Performance Analysis of Non-linear Generalised Pre-coding Aided Spatial Modulation

Rong Zhang, Lie-Liang Yang and Lajos Hanzo
Southampton Wireless, School of ECS, University of Southampton, SO17 1BJ, UK
Email: rz@ecs.soton.ac.uk, http://www-mobile.ecs.soton.ac.uk

***Abstract*—Developed from the recently emerged Generalised Pre-coding aided Spatial Modulation (GPSM) concept, a novel non-linear GPSM scheme based on the powerful Vector Perturbation (VP) philosophy is proposed, where a particular subset of *receive* antennas is activated and the specific activation pattern itself conveys useful implicit information in addition to the conventional modulated and perturbed symbols. Explicitly, both the infinite and the finite alphabet capacity are derived for the proposed non-linear GPSM scheme. The associated complexity, energy efficiency and error probability are also investigated. Our numerical results show that, as the only known non-linear realisation within the Spatial Modulation (SM) family, the proposed scheme constitutes to be an attractive solution to the flexible design of 'green' transceivers, since it is capable of striking a compelling compromise amongst the key performance indicators of throughput, energy consumption, complexity and performance. Particularly, in the challenging full-rank scenario, conveying information through RA indices exhibits a lower complexity, a higher energy efficiency and a better error resilience than that of the conventional arrangement.**

I. Introduction

1) Background: Multiple Input Multiple Output (MIMO) constitutes one of the most promising technical advances in wireless communications, since it facilitates high-throughput transmissions in the context of various standards [1]. Hence, it attracted substantial research interests, leading to the Vertical-Bell Laboratories Layered Space-Time (V-BLAST) [2] and Space Time Coding (STC) [3], etc. The point-to-point single-user MIMO systems are capable of offering diverse transmission functionalities in terms of multiplexing, diversity and beam-forming gains, while in a multi-user MIMO context [4], Space Division Multiple Access (SDMA) employed constitutes a beneficial building block both in the uplink and downlink. The basic benefits of MIMO have also been further exploited in the context of the Network MIMO [5], cooperative MIMO [6], massive MIMO [7] and interference-limited MIMO [8] concepts. Their applications have also been expanded to the areas of cognitive radio [9], physical-layer security [10], [11] and wireless powered communications [12].

Despite having a plethora of studies on classic MIMO systems, their practical constraints, such as their I/Q imbalance, transmitter and receiver complexity, the cost of using multiple Radio Frequency (RF) Power Amplifier (PA) chains as well as their Digital-Analogue / Analogue-Digital (DA/AD) converters have received limited attention. To circumvent these problems, low complexity alternatives to conventional MIMO transmission schemes have been proposed, which include the Antenna Selection (AS) [13], [14] and the Spatial Modulation (SM) philosophies [15], [16]. More specifically, SM constitutes a novel MIMO technique, which was conceived for providing a higher throughput than a single-antenna aided system, while maintaining both a lower complexity and a lower cost than that of the conventional MIMOs, since it may rely on a reduced number of RF up-conversion chains. To elaborate a little further, SM conveys extra information by mapping $\log_2(N_t)$ bits to the N_t Transmit Antenna (TA) indices, in addition to the classic modulation schemes [15].

By contrast, the family of Pre-coding aided Spatial Modulation (PSM) schemes is capable of conveying extra information by appropriately selecting the Receive Antenna (RA) indices, where the indices of the RA represent additional information in the spatial domain, as detailed in [17]. As a specific counterpart of the original SM, PSM benefits from both a low cost and a low complexity at the receiver side, therefore it may be considered to be eminently suitable for downlink transmissions. The further improved concept of Generalised PSM (GPSM) was proposed in [18], where comprehensive performance comparisons were carried out between the GPSM scheme as well as the conventional MIMO scheme and the associated detection complexity issues were discussed, along with the investigation of a range of practical issues. Analytical studies of GPSM were revealed in [19], which provided the upper bound of both the Symbol Error Ratio (SER) and Bit Error Ratio (BER) expressions of the GPSM scheme. Furthermore, the corresponding finite alphabet capacity as well as the achievable rate were quantified.

Since then several fundamental contributions have been made to extend the original scope of the GPSM scheme. In [20], the GPSM concept had been extended to the multi-user/multi-stream scenario, where both accurate BER expressions and the achievable diversity gains were revealed. In [21], low-complexity receive antenna subset selection was proposed in order to extend the original GPSM concept from low user-loaded MIMO settings to rank-deficient MIMO settings, complemented by a range of further improvements. In [22], the concept of dual-layered MIMO was proposed, which intrinsically amalgamated the classic spatial multiplexing and the GPSM scheme. In [23], a large-scale MIMO regime was proposed, where the user's sub-channel index was exploited for conveying extra implicit information. In [24], the

The financial support of the EPSRC projects (EP/N004558/1 and EP/N023862/1) and that of the European Research Council's (ERC) Advanced Fellow Grant is gratefully acknowledged.

conventional SM and the GPSM scheme were beneficially combined into hybrid SM for dual-hop relay-aided communications. In [25], a novel Pre-coding aided Differential Spatial Modulation (PDSM) was proposed, which is an extension of the Differential Spatial Modulation (DSM) proposed for the first time in [26]. Finally, various applications of the GPSM concept were also proposed, including the improvement of its usage in physical layer security [27] and its exploitation in wireless powered communications [28].

2) Scope: In contrast to all the existing GPSM schemes, which rely on linear pre-coding [29], it appears beneficial to explore the non-linear pre-coding aided GPSM scheme. To the best of the authors' knowledge, there has been no literature on this topic. It is a well-known fact that non-linear pre-coding is capable of outperforming its linear counterpart, especially in full-rank (even rank-deficient) scenarios. These challenging scenarios are becoming more and more prominent in the 5G era aiming for supporting the Internet of Things (IoT). Indeed, non-linear processing constitutes one of the recent research advances in the physical layer, as evidenced for example by the Non-Orthogonal Multiple Access (NOMA) air-interface concept [30] and non-linear vectoring in the next-generation ultra-fast copper techniques [31]. Historically speaking, non-linear pre-coding has been overlooked owing to its high implementation complexity and difficulty in analysis. With the recent advances in high-throughput fully parallel hardware architectures, the implementation of non-linear pre-coding has become much more affordable and hence the technique is now much closer to the market.

This motivates us to consider the proposed non-linear GPSM scheme employing Vector Perturbation (VP) [32], which subsumes the conventional VP as a special case. More explicitly, both the infinite and the finite alphabet capacity are characterized. Furthermore, tight upper bounds of the SER expressions are derived and finally, both its complexity as well as its energy efficiency are discussed. Our main finding is that the proposed scheme constitutes a promising solution for energy efficient transceivers, which is capable of striking a beneficial compromise amongst the bandwidth efficiency, energy efficiency, complexity and error resilience. More particularly, in the challenging full-rank scenario, conveying additional information through the RA indices exhibits a lower complexity, a higher energy efficiency and a better error resilience than that of its conventional counterpart.

The rest of the paper is organised as follows. In Section II, we introduce the underlying concept of the proposed non-linear GPSM scheme and discuss its transceiver structure. This is followed by our performance analysis in Section III, where both the capacity, the SER expressions and the complexity are discussed. Our simulation results are then provided in Section IV, and finally we conclude in Section V.

Notations: We use bold upper case to represent matrix and bold lower case to represent vector. We use $(\cdot)^T$, $(\cdot)^H$ and $(\cdot)^{-1}$ to represent the transpose, the conjugate transpose and the inverse of a matrix, respectively. Furthermore, we use $\lfloor\cdot\rfloor$ and $|\cdot|$ to represent the flooring operator and the cardinality of a set, respectively. Finally, we use $\mathcal{R}(\cdot)$ and $\mathcal{I}(\cdot)$ to represent the real and imaginary operators, respectively and correspondingly we use $(\cdot)^0$ to represent the real or imaginary part.

II. System Model

A. Concept

Consider a MIMO system equipped with N_t TAs and N_r RAs, where we have $N_t \geq N_r$. In GPSM schemes, a total of $N_a \leq N_r$ RAs are activated so as to facilitate the simultaneous transmission of N_a data streams, where the particular *pattern* of the N_a RAs activated conveys extra information in form of the so-called spatial symbols in addition to the information carried by the conventional modulated symbols. Hence, the number of bits conveyed by a spatial symbol of our GPSM scheme becomes $k_{ant} = \lfloor\log_2(|\mathcal{C}_t|)\rfloor$, where the set $\mathcal{C}_t$ contains all the combinations associated with choosing N_a activated RAs out of N_r RAs, although it is also possible to have non-integer values of k_{ant} with the aid of fractional bit encoding [33]. For example, when choosing 2 out of 4 RAs, a maximum of $\log_2 6$ bits of information can be conveyed. As a result, the total number of bits transmitted by one GPSM symbol is

$$k_{eff} = k_{ant} + N_a k_{mod}, \tag{1}$$

where $k_{mod} = \log_2(M)$ denotes the number of bits per symbol of a conventional M-ary Gray-mapped QAM scheme having an alphabet denoted by $\mathcal{A}$ and we consider even value of k_{mod} in this paper. Finally, it is plausible that a classic MIMO arrangement obeys $N_a = N_r$, where a maximum of $k_{eff} = N_r k_{mod}$ bits may be supported. For assisting our further discussions, we let $\mathcal{C}$ denote the selected set of RA activation pattens. We also let $\mathcal{C}_k$ and $\mathcal{C}_{k,i}$ denote the kth RA activation pattern and the ith activated RA in the kth activation pattern, respectively.

