Sum Discrete-Rate Maximization with Rate and Power Control in Layered Space-Time Coding

Patricia Layec, Raphaël Visoz, and Antoine O. Berthet

Abstract

This paper generalizes the information-theoretic optimality of minimum mean square error successive interference cancellation in layered space-time coding with rate and power control. Based on this derivation, a new concept relying on partial feedback is introduced, whose core idea is to exploit an additional degree of freedom relative to the partitioning of transmit antennas. Taking into account this additional degree of freedom, together with power control and decoding order, allows the reduction of the quantization noise induced by the use of finite discrete-rate sets at the transmitter. However, the simultaneous optimization of all those degrees of freedom proves to be computationally intensive and would result in a tremendous feedback load. Practical algorithms are thus proposed to achieve this optimization with a reasonable complexity and a limited amount of feedback. Monte-Carlo simulations show that those algorithms perform close to the theoretical limits.

Index Terms

MIMO with partial CSI at the transmitter, sum rate optimization

I. INTRODUCTION

Information theory states that the rich-scattering wireless channel can provide a significant increase of capacity, if properly exploited through the use of multiple transmit and receive antennas [1] [2]. When the channel matrix is perfectly known to the transmitter, the optimal strategy consists of spatially multiplexing independent data streams through the channel’s eigen-modes with optimal power and rate allocation across those modes, i.e., multiplexing in the so-called waterfilling coordinate system [2]. This approach basically relies on an instantaneous Channel State Information (CSI) feedback from the receiver to the transmitter, i.e., requires a

This work was presented in part at IEEE ICC’07 and IEEE VTC’07 Fall.
Patricia Layec and Raphaël Visoz are with France Telecom Research and Development, Issy Les Moulineaux, France.
Antoine O. Berthet is with Ecole Supérieure d’Electricité (SUPELEC), Department of Telecommunications, Gif-sur-Yvette, France.
closed-loop implementation. Without knowledge of the channel at the transmitter, the choice of the coordinate system in which the independent data streams are multiplexed has to be fixed arbitrarily. In conjunction with joint decoding at the receiver, spatial multiplexing, also referred to as open-loop V-BLAST in the literature [3] [7], achieves the capacity of the fast fading channel. On the contrary, for the slow fading channel, whose performance is characterized through the outage probability, it is well known that the open-loop V-BLAST fails to achieve the corresponding outage capacity. At least two solutions exist to overcome this weakness. If no closed-loop implementation is allowed by the system, the first solution consists in coding across the transmit antennas. Popular outage-achieving coding schemes based on this paradigm include D-BLAST [3] or Space-Time Bit-Interleaved Coded Modulations (STBICM) [4] [5] [6] with optional Linear Precoding (LP) and Iterative Decoding (ID). If a closed-loop implementation and a low-rate feedback is allowed, the V-BLAST architecture can be improved by adapting both the rate and the power at each transmit antenna so as to reach the maximum average capacity (which exceeds the open-loop capacity). This second solution, denoted as V-BLAST with Per-Antenna Rate Control (PARC), is the core subject of this paper.

In fact, the PARC concept has strong connections with the well-known Gaussian multiple access channel in network information theory [11, chapter 14]. One of the main results provided by this theory is that, in principle, the total sum capacity can be achieved at any corner point of the capacity region with (joint) optimum successive decoding [12]. For Gaussian signalling, the optimum receiver architecture reduces to a combination of MMSE estimation and Successive Interference Cancellation (MMSE-SIC). Unfortunately, several weak points prevent PARC from reaching the maximum capacity for the underlying idealized assumptions are not practically observed. First, real transmission techniques and systems requirements involve a finite-length coding with non-zero error rates (and certainly not an infinite-length Gaussian codebook with vanishingly small residual word error probability). Error propagation in the decoding process may come as a result. This detrimental effect can be solved by the concept of gap introduced in [13] at the price of a loss of capacity. Second, only rates and power levels from discrete (and not continuous) sets are feasible. This induces a quantization error (or distortion) which also directly translates into a loss of capacity. Power adaptation can be used as a first Degree of Freedom (DoF) and joint power and rate optimization performed, leading to a variety of algorithms described in [14, Sections VI, VII and VIII]. Notably, antenna selection [10], which can be assimilated as some kind of power quantization, can be particularly efficient to enhance the
PARC system for basically no overhead feedback. The combination is justified for the diversity gain it can potentially offer [9]. Optimum decoding order can also be activated as a second DoF and optimization strategies over all possible decoding orders considered [14, Section VIII].

We now briefly summarize the main contributions of this paper. The basic idea is to relax the stringent constraint of scalar coding in PARC and to consider layered multidimensional coding. Using information-theoretic arguments derived from Gaussian vector multiple access channels, we prove that (under mild conditions) such an enlarged view still achieves the maximum capacity. We show how it benefits from a third DoF, antenna partitioning (in adaptive mode), to increase the sum discrete-rate. We design original algorithms taking into account the new DoF at disposal and we highlight possible advantages it brings over conventional PARC. In particular, combining power adaptation and antenna partitioning comes as a generalization of [14, Section VIII], while combining transmit antenna selection and antenna partitioning (assuming equal power distribution) enriches the initial work of [15]. By designing such improved algorithms, our concern is to maintain a low computational complexity at the receiver side.

