The Achievable Rate Analysis of Generalized Quadrature Spatial Modulation and a Pair of Low-Complexity Detectors

Jiancheng An, Chao Xu, *Senior Member, IEEE*, Yusha Liu, Lu Gan, and Lajos Hanzo, *Life Fellow, IEEE*

***Abstract*—Generalized quadrature spatial modulation (GQSM) seamlessly amalgamates the generalized spatial modulation (GSM), quadrature spatial modulation (QSM) and vertical Bell Laboratories layered space-time (V-BLAST) techniques. In contrast to traditional multiple-input multiple-output (MIMO) schemes transmitting information bits only through the constellation defined on the complex plane, GQSM transmits additional information bits implicitly by selecting the transmit antenna (TA) activation pattern. The philosophy of QSM is that of separating the indices of the real and the imaginary parts of the transmit symbols, which increases the attainable throughput. In this paper, we analyze the inherent benefits of GQSM in terms of throughput. More specifically, we first derive the achievable rate expression of the recent GQSM scheme and unveil the condition for GQSM to approach the maximum rate. Furthermore, we compare the rate of the GQSM scheme to that of other transmission schemes and reveal the conditions for the GQSM's maximum throughput to exceed that of the benchmark schemes. Simulation results show that when the number of TAs is 40 and quadrature phase shift keying (QPSK) is adopted, the rate of the GQSM scheme may reach 150% of that of V-BLAST. Additionally, a pair of low-complexity detectors are conceived-one based on ordered successive interference cancellation (OSIC) and another one on the orthogonal matching pursuit (OMP) algorithm.**

***Index Terms*—Spatial modulation (SM), generalized spatial modulation (GSM), quadrature spatial modulation (QSM), generalized quadrature spatial modulation (GQSM), multiple-input multiple-output (MIMO), spatial multiplexing.**

I. Introduction

Multi-antenna systems constitute a key technology for next-generation wireless systems due to their capability substantially outperforming single-antenna systems [1], [2]. In traditional multi-antenna systems, a large number of radio-frequency (RF) chains are required both at the transmitter and receiver, which leads to excessive cost and signal processing complexity [3]. Therefore, how to design a multi-antenna system for striking a better trade-off between the throughput and the transceiver complexity has attracted considerable attention [4].

Spatial modulation (SM), which utilizes the index of the active transmit antenna (TA) as an extra dimension for implicitly transmitting extra information, is a low-cost single-RF multi-input multi-output (MIMO) transmission structure [5], [6]. More specifically, in SM, at any given time, only one out of N_t TAs is activated, while the other TAs remain inactive. Due to its advantages, SM has been extensively studied in [7]–[16]. However, activating a single RF chain substantially limits the transmission rate of the SM scheme. In order to increase the throughput of the SM scheme, multiple-TA activation based schemes have been proposed in [17], [18] as generalized SM (GSM). Explicitly, in GSM, N_α out of N_t TAs are activated to transmit the same phase-amplitude modulated symbol, and the index of the TA activation pattern is employed for conveying extra information bits. Furthermore, the generalized concept of index modulation (IM), has attracted extensive interest in both academia and industry [19], [20]. Recently, the coded IM philosophy was conceived for improving the spectrum- and energy-efficiency [21]–[23].

In order to increase the transmission rate further, combining GSM and the popular vertical Bell Laboratories layered space-time (V-BLAST) technique was proposed in [32], where the symbol transmitted by each activated TA was different, thus significantly improving the transmission rate of its GSM component. However, in large-scale MIMO systems, the number of information bits carried by the TA index is enormous, which makes the mapping of the bit to the TA activation pattern challenging. In order to address the mapping issue of large-scale SM-MIMO systems, a grouping-based GSM-VBLAST scheme was proposed in [33] for striking a flexible trade-off between the overall throughput and detection complexity, where the N_t TAs were divided into N_α groups and SM was applied within each group. Additionally, the combination of GSM and space-time block coding (STBC) to obtain both diversity as well as the throughput improvements was proposed in [34]–[36].

As a variant of SM, quadrature SM (QSM) proposed in [37] has also become a popular MIMO technology that exploits both the in-phase and quadrature-phase components for improving the overall throughput of conventional SM. In the QSM scheme, the indices of the real and imaginary parts are different, resulting in higher transmission rates than SM.

This work was supported by the China Scholarship Council. L. Hanzo would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council projects EP/P034284/1 and EP/P003990/1 (COALESCE) as well as of the European Research Council's Advanced Fellow Grant QuantCom (Grant No. 789028).

J. An, Y. Liu and L. Gan are with the School of Information and Communication Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu, Sichuan 611731, China. L. Gan is also with the Yibin Institute of UESTC, Yibin, Sichuan 644000, China. (E-mail: jiancheng_an@163.com; Yusha.liu@uestc.edu.cn; ganlu@uestc.edu.cn).

C. Xu and L. Hanzo are with the School of Electronics and Computer Science, University of Southampton, SO17 1BJ, UK. (E-mail: cx1g08@soton.ac.uk; lh@ecs.soton.ac.uk).

TABLE I
THE CONTRIBUTION OF OUR WORK ON GQSM COMPARED WITH EXISTING RESEARCH ON GQSM.

Contributions	$\star$	**[24]**	**[25]**	**[26]**	**[27]**	**[28]**	**[29]**	**[30]**	**[31]**
No TA grouping*	$\surd$	$\times$	$\surd$	$\surd$	$\times$	$\surd$	$\surd$	$\surd$	$\surd$
Optimum RF-chain configuration	$\surd$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
Multiple streams	$\surd$	$\surd$	$\surd$	$\times$	$\surd$	$\times$	$\surd$	$\surd$	$\surd$
Preprocessing exploiting CSIT	$\times$	$\times$	$\times$	$\surd$	$\times$	$\times$	$\surd$	$\surd$	$\times$
Throughput benefit elaboration	$\surd$	$\surd$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
Condition of throughput advantage	$\surd$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
Error performance analysis	$\times$	$\surd$	$\times$	$\surd$	$\surd$	$\surd$	$\times$	$\surd$	$\times$
Low-complexity demodulators	$\surd$	$\surd$	$\times$	$\times$	$\surd$	$\times$	$\times$	$\times$	$\surd$

$*$ In terms of throughput improvement, no TA grouping can significantly increase the achievable rate, but at the cost of detection complexity.

Given the rate advantages of QSM and GSM, the authors of [24] combine these two schemes for increasing the rate. More specifically, [24] divides N_t antennas into L groups and employs QSM in each group. However, the grouping reduces the number of available TA activation patterns, hence limiting the throughput improvement. In [25], the authors reconsider the combination of QSM, GSM, V-BLAST and propose the generalized quadrature spatial modulation (GQSM), which has the maximum throughput at the time of writing. Moreover, transmitter preprocessing exploiting the channel state information at the transmitter (CSIT) has also been conceived for further enhancing the performance of GQSM [27], [29]–[31], [38], [39]. Single-stream GQSM was proposed in [26]–[28] for eliminating the inter-antenna interference (IAI), but its rate improvement was far lower than that of the multi-stream GQSM of [25]. In [40], GQSM was amalgamated with the non-orthogonal multiple access (NOMA) technique as a potential solution for employment in cooperative vehicular networks. The reader might like to refer to [20], [41], [42] and the references therein for comprehensive surveys on the state-of-the-art progress of SM-family techniques.

Naturally, the rate advantage of GQSM heavily depends on the number of RF chains [25]. To the best of our knowledge, there is no open literature on the optimal number of RF chains for maximizing the achievable rate of the GQSM scheme. On the other hand, the detection complexity of GQSM is significantly increased compared to that of traditional SM, thus motivating the design of low-complexity detectors to strike a performance vs. detection complexity tradeoff. Against the above background, in this paper, we analyze the throughput benefits of GQSM and propose a pair of low-complexity detectors. We boldly and explicitly contrast our work to the existing research on GQSM in Table I, where some well-known features of the GQSM scheme are also included as part of the state-of-the-art. More explicitly, the contributions of this paper are summarized as follows:

- We analyze the achievable rate of GQSM in [25] and compare it to that of other MIMO schemes, including several single-stream SM, V-BLAST, G-QSM, and GSIM schemes. The conditions to be satisfied for attaining rate advantage for the GQSM scheme over its benchmark schemes are also provided.
- We answer the question of how many activated TAs (i.e. RF chains) are required for maximizing the rate of GQSM. Our results are also directly applicable to orthogonal frequency division multiplexing relying on index modulation (OFDM-IM) [43], and to generalized spatial index modulation (GSIM) scheme in [32], [44].
- Finally, a pair of low-complexity detectors are proposed for GQSM. Explicitly, our first detection scheme is based on a sophisticated amalgam of the spatial-domain maximum likelihood (ML) scheme and of the conventional ordered successive interference cancellation (OSIC) detector of V-BLAST systems, which we refer to as the SML-OSIC detector. Furthermore, to reduce the complexity of searching for the TA activation pattern in the SML-OSIC detector, the orthogonal matching pursuit (OMP) algorithm is invoked.

The remainder of this paper is organized as follows. Section II presents the GQSM system model. In Section III, we analyze the achievable rate of GQSM and give the approximate condition of achieving the maximum attainable rate. Our rate comparison between GQSM and other transmission schemes is presented in Section IV. In Section V, we design low-complexity GQSM detection schemes. Our simulation results are presented in Section VII. Finally, Section VIII concludes this paper.

Notation: We use upper (lower) bold face letters to indicate matrices (column vectors); $(\cdot)^T$ and $(\cdot)^H$ represent transpose and Hermitian transpose, respectively; $E\{\cdot\}$ stands for the expected value; $\|\cdot\|$ is the Frobenius norm of a complex vector; We denote the $N \times N$ identity matrix as $\mathbf{I}_N$; $\mathbf{0}$ and $\mathbf{1}$ denote an all-zero vector and an all-one vector, respectively, with appropriate dimensions; Furthermore, $\lfloor \cdot \rfloor$ is the floor operation; $\binom{n}{k}$ denotes the number of combinations of taking k out of n; $\Re(z)$ and $\Im(z)$ denote the real and imaginary part of a complex number z, respectively.

II. GQSM SYSTEM MODEL

A GQSM transmitter is shown in Fig. 1, consisting of N_t TAs and N_α transmit RF chains with $1 \leq N_\alpha \leq N_t$. For each channel use, N_α out of N_t TAs are activated for transmitting

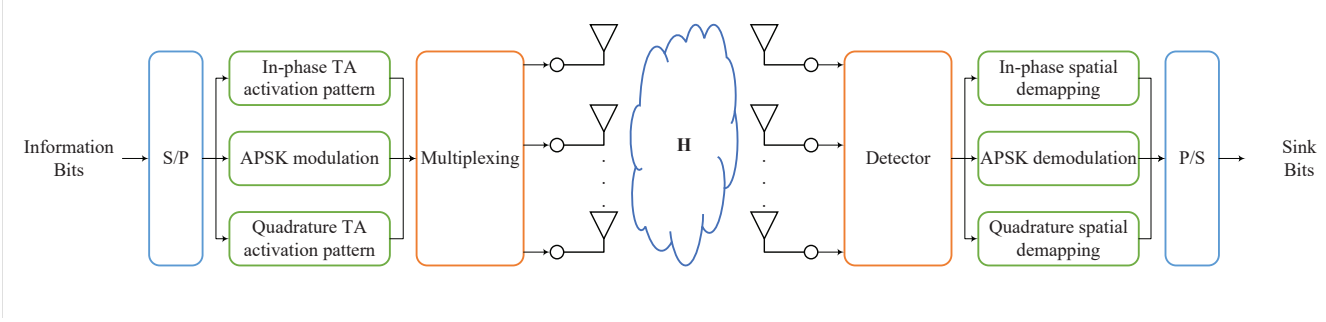

Fig. 1. The schematic of the GQSM system model.

the real part of the N_α M-ary PSK/QAM symbols. In a similar way, another N_α TAs are activated to transmit the imaginary part of the same N_α M-ary PSK/QAM symbols[1]. The rest of the TAs remain inactive. These inactive TAs can be viewed as transmitting the value zero.