B. Transmitter

Let $\boldsymbol{s}_m^k$ be an explicit representation of a super-symbol $\boldsymbol{s} \in \mathbb{C}^{N_r \times 1}$, indicating that the kth RA activation pattern $\mathcal{C}_k$ is employed and the mth conventional modulated symbols of the set $\boldsymbol{b}_m = [b_{m_1}, \cdots, b_{m_{N_a}}]^T$ is transmitted, where $\boldsymbol{b}_m \in \mathcal{A}_0^{N_a}$ and $\mathcal{A}_0 = \mathcal{A}/2\sqrt{M}$. In other words, we have

$$\boldsymbol{s}_m^k = \boldsymbol{G}_k \boldsymbol{b}_m, \tag{2}$$

where $\boldsymbol{G}_k = \boldsymbol{I}_{N_r}[:, \mathcal{C}_k]$ is constituted by the specifically selected columns determined by $\mathcal{C}_k$ of an identity matrix.

In this paper, we use the powerful VP concept of [32] for the proposed non-linear GPSM scheme. Instead of transmitting the conventional modulated symbols $\boldsymbol{b}_m$ as in its linear GPSM counterpart, the perturbed version is formed, namely we construct [32]

$$\tilde{\boldsymbol{b}}_m = \boldsymbol{b}_m + \boldsymbol{p}, \tag{3}$$

where $\boldsymbol{p} \in \mathbb{Z}^{N_a \times 1} + j\mathbb{Z}^{N_a \times 1}$ is an N_a dimensional complex-valued perturbation vector. Accordingly, (2) becomes

$$\tilde{\boldsymbol{s}}_m^k = \boldsymbol{G}_k \tilde{\boldsymbol{b}}_m. \tag{4}$$

When applying the transmit pre-coding matrix $\boldsymbol{P} \in \mathbb{C}^{N_t \times N_r}$, the resultant transmit signal $\tilde{\boldsymbol{x}} \in \mathbb{C}^{N_t \times 1}$ may be written as

$$\tilde{\boldsymbol{x}} = \sqrt{\beta/N_a}\boldsymbol{P}\tilde{\boldsymbol{s}}_m^k, \tag{5}$$

where the scaling factor β is given by

$$\beta = N_r/\mathbb{E}[||\boldsymbol{P}\tilde{\boldsymbol{s}}_m^k||^2], \tag{6}$$

which is introduced to maintain a constant average transmit power of $\mathbb{E}[||\tilde{\boldsymbol{x}}||^2] = 1$ after transmit pre-coding.

To elaborate a little further, when VP is employed, the transmit pre-coding matrix $\boldsymbol{P}$ may be formulated as [29]

$$\boldsymbol{P} = \boldsymbol{H}^H(\boldsymbol{H}\boldsymbol{H}^H)^{-1}, \tag{7}$$

where $\boldsymbol{H} \in \mathbb{C}^{N_r \times N_t}$ represents the MIMO channel encountered. We assume in this paper that each entry of the MIMO channel undergoes frequency-flat Rayleigh fading and it is uncorrelated between different super-symbol transmissions, while it remains constant within the duration of a super-symbol's transmission [1]. After formulating the transmit pre-coding matrix $\boldsymbol{P}$ as in (7), the optimal perturbation vector $\boldsymbol{p}^*$ for (3) may be obtained by solving the integer least square problem of [32]

$$\boldsymbol{p}^* = \arg\min_{\boldsymbol{p}} ||\boldsymbol{P}\tilde{\boldsymbol{s}}_m^k||^2 = \arg\min_{\boldsymbol{p}} ||\boldsymbol{P}\boldsymbol{G}_k(\boldsymbol{b}_m + \boldsymbol{p})||^2, \tag{8}$$

which may be achieved by employing various forms of sphere search algorithms detailed for example in [34].

Remark: Conventional VP constitutes a special case of our proposed non-linear GPSM scheme associated with $N_a = N_r$. Without VP, the scaling factor β of (6) may be very small, since the channel inversion aided linear pre-coding of (7) imposes a reduction in received signal power. Hence, the aim of VP is to minimise this reduction as seen in (8), where the underlying principle is to find the new symbol vector $\tilde{\boldsymbol{b}}_m$ that best matches the eigen-values of the equivalent channel $\boldsymbol{P}\boldsymbol{G}_k$. This is achieved by searching for the optimal integer perturbation vector $\boldsymbol{p}^*$. Because the new symbol vector is an integer shifted version of the original symbol vector $\boldsymbol{b}_m$, a simple modulo operation may be employed at the receiver to retrieve the information.

The optimal perturbation vector $\boldsymbol{p}^*$ may be obtained by solving the problem defined in (8), which is an integer least square problem. This problem can also be interpreted as finding the closest point in the associated lattice, which is a classic problem in number theory, that has predominantly found applications in the area of cryptography. There are mature solutions to this problem such as the Lattice Aided Search (LAS) of [34]. To elaborate a little further, LAS is a family of algorithms combining the lattice-basis reduction step followed by a search step, such as the widely used Lenstra-Lenstra-Lovász (LLL) lattice reduction combined with the classic Schnorr-Euchner search [34]. This particular LAS algorithm is also known as sphere decoding in the area of communications detection. Analogously, VP may be also referred to as sphere encoding, apart from the difference that sphere decoding treats a finite lattice, while sphere encoding treats an infinite lattice.

[1]This assumption has been commonly used in the literature to facilitate ergodic capacity analysis and average error probability analysis. Other assumptions are possible, but they may require different analytical methodologies.

C. Receiver

After few normalisations, the signal observed at the N_r RAs may be written as

$$\tilde{\boldsymbol{y}} = \boldsymbol{H}\boldsymbol{P}\tilde{\boldsymbol{s}}_m^k + \tilde{\boldsymbol{w}} = \tilde{\boldsymbol{s}}_m^k + \tilde{\boldsymbol{w}}, \tag{9}$$

where $\tilde{\boldsymbol{w}} \in \mathbb{C}^{N_r \times 1}$ is the circularly symmetric complex Gaussian noise vector with each entry having a zero mean and a variance of $\sigma_{\tilde{w}}^2 = \sigma^2 N_a/\beta$. We also assume that the super-symbols transmitted are statistically independent from the noise.

Consequently, we may employ the *joint* detection of both the conventional modulated symbols $\boldsymbol{b}_m$ and of the RA activation pattern index k. More explicitly, having $\mathcal{A}_0 = \mathcal{A}/2\sqrt{M}$ implies that both the real and the imaginary parts of $b_{m_i}, \forall i \in [1, N_a]$ are constrained within the range of $\zeta^0 = [-1/2, 1/2]$. Hence, after passing the received signal $\tilde{\boldsymbol{y}}$ through the modulo operation $\wedge_\zeta$ over the range of $\zeta = \zeta^0 + j\zeta^0$, the effect of perturbation can be removed, hence we have

$$\boldsymbol{y} = \wedge_\zeta[\tilde{\boldsymbol{y}}] = \wedge_\zeta[\boldsymbol{s}_m^k + \tilde{\boldsymbol{w}}], \tag{10}$$

where the modulo operation $\wedge_\zeta$ is applied to each entries of $\tilde{\boldsymbol{y}}$, namely for the real and the imaginary parts of $\tilde{\boldsymbol{y}}$, separately. Then the joint Maximum Likelihood (ML) detector of the RA activation pattern index and of the conventional modulated symbols may be formulated as

$$[\hat{\boldsymbol{b}}_m, \hat{k}] = \arg\min_{\boldsymbol{s}_n^\ell \in \mathcal{B}} ||\boldsymbol{y} - \boldsymbol{s}_n^\ell||^2, \tag{11}$$

where $\mathcal{B} = \mathcal{C} \times \mathcal{A}_0^{N_a}$ represents the super-symbol alphabet. Alternatively, their reduced-complexity *decoupled* detection may also be employed, which treats the detection of the conventional modulated symbols $\boldsymbol{b}_m$ and of the RA activation pattern index k, separately. In this reduced-complexity variant, we have

$$\hat{k} = \arg\max_{\ell \in [1, |\mathcal{C}|]} \sum_{i=1}^{N_a} |\tilde{y}_{\mathcal{C}_{\ell,i}}|^2, \tag{12}$$

$$\hat{b}_{m_i} = \arg\min_{b_{n_i} \in \mathcal{A}_0} |y_{\mathcal{C}_{\hat{k},i}} - b_{n_i}|^2. \tag{13}$$

Thus, correct detection is declared, when we have $\hat{k} = k$ and $\hat{b}_{m_i} = b_{m_i}, \forall i \in [1, N_a]$. To elaborate a little further, the complexity of the ML detection of (11) is on the order determined by the super-alphabet $\mathcal{B}$, hence obeying $\mathcal{O}(|\mathcal{C}|M^{N_a})$. By contrast, the complexity of the decoupled detection is imposed by detecting the N_a conventional modulated symbols of (13), plus the complexity (κ) imposed by the comparisons invoked for non-coherently detecting the spatial symbol of (12), which may be written as $\mathcal{O}(N_a M + \kappa)$. More details can be found in [18].

Remark: Since only integer valued perturbation is imposed at the transmitter, a simple modulo operation may be employed to remove the perturbation effect before demodulation. More explicitly, passing $\tilde{\boldsymbol{y}}$ through the modulo operation $\wedge_\zeta$ may be mathematically written as $\wedge_\zeta[\tilde{\boldsymbol{y}}] = \tilde{\boldsymbol{y}} - \lfloor \tilde{\boldsymbol{y}} + \mathbb{1}_{N_r}(1+j)/2 \rfloor$, where $\mathbb{1}_{N_r}$ represents the column vector of length N_r. Whilst totally removing the perturbation effect, the modulo operation imposes extra noise, since the distribution of the noise term $\tilde{\boldsymbol{w}}$ seen in (10) changes from Gaussian to folded Gaussian.

