Enabling Energy-Efficiency in Massive-MIMO: A Scalable Low-Complexity Decoder for Generalized Quadrature Spatial Modulation

Hyeon Seok Rou, Graduate Student Member, IEEE, Giuseppe Thadeu Freitas de Abreu, Senior Member, IEEE, David González G., Senior Member, IEEE, and Osvaldo Gonsa

Abstract—Generalized quadrature spatial modulation (GQSM) schemes are known to achieve high energy- and spectral- efficiencies by modulating information both in transmitted symbols and in coded combinatorial activations of subsets of multiple transmit antennas. A challenge of the approach is, however, the decoding complexity which scales with the efficiency of the scheme. In order to circumvent this bottleneck and enable high-performance and feasible GQSM in massive multiple-input multiple-output (mMIMO) scenarios, we propose a novel decoding algorithm which enjoys a complexity order that is independent of the combinatorial factor. This remarkable feature of the proposed decoder is a consequence of a novel vectorized Gaussian belief propagation (GaBP) algorithm, here contributed, whose message passing (MP) rules leverage both pilot symbols and the unit vector decomposition (UVD) of the GQSM signal structure. The effectiveness of the proposed UVD-GaBP method is illustrated via computer simulations including numerical results for systems of a size never before reported in related literature (up to 32 transmit antennas), which demonstrates the potential of the approach in paving the way towards high energy and spectral efficiency for wireless systems in a truly mMIMO setting.

Index Terms—Generalized quadrature spatial modulation (GQSM), massive multiple-input multiple-output (mMIMO), message passing (MP), enabling technology, low-complexity.

I. INTRODUCTION

Spatial modulation (SM) techniques [1] have been widely studied as a promising multiple-input multiple-output (MIMO) transmission scheme, which exploits multiple-antenna resources not only to gain spatial diversity, but also to encode "energy-free" (spatially-modulated) information in the form of a sparse activation of small subsets of the total available transmit antennas. This unique feature of SM schemes allow large data to be encoded with lower effective transmit energy.

Therefore, SM is an excellent candidate enabling technology for beyond fifth generation (B5G) and sixth generation (6G) wireless communication systems [2] because it offers simultaneously: high gains in energy efficiency (EE) and spectral efficiency (SE); a sparse utilization of radio frequency (RF) chains, which are highly appealing to millimeter-wave (mmWave) and Terahertz (THz) systems [3]; and great flexibility, since the symbols actually transmitted are not directly relevant to the antenna subset activation scheme itself. Motivated by the aforementioned advantages, a rapid development of enhanced SM designs aiming to maximize energyand spectral-efficiencies, and minimize error rates and decoding complexities can be found in recent literature [4]–[10]. An excellent example is the quadrature spatial modulation (QSM) design [4], in which the SM technique is applied independently to each in-phase and quadrature (IQ) component of the transmit symbols, resulting not only a two-fold increase in the rate of the spatially-modulated information, but also in improved bit error rate (BER) performance compared to the original SM scheme of [1]. Further examples in this line of work are the improved generalized quadrature spatial modulation (GQSM) designs incorporating space-time codings (STCs) and other techniques to optimize the antenna activation patterns, as proposed in [5], [6].

Since naive decoding of GQSM requires searching over a combinatorial space of size $\binom{\alpha N_T}{P}^2 \cdot M^P$ – where N_T is the number of transmit antennas, α is the codebook amplification factor ($\alpha = 1$ for the classic GQSM), P is the number of symbols transmitted, and M is the size of the complex symbol constellation – another line of work towards improving GQSM is to lower decoding complexity. One example is the compressive sensing (CS)-based method of [6], which can achieve a significantly low complexity, but may suffer from high spatial domain error since the elaborate GQSM codebook structure cannot be incorporated in the estimation rules. On the other hand, the sphere decoding (SD) approach of [7] and the reduced-search methods of [8], [9] were shown to approach maximum likelihood (ML) performance, but can only reduce the squared-combinatorial complexity order by a linear factor.

Recently, we proposed a near-ML message passing (MP) method for GQSM decoding [10], which was shown to eliminate the quadratic exponent in the combinatorial factor by employing a novel decomposition of the GQSM signal into two independent vectors. Following the latter work, we propose in this article a novel MP-based GQSM decoder based on a further decomposition of the GQSM signal model into independent unit vectors, which results in a complexity that is completely independent of the combinatorial space, while still incorporating the information of the structured GQSM codebook patterns.

The contributions of this article can be summarized as:

 A novel pilotted GQSM decoding algorithm is proposed, which achieves a combinatorial-free complexity order,

H. S. Rou and G. T. F. Abreu are with the School of Computer Science and Engineering, Constructor University, Campus Ring 1, 28759, Bremen, Germany (e-mails: [h.rou, g.abreu]@constructor.university).