The paper is organized as follows. In section II, we detail the proposed transceiver architecture and demonstrate its information-theoretic optimality. Section III addresses the joint power and rate optimization problem. Several practical algorithms are described to solve it in an efficient way. Sections IV and V are devoted to the cost analysis of the newly investigated DoF in terms of additional feedback load and computational complexity at the receiver side. Numerical results are discussed in Section VI. Section VII concludes the paper and opens up perspectives for future research.

**Notation**

- $[a, b]$ is the interval of integers or real numbers such that $a \leq x \leq b$.
- $\lceil x \rceil$ is the smallest integer $m$ such that $m \geq x$.
- $\lfloor x \rfloor$ is the largest integer $n$ such that $n \leq x$.
- The number of elements (resp. size) of any finite set (resp. block) $S$ is denoted $|S|$.
- Greek letters $\pi$ and $\sigma$ are used to designate a partition (of any given finite set of integers) and a decoding order, respectively. $\pi_{(S|K)}$ designates a partition of set $S$ into $K_{(S|\pi)}$ distinct blocks $\{\Delta_{1}(\pi), \ldots, \Delta_{K_{(S|\pi)}}(\pi)\}$. For the sake of notation simplicity, the references to $\pi$ or $S$ are omitted when implicit dependency is obvious. $\sigma$ is a $K$-tuple whose $k^{th}$ element designates the block $\sigma(k)$ to be decoded at the $k^{th}$ decoding stage, e.g., the natural ascending order is $\sigma = (1, \ldots, K)$. 

• Lower-case and capital bold letters are used to denote vectors and matrices.
• Let \( A \) be a matrix. \( a_i \) and \( a_j \) denote the \( i^{th} \) row and the \( j^{th} \) column respectively. \( a_{i,j} \) or \( [A]_{i,j} \) designates the value of \( A \) (or entry) at the pair \((i, j)\). \( I_n \) is the \( n \times n \) identity matrix.
• The superscripts \(^\top\) and \(^\dagger\) indicate transpose and Hermitian transpose.
• \( \det(A) \) is the determinant of \( A \), \( \text{tr}(A) \) is the trace of \( A \), \( \rho(A) \) is the rank of \( A \), \( \text{diag}(A) \) is the diagonal operator for \( A \). \( \text{diag}\{a_1, a_2, \ldots, a_n\} \) is the square diagonal matrix with entries \( a_1, a_2, \ldots, a_n \) on its diagonal.
• Let \( \mathbf{x} \) be a \( n \times 1 \) vector and \( \Delta \subseteq [1, n] \) a subset of \( |\Delta| \leq n \) integers \( \delta_1, \delta_2, \ldots, \delta_{|\Delta|} \). Then, \( x_\Delta \) denotes the \( |\Delta| \times 1 \) subvector \( [x_{\delta_1}, x_{\delta_2}, \ldots, x_{\delta_{|\Delta|}}]\)^\top.
• \( E_\Delta \) denotes the matrix of dimension \( n \times |\Delta| \) such that \( x_\Delta = E_\Delta^\dagger x \).
• Let \( A \) be a \( m \times n \) matrix. \( A_\Delta \) denotes the \( m \times |\Delta| \) submatrix such that \( A_\Delta = A E_\Delta \).
• Let \( \mathbf{x} \) be a random vector. \( \mu_\mathbf{x} = \mathbb{E}\{\mathbf{x}\}, \Theta_\mathbf{x} = \mathbb{E}\{(\mathbf{x} - \mu_\mathbf{x})(\mathbf{x} - \mu_\mathbf{x})^\dagger\} \).

II. LAYERED SPACE-TIME CODING WITH RATE CONTROL

A. Communication model

We consider a point-to-point communication scenario with a MIMO spatial multiplexing system under rate control. A coarse Channel Quality Indicator (CQI) is available at transmitter through the use of low-rate feedback link and perfect Channel State Information is available at receiver (perfect CSIR). Transmission occurs on a correlated narrowband MIMO Rayleigh flat fading channel with \( N_T \) transmit antennas and \( N_R \) receive antennas which remains constant during \( L \) consecutive channel uses (quasi-static assumption). Let \( \mathbf{H} \in \mathbb{C}^{N_R \times N_T} \) denote the channel matrix. A selection procedure ensures that only the best (in a sense to specify) \( n_T \) antennas out of the \( N_T \) available transmit antennas are activated during each transmission (subset \( \mathcal{A} \subseteq [1, N_T] \)). At any instant \( m \in [1, L] \), the receive samples are stacked in the vector \( \mathbf{y}_m \in \mathbb{C}^{N_R} \) given by

\[
\mathbf{y}_m = \widetilde{\mathbf{H}} \mathbf{x}_m + \mathbf{n}_m
\]

where \( \mathbf{x}_m \in \mathbb{C}^{n_T} \) denotes the transmitted vector, \( \widetilde{\mathbf{H}} \in \mathbb{C}^{N_R \times n_T} \) is the channel matrix limited to active antennas and \( \mathbf{n}_m \in \mathbb{C}^{N_R} \) is a zero-mean independent identically distributed (i.i.d.) circularly-symmetric complex Gaussian noise. Its covariance matrix is thus modelled as \( \Theta_\mathbf{n} = \sigma_n^2 I_{N_R} \). The covariance matrix of the transmit signal, denoted \( \Theta_\mathbf{x} \), satisfies the power constraint \( \text{tr}(\Theta_\mathbf{x}) \leq P \) where \( P \) is the total power available at transmitter. For simplicity, we assume that \( \sigma_n^2 = 1 \) and scale the channel appropriately.
B. Proposed transceiver architecture

