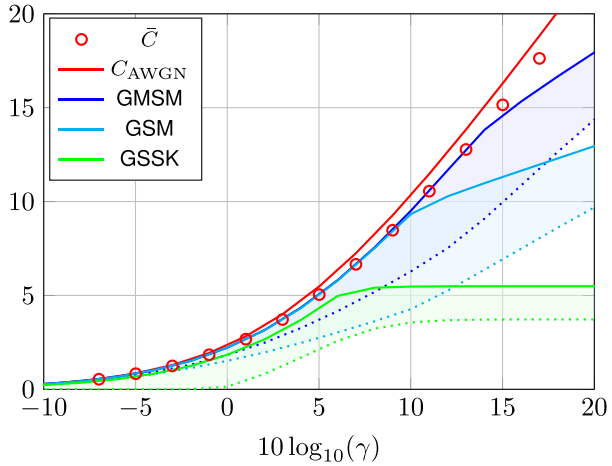

FIGURE 9. Channel capacities and upper and lower bounds on mutual information for $N_t = 10$, $N_a = 2$, $N_r = 4$ (solid: $\min(I_u(\mathcal{X};\mathcal{Y}), I_v(\mathcal{X};\mathcal{Y}))$, dotted: $I_l(\mathcal{X};\mathcal{Y})$).

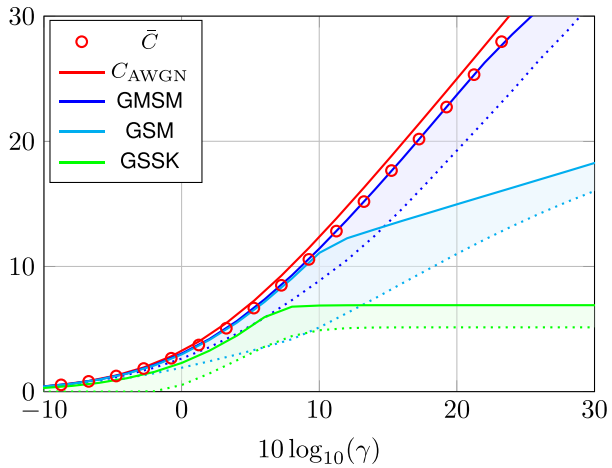

FIGURE 10. Channel capacities and upper and lower bounds on mutual information for $N_t = 10$, $N_a = 3$, $N_r = 4$ (solid: $\min(I_u(\mathcal{X};\mathcal{Y}), I_v(\mathcal{X};\mathcal{Y}))$, dotted: $I_l(\mathcal{X};\mathcal{Y})$).

5.49 bit/s/Hz for the antenna selection level can only be reached with GSM and GSSK while GMSM saturates at 1.5 bit/s/Hz.

As an example of a larger SM system, we consider the results in Fig. 8 for $N_t = 64$, $N_a = 1$, $N_r = 2$, and the SM variants GSM and GSSK. Please not that GMSM and GSM are identical for $N_a = 1$. Principally, the same observations can be made. At low SNR, the refined upper bound is very close to $I(\mathcal{L};\mathcal{Y})$ for both, GSM and GSSK. At medium SNRs, a significant gap occurs before the upper bounds become tight in the saturation area at high SNR. The variational upper bound again outperforms the refined upper bound at high SNR.

Astonishingly, upper and lower bounds are much tighter for GSSK than for GSM or GMSM in both scenarios. This is illustrated in Figs. 7 and 8 by the colored areas. We could not find an obvious reason for this observation. According to Section II-B, the statistical properties of GSSK on the one hand and GSM, as well as GMSM on the other hand, are quite different. While the GMM components of GSSK distinguish w.r.t. their means, the GMM components of GSM and GMSM have identical means but different covariance matrices. This might be the reason why the bounds have different accuracies.

B. COMPARISON OF GMSM, GSM, AND GSSK

As discussed in Section III-B, the level-specific mutual information depends on the SM variant and the signal constellation used for IQ modulation. Next, we compare GMSM, GSM, and GSSK for systems with more receive antennas. For these systems, accurate numerical integration becomes challenging due to the huge amount of required RAM for appropriately sampling the signal space.