At the transmitter, the real and imaginary parts of the GQSM transmit symbols can be expressed in vectorial form as

$$\mathbf{s}_\Re = [\underbrace{0,\cdots,0}_{l_\Re^1-1}, s_\Re^1, \underbrace{0,\cdots,0}_{l_\Re^i-l_\Re^1-1}, s_\Re^i, \underbrace{0,\cdots,0}_{l_\Re^{N_\alpha}-l_\Re^i-1}, s_\Re^{N_\alpha}, \underbrace{0,\cdots,0}_{N_t-l_\Re^{N_\alpha}}]^T, \tag{1}$$

and

$$\mathbf{s}_\Im = [\underbrace{0,\cdots,0}_{N_t-l_\Im^{N_\alpha}}, s_\Im^{N_\alpha}, \underbrace{0,\cdots,0}_{l_\Im^{N_\alpha}-l_\Im^i-1}, s_\Im^i, \underbrace{0,\cdots,0}_{l_\Im^i-l_\Im^1-1}, s_\Im^1, \underbrace{0,\cdots,0}_{l_\Im^1-1}]^T, \tag{2}$$

where $s_\Re^i$ and $s_\Im^i$ denote the real and imaginary parts of the i-th symbol s^i for $1 \le i \le N_\alpha$, respectively, while $l_\Re^i \in (1,\cdots,N_t)$ and $l_\Im^i \in (1,\cdots,N_t)$ represent the corresponding activated TA indices. Note that here we adopt the reverse order to arrange the indices associated with the imaginary part to indicate that the indices of the real and imaginary parts are independent of each other. Hence, the transmitted signal of the GQSM system can be expressed as

$$\mathbf{s} = \mathbf{s}_\Re + j\mathbf{s}_\Im. \tag{3}$$

Note that for the special case of BPSK, we should adopt the appropriate initial phase of $\pi/4$ or $3\pi/4$, in order to ensure the existence of both real and imaginary parts.

We assume a memoryless MIMO system having N_r receive antennas (RAs). Let $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ and $\mathbf{n} \in \mathbb{C}^{N_r \times 1}$ be the MIMO channel matrix and noise vector, whose entries are complex-valued Gaussian distributed, yielding $\mathcal{CN}(0,1)$ and $\mathcal{CN}(0,\sigma^2)$, respectively. Therefore, the received signal $\mathbf{y} \in \mathbb{C}^{N_r \times 1}$ can be written as

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}. \tag{4}$$

III. Achievable Rate Analysis of the GQSM System

The transmit vector in GQSM is formed by two parts: *1)* TA activation pattern selection bits and *2)* M-ary PSK/QAM modulation bits. The number of activation pattern selection bits is $2\left\lfloor \log_2 \binom{N_t}{N_\alpha} \right\rfloor$, while the number of M-ary PSK/QAM modulation bits is $N_\alpha \log_2 M$. By combining these two parts, the achievable rate of GQSM having N_t TAs, N_α transmit RF chains and M-ary PSK/QAM modulation can be written as

$$R_{\text{GQSM}} = 2\left\lfloor \log_2 \binom{N_t}{N_\alpha} \right\rfloor + N_\alpha \log_2 M. \tag{5}$$

Next, let us analyze the throughput R_{GQSM} of the GQSM scheme in (5) in more detail. In particular, we will examine how R_{GQSM} varies as a function of its variables: N_α, N_t, and M.

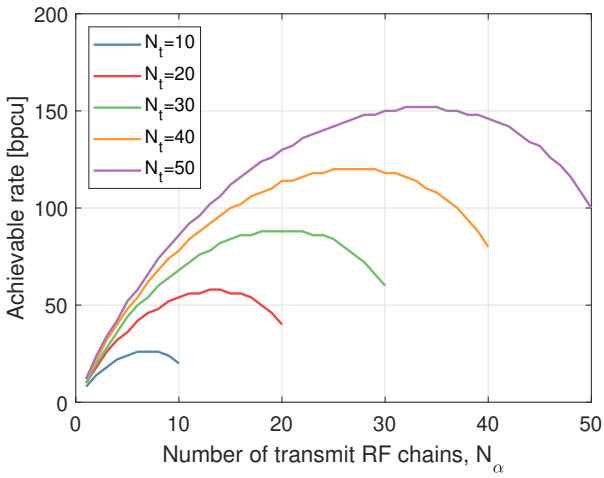

Fig. 2. The achievable rate of the GQSM scheme R_{GQSM} versus the number of transmit RF chains N_α under the assumption of different number of TAs. Note that the achievable rate is maximized when N_α reaches a value between $\frac{N_t}{2}$ and N_t.

A. R_{GQSM} versus N_α

Firstly, Fig. 2 shows the variation of R_{GQSM} as a function of N_α for different number of TAs $N_t = 10, 20, 30, 40, 50$, where QPSK is adopted. The value of N_α is varied from 1 to N_t. As mentioned earlier, $N_\alpha = N_t$ corresponds to the traditional V-BLAST scheme. The R_{GQSM} versus N_α plot for a given N_t shows an intriguing behavior. Explicitly, for a given N_t, there is an optimum N_α that maximizes the achievable rate R_{GQSM}.

[1]We note that the TA activation pattern of the imaginary part here can overlap with the TA activation pattern of the real part, which depends on the current bitstream.

Let $R^{\mathrm{m}}_{\mathrm{GQSM}}$ denote the maximum achievable rate of GQSM, i.e.,

$$R^{\mathrm{m}}_{\mathrm{GQSM}} = \max_{1 \le N_\alpha \le N_t} R_{\mathrm{GQSM}}. \tag{6}$$

In Fig. 2, it is interesting to see that $R^{\mathrm{m}}_{\mathrm{GQSM}}$ occurs at some $N_\alpha < N_t$. In other words, $R^{\mathrm{m}}_{\mathrm{GQSM}}$ can exceed the rate of V-BLAST scheme of $N_t \log_2 M$, when the first term in (5) exceeds $(N_t - N_\alpha)\log_2 M$. Hence, we are interested to find out, when the achievable rate of GQSM reaches its maximum value and in deriving the expression for the maximum rate $R^{\mathrm{m}}_{\mathrm{GQSM}}$.

Since the floor operation in (5) makes the analysis of R_{GQSM} versus N_α difficult, as an alternative, we relax R_{GQSM} in (5) as

$$\begin{aligned}\tilde{R}_{\mathrm{GQSM}} &= 2\log_2 \binom{N_t}{N_\alpha} + N_\alpha \log_2 M \\ &= 2\log_2 \frac{N_t!}{N_\alpha!\,(N_t - N_\alpha)!} + N_\alpha \log_2 M.\end{aligned} \tag{7}$$

According to the ubiquitous Stirling formula of [45], we have

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n. \tag{8}$$

Following this, the number of combinations $\binom{N_t}{N_\alpha}$ in (7) can be simplified to

$$\begin{aligned}\binom{N_t}{N_\alpha} &= \frac{\sqrt{2\pi N_t}\left(\frac{N_t}{e}\right)^{N_t}}{2\pi\sqrt{N_\alpha\,(N_t - N_\alpha)}\left(\frac{N_\alpha}{e}\right)^{N_\alpha}\left(\frac{N_t - N_\alpha}{e}\right)^{N_t - N_\alpha}} \\ &= \frac{(N_t)^{N_t + 0.5}}{\sqrt{2\pi}\,(N_\alpha)^{N_\alpha + 0.5}\,(N_t - N_\alpha)^{N_t - N_\alpha + 0.5}}.\end{aligned} \tag{9}$$

Substituting (9) into (7) and simplifying the result, we obtain

$$\begin{aligned}\tilde{R}_{\mathrm{GQSM}} &= N_\alpha \log_2 M \\ &\quad + 2\log_2\left\{\frac{(N_t)^{N_t + 0.5}}{\sqrt{2\pi}\,(N_\alpha)^{N_\alpha + 0.5}\,(N_t - N_\alpha)^{N_t - N_\alpha + 0.5}}\right\} \\ &= N_\alpha \log_2 M - (2N_t - 2N_\alpha + 1)\log_2(N_t - N_\alpha) \\ &\quad - (2N_\alpha + 1)\log_2 N_\alpha + (2N_t + 1)\log_2 N_t - \log_2(2\pi).\end{aligned} \tag{10}$$

By relaxing the floor operation and using Stirling's formula, we now convert the expression of R_{GQSM} in (5) to the approximate form seen in (10). However, it is important to note that, the value of N_α in (10) is discrete, i.e., $1 \le N_\alpha \le N_t$, $N_\alpha \in \mathbb{Z}^+$, which still limits our analysis to some extent; hence, we further relax $\tilde{R}_{\mathrm{GQSM}}$ in (10) as a function of the positive real number $\tilde{N}_\alpha$, yielding

$$\begin{aligned}\tilde{R}_{\mathrm{GQSM}} &= \tilde{N}_\alpha \log_2 M - \left(2N_t - 2\tilde{N}_\alpha + 1\right)\log_2\left(N_t - \tilde{N}_\alpha\right) \\ &\quad - \left(2\tilde{N}_\alpha + 1\right)\log_2 \tilde{N}_\alpha + (2N_t + 1)\log_2 N_t - \log_2(2\pi),\end{aligned} \tag{11}$$

where $1 \le \tilde{N}_\alpha \le N_t$, $\tilde{N}_\alpha \in \mathbb{R}^+$.

Then, we can readily find the first-order derivative of $\tilde{R}_{\mathrm{GQSM}}$ with respect to $\tilde{N}_\alpha$ as

$$\begin{aligned}\frac{\partial \tilde{R}_{\mathrm{GQSM}}}{\partial \tilde{N}_\alpha} &= \log_2 M - 2\log_2 \tilde{N}_\alpha - \frac{1}{\tilde{N}_\alpha \ln 2} \\ &\quad + 2\log_2\left(N_t - \tilde{N}_\alpha\right) + \frac{1}{\left(N_t - \tilde{N}_\alpha\right)\ln 2}.\end{aligned} \tag{12}$$

According to the convexity of $\tilde{R}_{\mathrm{GQSM}}$ over the interval of $1 \le \tilde{N}_\alpha \le N_t$, the optimal $\tilde{N}^{\mathrm{o}}_\alpha$ maximizing $\tilde{R}_{\mathrm{GQSM}}$ in (11) is equivalent to the solution, when the first-order derivative equals zero, i.e.

$$\tilde{N}^{\mathrm{o}}_\alpha = \arg_{\tilde{N}_\alpha}\left\{\frac{\partial \tilde{R}_{\mathrm{GQSM}}}{\partial \tilde{N}_\alpha} = 0\right\}. \tag{13}$$

The solution of (13) is not facile because of the logarithmic function in (12). Therefore, we harness the truncated Taylor approximation for $\frac{\partial \tilde{R}_{\mathrm{GQSM}}}{\partial \tilde{N}_\alpha}$ in (12) to obtain an equivalent expression of $\frac{\partial \tilde{R}_{\mathrm{GQSM}}}{\partial \tilde{N}_\alpha}$ with respect to $\tilde{N}_\alpha$. Please see Appendix I for a detailed derivation. According to Appendix I, the approximate expression of $\frac{\partial \tilde{R}_{\mathrm{GQSM}}}{\partial \tilde{N}_\alpha}$ in (12) can be rewritten as

$$\frac{\partial \tilde{R}_{\mathrm{GQSM}}}{\partial \tilde{N}_\alpha} = \log_2 M - \frac{4\,(N_t - 1)\left(2\tilde{N}_\alpha - N_t\right)}{N_t^2 \ln 2}. \tag{14}$$