III. Performance Analysis

We continue by investigating both the infinite and the finite alphabet capacity of the proposed non-linear GPSM scheme, when the joint ML detection of (11) is used. We then quantify its SER, when the more realistic reduced-complexity decoupled detection of (12) and (13) is employed. Finally, we discuss the complexity of the proposed non-linear GPSM scheme.

A. Infinite Alphabet Capacity

Let us now derive the infinite alphabet capacity of the proposed non-linear GPSM scheme dispensing with any specific discrete alphabet $\mathcal{A}_0$ and we use $\boldsymbol{b}$ to represent the conventional modulated symbols in a generic sense. Hence, the mutual information between the received signal $\boldsymbol{y}$ and the super-symbol $\boldsymbol{s}$ may be formulated as

$$I(\boldsymbol{s};\boldsymbol{y}) = I(\boldsymbol{b};\boldsymbol{y}|\mathcal{C}) + I(\mathcal{C};\boldsymbol{y}), \tag{14}$$

as a result of the chain rule of information theory [35].

Owing to the independence between of the RA activation patterns, the first term of the Right Hand Side (RHS) of (14) may be expanded as

$$I(\boldsymbol{b};\boldsymbol{y}|\mathcal{C}) = \frac{1}{|\mathcal{C}|}\sum_{k=1}^{|\mathcal{C}|} I(\boldsymbol{b};\boldsymbol{y}|\mathcal{C}_k), \tag{15}$$

where by closely investigating the kth summation term of (15), we have

$$I(\boldsymbol{b};\boldsymbol{y}|\mathcal{C}_k) = 2I(\boldsymbol{b}^0;\boldsymbol{y}^0|\mathcal{C}_k), \tag{16}$$

where the factor of 2 is introduced owing to the independence between the real and the imaginary parts of $\boldsymbol{b}$ and $\boldsymbol{y}$, where the mutual information for these two parts is identical in the infinite alphabet scenario. Elaborating a little further, we have

$$I(\boldsymbol{b}^0;\boldsymbol{y}^0|\mathcal{C}_k) = \sum_{i=1}^{N_a} I(b_i^0;y_{v_i}^0), \tag{17}$$

where $v_i = \mathcal{C}_{k,i}$ and (17) follows the fact that given $\mathcal{C}_k$, the received signals of the different RAs are independent and the mutual information is zero for the $(N_r - N_a)$ passive RAs. For the N_a activated RAs, the ith summation term of (17) can be expanded by definition as

$$I(b_i^0;y_{v_i}^0) = h(y_{v_i}^0) - h(y_{v_i}^0|b_i^0), \tag{18}$$

where the differential entropy of $y_{v_i}^0$ in the first term of the RHS of (18) is maximised when b_i^0 follows a uniform distribution. Hence $y_{v_i}^0$ obeys a uniform distribution over ζ^0 imposed by the modulo operation of (10). This is a direct result from the fact that the uniform distribution maximises the entropy for discrete inputs. Under this assumption, the distribution of $y_{v_i}^0$ is $f(y_{v_i}^0) = 1$ and hence we have $h(y_{v_i}^0) = 0$. Furthermore, the second term of the RHS of (18) can be expanded by definition as

$$h(y_{v_i}^0|b_i^0) = -\int_{\zeta^0} f(w_{v_i}^0)\log_2(f(w_{v_i}^0))dw_{v_i}^0, \tag{19}$$

where $w_{v_i}^0$ is the per dimensional folded Gaussian noise of $\wedge_\zeta(\tilde{w}_{v_i})$ and its distribution $f(w_{v_i}^0)$ may be formulated as [36]

$$f(w_{v_i}^0) = \sum_{z\in\mathbb{Z}} \frac{1}{\sqrt{\pi\sigma_{\tilde{w}}^2}} e^{-(w_{v_i}^0 - z)^2/\sigma_{\tilde{w}}^2}, \qquad w_{v_i}^0 \in \zeta^0. \tag{20}$$

Let us now focus on the second term of the RHS of (14), which can be expanded by definition as

$$I(\mathcal{C};\boldsymbol{y}) = \frac{1}{|\mathcal{C}|}\sum_{k=1}^{|\mathcal{C}|}\int_{\boldsymbol{y}} f(\boldsymbol{y}|\mathcal{C}_k)\log_2\left(\frac{f(\boldsymbol{y}|\mathcal{C}_k)}{f(\boldsymbol{y})}\right)d\boldsymbol{y}, \tag{21}$$

where owing to the independence of the real and the imaginary parts of $\boldsymbol{y}$, (21) may be further simplified as [2]

$$I(\mathcal{C};\boldsymbol{y}) = I(\mathcal{C};\boldsymbol{y}^0), \tag{22}$$

which may be explicitly written as

$$I(\mathcal{C};\boldsymbol{y}^0) = \frac{1}{|\mathcal{C}|}\sum_{k=1}^{|\mathcal{C}|}\int_{\boldsymbol{y}^0} f(\boldsymbol{y}^0|\mathcal{C}_k)\log_2\left(\frac{f(\boldsymbol{y}^0|\mathcal{C}_k)}{f(\boldsymbol{y}^0)}\right)d\boldsymbol{y}^0. \tag{23}$$

Following a few further manipulations, we have

$$\log_2\left(\frac{f(\boldsymbol{y}^0|\mathcal{C}_k)}{f(\boldsymbol{y}^0)}\right) = -\log_2\left(\frac{1}{|\mathcal{C}|}\sum_{\ell=1}^{|\mathcal{C}|}\frac{f(\boldsymbol{y}^0|\mathcal{C}_\ell)}{f(\boldsymbol{y}^0|\mathcal{C}_k)}\right), \tag{24}$$

where by substituting (24) into (23), we finally arrive at

$$I(\mathcal{C};\boldsymbol{y}^0) = \log_2(|\mathcal{C}|) - \frac{1}{|\mathcal{C}|}\sum_{k=1}^{|\mathcal{C}|}\mathbb{E}_{\boldsymbol{y}^0}\left[\log_2\left(\sum_{\ell=1}^{|\mathcal{C}|}\frac{f(\boldsymbol{y}^0|\mathcal{C}_\ell)}{f(\boldsymbol{y}^0|\mathcal{C}_k)}\right)\right], \tag{25}$$

where in (25), we have

$$f(\boldsymbol{y}^0|\mathcal{C}_k) = \prod_{v_i\in\mathcal{C}_k} f(y_{v_i}^0)\prod_{u_i\in\bar{\mathcal{C}}_k} f(w_{u_i}^0) = \prod_{u_i\in\bar{\mathcal{C}}_k} f(w_{u_i}^0), \tag{26}$$

where $\bar{\mathcal{C}}_k$ denotes the complementary set of $\mathcal{C}_k$, hosting the passive RA indices.

B. Finite Alphabet Capacity

When considering the finite alphabet capacity of the proposed non-linear GPSM scheme, the conventional modulated symbols obey a specific alphabet. In this paper, we only consider the scenario where the real and imaginary parts of the M-ary QAM are symmetric with an even value of k_{mod}. According to (14), (15), (16) and (22), the mutual information between the received signal $\boldsymbol{y}$ and the super-symbol $\boldsymbol{s}$ may be formulated as

$$I(\boldsymbol{s};\boldsymbol{y}) = \frac{1}{|\mathcal{C}|}\sum_{k=1}^{|\mathcal{C}|} 2I(\boldsymbol{b}^0;\boldsymbol{y}^0|\mathcal{C}_k) + I(\mathcal{C};\boldsymbol{y}^0). \tag{27}$$

[2] Note that, (22) is indeed correct, since the information embedded in the antenna pattern is one-dimensional. This is different from the information embedded in the conventional modulation, which is two-dimensional as suggested by (16).

Rearranging (27), we have

$$I(\boldsymbol{s};\boldsymbol{y}) = I(\boldsymbol{b}^0;\boldsymbol{y}^0|\mathcal{C}) + I(\mathcal{C};\boldsymbol{y}^0) + \frac{1}{|\mathcal{C}|}\sum_{k=1}^{|\mathcal{C}|} I(\boldsymbol{b}^0;\boldsymbol{y}^0|\mathcal{C}_k)$$
$$= \underbrace{I(\boldsymbol{s}^0;\boldsymbol{y}^0)}_{C_1} + \frac{1}{|\mathcal{C}|}\sum_{k=1}^{|\mathcal{C}|}\sum_{i=1}^{N_a}\underbrace{I(b_i^0;y_{v_i}^0)}_{C_2}, \tag{28}$$

where the last equation of (28) holds true according to (17).