D. González G. and O. Gonsa are with Wireless Communications Technologies, Continental AG, 65936 Frankfurt/Main, Germany (e-mails: david.gonzalez.g@ieee.org, osvaldo.gonsa@continental-corporation.com).

- Tailored MP rules are derived based on a unit vector decomposition (UVD) of the GQSM signal model,
- Simulation results are provided at unprecedented GQSM mMIMO scales, including system sizes up to $N_T = 32$.

II. GQSM SYSTEM MODEL

A. Transmit Signal Model

Given a MIMO transmitter equipped with N_T antenna elements, the GQSM transmit signal is described by

$$\mathbf{x} = \mathbf{x}^{\mathrm{R}} + j\mathbf{x}^{\mathrm{I}} \in \mathbb{C}^{N_T \times 1},\tag{1}$$

where $\mathbf{x}^{\mathrm{R}} \in \mathbb{C}^{N_T \times 1}$ and $\mathbf{x}^{\mathrm{I}} \in \mathbb{C}^{N_T \times 1}$ are respectively the real and imaginary parts of the GQSM transmit signal vector, respectively defined as

$$\mathbf{x}^{\mathrm{R}} = \begin{bmatrix} 0, s_{1}^{\mathrm{R}}, 0, & \cdots, 0, s_{p}^{\mathrm{R}}, 0, & \cdots, 0, s_{p}^{\mathrm{R}}, 0, & \cdots, 0, s_{p}^{\mathrm{R}}, 0 \end{bmatrix}^{\mathsf{T}}, (2a)$$
$$\mathbf{x}^{\mathrm{I}} = \begin{bmatrix} 0, \cdots, 0, s_{1}^{\mathrm{I}}, 0, \cdots, 0, s_{p}^{\mathrm{I}}, 0, \cdots, 0, s_{p}^{\mathrm{I}}, 0 \end{bmatrix}^{\mathsf{T}}, (2b)$$

$$[0, \cdots, 0, \underbrace{s_1}_{k_1^{\mathrm{I}} \text{-th position}}, 0, \cdots, 0, \underbrace{s_p}_{k_p^{\mathrm{I}} \text{-th position}}, 0, \cdots, 0, \underbrace{s_P}_{k_p^{\mathrm{I}} \text{-th position}}, 0] ,$$

which contain the real and imaginary parts of the P transmit symbols¹ $s_p \triangleq s_p^{\mathrm{R}} + js_p^{\mathrm{I}} \in \mathbb{C}$, with $p \in \{1, \dots, P\}$, selected from a discrete constellation S of cardinality M, where the positions of the symbol components are respectively described by the index vectors $\mathbf{k}^{\mathrm{R}} \triangleq [k_1^{\mathrm{R}}, \dots, k_p^{\mathrm{R}}, \dots, k_P^{\mathrm{R}}]$ and $\mathbf{k}^{\mathrm{I}} \triangleq [k_1^{\mathrm{I}}, \dots, k_p^{\mathrm{I}}, \dots, k_P^{\mathrm{I}}]$.

It is important to note that the positions of the symbols between \mathbf{x}^{R} and \mathbf{x}^{I} are independent, and that the position of a symbol component in the vector corresponds directly to the activated antenna element at the transmitter. By exploiting the possible combinatorial patterns of the symbol component positions, the GQSM scheme conveys information not only from the encoding of P symbols from S, but also from the selection of P positions out of N_T possible positions, respectively for both \mathbf{x}^{R} and \mathbf{x}^{I} .

In light of the above, the total information conveyed by the GQSM signal x is given by

$$B_{\rm GQSM} = 2B_{\rm Sp} + B_{\rm Dg} = 2\left\lfloor \log_2 \binom{N_T}{P} \right\rfloor + P \log_2(M), \quad (3)$$

where B_{Sp} is the number of bits spatially encoded by the antenna position selection of P symbol parts, and B_{Dg} is the number of bits *digitally* encoded by the transmit symbols.

B. Received Signal Model

The received signal vector $\mathbf{y} \in \mathbb{C}^{N_R \times 1}$ at the MIMO receiver equipped with N_R antennas, is described by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} \in \mathbb{C}^{N_R \times 1},\tag{4}$$

where $\mathbf{H} \in \mathbb{C}^{N_R \times N_T}$ is the wireless channel matrix between the receiver and transmitter antennas, and $\mathbf{w} \in \mathbb{C}^{N_R \times 1}$ is the additive white Gaussian noise (AWGN) vector with independent and identically distributed (i.i.d.) elements $w_n \sim \mathcal{CN}(0, N_0)$ for $n \in \{1, \dots, N_R\}$ where N_0 is the noise power.