Reconsidering the problem in (13) using (14), we can readily obtain $\tilde{N}^{\mathrm{o}}_\alpha$ as

$$\tilde{N}^{\mathrm{o}}_\alpha = \frac{N_t}{2} + \frac{\log_2 M N_t^2 \ln 2}{8\,(N_t - 1)}. \tag{15}$$

Since the number of RF chains must be an integer, the floor operation is applied to $\tilde{N}^{\mathrm{o}}_\alpha$ for finding the optimal N^{o}_α as

$$N^{\mathrm{o}}_\alpha = \left\lfloor \frac{N_t}{2} + \frac{\log_2 M N_t^2 \ln 2}{8\,(N_t - 1)} \right\rfloor. \tag{16}$$

Upon substituting (16) into (5), the maximum achievable rate $R^{\mathrm{m}}_{\mathrm{GQSM}}$ becomes:

$$\begin{aligned}R^{\mathrm{m}}_{\mathrm{GQSM}} &= 2\left\lfloor \log_2 \binom{N_t}{N^{\mathrm{o}}_\alpha} \right\rfloor + N^{\mathrm{o}}_\alpha \log_2 M \\ &= 2\left\lfloor \log_2 \binom{N_t}{\left\lfloor \frac{N_t}{2} + \frac{\log_2 M N_t^2 \ln 2}{8(N_t - 1)} \right\rfloor} \right\rfloor \\ &\quad + \left\lfloor \frac{N_t}{2} + \frac{\log_2 M N_t^2 \ln 2}{8\,(N_t - 1)} \right\rfloor \log_2 M.\end{aligned} \tag{17}$$

The formulas (16) and (17) provide the maximum achievable rate of the GQSM scheme and the condition for the maximum achievable rate, when both N_t and M are given. A number of beneficial conclusions can be drawn from (16) and (17). For example, the number of RF chains N^{o}_α required for achieving the maximum attainable rate increases with the PSK/QAM order. Furthermore, (17) gives a quantitative rate expression of the GQSM scheme, which facilitates its comparison to other transmission schemes. Additionally, our derivation can also be used for the rate evaluation of other

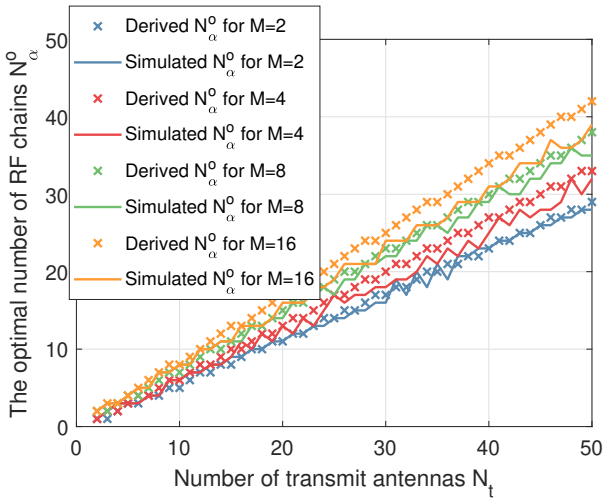

Fig. 3. Comparison of the derived $N_\alpha^{\rm o}$ and the simulated $N_\alpha^{\rm o}$ under different number of TAs.

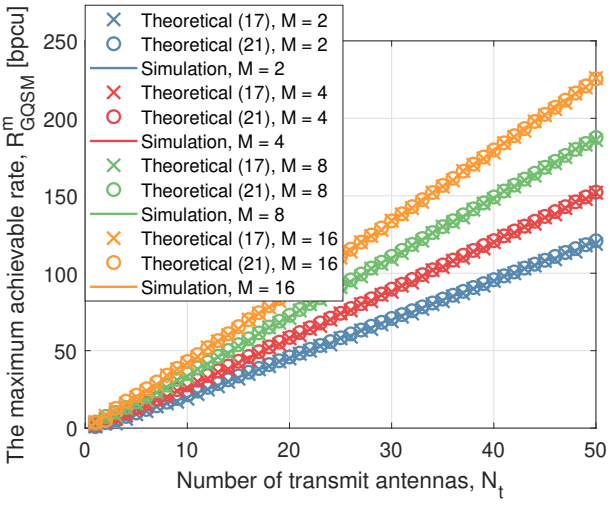

Fig. 4. Comparison of the achievable rate based on the derived $N_\alpha^{\rm o}$ and the maximum achievable rate under different numbers of TAs.

transmission schemes adopting a similar philosophy, such as OFDM-IM [43] and GSIM [44]. Fig. 3 compares the simulated and the analytical $N_\alpha^{\rm o}$ results for different M and N_t. Observe that the analytical $N_\alpha^{\rm o}$ is the upper bound of its simulated counterpart. The smaller M, the more accurate the analytical $N_\alpha^{\rm o}$ value becomes. Nevertheless, the maximum difference between the solution derived and the simulation results recorded for the optimal number of RF chains is less than 5 even for a large number of TAs. Furthermore, Fig. 4 compares the maximum achievable rate of the GQSM scheme with the theoretical maximum achievable rate of (17) being based on the $N_\alpha^{\rm o}$ derived. As seen from Fig. 4, the achievable rate based on the $N_\alpha^{\rm o}$ derived almost perfectly characterizes the maximum achievable rate of the GQSM scheme.

B. R_{GQSM} versus N_t & M

In this subsection, we will discuss the relationship of $R_{\rm GQSM}$ versus N_t and M. Similarly, we use the equivalent form of $\tilde{R}_{\rm GQSM}$ in (10) and relax N_t into a positive real number $\tilde{N}_t$.

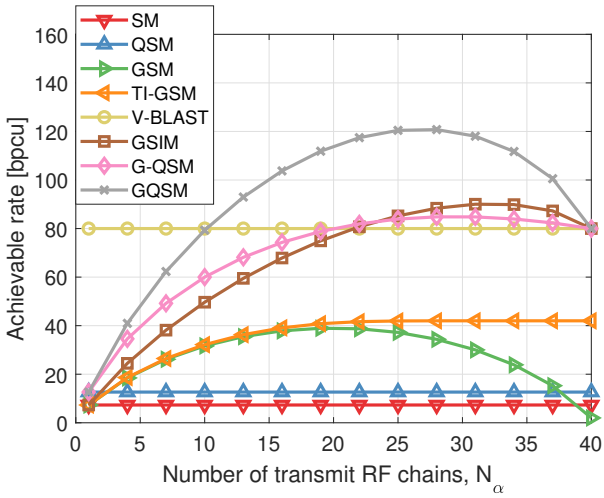

Fig. 5. The achievable rate versus the number of transmit RF chains N_α under different transmission schemes.

This allows us to get the derivative of $\tilde{R}_{\rm GQSM}$ with respect to $\tilde{N}_t$ as

$$\begin{aligned}\frac{\partial \tilde{R}_{\rm GQSM}}{\partial \tilde{N}_t} &= 2\log_2 \tilde{N}_t + \frac{1}{\tilde{N}_t \ln 2} \\ &\quad - 2\log_2\left(\tilde{N}_t - N_\alpha\right) - \frac{1}{\left(\tilde{N}_t - N_\alpha\right)\ln 2} \\ &= -2\log_2\left(1 - \frac{N_\alpha}{\tilde{N}_t}\right) - \frac{N_\alpha}{\tilde{N}_t\left(\tilde{N}_t - N_\alpha\right)\ln 2} \\ &\overset{(a)}{=} \frac{2N_\alpha}{\tilde{N}_t \ln 2} - \frac{N_\alpha}{\tilde{N}_t\left(\tilde{N}_t - N_\alpha\right)\ln 2} \\ &= \frac{N_\alpha}{\tilde{N}_t \ln 2}\left(2 - \frac{1}{\tilde{N}_t - N_\alpha}\right) > 0, \end{aligned} \tag{18}$$

where the equality '(a)' holds by using the equivalent expression of the logarithmic function.

Similarly, we can get the first-order derivative of $\tilde{R}_{\rm GQSM}$ with respect to $\tilde{M}$, namely to the real-valued expansion of M, as

$$\frac{\partial \tilde{R}_{\rm GQSM}}{\partial \tilde{M}} = \frac{N_\alpha}{\tilde{M}\ln 2} > 0. \tag{19}$$

From (18) and (19), we can conclude that when the number of transmit RF chains is limited, the achievable rate of the GQSM scheme will increase both with the number N_t of TAs and with the modulation order M, which is consistent with our intuition. In conclusion, the relationship between the achievable rate $R_{\rm GQSM}$ and the number of TAs N_t, the number of transmit RF chains N_α and the modulation order M has been characterized. Therefore, the appropriate choice of N_t, N_α and M maximizes the achievable rate. Moreover, the performance vs. complexity also has to be considered to strike a flexible trade-off.

TABLE II
COMPARISON OF ACHIEVABLE RATES FOR DIFFERENT MIMO TRANSMISSION SCHEMES.

Transmission scheme	Achievable rate	RF chains	Features
SM in [46]	$R_{\text{SM}} = \lfloor \log_2 N_t \rfloor + \log_2 M$	1	Single stream, IAI-free
QSM in [37]	$R_{\text{QSM}} = 2 \lfloor \log_2 N_t \rfloor + \log_2 M$	1	I/Q separation, IAI-free
GSM in [47]	$R_{\text{GSM}} = \left\lfloor \log_2 \binom{N_t}{N_\alpha} \right\rfloor + \log_2 M$	N_α	Multiple activated TA, IAI-free
TI-GSM in [48]	$R_{\text{TI-GSM}} = \left\lfloor \sum_{i=1}^{N_\alpha} \log_2 \binom{N_t}{i} \right\rfloor + \log_2 M$	N_α	Flexible activated TA, IAI-free
V-BLAST in [49]	$R_{\text{V-BLAST}} = N_t \log_2 M$	N_t	Multiple streams
GSIM in [44]	$R_{\text{GSIM}} = \left\lfloor \log_2 \binom{N_t}{N_\alpha} \right\rfloor + N_\alpha \log_2 M$	N_α	GSM+V-BLAST
G-QSM in [24]	$R_{\text{G-QSM}} = 2 \sum_{i=1}^{N_\alpha} \lfloor \log_2 N_i \rfloor + N_\alpha \log_2 M$ $s.t. \sum_{i=1}^{N_\alpha} N_i = N_t$	N_α	GSM+QSM+V-BLAST, TA grouping
GQSM in [25]	$R_{\text{GQSM}} = 2 \left\lfloor \log_2 \binom{N_t}{N_\alpha} \right\rfloor + N_\alpha \log_2 M$	N_α	GSM+QSM+V-BLAST, no TA grouping

IV. THE RATE COMPARISON BETWEEN THE GQSM SCHEME AND OTHER TRANSMISSION SCHEMES

In this section, we compare the achievable rate of GQSM to that of other popular MIMO schemes, including SM [46], QSM [37], GSM [47], TI-GSM [48], V-BLAST [49], G-QSM [24] and GSIM [44][2]. Note that for the G-QSM and GQSM schemes, the number of active antennas per time slot might exceed the number of RF chains, N_α, due to separate activation of the real and imaginary parts associated with the in-phase and quadrature components of the I/Q RF chains, respectively. Their specific rate expressions are shown in Table II. More specifically, we plot the achievable rate of the various transmission schemes in Fig. 5, where $N_t = 40$ and QPSK is adopted.