We partition the subset $\mathcal{A}$ of $n_T$ active transmit antennas into $K$ distinct groups (or blocks in the terminology of set partitioning) $\Delta_1, \ldots, \Delta_K$ of antenna indices with the constraint $\sum_{k=1}^{K} |\Delta_k| = n_T$. Let $\pi_{(A|K)}$ denote this partition. The information bit stream $\mathbf{d}$ is first serial-to-parallel converted into $K$ substreams $\{\mathbf{d}_1, \ldots, \mathbf{d}_K\}$. Substream $\mathbf{d}_k$ enters a specified Space-Time Modulation Coding Scheme (ST-MCS) whose output $\mathbf{X}_k$ is a space-time codeword which can always be put under the form of a complex matrix of dimension $|\Delta_k| \times L$. The parameter $L$ acts as the common length (in channel uses) of all produced space-time codewords. The final transmitted matrix $\mathbf{X}$ of dimension $n_T \times L$ is simply defined as

$$\mathbf{X} = \sum_{k=1}^{K} \mathbf{E}_{\Delta_k} \mathbf{X}_k. \quad (2)$$

This operation is referred to as layering and antenna switching. Discrete rates $\{R_k\}_{k=1}^{K}$ for all substreams are derived from the calculation of the corresponding capacities $\{C_k\}_{k=1}^{K}$ at the receiver (see next subsection) and may be different. They are regularly adapted by the reception of CQIs (one per layer). The transmit scheme is by essence adaptive, i.e., the set $\mathcal{A}$ of active antennas, the number of groups $K$ and the antenna partition $\pi_{(A|K)}$ vary from one channel outcome to another (or equivalently from one feedback to another). This adaptation defines the new DoF at disposal. It is controlled by the receiver thanks to an additional information feedback and impacts on both encoding and antenna switch functions. The overhead needed to update the partition depends on the number of possible partitions, which is a non-trivial function of the number of transmit and receive antennas. The transceiver architecture, referred to as Layered Space-Time Codes with Rate and Power Control (LSTC-RPC), is depicted in Fig. 1.

After simultaneous transmission over the air interface, the decoding is done (multidimensional) layer by (multidimensional) layer. The received samples are collected into a matrix $\mathbf{Y}$ of dimension $N_R \times L$. For the sake of clarity, the successive decoding procedure follows the ascending order of layer indices. Since the treatment is the same for all channel uses, we will omit the time index in the following. The first layer is decoded based on the MMSE estimate $\hat{x}_{\Delta_1}^{\text{mmse}}$ of $x_{\Delta_1}$ given by $\hat{x}_{\Delta_1}^{\text{mmse}} = \mathbf{F}_1 \mathbf{y}$ where $\mathbf{F}_1$ denotes the multidimensional Wiener filter (of dimension $|\Delta_1| \times N_R$) whose exact expression

$$\mathbf{F}_1 = \left[ \mathbf{I}_{|\Delta_1|} + \mathbf{B}_1 \right]^{-1} \mathbf{H}_{\Delta_1} \mathbf{H}_{\Delta_1}^{\dagger} \mathbf{\Theta}^{-1} \mathbf{z}_1 \quad (3)$$
involves the intermediate matrices

\[ B_1 = \Theta_{x_{\Delta_1}} H_{\Delta_1}^\dagger \Theta_{z_1}^{-1} H_{\Delta_1} \]

(4)

\[ \Theta_{z_1} = \sum_{j \geq 2} H_{\Delta_j} \Theta_{x_{\Delta_j}} H_{\Delta_j}^\dagger + I_{NR} \]

(for invertible matrix \( I_{|\Delta_1|} + B_1 \)). The decoded bits are re-encoded to produce \( \hat{x}_{\Delta_1}^{\text{dec}} \). Here, decoding of layer 1 (and subsequent layers) is assumed error-free which is legitimate for the analysis of capacity we conduct in the next subsection. The second layer is similarly decoded based on the MMSE estimate \( \hat{x}_{\Delta_2}^{\text{mmse}} \) of \( x_{\Delta_2} \) now given by \( \hat{x}_{\Delta_2}^{\text{mmse}} = F_2 \tilde{y}_2 \) with

\[ \tilde{y}_2 = y - H_{\Delta_1} \hat{x}_{\Delta_1}^{\text{dec}} \]

the new observation vector for layer 2 after subtraction of the effect of layer 1 and \( F_2 \) the multidimensional Wiener filter (of dimension \( |\Delta_2| \times N_R \)) whose exact expression

\[ F_2 = [I_{|\Delta_2|} + B_2]^{-1} \Theta_{x_{\Delta_2}} H_{\Delta_2}^\dagger \Theta_{z_2}^{-1} \]

(6)

now involves the intermediate matrices

\[ B_2 = \Theta_{x_{\Delta_2}} H_{\Delta_2}^\dagger \Theta_{z_2}^{-1} H_{\Delta_2} \]

(7)

\[ \Theta_{z_2} = \sum_{j \geq 3} H_{\Delta_j} \Theta_{x_{\Delta_j}} H_{\Delta_j}^\dagger + I_{NR} \]

(for invertible matrix \( I_{|\Delta_2|} + B_2 \)). The same procedure continues until all layers have been decoded.