The complex-valued system model in eq. (4) can be transformed into an equivalent system in the real domain as

$$\boldsymbol{y} \triangleq \begin{bmatrix} \Re\{\mathbf{y}\}\\ \Im\{\mathbf{y}\} \end{bmatrix} = \overbrace{\begin{bmatrix} \Re\{\mathbf{H}\} & -\Im\{\mathbf{H}\}\\ \Im\{\mathbf{H}\} & \Re\{\mathbf{H}\} \end{bmatrix}}^{\triangleq \boldsymbol{x}} \overbrace{\begin{bmatrix} \mathbf{x}^{\mathrm{R}}\\ \mathbf{x}^{\mathrm{I}} \end{bmatrix}}^{\triangleq \boldsymbol{x}} + \overbrace{\begin{bmatrix} \Re\{\mathbf{w}\}\\ \Im\{\mathbf{w}\} \end{bmatrix}}^{\triangleq \boldsymbol{w}}, \quad (5a)$$
$$= \boldsymbol{H}\boldsymbol{x} + \boldsymbol{w} \in \mathbb{R}^{2N_R \times 1}, \quad (5b)$$

where $\boldsymbol{y} \in \mathbb{R}^{2N_R \times 1}$, $\boldsymbol{H} \in \mathbb{R}^{2N_R \times 2N_T}$, $\boldsymbol{x} \in \mathbb{R}^{2N_T \times 1}$, and $\boldsymbol{w} \in \mathbb{R}^{2N_R \times 1}$ respectively denote the IQ-decoupled counterparts of \boldsymbol{y} , \boldsymbol{H} , \boldsymbol{x} , and \boldsymbol{w} ; $\boldsymbol{H} \triangleq [\boldsymbol{H}^{\mathrm{R}}, \boldsymbol{H}^{\mathrm{I}}]$ with $\boldsymbol{H}^{\mathrm{R}} \in \mathbb{R}^{2N_R \times N_T}$; and $\boldsymbol{H}^{\mathrm{I}} \in \mathbb{R}^{2N_R \times N_T}$ is introduced to denote the effective channel components for $\boldsymbol{x}^{\mathrm{R}}$ and $\boldsymbol{x}^{\mathrm{I}}$, respectively.

The IQ-decoupled system in eq. (5) is the basis of most state-of-the-art (SotA) QSM decoders, which estimates the effective transmit signal \boldsymbol{x} , in knowledge of \boldsymbol{y} and \boldsymbol{H} . While the linear recovery problem appears trivial, the challenge lies in the infeasible size of the discrete domain \mathcal{X} of $\boldsymbol{x} \in \mathbb{R}^{2N_T \times 1}$, with cardinality $Q \triangleq (\lfloor {N_T \choose P} \rfloor_2)^2 \cdot M^P$, where $\lfloor \cdot \rfloor_2 \triangleq 2^{\lfloor \log_2(\cdot) \rfloor}$ is the flooring operation to the nearest power of 2.

As can been seen from the cardinality of \mathcal{X} , the codebook size Q scales at a geometric rate on P, and at a squaredfactorial rate on N_T , such that complexity becomes prohibitive even for moderately large MIMO scenarios, which is why simulation results in the literature exist only up to $N_T \leq 10$, $P \leq 3$, even with lowest-complexity methods [6], [9], [10].

III. PROPOSED DECOUPLED VECTOR GABP DECODER

In light of the above, this article provides a solution to the combinatorial spatial domain search challenge, by first considering the <u>piloted</u> GQSM scenario, where all symbol component values are known at the receiver². The pilot symbols are assumed to be arbitrary and can be utilized for other functionalities such as authentication, radar, channel estimation, etc. However, even with the known pilot symbols, the challenge remains in estimating the unknown indices of the symbols in the combinatorial space, as seen in eq. (2).

A. Signal Reformulation - Unit Vector Decomposition (UVD)

First, noticing that each symbol component $s_p^{\rm R}$ and $s_p^{\rm I}$ occupies only a single position in the transmit vector (*i.e.*, only transmitted from a single antenna element), the GQSM transmit signal described by eq. (1) and eq. (2) can be more intuitively rewritten as a superposition of the symbol components multiplied by *activation* vectors, *i.e.*,

$$\mathbf{x} = \underbrace{\sum_{p=1}^{P} s_p^{\mathrm{R}} \mathbf{e}_{k_p^{\mathrm{R}}}}_{\triangleq \mathbf{x}^{\mathrm{R}}} + j \underbrace{\sum_{p=1}^{P} s_p^{\mathrm{I}} \mathbf{e}_{k_p^{\mathrm{I}}}}_{\triangleq \mathbf{x}^{\mathrm{I}}} \in \mathbb{C}^{N_T \times 1}, \qquad (6)$$

where an activation vector \mathbf{e}_t is the *t*-th column of a $N_T \times N_T$ identity matrix, which therefore, is a unit vector.