Since SM, QSM, GSM and TI-GSM transmit the same symbol from different TAs, their IAI-free structure constrains their rates compared to the other schemes combined with spatial multiplexing, but ensures a better error performance and lower complexity. For a moderate number of RF chains, we have

$$\begin{aligned} &R_{\text{SM}} < R_{\text{QSM}} < R_{\text{GSM}} < R_{\text{TI-GSM}} < \log_2 M + N_t \\ <&R_{\text{GQSM}} = 2 \left\lfloor \log_2 \binom{N_t}{N_\alpha} \right\rfloor + N_\alpha \log_2 M, \end{aligned} \tag{20}$$

which can be readily confirmed from Fig. 5.

Since the achievable rate of these four transmission schemes in (20) is not as high as that of the rest of transmission schemes, we will focus our attention on the achievable rates of V-BLAST [49], G-QSM [24], GSIM [44] and GQSM [25].

Firstly, we consider the maximum achievable rate of the proposed GQSM scheme seen in (17). Since the floor operation in $R^{\text{m}}_{\text{GQSM}}$ makes it complex to analyze, we use $\tilde{N}^{\text{o}}_\alpha$ in (15) and $\tilde{R}_{\text{GQSM}}$ in (11) to derive an approximate expression for $R^{\text{m}}_{\text{GQSM}}$ as follows:

$$\begin{aligned} &R^{\text{m}}_{\text{GQSM}} \\ &= \left(\frac{1}{2} + \frac{\log_2 M N_t \ln 2}{8 (N_t - 1)} \right) N_t \log_2 M \\ &- \log_2 N_t - \log_2 (2\pi) + 2 (N_t + 1) \log_2 (8 (N_t - 1)) \\ &- (N_t + 1) \log_2 \left(16 (N_t - 1)^2 - \log_2^2 M N_t^2 \ln^2 2 \right) \\ &+ \frac{\log_2 M N_t^2 \ln 2}{4 (N_t - 1)} \log_2 \frac{4 (N_t - 1) - \log_2 M N_t \ln 2}{4 (N_t - 1) + \log_2 M N_t \ln 2}. \end{aligned} \tag{21}$$

More specifically, Fig. 4 compares the maximum achievable rate in (17) to the approximate expression in (21). Observe from Fig. 4 that the achievable rate of the (21) is a tight approximation of the maximum achievable rate of (17), thus ensuring the applicability of the following lemmas for rate comparison.

Subsequently, the achievable rate of the V-BLAST scheme is given by [49]

$$R_{\text{V-BLAST}} = N_t \log_2 M, \tag{22}$$

while that of the G-QSM scheme in [24] is expressed as

$$R_{\text{G-QSM}} = 2 N_\alpha \log_2 \frac{N_t}{N_\alpha} + N_\alpha \log_2 M. \tag{23}$$

Using the same analysis method in this paper, we can obtain the optimal N_α for G-QSM as

$$N^{\text{o}}_{\alpha,\text{G-QSM}} = \left\lfloor \frac{\sqrt{M} N_t}{e} \right\rfloor, \tag{24}$$

and the maximum achievable rate of the G-QSM scheme [24] as

$$R^{\text{m}}_{\text{G-QSM}} = \frac{2\sqrt{M} N_t}{e \ln 2}. \tag{25}$$

Finally, the achievable rate of the GSIM scheme in [44] is given as

$$R_{\text{GSIM}} = \left\lfloor \log_2 \binom{N_t}{N_\alpha} \right\rfloor + N_\alpha \log_2 M. \tag{26}$$

[2]We note that single-stream GQSM is reported in [26], [28], and its rate is $R_{ss-GQSM} = 2 \left\lfloor \log_2 \binom{N_t}{N_\alpha} \right\rfloor + \log_2 M$. Since the achievable rate of single-stream GQSM is on the same order of magnitude as TI-GSM and GSM, and to avoid confusion with the two multi-stream GQSM schemes in this paper, we do not consider it next.

To elaborate a little further, in [44], the rate of the GSIM scheme is shown to be higher than that of V-BLAST if and only if $N_t \geq 2M$. However, [44] does not give the expression for the maximum rate of GSIM. As mentioned earlier, by leveraging our derivation in Section III, the $N^{\mathrm{o}}_{\alpha,\mathrm{GSIM}}$ maximizing the GSIM's rate can be formulated as

$$N^{\mathrm{o}}_{\alpha,\mathrm{GSIM}} = \frac{N_t}{2} + \frac{\log_2 MN_t^2 \ln 2}{4\left(N_t-1\right)}. \tag{27}$$

Similar to (21), the maximum achievable rate of GSIM can be obtained as

$$\begin{aligned} &R^{\mathrm{m}}_{\mathrm{GSIM}} \\ &= \left(\frac{1}{2} + \frac{\log_2 MN_t \ln 2}{4\left(N_t-1\right)}\right) N_t \log_2 M - 0.5\log_2 N_t \\ &-0.5\log_2\left(2\pi\right) + \left(N_t+1\right)\log_2\left(4\left(N_t-1\right)\right) \\ &-0.5\left(N_t+1\right)\log_2\left(4\left(N_t-1\right)^2 - \log_2^2 MN_t^2 \ln^2 2\right) \\ &+\frac{\log_2 MN_t^2 \ln 2}{4\left(N_t-1\right)} \log_2 \frac{2\left(N_t-1\right) - \log_2 MN_t \ln 2}{2\left(N_t-1\right) + \log_2 MN_t \ln 2}. \end{aligned} \tag{28}$$

Now that we have the maximum achievable rate for all MIMO transmission schemes considered, we will compare them one by one to the maximum rate $R^{\mathrm{m}}_{\mathrm{GQSM}}$ of the GQSM scheme in (21).

A. GQSM vs. V-BLAST

We first compare the rate of GQSM and V-BLAST. The rate comparison between GQSM and V-BLAST is summarized in ***Lemma 1***.

Lemma 1: For low-order PSK/QAM modulation with $M \leq 32$, the maximum achievable rate of the GQSM scheme $R^{\mathrm{m}}_{\mathrm{GQSM}}$ is higher than that of V-BLAST, i.e. $N_t \log_2 M$, if and only if $N_t \geq \frac{1}{1-\beta\log_2 M}$, where $\beta = \frac{\ln 2}{4}$.

Proof: For GQSM, the optimal number of RF chains corresponding to its maximum throughput is shown in (16). On the other hand, the rate of GQSM corresponding to N_t RF chains is equal to the rate of V-BLAST. Thus, the condition of ensuring that the maximum achievable rate of GQSM is higher than the rate of V-BLAST is equivalent to

$$\frac{N_t}{2} + \frac{\log_2 MN_t^2 \ln 2}{8\left(N_t-1\right)} < N_t. \tag{29}$$

By simplifying (29), we can arrive at:

$$N_t \geq \frac{1}{1-\beta\log_2 M}, \tag{30}$$

given that $M < 2^{\frac{1}{\beta}}$. Since M is always a power of 2, we take $M \leq 32$ in ***Lemma 1***.

Remark 1: Since (16) is the upper bound of the true $N^{\mathrm{o}}_{\alpha,\mathrm{GQSM}}$, the condition in *Lemma 1* is actually a tighter one. Under higher-order modulation, the modeling error in (16) will be high enough to render (30) invalid. Therefore, *Lemma 1* is only valid when M is less than 54 (i.e. $2^{\frac{1}{\beta}}$) and is strictly valid. On the other hand, the throughput improvement of GQSM compared to V-BLAST is extremely small for high-order PSK/QAM modulation. Nonetheless, our derivation is still of great value for practical applications. Although [44] suggests that N_t is higher than $2M$, our conclusion shows that the smallest N_t to ensure $R^{\mathrm{m}}_{\mathrm{GQSM}} > R_{\text{V-BLAST}}$ is less than $2M$. In other words, our derivation shows that the rate advantage of GQSM and that of its relatives over V-BLAST holds under a wider range of conditions, without requiring the number of TAs to be higher than $2M$. For example, when 16-QAM is adopted, only 10 TAs are required for ensuring that the rate of GQSM is higher than that of V-BLAST, while the conclusion given in [44] is 32.

B. GQSM vs. GSIM

In this subsection, we will compare the rates of GQSM and GSIM. The lemma describing the rate relationship between GQSM and GSIM is given as follows.

Lemma 2: The maximum rate of GQSM is strictly higher than that of GSIM for all N_t and APM modulation orders.

Proof: When the number of RF chains is the same for GQSM and GSIM, we can readily show from Table II that R_{GQSM} is higher than R_{GSIM}. Therefore, we have

$$R_{\mathrm{GQSM}}\left(N^{\mathrm{o}}_{\alpha,\mathrm{GSIM}}\right) > R_{\mathrm{GSIM}}\left(N^{\mathrm{o}}_{\alpha,\mathrm{GSIM}}\right) = R^{\mathrm{m}}_{\mathrm{GSIM}}. \tag{31}$$

On the other hand, the rate of GQSM is maximized when N_α equals $N^{\mathrm{o}}_{\alpha,\mathrm{GQSM}}$, so we have

$$R^{\mathrm{m}}_{\mathrm{GQSM}} = R_{\mathrm{GQSM}}\left(N^{\mathrm{o}}_{\alpha,\mathrm{GQSM}}\right) > R_{\mathrm{GQSM}}\left(N^{\mathrm{o}}_{\alpha,\mathrm{GSIM}}\right). \tag{32}$$

Combining (31) with (32), we obtain

$$R^{\mathrm{m}}_{\mathrm{GQSM}} > R^{\mathrm{m}}_{\mathrm{GSIM}}, \tag{33}$$

which holds for all N_t and M.

C. GQSM vs. G-QSM

In a similar way, we can formulate the Lemma comparing the rate of GQSM and G-QSM as follows.

Lemma 3: The maximum rate of GQSM is strictly higher than the maximum rate of G-QSM for all N_t and APM modulation orders.

Proof: Considering the same number of RF chains for GQSM and G-QSM, R_{GQSM} is higher than $R_{\text{G-QSM}}$, i.e.

$$R_{\mathrm{GQSM}}\left(N^{\mathrm{o}}_{\alpha,\text{G-QSM}}\right) > R_{\text{G-QSM}}\left(N^{\mathrm{o}}_{\alpha,\text{G-QSM}}\right) = R^{\mathrm{m}}_{\text{G-QSM}}. \tag{34}$$

Following this, we have

$$R^{\mathrm{m}}_{\mathrm{GQSM}} = R_{\mathrm{GQSM}}\left(N^{\mathrm{o}}_{\alpha,\mathrm{GQSM}}\right) > R_{\mathrm{GQSM}}\left(N^{\mathrm{o}}_{\alpha,\text{G-QSM}}\right). \tag{35}$$

Substituting (34) into (35), it can be readily shown that

$$R^{\mathrm{m}}_{\mathrm{GQSM}} > R^{\mathrm{m}}_{\text{G-QSM}}, \tag{36}$$

which holds for all N_t and M.

Lemmas 1 $\sim$ 3 demonstrate the rate advantage of GQSM [25] over V-BLAST [49], GSIM [44], and G-QSM [24]. The same conclusion can also be drawn from our simulations. In Fig. 5, we can see that GQSM has a significant throughput advantage over G-QSM, V-BLAST and GSIM. Specifically, the maximum achievable rate of GQSM, when the number of antennas is 40 and QPSK is adopted, can be increased by 50% compared to V-BLAST.