Let us now elaborate in more detail on the two terms C_1 and C_2 of (28), respectively. On one hand, C_1 may be written by definition as

$$C_1 = \max_{p(\boldsymbol{s}_\tau^0)} \sum_{\tau=1}^{|\mathcal{B}^0|}\int_{\boldsymbol{y}^0} p(\boldsymbol{y}^0,\boldsymbol{s}_\tau^0)\log_2\left(\frac{p(\boldsymbol{y}^0|\boldsymbol{s}_\tau^0)}{p(\boldsymbol{y}^0)}\right)d\boldsymbol{y}^0, \tag{29}$$

where $\mathcal{B}^0$ stands for the real or the imaginary part of $\mathcal{B}$ and (29) is maximized, when we have $p(\boldsymbol{s}_\tau^0) = 1/|\mathcal{B}^0|$, where $|\mathcal{B}^0| = |\mathcal{C}|M^{N_a/2}$. Hence, (29) may be explicitly written as

$$C_1 = \log_2(|\mathcal{B}^0|) - \frac{1}{|\mathcal{B}^0|}\sum_{\tau=1}^{|\mathcal{B}^0|}\mathbb{E}_{\boldsymbol{y}^0}\left[\log_2\sum_{\epsilon=1}^{|\mathcal{B}^0|}\frac{p(\boldsymbol{y}^0|\boldsymbol{s}_\epsilon^0)}{p(\boldsymbol{y}^0|\boldsymbol{s}_\tau^0)}\right], \tag{30}$$

where in (30), we have

$$p(\boldsymbol{y}^0|\boldsymbol{s}_\epsilon^0) \propto \prod_{i=1}^{N_r}\sum_{z\in\mathbb{Z}} e^{-(y_i^0 - s_{\epsilon,i}^0 - z)^2/\sigma_{\tilde{w}}^2}, \qquad y_i^0 \in \zeta^0. \tag{31}$$

On the other hand, C_2 may be written as,

$$C_2 = \max_{p(b_{m_i}^0)}\sum_{m_i=1}^{|\mathcal{A}_0^0|}\int_{y_{v_i}^0} p(y_{v_i}^0, b_{m_i}^0)\log_2\left(\frac{p(y_{v_i}^0|b_{m_i}^0)}{p(y_{v_i}^0)}\right)dy_{v_i}^0$$
$$= \log_2(|\mathcal{A}_0^0|) - \frac{1}{|\mathcal{A}_0^0|}\sum_{m_i=1}^{|\mathcal{A}_0^0|}\mathbb{E}_{y_{v_i}^0}\left[\log_2\sum_{n_i=1}^{|\mathcal{A}_0^0|}\frac{p(y_{v_i}^0|b_{n_i}^0)}{p(y_{v_i}^0|b_{m_i}^0)}\right], \tag{32}$$

where $\mathcal{A}_0^0$ stands for the real or for the imaginary part of $\mathcal{A}_0$. Furthermore, the second equation of (32) holds true, when we have $p(b_{m_i}^0) = 1/|\mathcal{A}_0^0|$ and in (32), we have

$$p(y_{v_i}^0|b_{n_i}^0) \propto \sum_{z\in\mathbb{Z}} e^{-(y_{v_i}^0 - b_{n_i}^0 - z)^2/\sigma_{\tilde{w}}^2}, \qquad y_{v_i}^0 \in \zeta^0. \tag{33}$$

C. Error Probability

Let e_{ant}^s and e_{mod}^s represent the SER of the spatial symbols and of the conventional modulated symbols, respectively. We further use $e_{mod}^{s,0}$ and $e_{mod}^{s,1}$ to represent the SER of the conventional modulated symbols in the *absence* and in the *presence* of spatial symbol errors, respectively.

1) Upper bound of e_{ant}^s: We commence our discussions by directly introducing Lemma III.1 for quantifying the upper bound of the SER of the spatial symbols.

Lemma III.1. *The upper bound of the SER of the spatial symbols for the proposed non-linear GPSM scheme may be formulated as*

$$e_{ant}^{s,ub} = 1 - \mathbb{E}_{\boldsymbol{\lambda}}\left[\prod_{i=1}^{N_a}\int_0^\infty F_{\chi_2^2}(g)^{N_{\bar{a}}} f_{\chi_2^2}(g;\lambda_{v_i})dg\right], \tag{34}$$

where $F_{\chi_2^2}(g)$ represents the Cumulative Distribution Function (CDF) of a chi-square distribution having two degrees of freedom, while $f_{\chi_2^2}(g;\lambda_{v_i})$ represents the Probability Distribution Function (PDF) of a non-central chi-square distribution having two degrees of freedom and non-centrality given by $\lambda_{v_i} = 2|\tilde{b}_{m_i}|^2/\sigma_{\tilde{w}}^2$. Finally, we have $\boldsymbol{\lambda} = [\lambda_{v_1},\cdots,\lambda_{v_{N_a}}]$, with $N_{\bar{a}} = N_r - N_a$ representing the number of passive RAs.

Proof. A broken down representation of (9) may be written as

$$\tilde{y}_{v_i} = \tilde{b}_{m_i} + \tilde{w}_{v_i}, \forall v_i \in \mathcal{C}(k); \quad \tilde{y}_{u_i} = \tilde{w}_{u_i}, \forall u_i \in \bar{\mathcal{C}}(k). \tag{35}$$

When the decoupled detection of (12) is employed, we have

$$|\tilde{y}_{v_i}|^2 \sim \mathcal{N}(\mathcal{R}(\tilde{b}_{m_i}),\sigma_{\tilde{w}}^2/2) + \mathcal{N}(\mathcal{I}(\tilde{b}_{m_i}),\sigma_{\tilde{w}}^2/2), \tag{36}$$
$$|\tilde{y}_{u_i}|^2 \sim \mathcal{N}(0,\sigma_{\tilde{w}}^2/2) + \mathcal{N}(0,\sigma_{\tilde{w}}^2/2). \tag{37}$$

As a result, after normalisation with respect to $\sigma_{\tilde{w}}^2/2$, we have the following observations

$$|\tilde{y}_{v_i}|^2 \sim \chi_2^2(g;\lambda_{v_i}); \quad |\tilde{y}_{u_i}|^2 \sim \chi_2^2(g), \tag{38}$$

where the non-centrality is given by $\lambda_{v_i} = 2|\tilde{b}_{m_i}|^2/\sigma_{\tilde{w}}^2$.

Recall from (12) that the correct decision concerning the spatial symbols occurs, when $\sum_{i=1}^{N_a}|\tilde{y}_{v_i}|^2$ is the maximum. By exploiting the fact that $\mathbb{E}_{\mathcal{C}(k)}[\Delta] = \Delta$, the correct detection probability of the spatial symbols Δ, when the RA pattern $\mathcal{C}(k)$ was activated may be lower bounded as in (39). More explicitly, equation (a) serves as the lower bound, since it sets the most strict condition for the correct detection, when each of the passive RA indices in $\bar{\mathcal{C}}(k)$ is lower than each of the activated RA indices in $\mathcal{C}(k)$. Furthermore, equation (b) follows from the fact that the N_a random variables $|\tilde{y}_{v_i}|^2$ are independent and similarly, equation (c) holds true, since the $N_{\bar{a}} = N_r - N_a$ random variables $|\tilde{y}_{u_i}|^2$ are also independent.

Finally, after taking the expectations over $\boldsymbol{\lambda} = [\lambda_{v_1},\ldots,\lambda_{v_{N_a}}]$, the SER of the spatial symbols for the proposed non-linear GPSM scheme may be upper bounded as in (34). $\square$

2) Upper bound of e_{mod}^s: The upper bound of the SER of the conventional modulated symbols in the *absence* of spatial symbol errors may be formulated as

$$e_{mod}^{s,0,ub} = 1 - (1 - e_{mod}^{b,0,ub})^{k_{mod}}, \tag{40}$$

where $e_{mod}^{b,0,ub}$ represents the upper bound of the BER of the conventional modulated symbols in the *absence* of spatial symbol errors, which may be formulated as [37]

$$e_{mod}^{b,0,ub} = \frac{4}{k_{mod}}\left[Q\left(\sqrt{\varphi}\right) + \left(\frac{k_{mod}}{2} - 1\right)Q\left(3\sqrt{\varphi}\right)\right], \tag{41}$$

where $\varphi = 1/2M\sigma_{\tilde{w}}^2$ and $Q(\cdot)$ denotes the Gaussian Q-function.

When taking into account of the spatial symbol errors, we can formulate Lemma III.2 for quantifying the upper bound of the SER of the conventional modulated symbols in the *presence* of spatial symbol errors.

Lemma III.2. *Given the kth activated RA patten, the upper bound of the SER of the conventional modulated symbols in the*

$$
\begin{aligned}
\Delta &\overset{a}{\geq} \int_0^\infty P(|\tilde{y}_{u_1}|^2 < g_{v_1}, \ldots, |\tilde{y}_{u_{N_{\bar{a}}}}|^2 < g_{v_1}, \ldots, |\tilde{y}_{u_1}|^2 < g_{v_{N_a}}, \ldots, |\tilde{y}_{u_{N_{\bar{a}}}}|^2 < g_{v_{N_a}}) \cdot \\
&\qquad P(|\tilde{y}_{v_1}|^2 = g_{v_1}, \ldots, |\tilde{y}_{v_{N_a}}|^2 = g_{v_{N_a}} | \lambda_{v_1}, \ldots, \lambda_{v_{N_a}}) dg_{v_1} \cdots dg_{v_{N_a}} \\
&\overset{b}{=} \prod_{i=1}^{N_a} \int_0^\infty P(|\tilde{y}_{u_1}|^2 < g_{v_i}, \ldots, |\tilde{y}_{u_{N_{\bar{a}}}}|^2 < g_{v_i}) P(|\tilde{y}_{v_i}|^2 = g_{v_i} | \lambda_{v_i}) dg_{v_i} \\
&\overset{c}{=} \prod_{i=1}^{N_a} \int_0^\infty F_{\chi_2^2}(g)^{N_{\bar{a}}} f_{\chi_2^2}(g; \lambda_{v_i}) dg
\end{aligned}
\tag{39}
$$

---

presence of spatial symbol errors for the proposed non-linear GPSM scheme may be formulated as

$$
e_{mod}^{s,1,ub} = \sum_{\ell \neq k} \frac{N_c e_{mod}^{s,0,ub} + N_d e_o^s}{N_a (2^{k_{ant}} - 1)}, \tag{42}
$$

where N_c and $N_d = (N_a - N_c)$ represent the number of common and different RAs between $\mathcal{C}(\ell)$ and $\mathcal{C}(k)$, respectively. Moreover, $e_o^s = (M-1)/M$ is the SER of using random guesses.