In light of the reformulation in eq. (6), the IQ-decoupled received signal vector in eq. (5) becomes

$$\boldsymbol{y} = \boldsymbol{H} \begin{bmatrix} \sum_{p=1}^{P} s_p^{\mathrm{R}} \mathbf{e}_{k_p^{\mathrm{R}}} \\ \sum_{p=1}^{P} s_p^{\mathrm{I}} \mathbf{e}_{k_p^{\mathrm{I}}} \end{bmatrix} + \boldsymbol{w} \in \mathbb{R}^{2N_R \times 1}, \quad (7)$$

where the linear recovery of x has been transformed into the joint estimation problem of 2P activation (unit) vectors.

¹Note that the classic QSM [4] is the basic case of the GQSM with P = 1.

²The natural subsequent extension to the full GQSM with unknown transmit symbols will be addressed in an upcoming journal article.



(a) Non-decoupled univariate model in eq. (5b) as the SotA [6]-[9].



(b) Proposed fully-decoupled 2*P*-multivariate UVD model in eq. (9). Fig. 1: Different FG representations of the GQSM system.

In light of the above, the random vector variable a is introduced to model the unit vectors, where the discrete uniform prior probability mass function (PMF) is given by

$$\mathbb{P}_{\mathbf{a}}(\mathbf{a}) \triangleq \frac{1}{N_T} \cdot \sum_{\mathbf{e}_t \in \mathcal{A}} \delta(\mathbf{e}_t - \mathbf{a}), \tag{8}$$

where a denotes an instance of a, $\mathcal{A} \triangleq \{\mathbf{e}_t\}_{t=1}^{N_T} \in [0, 1]^{N_T \times 1}$ is the event set of a, and $\delta(\cdot)$ denotes the unit impulse function where $\delta(\mathbf{x}) = 1$ if $|\mathbf{x}|_0 = 0$, and $\delta(\mathbf{x}) = 0$ otherwise.

Since the 2P vector variables are instances of the variable a, the estimation problem is rewritten into the UVD form as

$$\boldsymbol{y} = \boldsymbol{H} \begin{bmatrix} \sum_{p=1}^{P} s_p^{\mathrm{R}} \mathbf{a}_p^{\mathrm{R}} \\ \sum_{p=1}^{P} s_p^{\mathrm{I}} \mathbf{a}_p^{\mathrm{I}} \end{bmatrix} + \boldsymbol{w} \in \mathbb{R}^{2N_R \times 1}, \quad (9)$$

where the random variables $\mathbf{a}_1^{\mathrm{R}}, \cdots, \mathbf{a}_P^{\mathrm{R}}$ respectively model the unit vectors $\mathbf{e}_{k_1^{\mathrm{R}}}, \cdots, \mathbf{e}_{k_P^{\mathrm{R}}}$ and likewise for $\mathbf{a}_1^{\mathrm{I}}, \cdots, \mathbf{a}_P^{\mathrm{I}}$, which has been illustrated as a factor graph (FG) in Fig. 1b.

B. Vector-valued Gaussian Belief Propagation (GaBP)

In light of the above, this section provides the derivation of the purpose-fit vector-valued MP rules operating on the factor graph of Fig. 1b, based on the Gaussian belief propagation (GaBP) framework [10], [11], assuming perfect channel state information (CSI) at the receiver. This enables the joint estimation of the 2P activation vector variables (unit vectors) only within their respective signal domains of size N_T each.

First, the soft-replica vectors for the activation vector variables $\mathbf{a}_p^{\mathrm{R}}$ and $\mathbf{a}_p^{\mathrm{I}}$ for $p \in \{1, \dots, P\}$ are defined as $\hat{\mathbf{a}}_{p:n}^{\mathrm{R}}$ and $\hat{\mathbf{a}}_{p:n}^{\mathrm{I}}$ respectively for the *n*-th factor node with $n \in \{1, \dots, 2N_R\}$. The corresponding expected error covariance matrix of the soft-replica $\hat{\mathbf{a}}_{p:n}^{\mathrm{R}}$ is defined as

$$\boldsymbol{\Gamma}_{p:n}^{\mathrm{R}} \triangleq \mathbb{E}_{\mathbf{a}}[(\mathbf{a} - \hat{\mathbf{a}}_{p:n}^{\mathrm{R}})(\mathbf{a} - \hat{\mathbf{a}}_{p:n}^{\mathrm{R}})^{\mathsf{T}}].$$
(10)

Remark: Due to page limitations, the derivations are provided only for the real components (i.e., for $\mathbf{a}_p^{\mathrm{R}}$), as the expressions for the respective imaginary components are <u>identical</u>, except for the change of superscripts $(\cdot)^{\mathrm{R}}$ to $(\cdot)^{\mathrm{I}}$ and vice-versa.