TABLE III
ALGORITHM I

The SML-OSIC Detector Proposed for the GQSM Scheme

Input: $\tilde{\mathbf{y}}$, $\tilde{\mathbf{H}}$
for $l_{\Re} \in [1, N_s]$, **do**
  $\mathbf{l}_{\Re} = \mathrm{U}(l_{\Re})$;
  for $l_{\Im} \in [1, N_s]$, **do**
    $\mathbf{l}_{\Im} = \mathrm{U}(l_{\Im})$;
    $\tilde{\mathbf{H}}_l = \tilde{\mathbf{H}}_{:,\mathbf{l}_{\Re} \cup \{\mathbf{l}_{\Im}+N_t\}}$;
    $\mathbf{G} = \left(\tilde{\mathbf{H}}_l^H \tilde{\mathbf{H}}_l\right)^{-1} \tilde{\mathbf{H}}_l^H$;
    $\mathbf{h} = \mathrm{vecnorm}\left(\tilde{\mathbf{H}}_l\right)$;
    $\mathbf{p} = \mathrm{argsort}(\mathbf{h}, '\,\mathrm{descend}')$;
    for $k \in [1, 2N_\alpha]$, **do**
      $\tilde{s}'_{l,p(k)} = \mathbf{G}_{p(k),:}\tilde{\mathbf{y}}$;
      $\hat{\tilde{s}}'_{l,p(k)} = \mathbb{M}^{-1}\left(\tilde{s}'_{l,p(k)}\right)$;
      $\tilde{\mathbf{y}} \leftarrow \tilde{\mathbf{y}} - \tilde{\mathbf{H}}_l|_{:,p(k)}\hat{\tilde{s}}'_{l,p(k)}$;
    end for
    $\mathbf{d}_l = \left\|\tilde{\mathbf{y}} - \tilde{\mathbf{H}}_l\hat{\tilde{\mathbf{s}}}'_l\right\|^2$;
  end for
end for
$\left[\hat{l}_{\Re}, \hat{l}_{\Im}\right] = \arg\min_{l}\{\mathbf{d}\}$;
$\left(\hat{l}^1_{\Re}, \hat{l}^2_{\Re}, \cdots, \hat{l}^{N_\alpha}_{\Re}\right) = \mathrm{U}\left\{\hat{l}_{\Re}\right\}$;
$\left(\hat{l}^1_{\Im}, \hat{l}^2_{\Im}, \cdots, \hat{l}^{N_\alpha}_{\Im}\right) = \mathrm{U}\left\{\hat{l}_{\Im}\right\}$;
$\hat{\mathbf{s}}' = \hat{\tilde{\mathbf{s}}}_{\hat{l},1:N_\alpha} + j\hat{\tilde{\mathbf{s}}}_{\hat{l},1+N_\alpha:2N_\alpha}$.

V. Signal Detection for GQSM Systems

In this section, we will discuss the detection of GQSM signals. We first consider the optimal ML detector in Section V-A, whereas Section V-B derives the SML-OSIC detector. Then, based on this the ROMP-OSIC detector is developed in Section V-C. We assume that the receiver has perfect knowledge of the channels.

A. Maximum Likelihood Detector

We first introduce the classic ML detector for determining the best-case performance. The received signal in (4) can be rewritten in real-valued form as

$$\underbrace{\begin{bmatrix}\Re(\mathbf{y})\\ \Im(\mathbf{y})\end{bmatrix}}_{\tilde{\mathbf{y}}} = \underbrace{\begin{bmatrix}\Re(\mathbf{H}) & -\Im(\mathbf{H})\\ \Im(\mathbf{H}) & \Re(\mathbf{H})\end{bmatrix}}_{\tilde{\mathbf{H}}}\underbrace{\begin{bmatrix}\Re(\mathbf{s})\\ \Im(\mathbf{s})\end{bmatrix}}_{\tilde{\mathbf{s}}} + \underbrace{\begin{bmatrix}\Re(\mathbf{n})\\ \Im(\mathbf{n})\end{bmatrix}}_{\tilde{\mathbf{n}}}, \tag{37}$$

where $\tilde{\mathbf{y}} \in \mathbb{R}^{2N_r \times 1}$, $\tilde{\mathbf{H}} \in \mathbb{R}^{2N_r \times 2N_t}$, $\tilde{\mathbf{s}} \in \mathbb{R}^{2N_t \times 1}$ and $\tilde{\mathbf{n}} \in \mathbb{R}^{2N_r \times 1}$. Meanwhile, we note that there are only $2N_\alpha$ non-zero elements in $\tilde{\mathbf{s}}$, i.e., the sparsity of $\tilde{\mathbf{s}}$ is $2N_\alpha$.

It follows from (37) that the optimal ML detector can be formulated as

$$\hat{\tilde{\mathbf{s}}}_{ML} = \underset{\tilde{\mathbf{s}} \in \mathbb{S}^{N_t} \otimes \mathbb{S}^{N_t}}{\arg\min} \left\|\tilde{\mathbf{y}} - \tilde{\mathbf{H}}\tilde{\mathbf{s}}\right\|^2, \tag{38}$$

where we have $\mathbb{S} = \mathbb{A} \cup \{0\}$ and $|\mathbb{S}| = N_\alpha$. $\tilde{\mathbf{s}} \in \mathbb{S}^{N_t} \otimes \mathbb{S}^{N_t}$ means that both the first N_t elements and the last N_t elements of $\tilde{\mathbf{s}}$ contain N_α non-zero elements.

B. Spatial ML and Ordered Successive Interference Cancellation

Since the ML detector has to evaluate N_s^2 TA combinations and M^{N_α} possible symbol vectors, its detection complexity order is $\mathcal{O}\left(N_s^2 M^{N_\alpha}\right)$, which becomes excessive for large-scale antenna arrays, and high-order constellations. Therefore, we propose a low-complexity detector based on the traditional ordered successive interference cancellation method used for V-BLAST [50]. More explicitly, for each specific TA activation pattern, the GQSM detection in (38) can be reduced to

$$\tilde{\mathbf{s}}'_l = \underset{\tilde{\mathbf{s}}' \in \mathbb{A}^{2N_\alpha}}{\arg\min} \left\|\tilde{\mathbf{y}} - \tilde{\mathbf{H}}_l\tilde{\mathbf{s}}'\right\|^2, \tag{39}$$

with $l = [l_{\Re}, l_{\Im}]$ and $\tilde{\mathbf{H}}_l = \tilde{\mathbf{H}}_{:,\mathbf{l}_{\Re} \cup \{\mathbf{l}_{\Im}+N_t\}}$, where $\mathbf{l}_{\Re}$ and $\mathbf{l}_{\Im}$ represent the $l_{\Re}$-th active TA pattern corresponding to the real part of $\mathbf{s}$ and the $l_{\Im}$-th active TA pattern corresponding to the imaginary part of $\mathbf{s}$, respectively.

Since the detection problem of (39) is equivalent to the detection of the traditional V-BLAST scheme, numerous effective detectors can be found in the literature [51]–[53]. In this paper, we adopt the OSIC detector of [53] to address (39). The philosophy of the OSIC detector will not be detailed in this paper, please refer to [53]. As the detector progresses, a simple rounding-based hard decision is presented when the k-th data stream is obtained, $1 \le k \le 2N_\alpha$. Based on [54], hard decisions can be carried out for M-QAM, yielding:

$$\hat{\tilde{s}}'_{k|l} = \frac{1}{\alpha}\min\left[\max\left(2\left\lfloor\frac{\alpha\tilde{s}'_k + 1}{2}\right\rfloor - 1, -M + 1\right), M - 1\right], \tag{40}$$

where $\alpha = 2\sqrt{\frac{\sum_{m=0}^{\frac{\sqrt{M}}{2}-1}\left(\sqrt{M}-2m-1\right)^2}{\sqrt{M}}}$ denotes the normalization factor of the M-QAM constellation. The corresponding rounding operation for M-PSK can be found in [55].

When all $2N_\alpha$ symbols are detected, we have

$$\hat{\tilde{\mathbf{s}}}'_l = \left[\hat{\tilde{s}}'_{k_1|l}, \hat{\tilde{s}}'_{k_2|l}, \cdots, \hat{\tilde{s}}'_{k_{2N_\alpha}|l}\right]^T, \tag{41}$$

where $k_1, k_2, \cdots, k_{2N_\alpha}$ represents the index of the original sequence corresponding to the sorted k.

Since we have determined the $\hat{\tilde{\mathbf{s}}}'_l$ associated with the l-th antenna combination, the estimate of the TA activation pattern l is given by:

$$\hat{l} = \underset{l}{\arg\min} \left\|\tilde{\mathbf{y}} - \tilde{\mathbf{H}}_l\hat{\tilde{\mathbf{s}}}'_l\right\|^2. \tag{42}$$

Once the estimate of l is obtained, the estimate of $\tilde{\mathbf{s}}'$ can also be readily found as $\hat{\tilde{\mathbf{s}}}'_{\hat{l}}$.

We term this detector SML-OSIC because it still uses ML detection in the spatial domain, but employs OSIC for multi-stream detection. The detailed procedure of the SML-OSIC is listed in Table V-A, where $\mathbb{M}^{-1}$ represents the hard decision expressed in (40), and $\mathbf{l}_{\Re/\Im} = \mathrm{U}\left(l_{\Re/\Im}\right)$ denotes the $l_{\Re/\Im}$-th TA activation pattern.

TABLE IV
ALGORITHM II

The ROMP-OSIC Detector Proposed for the GQSM Scheme

Input: $\tilde{\mathbf{r}}$, $\mathbf{V}$
Iterations: $\hat{l} = \arg\max_{l} (|\mathbf{V}\tilde{\mathbf{r}}|)$;
$\mathbf{c} = \tilde{\mathbf{r}}$;
$\mathbf{K} = \mathbf{V}^H$;
for $k \in [1, 2N_v]$, **do**
  $\mathbf{P}_{:,k} = \mathbf{K}_{:,\hat{l}}$;
  $\mathbf{t}_k = \hat{l}$;
  $\mathbf{K}_{:,\hat{l}} = \mathbf{0}$;
  $\mathbf{z} = \mathbf{K}^H\mathbf{c}$;
  if $\hat{l} \le N_t$, **do**
    $\hat{l} = \arg\max_{l} (|\mathbf{z}_{1+N_t:2N_t}|) + N_t$;
  else
    $\hat{l} = \arg\max_{l} (|\mathbf{z}_{1:N_t}|)$;
  end if
  $\mathbf{a} = \left(\mathbf{P}_{:,1:k}^H \mathbf{P}_{:,1:k}\right)^{-1} \mathbf{P}_{:,1:k}^H \tilde{\mathbf{r}}$;
  $\mathbf{c} = \tilde{\mathbf{r}} - \mathbf{P}_{:,1:k}\mathbf{a}$;
end for
$\mathbf{m} = \text{sort}(\mathbf{t})$;
$S_{\Re} = U(l_{\Re}) \cap S_{N_\alpha} \{\mathbf{m}_1, \cdots, \mathbf{m}_{N_v}\}$;
$S_{\Im} = U(l_{\Im}) \cap S_{N_\alpha} \{\mathbf{m}_{1+N_v}, \cdots, \mathbf{m}_{2N_v}\}$;
for $l_{\Re} \in [1, |S_{\Re}|]$, **do**
  $\mathbf{l}_{\Re} = S_{\Re}(l_{\Re})$;
  for $l_{\Im} \in [1, |S_{\Im}|]$, **do**
    $\mathbf{l}_{\Im} = S_{\Im}(l_{\Im})$;
    $\mathbf{V}_l = \mathbf{V}^H_{:,\mathbf{l}_{\Re} \cup \{\mathbf{l}_{\Im} + N_t\}}$;
    $\mathbf{G} = \left(\mathbf{V}_l^H \mathbf{V}_l\right)^{-1} \mathbf{V}_l^H$;
    $\mathbf{h} = \text{vecnorm}(\mathbf{V}_l)$;
    $\mathbf{p} = \text{argsort}(\mathbf{h}, '\text{descend}')$;
    for $k \in [1, 2N_\alpha]$, **do**
      $\tilde{s}'_{p(k)} = \mathbf{G}_{p(k),:}\tilde{\mathbf{r}}$;
      $\hat{\tilde{s}}'_{p(k)} = \mathbb{M}^{-1}\left(\tilde{s}'_{p(k)}\right)$;
      $\tilde{\mathbf{r}} \leftarrow \tilde{\mathbf{r}} - \mathbf{V}_l|_{:,p(k)} \hat{\tilde{s}}'_{p(k)}$;
    end for
    $\mathbf{d}_l = \left\|\tilde{\mathbf{r}} - \mathbf{V}_l \hat{\tilde{\mathbf{s}}}'\right\|^2$;
  end for
end for
$\left[\hat{l}_{\Re}, \hat{l}_{\Im}\right] = \arg\min_{l} \{\mathbf{d}\}$;
Output: $\left(\hat{l}^1_{\Re}, \hat{l}^2_{\Re}, \cdots, \hat{l}^{N_\alpha}_{\Re}\right) = S_{\Re}\left\{\hat{l}_{\Re}\right\}$;
$\left(\hat{l}^1_{\Im}, \hat{l}^2_{\Im}, \cdots, \hat{l}^{N_\alpha}_{\Im}\right) = S_{\Im}\left\{\hat{l}_{\Im}\right\}$;
$\hat{\mathbf{s}}' = \hat{\tilde{\mathbf{s}}}_{\hat{l},1:N_\alpha} + j\hat{\tilde{\mathbf{s}}}_{\hat{l},1+N_\alpha:2N_\alpha}$;