C. Information-theoretic optimality (chain rule)

In this subsection, we prove that LSTC-RPC combined with MMSE-SIC achieves the maximum capacity if a capacity-achieving multidimensional Gaussian codebook is employed for each layer. This proof can be seen as a vectorial generalization of [12]. Using matrix manipulations and taking into account that the vectors \( x_{\Delta_k} \) are mutually independent, it is easy to demonstrate that the maximum capacity

\[ C \triangleq \log \det(I_{NR} + H\Theta_{x} H^\dagger) \]

(8)

can be split into a sum of \( K \) terms \( \{C_k\}_{k=1}^{K} \) (one per layer), each of them being expressed as

\[ C_k = \log \det(I_{|\Delta_k|} + B_k) \]

(9)

with, as previously,

\[ B_k = \Theta_{x_{\Delta_k}} H_{\Delta_k}^\dagger \Theta_{z_k}^{-1} H_{\Delta_k} \]

\[ \Theta_{z_k} = \sum_{j \geq k+1} H_{\Delta_j} \Theta_{x_{\Delta_j}} H_{\Delta_j}^\dagger + I_{NR} \]

(10)
This result requires $\Theta_{z_k}$ to be \textit{invertible} (which is always the case). If we now consider the generic vector channel model with additive colored noise

$$
\tilde{y}_k = H_{\Delta_k} x_{\Delta_k} + \sum_{j=k+1}^K H_{\Delta_j} x_{\Delta_j} + n \tag{11}
$$

and we further assume that $x_{\Delta_k}$ and $z_k$ are independent and Gaussian, we can show that

$$
I(x_{\Delta_k}; \tilde{y}_k) = \log \det(I_{|\Delta_k|} + B_k) = C_k \tag{12}
$$

and that the MMSE filter $F_k$ for layer $k$ not only minimizes the MSE but also provides a sufficient statistic to detect $x_{\Delta_k}$, i.e., is information lossless:

$$
I(x_{\Delta_k}; \tilde{y}_k) = I(x_{\Delta_k}; \hat{x}_{\Delta_k}^{\text{mmse}}). \tag{13}
$$

Note that this last equality requires $F_k \Theta_{z_k} F_k^\dagger$ to be \textit{invertible}. Sylvester’s inequality yields

$$
\rho(F_k \Theta_{z_k} F_k^\dagger) \leq \rho(\Theta_{z_k}) = N_R. \tag{14}
$$

As a result, the matrix $F_k \Theta_{z_k} F_k^\dagger$ of dimension $|\Delta_k| \times |\Delta_k|$ is non invertible if $|\Delta_k| > N_R$. Thus, a partition block cardinality should not exceed the number of receive antennas. This partition guideline is strictly respected in the following. Finally, from the chain rule of mutual information, we can rewrite $I(x; y)$ as

$$
I(x; y) = \sum_{k=1}^K I\left(x_{\Delta_k}; y \mid x_{\Delta_1}, \ldots, x_{\Delta_{k-1}}\right). \tag{15}
$$

By invoking the perfect decoding assumption of previous layers, i.e.,

$$
I\left(x_{\Delta_k}; y \mid x_{\Delta_1}, \ldots, x_{\Delta_{k-1}}\right) = I(x_{\Delta_k}; \tilde{y}_k) \tag{16}
$$

and by taking into account (13), we come up to the equality

$$
I\left(x_{\Delta_k}; y \mid x_{\Delta_1}, \ldots, x_{\Delta_{k-1}}\right) = I(x_{\Delta_k}; \hat{x}_{\Delta_k}^{\text{mmse}}) \tag{17}
$$

and conclude that the rate achieved in the $k^{th}$ stage of the MMSE-SIC receiver is $I\left(x_{\Delta_k}; y \mid x_{\Delta_1}, \ldots, x_{\Delta_{k-1}}\right)$. This completes the proof.
III. JOINT POWER AND RATE OPTIMIZATION

A. Problem formulation

For any given partition $\pi(A|K)$, an additional power adaptation enables the achievement of a larger part of the closed-loop MIMO capacity. Clearly, the sum capacity optimization in LSTC-RPC is similar to the multi-user Multiple Access Channel (MAC) problem [22] with a sum power constraint [23]. It has been demonstrated that the optimal transmit covariance matrix can be found by the use of a sum power iterative waterfilling [24, Theorem 1]. The objective function of this algorithm is

$$C = \max_{\Theta_x \Delta_k} \left\{ \log \det \left( I_{N_R} + \sum_{k=1}^{K} H_{\Delta_k} \Theta_x \Delta_k H_{\Delta_k}^H \right) \right\}$$

subject to: $\Theta_x \Delta_k \geq 0$, $\forall k \in [1,K]$ and $\sum_{k=1}^{K} \text{tr}(\Theta_x \Delta_k) \leq P$ (18)

where $P$ is the total power available at the transmitter. If $K = 1$, the optimal solution of (18) is simply given by the waterfilling algorithm [2]. To satisfy the assumption of limited feedback available at transmitter, the power adaptation is restricted to diagonal covariance matrices only. With such an assumption, optimal derivations were addressed in [14] and [16] for PARC systems, and are still valid for LSTC-RPC. From the approximations of various SNR values in [16], general trends can be observed. At low SNR, optimal power allocation is antenna selection while at high SNR an equidistribution of power is optimum. We further assume that each antenna belonging to a given block is allocated the same power, i.e.,

$$\Theta_x = \sum_{k=1}^{K} P_k E_{\Delta_k} E_{\Delta_k}^H$$

subject to: $\text{tr}(\Theta_x) \leq P$ (19)