In hand of the soft-replicas, the factor nodes perform softinterference cancellation (IC) on the received signals y_n as $\bar{y}_{p:n}^{\mathrm{R}} = y_n - \mathbf{h}_n^{\mathrm{R}^{\mathrm{T}}} \sum_{p'\neq p}^{P} (s_{p'}^{\mathrm{R}} \hat{\mathbf{a}}_{p':n}^{\mathrm{R}}) - \mathbf{h}_n^{\mathrm{I}^{\mathrm{T}}} \sum_{p'=1}^{P} (s_{p'}^{\mathrm{I}} \hat{\mathbf{a}}_{p':n}^{\mathrm{I}})$ (11) $= \underbrace{\mathbf{h}_n^{\mathrm{R}^{\mathrm{T}}} s_p^{\mathrm{R}} \mathbf{a}_p^{\mathrm{R}}}_{p \neq p} + \mathbf{h}_n^{\mathrm{R}^{\mathrm{T}}} \sum_{p'\neq p}^{P} s_{p'}^{\mathrm{R}} (\hat{\mathbf{a}}_{p'}^{\mathrm{R}} - \hat{\mathbf{a}}_{p':n}^{\mathrm{R}}) + \mathbf{h}_n^{\mathrm{I}^{\mathrm{T}}} \sum_{p'=1}^{P} s_{p'}^{\mathrm{I}} (\hat{\mathbf{a}}_{p'}^{\mathrm{I}} - \hat{\mathbf{a}}_{p':n}^{\mathrm{I}}) + w_n,$

where
$$\mathbf{h}_{n}^{\mathrm{R}^{\mathsf{T}}} \in \mathbb{R}^{1 \times N_{T}}$$
 and $\mathbf{h}_{n}^{\mathrm{I}^{\mathsf{T}}} \in \mathbb{R}^{1 \times N_{T}}$ respectively denote the
n-th rows of the channel components $\boldsymbol{H}^{\mathrm{R}} \in \mathbb{R}^{2N_{T} \times N_{T}}$ and
 $\boldsymbol{H}^{\mathrm{I}} \in \mathbb{R}^{2N_{T} \times N_{T}}$ which are defined from $\boldsymbol{H} \triangleq [\boldsymbol{H}^{\mathrm{R}}, \boldsymbol{H}^{\mathrm{I}}]$.

The sum of the latter error terms and AWGN term w_n , excluding the true symbol part, is approximated as a Gaussian scalar via the central limit theorem (CLT), which yield the conditional probability density functions (PDFs) of the soft-IC symbols with respect to a given activation vector $\mathbf{a}_n^{\mathrm{R}}$ as

$$\mathbb{P}(\bar{y}_{p:n}^{\mathrm{R}}|\mathbf{a}_{p}^{\mathrm{R}}) \propto \exp\left(-\frac{|\bar{y}_{p:n}^{\mathrm{R}} - s_{p}^{\mathrm{R}}\mathbf{h}_{n}^{\mathrm{R}}\mathbf{a}_{p}^{\mathrm{R}}|^{2}}{\nu_{p:n}^{\mathrm{R}}}\right), \qquad (12)$$

where the conditional variance $\nu_{p:n}^{R}$ is obtained by

$$\nu_{p:n}^{\mathrm{R}} = \mathbb{E}\left[|\bar{y}_{p:n}^{\mathrm{R}} - s_{p}^{\mathrm{R}}\mathbf{h}_{n}^{\mathrm{R}}\mathbf{a}_{p}^{\mathrm{R}}|^{2}\right]$$
$$= \nu_{n} - \mathbf{h}_{n}^{\mathrm{R}^{\mathrm{T}}}\left(|s_{p}^{\mathrm{R}}|^{2} \cdot \boldsymbol{\Gamma}_{p:n}^{\mathrm{R}}\right)(\mathbf{h}_{n}^{\mathrm{R}})^{*} + \frac{N_{0}}{2}, \qquad (13)$$
$$\triangleq \mathbf{h}^{\mathrm{R}^{\mathrm{T}}}\left(\sum_{i=1}^{P} |s_{i}^{\mathrm{R}}|^{2} \cdot \boldsymbol{\Gamma}_{i}^{\mathrm{R}}\right)\mathbf{h}^{\mathrm{R}^{*}} + \mathbf{h}^{\mathrm{I}^{\mathrm{T}}}\left(\sum_{i=1}^{P} |s_{i}^{\mathrm{R}}|^{2} \cdot \boldsymbol{\Gamma}_{i}^{\mathrm{R}}\right)\mathbf{h}^{\mathrm{I}^{*}}.$$