C. Relaxed OMP Aided Detector

In this subsection, we will introduce a detector based on a sparse recovery algorithm. Since there are only $2N_\alpha$ non-zero elements in the $2N_t \times 1$ vector $\tilde{\mathbf{s}}$, the classic OMP algorithm [56] is first invoked for determining the index of the TA activation pattern. However, due to the flat effect of the OMP algorithm when used for detection [18], we relax the sparsity of $\tilde{\mathbf{s}}$. More explicitly, when we employ the OMP algorithm, we assume that the sparsity of the transmit vector is $2N_v$, where $N_\alpha \le N_v \le N_t$. In this way, we end up with a relaxed index set of TA activation patterns compared to N_s. Following this, we can reuse Algorithm 1 to acquire the final estimate.

However, we note that the performance of OMP depends heavily on the orthogonality of $\tilde{\mathbf{H}}$'s columns. Therefore, the singular value decomposition is invoked to transform $\tilde{\mathbf{H}}$ into an equivalent channel matrix having better column orthogonality. The singular value decomposition of $\tilde{\mathbf{H}}$ can be formulated as

$$\tilde{\mathbf{H}} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H, \tag{43}$$

where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices and $\mathbf{\Lambda}$ is a diagonal matrix containing the singular values on its diagonal.

According to (43), the ML estimation in (38) can be converted into

$$\hat{\tilde{\mathbf{s}}}_{ML} = \underset{\tilde{\mathbf{s}} \in \mathbb{S}^{N_t} \otimes \mathbb{S}^{N_t}}{\arg\min} \left\|\tilde{\mathbf{r}} - \mathbf{V}^H\tilde{\mathbf{s}}\right\|^2, \tag{44}$$

where we have $\tilde{\mathbf{r}} = \mathbf{\Lambda}^{-1}\mathbf{U}^H\tilde{\mathbf{y}}$. Since $\mathbf{V}^H$ exhibits better column orthogonality than $\tilde{\mathbf{H}}$, this improves the performance of OMP.

Since this algorithm uses the relaxed OMP algorithm to estimate the index of the TA activation pattern, we term this algorithm as the ROMP-OSIC detector. The detailed steps of ROMP-OSIC are summarized in Table V-B. The Line 19 in Table V-B means that only the antenna combinations consisting of TAs $\mathbf{m}_1, \mathbf{m}_2, \cdots, \mathbf{m}_{N_v}$ have to be considered from the predesigned legitimate TA activation pattern set. The same interpretation also applies to Line 20. Upon appropriately adjusting the value of N_v, the performance of the ROMP-OSIC detector approaches that of the SML-OSIC and ML detectors, as verified in Section VII.

VI. Computational Complexity Analysis

In this section, we will analyze the computational complexity of the detectors proposed for the GQSM scheme[3]. The detection complexity is evaluated in terms of the number of real-valued multiplications involved. Specifically, the detection complexity of the full-search based ML detector for the GQSM scheme is given by

$$\mathcal{C}_{\text{ML}} = N_s^2 M^{N_\alpha} \left(4N_\alpha N_r + 2N_r\right), \tag{45}$$

where $N_s = 2^{\left\lfloor \log_2 \binom{N_t}{N_\alpha} \right\rfloor}$ represents the number of legitimate antenna activation patterns, while $(4N_\alpha N_r + 2N_r)$ represents the complexity of calculating $\left\|\widetilde{\mathbf{y}} - \widetilde{\mathbf{H}}\widetilde{\mathbf{s}}\right\|^2$, given each feasible $\widetilde{\mathbf{s}}$.

When the SML-OSIC detector of Table III is adopted, the GQSM scheme's detection complexity becomes

$$\begin{aligned}
&\mathcal{C}_{\text{SML - OSIC}} \\
=&N_s^2 \left[\underbrace{8N_r N_\alpha^2}_{L-7} + \underbrace{2N_r}_{L-8} + 2N_\alpha \left(\underbrace{2N_r}_{L-11} + \underbrace{2N_r}_{L-13}\right)\right] \\
=&N_s^2 \left(8N_r N_\alpha^2 + 2N_r + 8N_r N_\alpha\right),
\end{aligned} \tag{46}$$

where $L-i$ represents the complexity of carrying out the step in the i-th line of Algorithm I.

[3]The theoretical error performance of the GQSM scheme using an ML detector has been analyzed in [24].

Moreover, when the heuristic ROMP-OSIC detector of Table IV is employed, the GQSM scheme's detection complexity can be further reduced to

$$\begin{aligned}
&\mathcal{C}_{\text{ROMP - OSIC}} \\
&= \underbrace{4N_mN_t}_{L-2} \\
&+ \sum_{k=1}^{2N_v} \left[\underbrace{2N_m\left(2N_t-k\right)}_{L-9} + \underbrace{2N_mk+k^2}_{L-15} + \underbrace{2N_mk}_{L-16} \right] \\
&+ \binom{N_v}{N_\alpha}^2 \left[\underbrace{8N_mN_\alpha^2}_{L-26} + \underbrace{4N_mN_\alpha}_{L-27} \right. \\
&\left. + 2N_\alpha \left(\underbrace{4N_mN_\alpha}_{L-30} + \underbrace{2N_m}_{L-32} \right) + \underbrace{2N_m}_{L-34} \right],
\end{aligned} \tag{47}$$

where we have $N_m = \min\left(N_t, N_r\right)$, while $L-i$ represents the complexity of performing the i-th step of Algorithm II.

For the sake of simplicity, we consider the case of $N_v = N_\alpha$. Therefore, the complexity of the ROMP-OSIC detector described in (47) can be readily simplified as

$$\begin{aligned}
\mathcal{C}_{\text{ROMP - OSIC}} &= 4N_mN_t + 8N_mN_tN_\alpha \\
&+ 10N_mN_\alpha + 20N_mN_\alpha^2 + 2N_m \\
&+ \frac{N_\alpha\left(2N_\alpha+1\right)\left(4N_\alpha+1\right)}{3}.
\end{aligned} \tag{48}$$

Note that the complexity of the SML-OSIC and ROMP-OSIC detectors does not increase with the modulation order, which is a significant benefit compared to their ML counterpart. The specific detection complexity will be evaluated by numerical simulations in Section VII.

VII. Simulation Results

In this section, we will characterize the error performance of the GQSM scheme through numerical simulation, assuming that the normalized Rayleigh MIMO channel is adopted. In our simulations, the number of RF chains is set to 3. At each SNR, 10,000 independent experiments are performed to obtain the average error performance.

Firstly, Fig. 6 shows the BER performance of GQSM for different detectors, where the number of TAs is $N_t = 5$ and QPSK modulation is considered. In order to obtain a high performance for the ROMP-OSIC detector, we adopt $N_v = N_t$. As seen from Fig. 6, when the number of RAs is $N_r = 5$, both SML-OSIC and ROMP-OSIC exhibit a performance loss of 1.5dB compared to the ML detector at BER of 10^{-3}, which is due to the fact that the SML-OSIC and ROMP-OSIC detectors rely heavily on the orthogonality between different channels for inferring the correct active antenna indices. As the number of RAs increases, the performance gap between ROMP-OSIC and ML will be gradually narrowed due to the improved column orthogonality. When the number of RAs increases to $N_r = 10$, SML-OSIC and ROMP-OSIC perform almost as well as ML, which is reminiscent of the trend seen for massive MIMO uplink detection due to the channel hardening phenomenon. The overall detection performance of $N_r = 10$ has been significantly improved due to the diversity gain attained. Moreover, Fig. 7 exhibits the error performance of GQSM considering 16QAM. Observe from Fig. 7 that the BER of the GQSM scheme was severely degraded for high-order modulation. Specifically, when the number of RAs is $N_r = 5$, SML-OSIC and ROMP-OSIC have a performance loss of about 1dB compared to the ML detector at the BER of 10^{-2} and they approach the ML performance, when the number of RAs increases to $N_r = 10$.

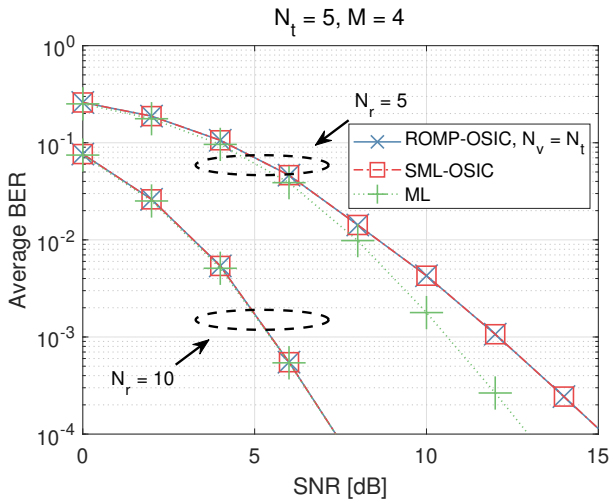

Fig. 6. The BER of GQSM versus SNR for different detectors, where $N_t = 5$ and QPSK are adopted. The throughput is 12 bits/channel use (bpcu).