To this point, the power allocation policy has been carried out irrespective of the available rate set, which is an inherent feature of rate control. The prior maximization problem must be redefined into the following: maximize the sum discrete-rate subject to (i) discrete rate sets, (ii) chain rule constraints, and (iii) quantized transmit power allocation for each group of antennas. All DoFs at disposal should be used to achieve that purpose namely antenna partitioning, decoding order and power adaptation subject to both rate and power quantization (transmit antenna selection being viewed as a special case of power quantization). Unfortunately, we are not able to find a closed-form solution and exhaustive search would be too intensive. Hence, by taking advantage of the layered architecture, we design sub-optimal algorithms to perform a joint optimization between identified DoFs under a low-rate feedback assumption. To this end, power allocation is handled in the next two sections: firstly, by assuming additional feedback bits with
quantized power levels, and secondly, by adopting the antenna selection strategy. The antenna selection strategy consists of switching off one or several antennas while balancing equally the total power budget across all active antennas. This latter strategy does not induce any feedback overhead which, of course, makes it particularly attractive.

B. Strategy 1: Power quantization

In this Section, antenna selection is not activated, i.e., $A = \{1, \ldots, N_T\}$. The sum discrete-rate optimization leads to the optimal ordered partition $(\pi^*, \sigma^*)$ along with the set $P^*$ made of the $K_{(A|\pi^*)}$ optimal quantized powers $\{P^*_k\}_{k=1}^K$. These parameters are solutions of the following problem:

$$\{P^*, \pi^*, \sigma^*\} = \arg \max_{\{P, \pi, \sigma\}} R_{\text{sum}}(P, \pi, \sigma)$$

(20)

where $R_{\text{sum}}(P, \pi, \sigma)$ is the sum discrete-rate defined as

$$R_{\text{sum}}(P, \pi, \sigma) = \sum_{k=1}^{K_{(A|\pi)}} R_k(P, \pi, \sigma)$$

(21)

Each group (or layer) $\Delta_k$ reckons on a specific (finite) discrete-rate distribution $D_{|\Delta_k|$ which is assumed to match the (true) layered capacities. Note that this distribution only depends on the group cardinality. The best discrete rate for group $\Delta_k$ is picked up as

$$R_k(P, \pi, \sigma) = \arg \min_{R \in D_{|\Delta_k|}} |C_k(P, \pi, \sigma) - R|$$

subject to (chain rule): $R \leq C_k(P, \pi, \sigma)$

(22)

where $C_k(P, \pi, \sigma)$ is defined in (9). Clearly, the algorithm efficiency depends on the number of power levels. In the sequel, several efficient algorithms are described.

1) Vectorial successive rates and power quantization: As seen previously, at high SNR an equal power allocation may result in close to optimal performance. To this end, a sub-optimal power allocation, inspired by [14, section 8], deviates from the equidistribution to increase the discrete bit loading process. For a given ordered partition $(\pi_{(A|K)}, \sigma)$, a vectorial version of Successive Rate and Power Quantization (SRPQ) is proposed. Algorithm 1 depicts the procedure. From this algorithm, two strategies based on exhaustive search are considered, namely, SRPQ1 which optimizes the partition DoF for a fixed decoding order $\sigma = (1, \ldots, K)$ and SRPQ2 which optimizes jointly the decoding order and the partition DoFs.

After allocation of the discrete rate $R_{\sigma(k)}$, the minimal amount of power $P_{\sigma(k)}$ required so as to transmit data without error satisfies the following equation

$$R_{\sigma(k)} = \log \det \left( I_{|\Delta_k|} + P_{\sigma(k)} H_{\Delta_k}^H \Theta_{z_k}^{-1} H_{\Delta_k} \right)$$

(23)
with the constraint $P_{\sigma(k)} \in [0, P]$. A closed-form solution exists as long as $|\Delta_k| \leq 4$ since the determinant can be expanded as a polynomial in $P_{\sigma(k)}$ given by

$$\det \left( I_{|\Delta_k|} + P_{\sigma(k)} H_{\Delta_k}^\dagger \Theta_{\Delta_k}^{-1} H_{\Delta_k} \right) = 1 + \sum_{j=1}^{|\Delta_k|} \sum_{i_1=1}^{|\Delta_k|} \sum_{i_2>i_1}^{|\Delta_k|} \sum_{i_3>i_2}^{|\Delta_k|} \ldots \sum_{i_j>i_{j-1}}^{|\Delta_k|} (P_{\sigma(k)})^j \det \left( H_{\Delta_k} (i_1, \ldots, i_j) \Theta_{\Delta_k}^{-1} H_{\Delta_k} (i_1, \ldots, i_j) \right)$$

(24)

with $H_{\Delta_k} (i_1, \ldots, i_j) = [h_{\Delta_{i_1}} h_{\Delta_{i_2}} \cdots h_{\Delta_{i_j}}]$ the $N_R \times j$ submatrix. For $|\Delta_k| > 4$, efficient numerical methods are helpful. Powers are quantized by taking $2^q$ equidistributed values in $[0, P]$ where rounding up is clearly obvious.