with $\nu_n \triangleq \mathbf{h}_n^{\mathrm{R}^{\mathrm{I}}} \left(\sum_{p=1} |s_p^{\mathrm{R}}|^2 \cdot \boldsymbol{\Gamma}_{p:n}^{\mathrm{R}} \right) \mathbf{h}_n^{\mathrm{R}^{\mathrm{T}}} + \mathbf{h}_n^{\mathrm{I}^{\mathrm{I}}} \left(\sum_{p=1} |s_p^{\mathrm{I}}|^2 \cdot \boldsymbol{\Gamma}_{p:n}^{\mathrm{I}} \right) \mathbf{h}_n^{\mathrm{I}^{\mathrm{T}}}.$ Then, each variable node aggregates the conditional PDFs

from the connected factor nodes to compute the extrinsic belief $b_{p:n}^{\rm R}$ with self-interference cancellation, following

$$\mathbb{P}(b_{p:n}^{\mathrm{R}}|\mathbf{a}_{p}^{\mathrm{R}}) = \prod_{n'\neq n}^{2N_{R}} \mathbb{P}(\bar{y}_{p:n'}^{\mathrm{R}}|\mathbf{a}_{p}^{\mathrm{R}}) \propto e^{(\boldsymbol{\eta}_{p:n}^{\mathrm{R}\mathsf{T}}\mathbf{a}_{p}^{\mathrm{R}} - \frac{1}{2}\mathbf{a}_{p}^{\mathrm{R}\mathsf{T}}\boldsymbol{\Lambda}_{p:n}^{\mathrm{R}}\mathbf{a}_{p}^{\mathrm{R}})},$$
(14)

where $\eta_{p:n}^{\text{R}}$ and $\Lambda_{p:n}^{\text{R}}$ are the information vector and the precision matrix of the extrinsic belief $b_{p:n}^{\text{R}}$, given by

$$\boldsymbol{\eta}_{p:n}^{\mathrm{R}} = s_{p}^{\mathrm{R}} \sum_{n' \neq n}^{2N_{R}} \frac{\bar{y}_{p:n'}^{\mathrm{R}}}{\nu_{p:n'}^{\mathrm{R}}} \mathbf{h}_{n'}^{\mathrm{R}} \text{ and } \boldsymbol{\Lambda}_{p:n}^{\mathrm{R}} = |s_{p}^{\mathrm{R}}|^{2} \sum_{n' \neq n}^{2N_{R}} \frac{\mathbf{h}_{n'}^{\mathrm{R}} \mathbf{h}_{n'}^{\mathrm{R}}}{\nu_{p:n'}^{\mathrm{R}}}.$$
 (15)

In turn, the posterior Bayes-optimal soft-replicas are computed from the extrinsic beliefs via

$$\hat{\mathbf{a}}_{p:n}^{\mathrm{R}} = \frac{\mathbb{E}_{\mathsf{a}}[\mathbf{a} \cdot \mathbb{P}(b_{p:n}^{\mathrm{R}} | \mathbf{a})]}{\mathbb{E}_{\mathsf{a}}[\mathbb{P}(b_{p:n}^{\mathrm{R}} | \mathbf{a})]} = \frac{\sum_{q=1}^{N_{T}} \mathbf{a}_{q} \cdot \mathbb{P}(b_{p:n}^{\mathrm{R}} | \mathbf{a}_{q}) \cdot \mathbb{P}(\mathbf{a}_{q})}{\sum_{q'=1}^{N_{T}} \mathbb{P}(b_{p:n}^{\mathrm{R}} | \mathbf{a}_{q'}) \cdot \mathbb{P}(\mathbf{a}_{q'})}, (16)$$

while the corresponding error covariance matrix is given by

$$\boldsymbol{\Gamma}_{p:n}^{\mathrm{R}} \triangleq \mathbb{E}_{\mathsf{a}}[(\mathbf{a} - \hat{\mathbf{a}}_{p:n}^{\mathrm{R}})(\mathbf{a}_{p} - \hat{\mathbf{a}}_{p:n}^{\mathrm{R}})^{\mathsf{H}}].$$
(17)

Equations (11)-(17) describe the steps of one MP iteration to estimate the 2P activation vectors of the GQSM reformulated as eq. (9), which yields the refined posterior soft-replica vectors and the corresponding error covariance matrices. In addition, at the end of such τ -th MP iteration, the soft-replica vectors and the error covariance matrices are updated with damping [12] to prevent an early convergence to a local optima [13], [14], following $x^{[\tau+1]} \leftarrow \rho x^{[\tau]} + (1 - \rho) x^{[\tau+1]}$ where $\rho \in [0, 1]$ is the damping factor, and τ is the iteration number.