We consider a scenario with $N_t = 10$ transmit and $N_r = 4$ receive antennas but different N_a. Fig. 9 and Fig. 10 illustrate upper bounds on the mutual information as well as the channel capacities $\bar{C}$ and C_{AWGN}. It can be seen that GSSK performs close to GSM and GMSM until $I(\mathcal{L};\mathcal{Y})$ saturates. This is remarkable as no IQ symbols are transmitted at all. A comparison with the channel capacity $\bar{C}$ from (39) demonstrate the high performance of SM at least in the SNR region where $I(\mathcal{L};\mathcal{Y})$ not saturates. The upper bounds of GMSM and GSM systems even meet the ergodic capacity $\bar{C}$ which is itself close to the asymptotic capacity C_{AWGN}.

C. SM PERFORMANCE FOR MASSIVE MIMO SYSTEMS

Massive MIMO systems are usually considered in multi-user environments where the benefit comes from the channel hardening effect which allows to keep the signal processing very simple. However, we consider only point-to-point SM communication. For the single-user point-to-point communication, the comparison with the ergodic MIMO $\bar{C}$ is provided in the figures.

For large arrays with many possible antenna patterns $I(\mathcal{L};\mathcal{Y})$ contributes a larger share to the overall mutual information, especially for small N_a. In these regimes, SM becomes attractive as using only a few active antennas from

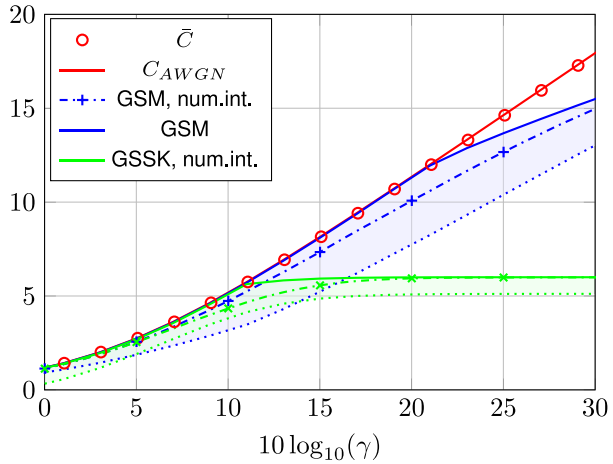

FIGURE 11. Channel capacities and upper and lower bounds on mutual information for $N_t = 64$, $N_a = 1$, $N_r = 2$ (solid: $\min(I_u(\mathcal{X};\mathcal{Y}), I_v(\mathcal{X};\mathcal{Y}))$, dotted: $I_l(\mathcal{X};\mathcal{Y})$).

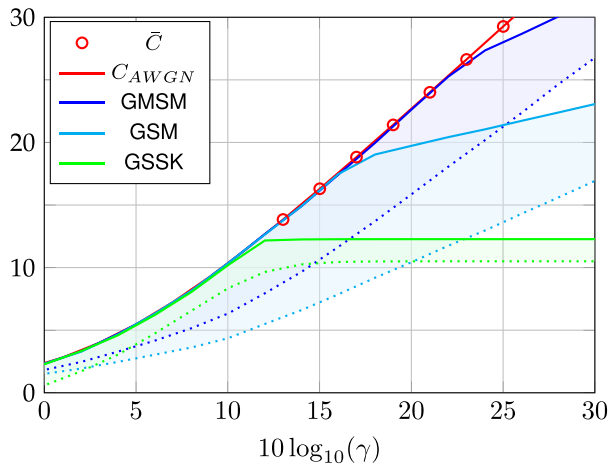

FIGURE 12. Channel capacities and upper and lower bounds on mutual information for $N_t = 100$, $N_a = 2$, $N_r = 4$ (solid: $\min(I_u(\mathcal{X};\mathcal{Y}), I_v(\mathcal{X};\mathcal{Y}))$, dotted: $I_l(\mathcal{X};\mathcal{Y})$).

a large set of available antennas saves hardware costs at the transmitter (only a few RF frontends are required).