Proof. The SER, when the detection of the spatial symbols is erroneous, may be expressed as

$$
e_{mod}^{s,1} = \sum_{\ell \neq k} P_{k \mapsto \ell} E_{k \mapsto \ell}, \tag{43}
$$

where $P_{k \mapsto \ell}$ represents the probability of the RA activation pattern $\mathcal{C}(k)$ being erroneously detected as another legitimate RA activation pattern $\mathcal{C}(\ell) \in \mathcal{C}, \ell \neq k$. Furthermore, we use $E_{k \mapsto \ell}$ to represent the error rate, when $\mathcal{C}(k)$ is erroneously detected as $\mathcal{C}(\ell)$. More explicitly, $E_{k \mapsto \ell}$ is constituted by the error rate of $e_{mod}^{s,0}$ for those N_c RAs, which are common between $\mathcal{C}(\ell)$ and $\mathcal{C}(k)$, plus the error rate of e_o^s for those $N_d = (N_a - N_c)$ RAs that are exclusively hosted by $\mathcal{C}(\ell)$. Hence we have

$$
E_{k \mapsto \ell} = \frac{N_c e_{mod}^{s,0} + N_d e_o^s}{N_a}, \tag{44}
$$

where $e_o^s = (M-1)/M$ simply represents the SER as a result of random guesses, since only random noise may be received by those N_d RAs in $\mathcal{C}(\ell)$.

To elaborate, since $e_{mod}^{s,1}$ is a monotonic function of $e_{mod}^{s,0}$, (43) can be upper bounded upon replacing $e_{mod}^{s,0}$ by $e_{mod}^{s,0,ub}$. Moreover, although it is natural that patterns associated with a higher N_c would be more likely to cause an erroneous detection, we assume an equal probability of $P_{k \mapsto \ell} = 1/(2^{k_{ant}} - 1)$. The equal probability assumption thus puts more weight on the patterns having higher N_d. Since e_o^s represents the worst-case SER, (42) can be readily established. □

Finally, the overall SER of the conventional modulated symbols e_{mod}^s is constituted by the SER, when the detection of the spatial symbols is correct, which has a probability of $(1 - e_{ant}^s)$, plus the SER, when the detection of the spatial symbols is erroneous, which has a probability of e_{ant}^s. This is expressed as

$$
e_{mod}^s = (1 - e_{ant}^s) e_{mod}^{s,0} + e_{ant}^s e_{mod}^{s,1}, \tag{45}
$$

and its corresponding upper bound may be expressed as

$$
e_{mod}^{s,ub} = (1 - e_{ant}^{s,ub}) e_{mod}^{s,0,ub} + e_{ant}^{s,ub} e_{mod}^{s,1,ub}, \tag{46}
$$

where the entries of the RHS of (46) are provided by (34), (40) and (42).

3) Upper bound of e_{eff}^s: Let N_{ant}^e and N_{mod}^e represent the number of symbol errors in the spatial symbols and in the conventional modulated symbols, respectively. Then we have $e_{ant}^s = N_{ant}^e / N_s$ and $e_{mod}^s = N_{mod}^e / N_a N_s$, where N_s is the total number of GPSM symbols. Hence, the SER of the proposed non-linear GPSM scheme denoted by e_{eff}^s may be expressed as

$$
e_{eff}^s = \frac{N_{ant}^e + N_{mod}^e}{(1 + N_a) N_s} = \frac{e_{ant}^s + N_a e_{mod}^s}{1 + N_a}. \tag{47}
$$

Hence, by substituting (34) and (46) into (47), we arrive at the upper bound of the SER of the proposed non-linear GPSM scheme, which is given by

$$
e_{eff}^{s,ub} = \frac{e_{ant}^{s,ub} + N_a e_{mod}^{s,ub}}{1 + N_a}. \tag{48}
$$

D. Complexity

Finally, let us discuss the complexity imposed by the sphere search of (8) at the transmitter for the proposed non-linear GPSM scheme. Typically, the complexity of sphere search is proportional both to the volume of the search space and to the number of lattice points visited during the search.

More explicitly, the sphere search starts from the top search layer associated with an infinite search radius, which is then gradually reduced on a per-layer basis. At the ith search layer, the search space constitutes a hypersphere with radius upper bounded by [38]

$$
\alpha_i = \frac{1}{2} \sqrt{r_1^2 + \cdots + r_i^2}, \tag{49}
$$

where r_i is the ith diagonal entry of the $\boldsymbol{R}$ matrix obtained from the QR decomposition of $\boldsymbol{PG}_1$. The expected number of lattice points within this hypersphere at the ith search layer with radius of α_i is given by [38]

$$
n_p(i) = \sum_{q=0}^{\infty} S_i(q) \int_0^{\alpha_i^2 / 2(\sigma^2 + q)} \frac{\rho^{i/2-1} e^{-\rho}}{\Gamma(i/2)} d\rho, \tag{50}
$$

where $\Gamma(\cdot)$ stands for the gamma function and $S_i(q)$ denotes the number of possible ways to represent the integer q as a sum of i squares. The closed-form expression of the resultant sum of squares function $S_i(q)$ is the solution of a number theory problem, where rich literatures exist. One of the classic solutions is due to Euler, where $S_i(q)$ is given by the coefficient of t^q in the following expansion [38]

$$\left(1 + 2\sum_{j=1}^{\infty} t^{j^2}\right)^i = 1 + \sum_{q=1}^{\infty} S_i(q) t^q. \quad (51)$$

Note that (50) implicitly assumes that the search lattice has an infinite radius. Since there is no closed form expression for the expected number of lattice points within a finite limit, we use (50) as a generic complexity expression. Furthermore, the number of numerical operations to be carried out per lattice point at the ith search layer is given by [38] $n_o(i) = 2i + 11$. Hence, for the $2N_a$ dimensional search space of (8) for both the real and the imaginary parts of $\boldsymbol{p}$, the total number of numerical operations after averaging over channel realisations may be expressed as

$$N_o = \mathbb{E}\left[2\sum_{i=1}^{N_a} n_o(i) n_p(i)\right]. \quad (52)$$

Note that the above complexity discussion concerning the sphere search algorithm is focused on the tree search stage, while the preparation stages of the sphere search algorithm are neglected, which is commonly assumed in the literature [39].

IV. Numerical Results

We now provide numerical results for characterizing both the infinite and the finite alphabet capacity as well as the, complexity, efficiency and performance of the proposed non-linear GPSM scheme. Furthermore, we define the Signal to Noise Ratio (SNR) as $1/\sigma^2$ and the SNR per bit as $N_a/k_{eff}\sigma^2$.

A. Capacity

1) Infinite Alphabet: Fig 1 shows the infinite alphabet capacity as a function of the SNR for the proposed non-linear GPSM scheme having two different $[N_t, N_r]$ settings combined with various values of $N_a = [4, 6, 8]$. It can be seen from Fig 1 that two distinct observation regions may be found for $[N_t, N_r] = [16, 8]$. More explicitly, when the SNR is low, activating less RAs results into a slightly higher capacity. This is owing to the power gain exhibited by a smaller value of N_a. Hence this capacity region may be referred to as the power limited region. Upon increasing the SNR, activating less RAs ultimately results into a lower capacity. This is due to the multiplexing loss associated with a smaller value of N_a. Hence this capacity region may be referred to as the Degree-of-Freedom (DoF) limited region. Note that, although conveying information by means of spatial symbols does contribute to the total capacity, it is capped at $\log_2(|\mathcal{C}|)$, as suggested by (25). Finally, in contrast to $[N_t, N_r] = [16, 8]$, no power limited region is observed for $[N_t, N_r] = [8, 8]$ from Fig 1, when the system operates at $N_t/N_r = 1$. In this case, although no capacity benefit is gleaned from having a smaller value of N_a, its energy efficiency actually becomes higher, as it will be shown in the left subplot of Fig 4.

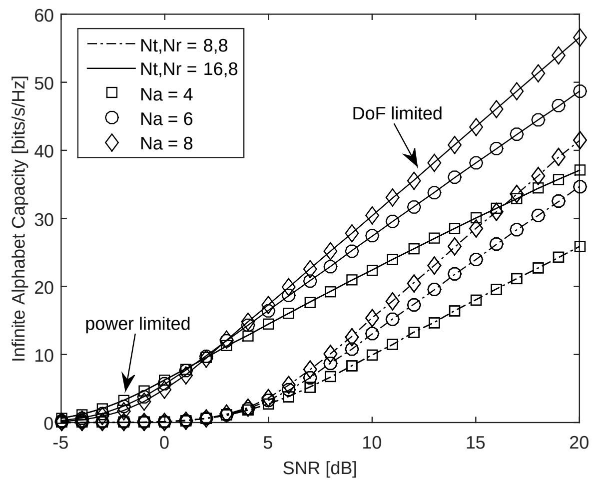


Fig. 1. Infinite alphabet capacity as function of the SNR for the proposed non-linear GPSM scheme having two different $[N_t, N_r]$ settings combined with various values of $N_a = [4, 6, 8]$.

2) Finite Alphabet: Fig 2 shows the finite alphabet capacity as a function of the SNR for the proposed non-linear GPSM scheme having two different $[N_t, N_r]$ settings combined with various values of $N_a = [4, 6, 8]$ using 4QAM. It can be seen from Fig 2 that two distinct observation regions, namely the power limited region and the DoF limited region, may be found for $[N_t, N_r] = [16, 8]$. In contrast to the infinite alphabet capacity shown in Fig 1, no multiplexing loss is encountered for $N_a = 6$, when compared to $N_a = 8$, since the ultimately achievable capacity using 4QAM in both cases is capped at 16 bits/s/Hz. This observation implies that, when 4QAM is employed, conveying part of the information using spatial symbols is more power efficient than relying on conventional modulated symbols. This is because spatial modulation belongs to the family of orthogonal signalling, which is naturally more power efficient. Finally, in contrast to $[N_t, N_r] = [16, 8]$, no power limited region exists observed for $[N_t, N_r] = [8, 8]$ in Fig 2, when $N_t/N_r = 1$. Again, although there is no capacity benefit of having a smaller value of N_a, the associated energy efficiency is actually higher, as it will be shown in the right subplot of Fig 4.