Next, to obtain the hard-decisions on the 2P activation vectors, a the information is aggregated between all $2N_R$ factor nodes to compute the consensus beliefs (*i.e.*, eq. (14) without the self-interference cancellation), which yields the extrinsic consensus PDFs of the belief b_n^R .

Finally, the optimal activation vector estimate is selected by evaluating the PDFs for the N_T valid states of $\mathbf{a} \in \mathcal{A}$ *i.e.*,

$$\tilde{\mathbf{a}}_{p}^{\mathrm{R}} = \underset{\mathbf{a}\in\mathcal{A}}{\operatorname{argmax}} \ \mathbb{P}(b_{p}^{\mathrm{R}}|\mathbf{a}).$$
(18)

The proposed UVD-GaBP decoder for pilotted GQSM, described by eq. (10)-(18), is summarized in Algorithm 1.

Algorithm 1: Proposed UVD-GaBP GQSM Decoder

IV. PERFORMANCE EVALUATION

A. Simulation Results

In Fig. 2, GQSM simulation results are provided for mMIMO systems with $N_T = 16$ and 32, and varying values of P, where the performance is evaluated in terms of the BER against the E_b/N_0 (signal-to-noise ratio (SNR) per bit).

Note the extremely low E_b/N_0 ranges of the GQSM, which benefits from the fact that most information is encoded without using any transmission power, corroborating the original motivation of enabling energy- and spectral-efficient mMIMO.

Since the computational complexity of the brute-force ML and SotA decoders are prohibitive in the considered system scales, a Genie-aided matched filter bound (MFB) [15] is introduced instead as an absolute performance bound, which is obtained by providing the UVD-GaBP method with perfect prior knowledge of the activation vectors and pilot symbols.

Fig. 2 demonstrates the efficient demodulation capability of the proposed UVD-GaBP decoder for high-rate GQSM signals in 16×16 and 32×32 mMIMO³ setups⁴, even with the affordable computational power of an average personal computer. It can be observed that optimal performance is achieved by the UVD-GaBP for P = 1 in both scenarios, and a slight performance loss of about $1 \sim 2$ dB and an errorfloor is exhibited at high E_b/N_0 , for increasing $P \ge 1$. However, notice that the negative effect in both the BER performance and the error-floor is reduced in the larger system with $N_T = 32$, which benefits from the increased sparsity in the system and the consequently increased orthogonality in the unit vector random variables.

Further improvements including the elimination of errorfloors can be expected if the joint distributions of the activation vectors and their non-uniform prior distributions are introduced in the MP design, and a non-piloted variation of the method is also under investigation, which will be addressed in an upcoming journal article.

³Unbalanced MIMO scenarios will be investigated in future works.

⁴Fixed MP parametrization have been used for all scenarios, with damping factor $\rho = 0.5$ and number of MP iterations $\tau = 100$.

BER Performance of GQSM with the Proposed UVD-GaBP Decoder $N_T = 16, N_R = 16$ $N_T = 32, N_R = 32$



Fig. 2: BER performance of the proposed (Prop.) UVD-GaBP decoder and its Genie-aided matched filter bound (MFB).

B. Complexity Analysis

Table I compares the decoding complexity of the proposed UVD-GaBP method against few SotA algorithms⁵, where it can be seen that the SotA methods have reduced the squared-combinatorial term appearing in the brute-force ML search.

Namely, the relaxed orthogonal matching pursuit-based ordered successive IC (ROMP-OSIC) decoder proposed in [9] relaxes the upper index of the binomial coefficient by a factor N_v with $P \leq N_v \leq N_T$, whereas the IQ-decoupled GaBP decoder proposed in [10] eliminates the quadratic factor on the binomial coefficient, and τ_{max} is the number of MP iterations. However, the SotA methods still retain the binomial coefficient which is not scalable to mMIMO system of consideration.

On the other hand, the proposed UVD-GaBP decoder enjoys a significantly reduced complexity⁶ which is completely independent of the binomial coefficient, enabling the decodability of GQSM schemes in significantly larger mMIMO systems, as verified in the performance evaluation results.