Figs. 11 and 12 present results for systems with $N_t = 64$, $N_a = 1$, $N_r = 2$ and $N_t = 100$, $N_a = 2$, $N_r = 4$, respectively. First of all, the upper and lower bounds of the GSSK system are much tighter than those of GMSM and GSM. This allows a rather accurate prediction of the true mutual information for GSSK. For GMSM and particularly for GSM, there are still big gaps between upper and lower bounds and a significant uncertainty remains.

Second, with a growing number of transmit antennas, the upper bounds of all SM variants approach the capacity limits up to the saturation SNR of the antenna selection level. For $N_t = 64$, $I_u(\boldsymbol{\mathcal{X}};\boldsymbol{\mathcal{Y}})$ of GSSK still has a small gap to $\bar{C}$, which vanishes for $N_t = 100$. Here, all upper bounds are very close to the capacity curves $\bar{C}$ and C_{AWGN}, which are nearly identical as argued in Section IV-C. The mutual information of GSSK saturates at $\log_2(N_p) \approx 12.3$ bit/s/Hz at about 12 dB. Up to this SNR value, GSSK performs within a gap of at most 2 dB to the capacity limit. For GMSM and GSM, $I_u(\mathcal{L};\boldsymbol{\mathcal{Y}})$ saturates at about 17 dB and 24 dB, respectively. Notably, $I(\mathcal{L};\boldsymbol{\mathcal{Y}})$ provides a large portion of the overall mutual information. In the high SNR regime, GMSM and GSM clearly outperform GSSK. The superior performance of GMSM comes at the expense of a higher decoding complexity because after detecting the active antennas, the data symbols need to be separated at the receiver. Therefore, GSM transmitting the same IQ data symbol over all active antennas might be a good trade-off between performance and complexity.

VI. CONCLUSION

In this work, we have investigated the capacity of SM systems. We have considered a capacity-achieving two-level coding approach for SM. With this coding scheme, a multi-stage detection and decoding approach can achieve the total channel capacity, provided that the code rates are chosen according to the capacity rule.

The presented numerical results show that SM becomes attractive for large antenna arrays with many possible antenna patterns, where the mutual information for the antenna selection level provides a large share of the overall mutual information. For such systems, using only a few active antennas from a large set of available antennas saves hardware costs at the transmitter and only a small number of receive antennas is required.

In general, GMSM with Gaussian input symbols leads to the best possible performance. On the other hand, the large peak-to-average power ratio of Gaussian input symbols affects the antenna selection layer because small magnitudes make antenna detection difficult. However, for small and medium SNRs, the Gaussian input symbols are not required, because finite symbol alphabets can achieve a comparable performance. In this regime, even GSSK approaches the capacity limit until the saturation starts due to the finite number of antenna patterns. Finally, the partitioning of the rates for the antenna selection level and the IQ modulation level are clearly influenced by the signal constellation.

The accuracy of the bounds requires further work. Generally, upper and lower bounds do not allow a tight prediction of the achievable rates. Particularly, the rate allocation for the multi-level coding approach requires more accurate predictions.

REFERENCES

[1] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996. [Online]. Available: https://ieeexplore.ieee.org/document/6770094

[2] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[3] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[4] A. Younis, N. Serafimovski, R. Mesleh, and H. Haas, "Generalised spatial modulation," in *Proc. 44th Asilomar Conf. Signals Syst. Comput.*, Nov. 2010, pp. 1498–1502.

[5] K. Ntontin, M. D. Renzo, A. Perez-Neira, and C. Verikoukis, "Performance analysis of multistream spatial modulation with maximum-likelihood detection," in *Proc. IEEE Global Conf. Telecommun. (Globecom)*, Dec. 2013, pp. 1590–1594.

[6] C.-C. Cheng, H. Sari, S. Sezginer, and Y. Su, "Enhanced spatial modulation with multiple signal constellations," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2237–2248, Jun. 2015.