B. Per-bit Complexity

Fig 3 shows the normalised per-bit complexity as a function of the SNR for the proposed non-linear GPSM scheme having two different $[N_t, N_r]$ settings combined with various values of $N_a = [4, 6]$ assuming an infinite alphabet (left) and 4QAM (right). More explicitly, the per-bit complexity is calculated as the ratio between the total number of numerical operations required for the sphere search as defined in (52) and the infinite/finite alphabet capacity. The per-bit complexity of $N_a = [4, 6]$ is then normalised with respect to that of $N_a = 8$.

It can be seen from Fig 3 that having fewer activated RAs exhibits a lower normalised per-bit complexity for both $[N_t, N_r]$ settings and for both the infinite and finite alphabet scenarios. Moreover, a lower normalised per-bit complexity may be observed in the low SNR region for all the scenarios

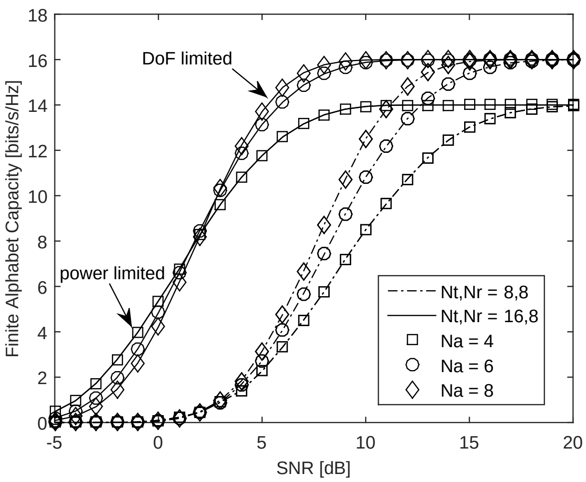


Fig. 2. Finite alphabet capacity as a function of the SNR for the proposed non-linear GPSM scheme having two different $[N_t, N_r]$ settings combined with various values of $N_a = [4, 6, 8]$ using 4QAM.

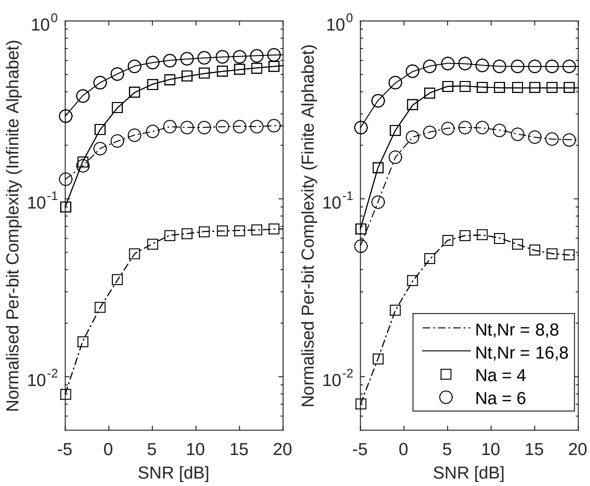


Fig. 3. The normalised per-bit complexity as a function of the SNR for the proposed non-linear GPSM scheme having two different $[N_t, N_r]$ settings combined with various values of $N_a = [4, 6]$ assuming an infinite alphabet (left) and 4QAM (right).

investigated. Quantitatively, for $[N_t, N_r] = [16, 8]$ and for both the infinite and finite alphabet scenarios, about half of the normalised per-bit complexity may be observed for $N_a = [4, 6]$ when compared to that of $N_a = 8$. More noticeably, the achievable normalised per-bit complexity becomes even lower, when having a system setting of $[N_t, N_r] = [8, 8]$, for both the infinite and finite alphabet scenarios. For example, having $N_a = 4$ in $N_t/N_r = 1$ system setting exhibits over an order of magnitude lower normalised per-bit complexity than that of $N_a = 8$ for both the infinite and finite alphabet scenarios.

C. Energy Efficiency

Fig 4 shows the energy efficiency as a function of the transmit power for the proposed non-linear GPSM scheme having two different $[N_t, N_r]$ settings combined with various values of $N_a = [4, 6, 8]$ assuming an infinite alphabet (left) and

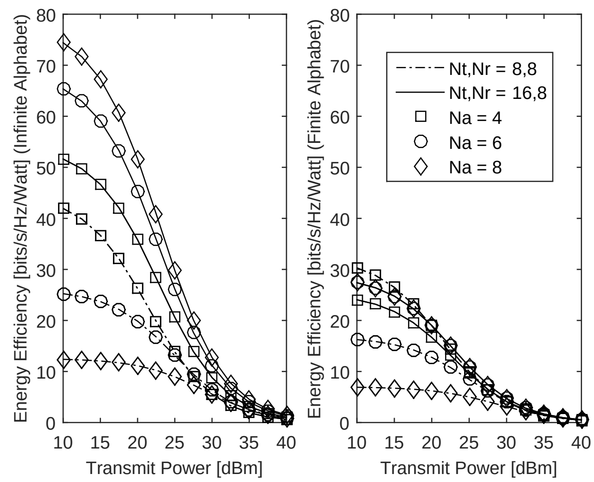


Fig. 4. Energy efficiency as a function of the transmit power for the proposed non-linear GPSM scheme having two different $[N_t, N_r]$ settings combined with various values of $N_a = [4, 6, 8]$ assuming an infinite alphabet (left) and 4QAM (right) under SNR at 15 dB.

4QAM (right) under SNR at 15 dB. More explicitly, the energy efficiency is calculated as the ratio between the infinite/finite alphabet capacity and the total power consumption P_t. To elaborate a little further, the total power consumption is comprised of the circuit-induced as well as the computation-induced power consumption [40], which may be written as

$$P_t = \underbrace{P_{PA} + P_{mix} + P_{filter} + P_{DA}}_{circuit} + \underbrace{P_o N_o}_{computation}. \quad (53)$$

In (53), $P_{PA} = P_{TX}/\eta$ is the power consumption of PA, where η is the efficiency of the PA and P_{TX} is the associated transmit power. Furthermore, P_{mix}, P_{filter} and P_{DA} represent the power consumption of mixers, filters and DA converters, respectively. When considering the computation-induced power consumption, P_o is the power consumption per numerical operation and N_o is the total number of numerical operations required for the sphere search defined in (52). Finally, we use the following practical values [40] in Fig 4: $\eta = 0.35$, $P_{mix} = 30.3$ mW, $P_{filter} = 2.5$ mW, $P_{DA} = 1.6$ mW and $P_o = 5$ mW/KOps.

It can be seen from Fig 4 that when having a system setting of $[N_t, N_r] = [8, 8]$ and for both the infinite and finite alphabet scenarios, having less activated RAs results into a higher energy efficiency. This is because the computation-induced power consumption dominates, when having $[N_t, N_r] = [8, 8]$. By contrast, when having $[N_t, N_r] = [16, 8]$ and for both the infinite and finite alphabet scenarios, fewer activated RAs tend to have a lower energy efficiency, since the circuit-induced power consumption dominates. Moreover, it is worth mentioning that for both the $[N_t, N_r]$ settings considered and for both the infinite and finite alphabet scenarios, the energy efficiency associated with various values of N_a tends to remain constant upon increasing the transmit power. Finally, the observation of Fig 4 suggests that the proposed non-linear GPSM scheme constitutes a flexible design for 'green' transceivers.

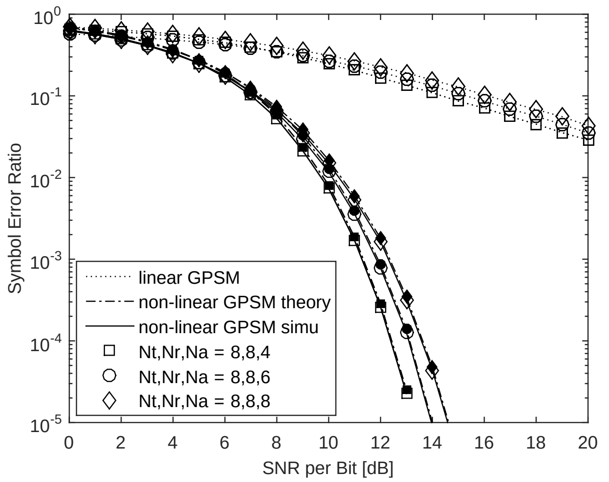


Fig. 5. SER as a function of the SNR per bit for the proposed non-linear GPSM scheme having $[N_t, N_r] = [8, 8]$ combined with various values of $N_a = [4, 6, 8]$ and using 4QAM, where the previously proposed linear GPSM scheme is also included for benchmarking.