2) Ordered rates and power quantization: While vectorial SRPQ exploits exhaustively the decoding order DoF, we propose the Ordered Rates and Power Quantization procedure (ORPQ) that rather considers one single (chosen) decoding order $\sigma$. This parameter is to be defined for a given partition $\pi(A|K)$ in a pre-processing stage, sorting groups from best to worst (in a sense to be defined). To benefit from the acquired knowledge, we suggest modifying the way of allocating power, i.e., unlike in (vectorial) SRPQ, we will now assign the total remaining budget to the best group. Afterwards, quantized rates and powers are allocated successively.

As said, blocks are sorted from best $i_1$ to worst $i_K$ link capacity, whose definition for group $i_k$ relies on the matrix $D_{i_k}$ out of a vectorial MMSE filter expressed as

$$D_{i_k} = \left[ E_{\Delta_{i_k}}^\dagger \left( I_{N_T} + \Theta_x H^\dagger H \right)^{-1} E_{\Delta_{i_k}} \right]^{-1} - I_{|\Delta_{i_k}|}$$

(25)

In order to prevent any favoritism, $D_{i_k}$ is determined with an equal power distribution between transmit antennas. The ordering rule is given by

$$\det \left( I_{|\Delta_{i_1}|} + D_{i_1} \right) \geq \cdots \geq \det \left( I_{|\Delta_{i_K}|} + D_{i_K} \right)$$

(26)

Note that this definition does not take into account SIC processing. With the ordering $\sigma = (i_K, \ldots, i_1)$, the best layer is preserved from interference. As a result, the layer $i_1$ achieves the highest discrete rate over all other orderings with the minimum amount of power while subsequent layers benefit from a large remaining power budget. The ORPQ procedure is given in Algorithm 2.

C. Strategy 2: Transmit antenna selection

Antenna selection can enhance the sum capacity by balancing the power budget equally across all active antennas, while filling the gap between the open-loop and the closed-loop capacity. A diversity gain can be earned either for correlated channels or for uncorrelated channels with
fewer receive than transmit antennas [17] [18]. To preserve a low-rate feedback, antenna selection decision should be made at the receiver side and then conveyed to the transmitter.

1) Joint transmit antenna selection and sum discrete-rate maximization: The selection procedure ends with the optimal subset of \( n_T \) active antennas \( \mathcal{A}^* \), the optimal set of discrete rates \( \{ R_k \}_{k=1}^K \) and the optimal ordered partition \( (\pi^*, \sigma^*) \) of \( \mathcal{A}^* \). All these parameters are jointly optimized according to the following rules:

\[
\{ \mathcal{A}^*, \pi^*, \sigma^* \} = \arg \max_{\{ \mathcal{A}, \pi, \sigma \}} R_{\text{sum}}(\mathcal{A}, \pi, \sigma) \tag{27}
\]

where \( R_{\text{sum}}(\mathcal{A}, \pi, \sigma) \) is the sum discrete-rate defined as

\[
R_{\text{sum}}(\mathcal{A}, \pi, \sigma) = \sum_{k=1}^{K(\mathcal{A}|\pi)} R_k(\mathcal{A}, \pi, \sigma) \tag{28}
\]

The best discrete rate for group \( \Delta_k \) is picked up as

\[
R_k(\mathcal{A}, \pi, \sigma) = \arg \min_{R \in D|\Delta_k} |C_k(\mathcal{A}, \pi, \sigma) - R| \tag{29}
\]

subject to (chain rule): \( R \leq C_k(\mathcal{A}, \pi, \sigma) \)

where \( C_k(\mathcal{A}, \pi, \sigma) \) is defined in (9) on the basis of the channel matrix \( \tilde{\mathbf{H}} \) corresponding to the active antenna subset \( \mathcal{A} \) and for the ordered partition \((\pi, \sigma)\). The exhaustive search considers the \( \sum_{n=1}^{N_T} \binom{N_T}{n} \) possible active antenna set \( \mathcal{A} \). Then it scans all ordered partitions computing (28) for each. Complexity grows quickly with the number of transmit antennas, and is mainly driven by the number \( K(\mathcal{A}|\pi)! \) of permutations in the partitions (see Section V for details). To manage both the computational complexity and the storage requirements in a practical implementation perspective, the searching space needs to be reduced. To this end, we propose a joint iterative antenna selection procedure depicted in more details in the following subsection.