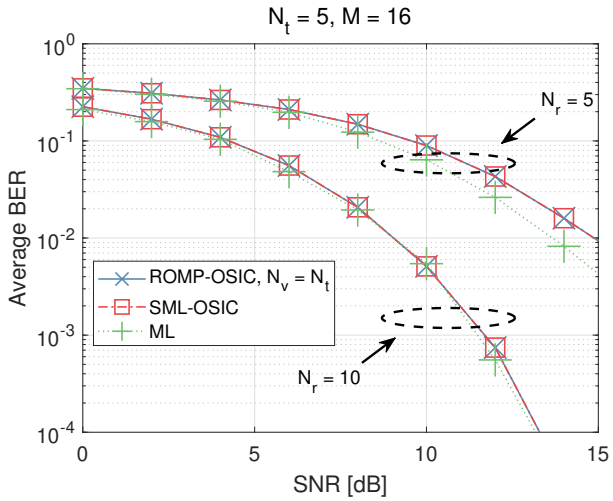

Fig. 7. The BER of GQSM versus SNR for different detectors, where $N_t = 5$ and 16QAM are adopted. The throughput is 18 bits/channel use (bpcu).

Figs. 8 and 9 compare the error performance of different detectors proposed for the GQSM scheme considering $N_t = 7$, where QPSK and 16QAM are adopted, respectively. In order to evaluate the effect of N_v in the ROMP-OSIC detector, we consider the setups of $N_v = N_t$ and $N_v = N_\alpha$. As seen from Fig. 8, when considering $N_v = N_\alpha$, ROMP-OSIC will exhibit an error floor, since the detection performance of ROMP-OSIC is highly dependent on the correctness of the antenna

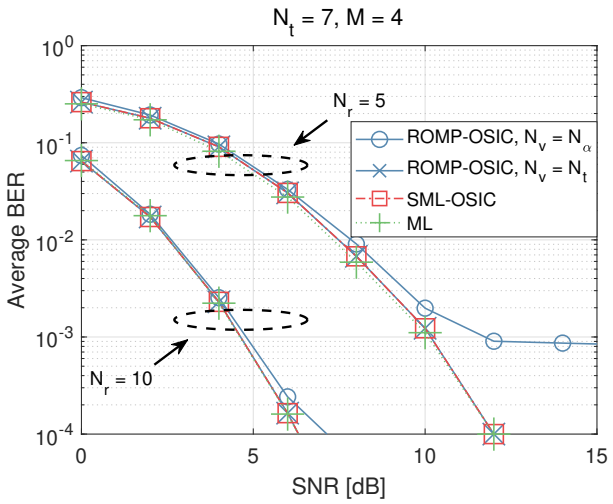

Fig. 8. The BER of GQSM versus SNR for different detectors, where $N_t = 7$ and QPSK are adopted. The throughput is 16 bits/channel use (bpcu).

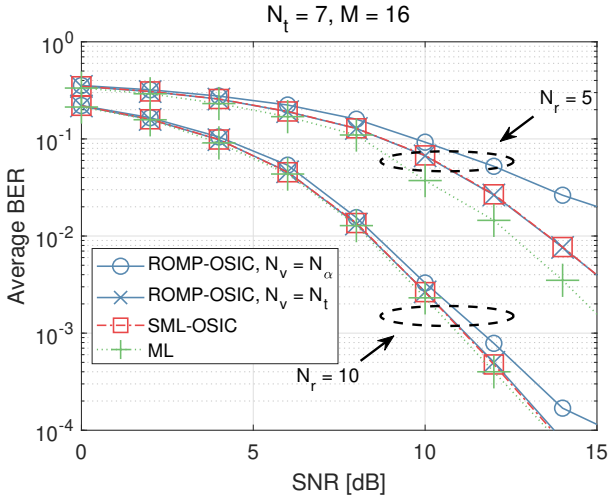

Fig. 9. The BER of GQSM versus SNR for different detectors, where $N_t = 7$ and 16QAM are adopted. The throughput is 22 bits/channel use (bpcu).

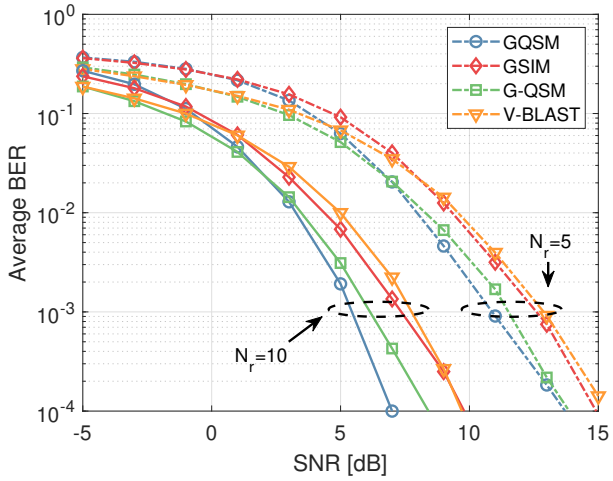

Fig. 10. Error performance of GQSM compared with other transmission schemes. In order to maintain the same spectral efficiency of 12bpcu, $N_t = 5$, $N_\alpha = 3$ and QPSK are adopted in GQSM, while 8QAM is adopted in GSIM and G-QSM and 16QAM is used in V-BLAST.

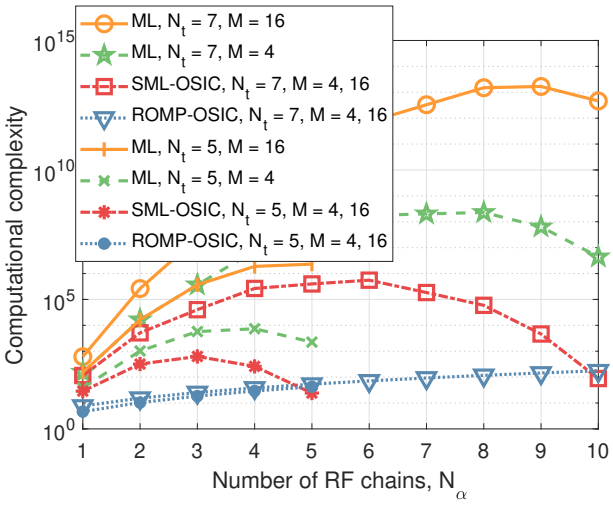

Fig. 11. The complexity comparison of three detectors proposed for the GQSM scheme, where we have $N_r = 10$.

indices recovered. Upon increasing N_v, the performance of the ROMP-OSIC detector will be gradually improved at the cost of increased complexity. Therefore, ROMP-OSIC strikes a flexible tradeoff between the detection complexity and error performance by appropriately increasing N_v. Additionally, the error floor of ROMP-OSIC will be gradually reduced, as the number of RAs increases. Explicitly, when $N_t = 7$, $M = 16$ and $N_r = 5$ are considered, ROMP-OSIC having $N_v = N_t$ has a performance loss of about 1dB compared to ML, as shown in Fig. 9. By contrast, ROMP-OSIC having $N_v = N_\alpha$ has a performance loss of about 5dB, but imposes the lowest complexity, which will be verified later. Furthermore, with the number of RAs increased from $N_r = 5$ to $N_r = 10$, the ROMP-OSIC detector having $N_v = N_t$ achieves near-optimal performance, while ROMP-OSIC having $N_v = N_\alpha$ still suffers from about 2dB performance loss at BER of 10^{-4}.

Fig. 10 compares the error performance of GQSM, GSIM, G-QSM and V-BLAST. In order to maintain the same spectral efficiency of 12bpcu, 8QAM is adopted in GSIM and G-QSM, while 16QAM is used in V-BLAST. Furthermore, we consider the same amount of total transmit power per slot for maintaining the same energy efficiency. It can be seen that, at moderate and high SNRs, GQSM has the best BER performance, followed by G-QSM and GSIM, with about 1dB SNR penalty. The BER performance of the V-BLAST is worst because of the severe IAI. The BER performance improvement of GQSM benefits from its better exploitation of spatial constellations.

Fig. 11 compares the computational complexity of three detectors proposed for the GQSM scheme, where the number of RAs is $N_r = 10$. As seen from Fig. 11, the complexity of all detectors increases as the number of TAs increases from $N_t = 5$ to $N_t = 7$. Additionally, the SML-OSIC and ROMP-OSIC detectors are not required to search through all feasible sets, thus significantly reducing the detection complexity com-

pared to the ML detector. More specifically, the complexity of the ROMP-OSIC detector barely increases upon increasing the number of TAs and RF chains, but it suffers from some performance loss, as shown in Figs. 6 $\sim$ 9. Nevertheless, the ROMP-OSIC detector is capable of striking attractive trade-offs between the complexity and error performance upon increasing N_v. Furthermore, the complexity of the SML-OSIC and ROMP-OSIC detectors is independent of the modulation order due to the employment of demodulation slicing. For the scenario having a large number of TAs and high-order modulation, e.g., $N_t = 7$ and $M = 16$, it is impractical to use an ML detector due to its excessive complexity, while the ROMP-OSIC scheme operates at a low detection complexity.

VIII. Conclusion

In this paper, we analyzed the benefits of the GQSM scheme inheriting the throughput advantage of GSM, QSM and V-BLAST. We derived the achievable rate expression of GQSM, and most importantly, the corresponding conditions of approaching its maximum throughput. Our derivation can also be used for other transmission schemes, such as OFDM-IM and GSIM. Compared to other transmission schemes, the rate advantage of the proposed GQSM is illustrated and verified by our numerical simulations. Finally, we conceived a pair of low-complexity detectors, namely SML-OSIC and ROMP-OSIC, based on the traditional V-BLAST detectors and on the sparse recovery algorithm, respectively. The simulation results verify the benefits of the proposed low-complexity detectors.

Appendix A
Second Order Taylor Approximation of (12)

We firstly define f to be

$$\begin{aligned} f &= \frac{\partial \tilde{R}_{\text{GQSM}}}{\partial \tilde{N}_\alpha} \\ &= \log_2 M - 2\log_2 \tilde{N}_\alpha - \frac{2\tilde{N}_\alpha + 1}{\tilde{N}_\alpha \ln 2} \\ &\quad + 2\log_2\left(N_t - \tilde{N}_\alpha\right) + \frac{2N_t - 2\tilde{N}_\alpha + 1}{\left(N_t - \tilde{N}_\alpha\right)\ln 2}. \end{aligned} \tag{49}$$

Then, the first-order derivative and the second-order derivative of f with respect to $\tilde{N}_\alpha$ can be formulated as

$$\begin{aligned} \frac{\partial f}{\partial \tilde{N}_\alpha} &= \frac{1}{\tilde{N}_\alpha^2 \ln 2} - \frac{2}{\tilde{N}_\alpha \ln 2} \\ &\quad + \frac{1}{\left(N_t - \tilde{N}_\alpha\right)^2 \ln 2} - \frac{2}{\left(N_t - \tilde{N}_\alpha\right)\ln 2}, \end{aligned} \tag{50}$$

and

$$\begin{aligned} \frac{\partial^2 f}{\partial \tilde{N}_\alpha^2} &= \frac{2}{\tilde{N}_\alpha^2 \ln 2} - \frac{2}{\left(N_t - \tilde{N}_\alpha\right)^2 \ln 2} \\ &\quad + \frac{2}{\left(N_t - \tilde{N}_\alpha\right)^3 \ln 2} - \frac{2}{\tilde{N}_\alpha^3 \ln 2}. \end{aligned} \tag{51}$$

Upon substituting $\tilde{N}_\alpha = \frac{N_t}{2}$ into (49), (50) and (51), respectively, we have

$$f|_{\tilde{N}_\alpha = \frac{N_t}{2}} = \log_2 M, \tag{52}$$

$$\frac{\partial f}{\partial \tilde{N}_\alpha}|_{\tilde{N}_\alpha = \frac{N_t}{2}} = -\frac{8\left(N_t - 1\right)}{N_t^2 \ln 2}, \tag{53}$$

$$\frac{\partial^2 f}{\partial \tilde{N}_\alpha^2}|_{\tilde{N}_\alpha = \frac{N_t}{2}} = 0. \tag{54}$$

Therefore, we can represent f as the second order Taylor expansion at $\tilde{N}_\alpha = \frac{N_t}{2}$ as

$$f = \frac{\partial \tilde{R}_{\text{GQSM}}}{\partial \tilde{N}_\alpha} = \log_2 M - \frac{4\left(N_t - 1\right)\left(2\tilde{N}_\alpha - N_t\right)}{N_t^2 \ln 2}. \tag{55}$$

References

[1] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, 2012.