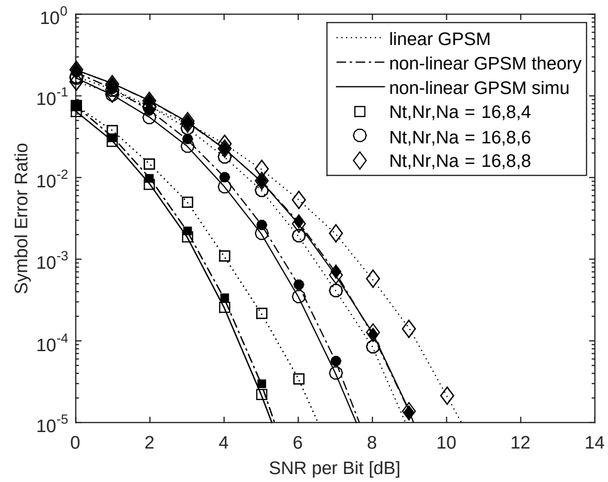


Fig. 6. SER as a function of the SNR per bit for the proposed non-linear GPSM scheme having $[N_t, N_r] = [16, 8]$ combined with various values of $N_a = [4, 6, 8]$ and using 4QAM, where the previously proposed linear GPSM scheme is also included for benchmarking.

D. Error Probability

Fig 5 and Fig 6 show the SER as a function of the SNR per bit, for the proposed non-linear GPSM scheme having $[N_t, N_r] = [8, 8]$ and $[N_t, N_r] = [16, 8]$, respectively combined with various values of $N_a = [4, 6, 8]$ and using 4QAM. Additional, the previously proposed linear GPSM scheme of [18] is also included in both figures for benchmarking. It can be seen from both figures that our theoretical analysis forms tight upper bounds of the simulation results across all the values of N_a investigated. Furthermore, we observe that, for both figures, the fewer number the activated RAs are, the better the SER performance becomes. Specifically, when 4QAM is considered, the SER performance of $N_a = 6$ is better than that of $N_a = 8$. Note that as seen in Fig 2, this SER performance improvement is achieved, while still maintaining the same finite alphabet capacity of 16 bits/s/Hz and exhibiting a lower per-bit complexity as well as a higher energy efficiency, as seen in the right subplots of both Fig 3 and Fig 4. Finally, as expected, the proposed non-linear GPSM scheme performs consistently better than its previously proposed linear GPSM counterpart for both figures, where the improvement becomes especially prominent for the setting of $[N_t, N_r] = [8, 8]$, associated with $N_t/N_r = 1$.

E. Comparisons

Fig 7 shows both the infinite alphabet capacity (left) as a function of the SNR and the energy efficiency (right) as a function of the transmit power assuming an infinite alphabet under SNR at 15 dB for both the proposed non-linear GPSM scheme and its previously proposed linear GPSM counterpart, when having $[N_t, N_r] = [8, 8]$ combined with various values of $N_a = [4, 6, 8]$. It can be seen from the left subplot of Fig 7 that the proposed non-linear GPSM scheme is capable of achieving a consistently higher infinite alphabet capacity than its linear GPSM counterpart. The capacity superiority of the proposed non-linear GPSM scheme out-weights its extra complexity, which is a benefit of our efficient sphere search, hence resulting in a consistently higher energy efficiency than its linear GPSM counterpart, as seen in the right subplot of Fig 7.

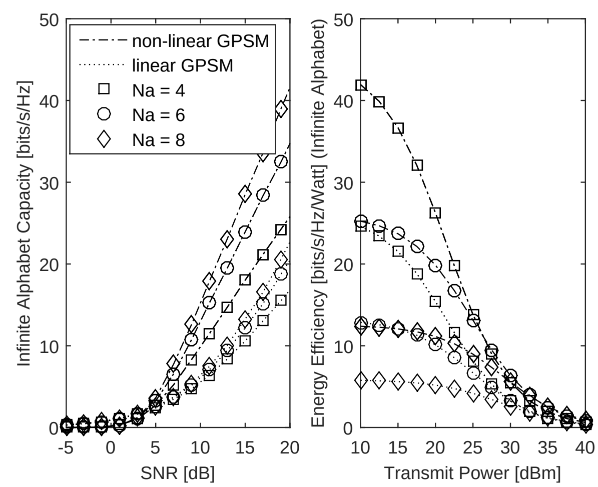


Fig. 7. Infinite alphabet capacity (left) as a function of the SNR and the energy efficiency (right) as a function of the transmit power assuming an infinite alphabet under SNR at 15 dB for both the proposed non-linear GPSM scheme and its previously proposed linear GPSM counterpart, when having $[N_t, N_r] = [8, 8]$ combined with various values of $N_a = [4, 6, 8]$.

Fig 8 shows the infinite alphabet capacity as a function of the number of transmit/receive antennas ($N_t = N_r$) under SNR at 10 dB for $N_a = N_r/2$, $N_a = N_r - 1$ and $N_a = N_r$, for the proposed non-linear GPSM scheme, its linear GPSM counterpart and the Dirty Paper Coding scheme, serving as the upper bound of transmit pre-coding. It can be seen from Fig 8 that the infinite alphabet capacity of the proposed non-linear GPSM scheme is steadily increasing upon increasing the number of transmit/receive antennas. Furthermore, the infinite alphabet capacity increase is much steeper for a higher

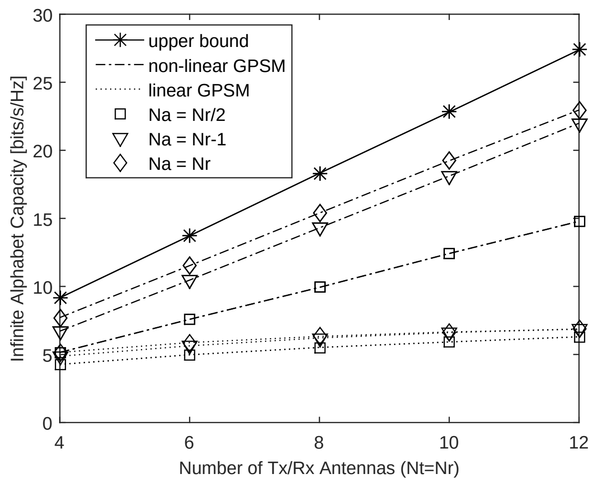


Fig. 8. Infinite alphabet capacity as a function of the number of transmit/receive antennas ($N_t = N_r$) under SNR at 10 dB combined with various values of $N_a = N_r/2$, $N_a = N_r - 1$ and $N_a = N_r$, for the proposed non-linear GPSM scheme, the linear GPSM counterpart and the Dirty Paper Coding scheme, serving as the upper bound of transmit pre-coding.

value of N_a, approaching a similar slope as that of the upper bound. By contrast, the infinite alphabet capacity of the linear GPSM scheme is almost flat across the range of the number of transmit/receive antennas. Finally, the infinite alphabet capacity of both the non-linear GPSM scheme and of its linear GPSM counterpart may be further substantially improved with the aid of optimal power/rate allocation, in order to reduce their gap with respect to the upper bound. This is left for our future work together with the impact of antenna switching time on capacity [41].

V. Conclusions

In this paper, we introduced the novel concept of non-linear GPSM relying on the powerful VP and carried out its rigorous performance analysis, where the infinite and the finite alphabet capacity, the SER expressions as well as the complexity were discussed. Our numerical results showed that the proposed non-linear GPSM scheme constitutes a promising solution for energy efficient transceivers, striking a flexibly reconfigurable balance amongst the bandwidth efficiency, energy efficiency, complexity and error resilience. More particularly, in the challenging full-rank scenario, conveying implicit information through the RA indices exhibits a lower complexity, a higher energy efficiency and a better error resilience than that of the conventional arrangement. Hence, we envision a high potential for the proposed non-linear GPSM scheme both in multi-user settings and in massive MIMO regimes, whilst enhancing the physical layer security.

References

[1] R. Zhang and L. Hanzo, “Wireless cellular networks,” *IEEE Vehicular Technology Magazine*, vol. 5, no. 4, pp. 31–39, 2010.

[2] P. Wolniansky, G. Foschini, G. Golden, and R. Valenzuela, “V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel,” in *1998 URSI International Symposium on Signals, Systems, and Electronics*, 1998, pp. 295–300.

[3] S. Alamouti, “A simple transmit diversity technique for wireless communications,” *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, 1998.

[4] D. Gesbert, M. Kountouris, R. Heath, C.-B. Chae, and T. Salzer, “Shifting the MIMO paradigm,” *IEEE Signal Processing Magazine*, vol. 24, no. 5, pp. 36–46, 2007.

[5] H. Zhang and H. Dai, “Cochannel interference mitigation and cooperative processing in downlink multicell multiuser MIMO networks,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2004, no. 2, pp. 222–235, 2004.

[6] R. Zhang and L. Hanzo, “Cooperative downlink multicell preprocessing relying on reduced-rate back-haul data exchange,” *IEEE Transactions on Vehicular Technology*, vol. 60, no. 2, pp. 539–545, 2011.

[7] F. Rusek, D. Persson, B. K. Lau, E. Larsson, T. Marzetta, O. Edfors, and F. Tufvesson, “Scaling up MIMO: Opportunities and challenges with very large arrays,” *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, 2013.

[8] V. Cadambe and S. Jafar, “Interference alignment and degrees of freedom of the K-user interference channel,” *IEEE Transactions on Information Theory*, vol. 54, no. 8, pp. 3425–3441, 2008.

[9] G. Scutari, D. Palomar, and S. Barbarossa, “Cognitive MIMO radio,” *IEEE Signal Processing Magazine*, vol. 25, no. 6, pp. 46–59, 2008.

[10] A. Khisti and G. W. Wornell, “Secure transmission with multiple antennas part I: The MISOME wiretap channel,” *IEEE Transactions on Information Theory*, vol. 56, no. 7, pp. 3088–3104, 2010.

[11] ——, “Secure transmission with multiple antennas part II: The MIMOME wiretap channel,” *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5515–5532, 2010.

[12] R. Zhang, R. Maunder, and L. Hanzo, “Wireless information and power transfer: from scientific hypothesis to engineering practice,” *IEEE Communications Magazine*, vol. 53, no. 8, pp. 99–105, 2015.