2) Reduced-complexity approach: The proposed algorithm reduces the overall complexity in two steps. First, by discarding one antenna at a time (decremental mode), the iterative selection does not scan all possible active antenna set \( \mathcal{A} \). Second, the decoding order is initialized for a given partition. A pre-processing step sorts antennas from best to worst link capacity. The selection rule for antenna \( t \in [1, N_T] \) is based on the SINR output of a MMSE filter expressed as

\[
\gamma_t = \left( \mathbf{I}_{N_T} + \Theta_{\mathbf{x}} \mathbf{H}^\dagger \mathbf{H} \right)^{-1}_{tt} - 1 . \tag{30}
\]

Note that this definition does not take into account SIC processing. According to this rule, antennas are sorted in the decreasing order of the corresponding link capacity. The initial \( N_T \)-tuple \( S = (t_1, \ldots, t_{N_T}) \) is thus defined so as to satisfy the ordering rule:

\[
\gamma_{t_1} \geq \cdots \geq \gamma_{t_{N_T}} . \tag{31}
\]
The iterative process starts with $S$. The worst element is taken off so as to obtain the next subset. Indeed, an antenna selection gain exists since energy is concentrated on limited (but contributory) antennas. One iteration (indice $l$) of this algorithm, with $S_l \subseteq S$, involves the use of DoFs at disposal to optimize antenna subset and discrete rate selection. As stated previously, the largest computational effort is due to the optimal decoding order search. A substantial saving in complexity is achieved through the initialization of $\sigma$. In relation to (26), we coin reverse peeling order $\sigma_{\text{rev}} = (i_K, \ldots, i_1)$ and the forward peeling order $\sigma_{\text{fwd}} = (i_1, \ldots, i_K)$. Owing to the SIC processing, it follows that the forward peeling ensures a more compact set of capacities $\{C_{i_k}^{\text{fwd}}\}_{k=1}^{K}$. As a result, each element of $\{C_{i_k}^{\text{fwd}}\}_{k=1}^{K}$ is contained in $[C_{i_K}^{\text{rev}}, C_{i_1}^{\text{rev}}]$ and the following set of inequalities is obtained

$$
\begin{align*}
C_{i_1}^{\text{fwd}} &\leq C_{i_1}^{\text{rev}} \\
C_{i_K}^{\text{rev}} &\leq C_{i_K}^{\text{fwd}}.
\end{align*}
$$

(32)

Let us next assume that the best layer $C_{i_1}^{\text{fwd}}$ is lower bounded by $R_{\text{max}}$, the largest available rate. The first inequality of (32) yields $R_{i_1}^{\star} = R_{\text{max}}$ whatever applied ordering. When $K = 2$, we can thus conclude from the second inequality that the sum discrete-rate is maximized with forward peeling. When $K > 2$, finding out the optimal decoding order is intractable. However, the prior result suggests that the sum discrete-rate can be increased by preserving weaker groups from interference (achieved by forward peeling). Based on this observation, a sub-optimal initialization for $\sigma$ consists in sorting the groups according to their best antenna SINR. More precisely, let $\gamma_k(\pi) = \max_{t \in \Delta_k} \gamma_t$, $\forall k \in [1, K(\rho, A|\pi)]$ where the $\gamma_t$’s are those previously computed for the construction of $S$. By a slight abuse of notation, we define similarly in the context of the proposed joint selection algorithm the forward ordering $\sigma_{\text{fwd}} = (i_1, \ldots, i_K)$ and the reverse ordering $\sigma_{\text{rev}} = (i_K, \ldots, i_1)$ which satisfy the ordering rule:

$$
\gamma_{i_1}(\pi) \geq \cdots \geq \gamma_{i_K}(\pi).
$$

(33)

The proposed joint iterative selection procedure is detailed in Algorithm 3.

**IV. FEEDBACK LOAD**

Antenna partitioning in LSTC-RPC induces a novel structure of feedback able to indicate the optimal partition. This new DoF has a cost in terms of feedback load whose rigorous characterization resorts to the theory of partitions [28, Chapter 13]. For the sake of simplicity, it is assumed that the number $N_{\text{mcs}}$ of discrete rates for a group of transmit antennas (partition block) is constant whatever its size. The null index MCS is employed so as to identify the transmit antennas that
can be switched off. The way of partitioning a set of \( n \) elements encompasses all partitions of a smaller set. The antenna selection strategy is thus feedback free, considering for instance each idle antenna as a group. In contrast, power quantization strategy costs additional feedback bits. In the sequel, power quantization is *uniform* and the number \( N_{\text{pwr}} \) of quantized power levels for a group of antennas (partition block) is *constant* whatever its size. Conventionally, \( N_{\text{pwr}} = 1 \) means that power control DoF is discarded. We denote \( S_n^{(k, \leq b)} \) the number of non-ordered partitions of \( \{1, \ldots, n\} \) into \( k \) non-empty blocks of maximal size \( b \) and \( S_n^{(\leq b)} = \sum_{k=1}^n S_n^{(k, \leq b)} \) the number of non-ordered partitions of \( \{1, \ldots, n\} \) into non-empty blocks of maximal size \( b \). \( S_n^{(k, \leq n)} = S_n^{(k)} \) corresponds to the Stirling number of the second kind (or Stirling "partition" number) and \( S_n^{(\leq n)} = \sum_{k=1}^n S_n^{(k)} \) is the Bell number. As well known, an explicit form exists for Stirling partition numbers \([29]\)

\[
S_n^{(k)} \equiv \left\{ \binom{n}{k} \right\} = \frac{1}{k!} \sum_{l=0}^k \binom{k}{l} (-1)^l (k-l)^n .
\]  