[7] M. D. Renzo, H. Haas, and P. M. Grant, "Spatial modulation for multiple-antenna wireless systems: A survey," *IEEE Commun. Mag.*, vol. 49, no. 12, pp. 182–191, Dec. 2011.

[8] 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.

[9] P. Yang, M. D. Renzo, Y. Xiao, S. Li, and L. Hanzo, "Design guidelines for spatial modulation," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 6–26, 1st Quart., 2015.

[10] 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.

[11] C. Li, Y. Huang, M. D. Renzo, J. Wang, and Y. Cheng, "Low-complexity ML detection for spatial modulation MIMO with APSK constellation," *IEEE Trans. Veh. Technol.*, vol. 64, no. 9, pp. 4315–4321, Sep. 2015.

[12] M. Maleki, H. R. Bahrami, and A. Alizadeh, "Constellation design for spatial modulation," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 2739–2743.

[13] J. Freudenberger and S. Shavgulidze, "Signal constellations based on Eisenstein integers for generalized spatial modulation," *IEEE Commun. Lett.*, vol. 21, no. 3, pp. 556–559, Mar. 2017.

[14] J. Freudenberger, D. Rohweder, and S. Shavgulidze, "Generalized multistream spatial modulation with signal constellations based on Hurwitz integers and low-complexity detection," *IEEE Wireless Commun. Lett.*, vol. 7, no. 3, pp. 412–415, Jun. 2018.

[15] N. Ugrelidze, S. Shavgulidze, and M. Sordia, "New generalized multistream spatial modulation for wireless communications," in *Proc. 11th Wireless Days (WD)*, Apr. 2019, pp. 1–7.

[16] F. Huang and Y. Zhan, "Design of spatial constellation for spatial modulation," *IEEE Wireless Commun. Lett.*, vol. 9, no. 7, pp. 1097–1100, Jul. 2020.

[17] D. Rohweder, S. Stern, R. F. H. Fischer, S. Shavgulidze, and J. Freudenberger, "Low-complexity detection for multi-dimensional spatial modulation schemes," in *Proc. 24th Int. ITG Workshop Smart Antennas (WSA)*, Feb. 2020, pp. 1–6.

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

[19] J. Wang, S. Jia, and J. Song, "Signal vector based detection scheme for spatial modulation," *IEEE Commun. Lett.*, vol. 16, no. 1, pp. 19–21, Jan. 2012.

[20] N. R. Naidoo, H. J. Xu, and T. A. M. Quazi, "Spatial modulation: Optimal detector asymptotic performance and multiple-stage detection," *IET Commun.*, vol. 5, no. 10, pp. 1368–1376, Jul. 2011.

[21] J. Wang, S. Jia, and J. Song, "Generalised spatial modulation system with multiple active transmit antennas and low complexity detection scheme," *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1605–1615, Apr. 2012.

[22] Y. Xiao, Z. Yang, L. Dan, P. Yang, L. Yin, and W. Xiang, "Low-complexity signal detection for generalized spatial modulation," *IEEE Commun. Lett.*, vol. 18, no. 3, pp. 403–406, Mar. 2014.

[23] 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.

[24] E. Sourour, "Codebook-based precoding for generalized spatial modulation with diversity," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, p. 229, Sep. 2019. [Online]. Available: https://jwcn-eurasipjournals.springeropen.com/articles/10.1186/s13638-019-1545-7

[25] D. Bailon, J. Freudenberger, and V. Kühn, "Constant-weight convolutional codes for index modulation," in *Proc. 25th Int. Workshop Smart Antennas*, Nice, France, Nov. 2021, pp. 1–6.

[26] A. Sumathi, S. K. Mohideen, and A. Anitha, "Performance analysis of space time block coded spatial modulation," in *Proc. Int. Conf. Softw. Eng. Mobile Appl. Model. Develop. (ICSEMA)*, 2012, pp. 1–7.