GQSM Decoding Algorithm	Decoding Complexity Order
Brute-force ML Search [4]	$\mathcal{O}\left[\binom{N_T}{P}^2 P N_R\right]$
ROMP-OSIC [9]	$\mathcal{O}\big[\binom{N_v}{P}^2 N_v{}^2 P N_T N_R\big]$
IQ-decoupled GaBP [10]	$\mathcal{O}\left[au_{\max} \cdot {N_T \choose P} N_T^2 N_R ight]$
Proposed UVD-GaBP	$\mathcal{O}[\tau_{\max} \cdot P N_T^2 N_R]$

TABLE I: Complexity orders of various GQSM decoders.

V. CONCLUSION

We paved the way towards feasible energy- and spectralefficient mMIMO systems with high-performance GQSM, by proposing a novel GaBP-based decoder exploiting a UVD of the GQSM signal model and pilots, which is shown to achieve a complexity order that is independent of combinatorial factors. Simulation results verify the effectiveness of the method.

In addition, the multi-user scenario should also be considered to support the B5G mMIMO access expectations.

⁵For fairness, the complexity of the symbol-level detection have been disregarded for the SotA methods, since a fully pilotted scenario is considered.

⁶Since the variables \mathbf{a}_p are unit vectors, the corresponding soft-replicas and covariance matrices reach high sparsity at convergence, such that the practical computational complexity is further reduced with each MP iteration.

REFERENCES

- 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.
- [2] Z. Zhang, Y. Xiao, Z. Ma, M. Xiao, Z. Ding, X. Lei, G. K. Karagiannidis, and P. Fan, "6G wireless networks: Vision, requirements, architecture, and key technologies," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 28–41, 2019.
- [3] T. S. Rappaport, Y. Xing, O. Kanhere, S. Ju, A. Madanayake, S. Mandal, A. Alkhateeb, and G. C. Trichopoulos, "Wireless communications and applications above 100 GHz: Opportunities and challenges for 6G and beyond," *IEEE Access*, vol. 7, pp. 78729–78757, 2019.
- [4] R. Mesleh, S. S. Ikki, and H. M. Aggoune, "Quadrature spatial modulation," *IEEE Trans. Veh. Technol.*, vol. 64, no. 6, pp. 2738–2742, Jun. 2015.
- [5] E. Başar, U. Aygölü, E. Panayirci, and H. V. Poor, "Space-time block coded spatial modulation," *IEEE Transactions on Communications*, vol. 59, no. 3, pp. 823–832, 2011.
- [6] H. S. Rou, G. T. F. Abreu, H. Iimori, D. González G., O. Gonsa, "Scalable quadrature spatial modulation," 2022, submitted to Trans. Wireless Commun.
- [7] L. Wang and Z. Chen, "Enhanced diversity-achieving quadrature spatial modulation with fast decodability," *IEEE Trans. Veh. Technol.*, vol. 69, no. 6, pp. 6165–6177, 2020.
- [8] J. Li, X. Jiang, Y. Yan, W. Yu, S. Song and M. Lee, "Low complexity detection for quadrature spatial modulation systems," *Wireless Person. Commun.*, vol. 95, no. 4, pp. 4171–4183, 2017.
- [9] J. An, C. Xu, Y. Liu, L. Gan, and L. Hanzo, "The achievable rate analysis of generalized quadrature spatial modulation and a pair of low-complexity detectors," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 5, pp. 5203–5215, 2022.
- [10] H. S. Rou, G. T. F. de Abreu, and T. Takahashi, "An efficient vectorvalued belief propagation decoder for quadrature spatial modulation," in 2022 56th Asilomar Conference on Signals, Systems, and Computers, 2022, pp. 27–31.
- [11] D. Bickson, "Gaussian belief propagation: Theory and aplication," arXiv preprint arXiv:0811.2518, 2008.
- [12] P. Som, T. Datta, A. Chockalingam and B. S. Rajan, "Improved largemimo detection based on damped belief propagation," in *IEEE Inf. Theory Workshop on Inf. Theory*, 2010, pp. 1–5.
- [13] Q. Su and Y.-C. Wu, "On convergence conditions of gaussian belief propagation," *IEEE Transactions on Signal Processing*, vol. 63, no. 5, pp. 1144–1155, 2015.
- [14] J. Du, S. Ma, Y.-C. Wu, S. Kar, and J. M. Moura, "Convergence analysis of distributed inference with vector-valued gaussian belief propagation." *J. Mach. Learn. Res.*, vol. 18, no. 1, pp. 6302–6339, 2017.
- [15] T. Takahashi, S. Ibi, and S. Sampei, "Design of adaptively scaled belief in multi-dimensional signal detection for higher-order modulation," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 1986–2001, 2019 2019.