[13] A. Molisch and M. Win, “MIMO systems with antenna selection,” *IEEE Microwave Magazine*, vol. 5, no. 1, pp. 46–56, 2004.

[14] A. Mohammadi and F. Ghannouchi, “Single RF front-end MIMO transceivers,” *IEEE Communications Magazine*, vol. 49, no. 12, pp. 104–109, 2011.

[15] R. Mesleh, H. Haas, S. Sinanovic, C. W. Ahn, and S. Yun, “Spatial modulation,” *IEEE Transactions on Vehicular Technology*, vol. 57, no. 4, pp. 2228–2241, 2008.

[16] M. Di Renzo, H. Haas, and P. M. Grant, “Spatial modulation for multiple-antenna wireless systems: a survey,” *IEEE Communications Magazine*, vol. 49, no. 12, pp. 182–191, 2011.

[17] L.-L. Yang, “Transmitter preprocessing aided spatial modulation for multiple-input multiple-output systems,” in *IEEE 73rd Vehicular Technology Conference (VTC Spring)*, 2011, pp. 1–5.

[18] R. Zhang, L.-L. Yang, and L. Hanzo, “Generalised pre-coding aided spatial modulation,” *IEEE Transactions on Wireless Communications*, vol. 12, no. 11, pp. 5434–5443, 2013.

[19] ——, “Error probability and capacity analysis of generalised pre-coding aided spatial modulation,” *IEEE Transactions on Wireless Communications*, vol. 14, no. 1, pp. 364–375, 2015.

[20] A. Stavridis, M. Renzo, and H. Haas, “Performance analysis of multistream receive spatial modulation in the MIMO broadcast channel,” *IEEE Transactions on Wireless Communications (Early Access)*, 2015.

[21] J. Zheng, “Fast receive antenna subset selection for pre-coding aided spatial modulation,” *IEEE Wireless Communications Letters*, vol. 4, no. 3, pp. 317–320, 2015.

[22] C. Masouros and L. Hanzo, “Dual layered MIMO transmission for increased bandwidth efficiency,” *IEEE Transactions on Vehicular Technology (Early Access)*, 2015.

[23] K. M. Humadi, A. I. Sulyman, and A. Alsanie, “Spatial modulation concept for massive multiuser MIMO systems,” *International Journal of Antennas and Propagation*, vol. 2014, article ID 563273, 9 pages.

[24] A. Stavridis, D. Basnayaka, S. Sinanovic, M. Di Renzo, and H. Haas, “A virtual MIMO dual-hop architecture based on hybrid spatial modulation,” *IEEE Transactions on Communications*, vol. 62, no. 9, pp. 3161–3179, 2014.

[25] M. Zhang, M. Wen, X. Cheng, and L. Yang, “Pre-coding aided differential spatial modulation,” in *2015 IEEE Global Communications Conference (GLOBECOM)*, Dec 2015, pp. 1–6.

[26] Y. Bian, X. Cheng, M. Wen, L. Yang, H. V. Poor, and B. Jiao, “Differential spatial modulation,” *IEEE Transactions on Vehicular Technology*, vol. 64, no. 7, pp. 3262–3268, 2015.

[27] F. Wu, R. Zhang, L.-L. Yang, and W. Wang, “Transmitter precoding aided spatial modulation for secrecy communications,” *IEEE Transactions on Vehicular Technology (Early Access)*, 2015.

[28] R. Zhang, L.-L. Yang, and L. Hanzo, "Energy pattern aided simultaneous wireless information and power transfer," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 8, pp. 1492–1504, 2015.

[29] Q. Spencer, A. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 461–471, 2004.

[30] G. Wunder, P. Jung, M. Kasparick, T. Wild, F. Schaich, Y. Chen, S. T. Brink, I. Gaspar, N. Michailow, A. Festag, L. Mendes, N. Cassiau, D. Ktenas, M. Dryjanski, S. Pietrzyk, B. Eged, P. Vago, and F. Wiedmann, "5GNOW: non-orthogonal, asynchronous waveforms for future mobile applications," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 97–105, 2014.

[31] J. Maes and C. Nuzman, "The past, present, and future of copper access," *Bell Labs Technical Journal*, vol. 20, pp. 1–10, 2015.

[32] B. Hochwald, C. Peel, and A. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part ii: perturbation," *IEEE Transactions on Communications*, vol. 53, no. 3, pp. 537–544, 2005.

[33] N. Serafimovski, M. D. Renzo, S. Sinanovic, R. Y. Mesleh, and H. Haas, "Fractional bit encoded spatial modulation (fbe-sm)," *IEEE Communications Letters*, vol. 14, no. 5, pp. 429–431, 2010.

[34] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Transactions on Information Theory*, vol. 48, no. 8, pp. 2201–2214, 2002.

[35] J. A. T. Thomas M. Cover, *Elements of Information Theory*, 2nd ed. Wiley-Interscience, 2006.

[36] A. Razi, D. Ryan, I. Collings, and J. Yuan, "Sum rates, rate allocation, and user scheduling for multi-user MIMO vector perturbation precoding," *IEEE Transactions on Wireless Communications*, vol. 9, no. 1, pp. 356–365, 2010.

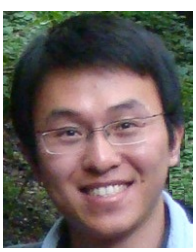
[37] D. Ryan, I. Collings, I. Clarkson, and R. Heath, "Performance of vector perturbation multiuser MIMO systems with limited feedback," *IEEE Transactions on Communications*, vol. 57, no. 9, pp. 2633–2644, 2009.

[38] B. Hassibi and H. Vikalo, "On the sphere-decoding algorithm I. expected complexity," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 2806–2818, 2005.

[39] J. Jalden and B. Ottersten, "On the complexity of sphere decoding in digital communications," *IEEE Transactions on Signal Processing*, vol. 53, no. 4, pp. 1474–1484, 2005.

[40] S. Cui, A. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Transactions on Wireless Communications*, vol. 4, no. 5, pp. 2349–2360, 2005.

[41] E. Soujeri and G. Kaddoum, "The impact of antenna switching time on spatial modulation," *IEEE Wireless Communications Letters*, early access, 2016.



Rong Zhang (M'09) received his PhD (Jun 09) from Southampton University, UK and his BSc (Jun 03) from Southeast University, China. Before doctorate, he was an engineer (Aug 03-July 04) at China Telecom and a research assistant (Jan 06-May 09) at Mobile Virtual Center of Excellence (MVCE), UK. After being a post-doctoral researcher (Aug 09-July 12) at Southampton University, he took industrial consulting leave (Aug 12-Jan 13) for Huawei Sweden R& D as a system algorithms specialist. Since Feb 13, he has been appointed as an Assistant Professor at Southampton Wireless group of ECS, Southampton University. He has 40+ journals in prestigious publication avenues (e.g. IEEE, OSA) and many more in major conference proceedings. He regularly serves as reviewer for IEEE transactions/journals and has been several times as TPC member/invited session chair of major conferences. He is the recipient of joint funding of MVCE and EPSRC and is also a visiting researcher under Worldwide University Network (WUN). More details can be found at http://www.ecs.soton.ac.uk/people/rz

Lie-Liang Yang (M'98, SM'02) received his BEng degree in communications engineering from Shanghai TieDao University, Shanghai, China in 1988, and his MEng and PhD degrees in communications and electronics from Northern (Beijing) Jiaotong University, Beijing, China in 1991 and 1997, respectively. From June 1997 to December 1997 he was a visiting scientist of the Institute of Radio Engineering and Electronics, Academy of Sciences of the Czech Republic. Since December 1997, he has been with the University of Southampton, United Kingdom, where he is the professor of wireless communications in the School of Electronics and Computer Science. Dr. Yang's research has covered a wide range of topics in wireless communications, networking and signal processing. He has published over 290 research papers in journals and conference proceedings, authored/co-authored three books and also published several book chapters. The details about his publications can be found at http://www-mobile.ecs.soton.ac.uk/lly/. He is a fellow of the IET, served as an associate editor to the *IEEE Trans. on Vehicular Technology* and *Journal of Communications and Networks* (JCN), and is currently an associate editor to the *IEEE Access* and the *Security and Communication Networks (SCN) Journal*.

Lajos Hanzo FREng, FIEEE, FIET, Fellow of EURASIP, DSc received his degree in electronics in 1976 and his doctorate in 1983. In 2009 he was awarded the honorary doctorate "Doctor Honoris Causa" by the Technical University of Budapest. During his 38-year career in telecommunications he has held various research and academic posts in Hungary, Germany and the UK. Since 1986 he has been with the School of Electronics and Computer Science, University of Southampton, UK, where he holds the chair in telecommunications. He has successfully supervised 80+ PhD students, co-authored 20 John Wiley/IEEE Press books on mobile radio communications totalling in excess of 10 000 pages, published 1400+ research entries at IEEE Xplore, acted both as TPC and General Chair of IEEE conferences, presented keynote lectures and has been awarded a number of distinctions. Currently he is directing a 100-strong academic research team, working on a range of research projects in the field of wireless multimedia communications sponsored by industry, the Engineering and Physical Sciences Research Council (EPSRC) UK, the European Research Council's Advanced Fellow Grant and the Royal Society's Wolfson Research Merit Award. He is an enthusiastic supporter of industrial and academic liaison and he offers a range of industrial courses. He is also a Governor of the IEEE VTS. During 2008 - 2012 he was the Editor-in-Chief of the IEEE Press and a Chaired Professor also at Tsinghua University, Beijing. His research is funded by the European Research Council's Senior Research Fellow Grant. For further information on research in progress and associated publications please refer to http://www-mobile.ecs.soton.ac.uk

Lajos has 20 000+ citations.