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

[3] C. Xu, N. Ishikawa, R. Rajashekar, S. Sugiura, R. G. Maunder, Z. Wang, L.-L. Yang, and L. Hanzo, "Sixty years of coherent versus non-coherent tradeoffs and the road from 5G to wireless futures," *IEEE Access*, vol. 7, pp. 178246–178299, 2019.

[4] P. Yang, M. Di Renzo, Y. Xiao, S. Li, and L. Hanzo, "Design guidelines for spatial modulation," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 6–26, 2014.

[5] M. Di Renzo, H. Haas, A. Ghrayeb, S. Sugiura, and L. Hanzo, "Spatial modulation for generalized MIMO: Challenges, opportunities, and implementation," *Proc. of the IEEE*, vol. 102, no. 1, pp. 56–103, 2013.

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

[7] J. Jeganathan, A. Ghrayeb, L. Szczecinski, and A. Ceron, "Space shift keying modulation for MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3692–3703, 2009.

[8] L. Xiao, P. Yang, Y. Xiao, S. Fan, M. Di Renzo, W. Xiang, and S. Li, "Efficient compressive sensing detectors for generalized spatial modulation systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1284–1298, 2016.

[9] N. Ishikawa, S. Sugiura, and L. Hanzo, "50 years of permutation, spatial and index modulation: From classic RF to visible light communications and data storage," *IEEE Commun. Surveys Tuts*, vol. 20, no. 3, pp. 1905–1938, 2018.

[10] C. Xu, P. Zhang, R. Rajashekar, N. Ishikawa, S. Sugiura, Z. Wang, and L. Hanzo, ""Near-perfect" finite-cardinality generalized space-time shift keying," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 9, pp. 2146–2164, 2019.

[11] C. Xu, P. Zhang, R. Rajashekar, N. Ishikawa, S. Sugiura, L. Wang, and L. Hanzo, "Finite-cardinality single-RF differential space-time modulation for improving the diversity-throughput tradeoff," *IEEE Trans. Commun.*, vol. 67, no. 1, pp. 318–335, 2018.

[12] L. Xiao, Y. Xiao, L. You, P. Yang, S. Li, and L. Hanzo, "Single-RF and twin-RF spatial modulation for an arbitrary number of transmit antennas," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 6311–6324, 2018.

[13] L. Xiao, P. Xiao, Y. Xiao, I. Hemadeh, A. Mohamed, and L. Hanzo, "Bayesian compressive sensing assisted space–time block coded quadrature spatial modulation," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 10044–10048, 2018.

[14] Y. Liu, L.-L. Yang, and L. Hanzo, "Spatial modulation aided sparse code-division multiple access," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1474–1487, 2017.

[15] Y. Liu, L.-L. Yang, P. Xiao, H. Haas, and L. Hanzo, "Spatial modulated multicarrier sparse code-division multiple access," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 610–623, 2020.

[16] J. An, C. Xu, Y. Liu, L. Gan, and L. Hanzo, "Low-complexity improved-throughput generalised spatial modulation: Bit-to-symbol mapping, detection and performance analysis," *IEEE Trans. Veh. Technol.*, vol. 71, pp. 1060–1065, Jan. 2021.

[17] T. Datta and A. Chockalingam, "On generalized spatial modulation," in *2013 IEEE Wireless Commun. Netw. Conf. (WCNC)*, pp. 2716–2721, IEEE, 2013.

[18] L. Xiao, P. Xiao, Z. Liu, W. Yu, H. Haas, and L. Hanzo, "A compressive sensing assisted massive SM-VBLAST system: Error probability and capacity analysis," *IEEE Trans. Wireless Commun.*, vol. 19, pp. 1990–2005, Mar. 2020.

[19] E. Basar, "Index modulation techniques for 5G wireless networks," *IEEE Commun. Mag.*, vol. 54, no. 7, pp. 168–175, 2016.

[20] T. Mao, Q. Wang, Z. Wang, and S. Chen, "Novel index modulation techniques: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, pp. 315–348, First quarter 2019.

[21] X. Cai, W. Xu, D. Wang, S. Hong, and L. Wang, "An m-ary orthogonal multilevel differential chaos shift keying system with code index modulation," *IEEE Trans. Commun.*, vol. 67, no. 7, pp. 4835–4847, 2019.

[22] E. Aydin, F. Cogen, and E. Basar, "Code-index modulation aided quadrature spatial modulation for high-rate MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 10, pp. 10257–10261, 2019.

[23] F. Cogen, E. Aydin, N. Kabaoglu, E. Basar, and H. Ilhan, "Generalized code index modulation and spatial modulation for high rate and energy-efficient MIMO systems on rayleigh block-fading channel," *IEEE Systems J.*, vol. 15, no. 1, pp. 538–545, 2021.

[24] L. Xiao, P. Xiao, Y. Xiao, H. Haas, A. Mohamed, and L. Hanzo, "Compressive sensing assisted generalized quadrature spatial modulation for massive MIMO systems," *IEEE Trans. Commun.*, vol. 67, pp. 4795–4810, Jul. 2019.

[25] K. Gunde and K. Hari, "Modified generalised quadrature spatial modulation," in *2019 National Conf. Commun. (NCC)*, pp. 1–5, IEEE, 2019.

[26] Y. Celik and S. A. Çolak, "Generalized quadrature spatial modulation techniques for VLC," *Opt. Commun.*, p. 125905, 2020.

[27] F. R. Castillo-Soria, J. Cortez-González, R. Ramirez-Gutierrez, F. M. Maciel-Barboza, and L. Soriano-Equigua, "Generalized quadrature spatial modulation scheme using antenna grouping," *ETRI Journal*, vol. 39, no. 5, pp. 707–717, 2017.

[28] R. Mesleh, O. Hiari, and A. Younis, "Generalized space modulation techniques: Hardware design and considerations," *Phy. Commun.*, vol. 26, pp. 87–95, 2018.

[29] S. Guo, H. Zhang, P. Zhang, S. Dang, C. Liang, and M.-S. Alouini, "Signal shaping for generalized spatial modulation and generalized quadrature spatial modulation," *IEEE Trans. Wireless Commun.*, vol. 18, pp. 4047–4059, Aug. 2019.

[30] J. Li, M. Wen, X. Cheng, Y. Yan, S. Song, and M. H. Lee, "Generalized precoding-aided quadrature spatial modulation," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1881–1886, 2016.

[31] R. Neelakandan, "Sub-optimal low-complexity detector for generalised quadrature spatial modulation," *Electronics Lett.*, vol. 54, no. 15, pp. 941–943, 2018.

[32] 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, 2012.

[33] L. Xiao, Y. Xiao, C. Xu, X. Lei, P. Yang, S. Li, and L. Hanzo, "Compressed-sensing assisted spatial multiplexing aided spatial modulation," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 794–807, 2017.

[34] E. Basar, U. Aygolu, E. Panayirci, and H. V. Poor, "Space-time block coded spatial modulation," *IEEE Trans. Commun.*, vol. 59, no. 3, pp. 823–832, 2010.

[35] C. Xu, R. Rajashekar, N. Ishikawa, S. Sugiura, and L. Hanzo, "Single-RF index shift keying aided differential space–time block coding," *IEEE Trans. Signal Process.*, vol. 66, no. 3, pp. 773–788, 2017.

[36] L. Xiao, Y. Xiao, P. Yang, J. Liu, S. Li, and W. Xiang, "Space-time block coded differential spatial modulation," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 8821–8834, 2017.

[37] R. Mesleh, S. S. Ikki, and H. M. Aggoune, "Quadrature spatial modulation," *IEEE Trans. Veh. Technol.*, vol. 64, no. 6, pp. 2738–2742, 2014.

[38] R. Ramanathan, P. S. Mallick, and T. S. Jayaraman, "Low complexity compressive sensing greedy detection of generalized quadrature spatial modulation," *IEICE Trans. Fundamentals of Electronics, Commun. and Comp. Sci.*, vol. 101, no. 3, pp. 632–635, 2018.

[39] M. Mohaisen, "Generalized complex quadrature spatial modulation," *Wireless Commun. and Mobile Comp.*, vol. 2019, 2019.

[40] J. Li, S. Dang, Y. Yan, Y. Peng, S. Al-Rubaye, and A. Tsourdos, "Generalized quadrature spatial modulation and its application to vehicular networks with NOMA," *IEEE Trans. Intell. Trans. Syst.*, vol. 22, pp. 4030–4039, Jul. 2021.

[41] M. Wen, B. Zheng, K. J. Kim, M. Di Renzo, T. A. Tsiftsis, K.-C. Chen, and N. Al-Dhahir, "A survey on spatial modulation in emerging wireless systems: Research progresses and applications," *IEEE J. Sel. Areas Commun.*, vol. 37, pp. 1949–1972, Sept. 2019.

[42] M. S. Kuran, H. B. Yilmaz, I. Demirkol, N. Farsad, and A. Goldsmith, "A survey on modulation techniques in molecular communication via diffusion," *IEEE Communications Surveys & Tutorials*, vol. 23, pp. 7–28, First quarter 2021.

[43] E. Başar, Ü. Aygölü, E. Panayırcı, and H. V. Poor, "Orthogonal frequency division multiplexing with index modulation," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5536–5549, 2013.

[44] T. Datta, H. S. Eshwaraiah, and A. Chockalingam, "Generalized space-and-frequency index modulation," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 4911–4924, 2015.

[45] J. Dutka, "The early history of the factorial function," *Archive for history of exact sciences*, pp. 225–249, 1991.

[46] 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, 2008.

[47] A. Younis, N. Serafimovski, R. Mesleh, and H. Haas, "Generalised spatial modulation," in *2010 Conf. Record of the Forty Fourth Asilomar Conf. Signals, Syst. and Comp.*, pp. 1498–1502, IEEE, 2010.

[48] J. An, C. Xu, Y. Liu, L. Gan, and L. Hanzo, "Low-complexity improved-throughput generalised spatial modulation: Bit-to-symbol mapping, detection and performance analysis," Jul. 2021. [Online]. Available: https://arxiv.org/abs/2107.12630.

[49] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *1998 URSI Int. Sym. Signals, Syst., Electronics. Conf. Proc. (Cat. No. 98EX167)*, pp. 295–300, IEEE, 1998.

[50] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.

[51] S. W. Kim and K. P. Kim, "Log-likelihood-ratio-based detection ordering in V-BLAST," *IEEE Trans Commun*, vol. 54, no. 2, pp. 302–307, 2006.

[52] B. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389–399, 2003.

[53] S. Yang and L. Hanzo, "Fifty years of MIMO detection: The road to large-scale MIMOs," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 1941–1988, 2015.

[54] R. Rajashekar, K. Hari, and L. Hanzo, "Reduced-complexity ML detection and capacity-optimized training for spatial modulation systems," *IEEE Trans. Commun.*, vol. 62, no. 1, pp. 112–125, 2013.

[55] C. Xu, S. Sugiura, S. X. Ng, P. Zhang, L. Wang, and L. Hanzo, "Two decades of MIMO design tradeoffs and reduced-complexity MIMO detection in near-capacity systems," *IEEE Access*, vol. 5, pp. 18564–18632, 2017.

[56] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.