[27] H. Tu, C. Lee, and I. Lai, "Low-complexity maximum likelihood (ML) decoder for space-time block coded spatial permutation modulation (STBC-SPM)," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst. (ISPACS)*, 2019, pp. 1–2.

[28] I. Lai *et al.*, "Spatial permutation modulation for multiple-input multiple-output (MIMO) systems," *IEEE Access*, vol. 7, pp. 68206–68218, 2019.

[29] R. Mesleh, M. D. Renzo, H. Haas, and P. M. Grant, "Trellis coded spatial modulation," *IEEE Trans. Wireless Commun.*, vol. 9, no. 7, pp. 2349–2361, Jul. 2010.

[30] E. Basar, U. Aygolu, E. Panayirci, and H. V. Poor, "New trellis code design for spatial modulation," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2670–2680, Aug. 2011.

[31] P. Henarejos and A. I. Pérez-Neira, "Capacity analysis of index modulations over spatial, polarization, and frequency dimensions," *IEEE Trans. Commun.*, vol. 65, no. 12, pp. 5280–5292, Dec. 2017.

[32] A. A. I. Ibrahim, T. Kim, and D. J. Love, "On the achievable rate of generalized spatial modulation using multiplexing under a Gaussian mixture model," *IEEE Trans. Commun.*, vol. 64, no. 4, pp. 1588–1599, Apr. 2016.

[33] A. Younis, N. Abuzgaia, R. Mesleh, and H. Haas, "Quadrature spatial modulation for 5G outdoor millimeter–wave communications: Capacity analysis," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2882–2890, May 2017.

[34] T. L. Narasimhan and A. Chockalingam, "On the capacity and performance of generalized spatial modulation," *IEEE Commun. Lett.*, vol. 20, no. 2, pp. 252–255, Feb. 2016.

[35] H. Yonghong, W. Pichao, W. Xiang, Z. Xiaoming, and H. Chunping, "Ergodic capacity analysis of spatially modulated systems," *China Commun.*, vol. 10, no. 7, pp. 118–125, Jul. 2013.

[36] Z. An, J. Wang, J. Wang, S. Huang, and J. Song, "Mutual information analysis on spatial modulation multiple antenna system," *IEEE Trans. Commun.*, vol. 63, no. 3, pp. 826–843, Mar. 2015.

[37] X. Wang, J. Wang, L. He, and J. Song, "Spectral efficiency analysis for downlink NOMA aided spatial modulation with finite alphabet inputs," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 10562–10566, Nov. 2017.

[38] Z. An, J. Wang, J. Wang, and J. Song, "Mutual information and error probability analysis on generalized spatial modulation system," *IEEE Trans. Commun.*, vol. 65, no. 3, pp. 1044–1060, Mar. 2017.

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

[40] W. Wang and W. Zhang, "Huffman coding-based adaptive spatial modulation," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5090–5101, Aug. 2017.

[41] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1361–1391, Jul. 1999.

[42] H. Imai and S. Hirakawa, "A new multilevel coding method using error-correcting codes," *IEEE Trans. Inf. Theory*, vol. 23, no. 3, pp. 371–377, May 1977.

[43] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley, 2006.

[44] M. F. Huber, T. Bailey, H. Durrant-Whyte, and U. D. Hanebeck, "On entropy approximation for Gaussian mixture random vectors," in *Proc. IEEE Int. Conf. Multisensor Fusion Integr. Intell. Syst.*, Aug. 2008, pp. 181–188.

[45] P. Yang, Y. Xiao, B. Zhang, S. Li, M. El-Hajjar, and L. Hanzo, "Star-QAM signaling constellations for spatial modulation," *IEEE Trans. Veh. Technol.*, vol. 63, no. 8, pp. 3741–3749, Oct. 2014.

[46] J. R. Hershey and P. Olsen, "Approximating the Kullback Leibler divergence between Gaussian mixture models," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2007, pp. 317–320.

[47] A. Younis and R. Mesleh, "Information-theoretic treatment of space modulation MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 6960–6969, Aug. 2018.