(34)

Let \( e_b(z) = 1 + z + \ldots + \frac{z^b}{b!} \) be the truncated Exponential Function (EF). The \( S_n^{(\leq b)} \)'s can be recovered from the Exponential Generating Function (EGF)

\[
S^{(\leq b)}(z) = e^{e_b(z)} - 1
\]  

(35)

as \( S_n^{(\leq b)}/n! = \lfloor z^n \rfloor S^{(\leq b)}(z) \) \([29]\). Defining the constrained set of indices

\[
\mathcal{T}_n^{(k, \leq b)} = \left\{ \{i_1, \ldots, i_k\} \in [1, b]^k : \sum_{l=1}^k i_l = n \right\},
\]

(36)

we can also express \( S_n^{(k, \leq b)} \) explicitly as

\[
S_n^{(k, \leq b)} = \frac{1}{k!} \sum_{\{i_1, \ldots, i_k\} \in \mathcal{T}_n^{(k, \leq b)}} \binom{n}{i_1, \ldots, i_k} .
\]  

(37)

We envisage two distinct feedback structures, both of them indicating the optimal set of active antennas \( A^* \), the optimal partition \( \pi^*_{(A^*|K)} \) into \( K \) non-empty blocks and the discrete rates \( \{R^*_k\}_{k=1}^K \). Conveying the partition number and the \( K \) values of discrete rates and quantized power levels in separate (e.g., consecutive) slots constitutes a first strategy. The average number of required feedback bits is expressed as

\[
\bar{N}^{(1)}_{\text{fdbk}} = \mathbb{E}_\pi \left\{ N^{(1)}_{\text{fdbk}, \pi} \right\}
\]  

(38)

where expectation is taken over all acceptable partitions and where \( N^{(1)}_{\text{fdbk}, \pi} \) denotes the number of feedback bits for the specific partition \( \pi \). Let \( \pi_i \) be the \( i^{th} \) possible partition. Assuming that all partitions arise with equal probability, we have

\[
\bar{N}^{(1)}_{\text{fdbk}} = \left\lfloor \log_2 S_n^{(\leq N_T)} \right\rfloor + \frac{1}{S_n^{(\leq N_T)}} \sum_{i=1}^{S_n^{(\leq N_T)}} \left[ K(\pi_i) \log_2 (N_{\text{mcs}} N_{\text{pwr}}) \right]
\]  

(39)
or equivalently

\[
N_{\text{fdbk}}^{(1)} = \left\lceil \log_2 S_{N_T}^{(\leq N_R)} \right\rceil + \sum_{k=1}^{N_T} \frac{S_{N_T}^{(k, \leq N_R)}}{S_{N_T}^{(\leq N_R)}} \left\lceil k \log_2 (N_{\text{mcs}} N_{\text{pwr}}) \right\rceil
\]  

(40)

Although minimizing the average feedback load, this signalling introduces a variable-length feedback which, from a practical implementation viewpoint, seems inappropriate. Moreover, the actual partition probability distribution may differ from the uniform distribution and is not known a priori. Indeed, the feedback link would certainly be constrained by the maximum amount of feedback bits (worst case):

\[
N_{\text{fdbk}}^{(1)} = \left\lceil \log_2 S_{N_T}^{(\leq N_R)} \right\rceil + \left\lceil N_T \log_2 (N_{\text{mcs}} N_{\text{pwr}}) \right\rceil
\]  

(41)

making this strategy not very interesting. A second scenario consists in associating to any acceptable partition all possible discrete rates and quantized power levels and counting all resulting combinations. The feedback of each such combination therefore requires a constant amount of feedback bits, expressed as

\[
N_{\text{fdbk}}^{(2)} = \left\lceil \log_2 \sum_{k=1}^{N_T} S_{N_T}^{(k, \leq N_R)} (N_{\text{mcs}} N_{\text{pwr}})^k \right\rceil
\]  

(42)

Clearly, the worst case \(N_{\text{fdbk}}^{(1)}\) always exceeds \(N_{\text{fdbk}}^{(2)}\). An obvious way to further reduce the overhead feedback is to fix the partition of transmit antennas once and for all. Of course, the saving in feedback depends on the partition choice. It is shown in [25] that this non-adaptive option may compete with conventional PARC. Finally, the number of feedback bits in conventional PARC can be easily computed from the definition \(N_{\text{fdbk}}^{(2)}\) as

\[
N_{\text{fdbk}}^{(\text{parc})} = \left\lceil N_T \log_2 (N_{\text{mcs}} N_{\text{pwr}}) \right\rceil
\]  

(43)

In Section VI, we show that scenario 2 turns out to be an efficient design.

V. COMPLEXITY ANALYSIS

This section analyzes the computational complexity of the proposed transceiver, i.e., the way of computing capacities \(\{C_k\}_{k=1}^K\). In the sequel, we assume that the decoding order follows the natural order of layer indices. We focus on the number of floating point complex multiply operations as a crude approach to the measuring of computational complexity (for large enough matrix dimensions). The inversion of an \(n \times n\) Hermitian definite positive complex matrix involves \(n^3/3\) floating point complex multiply operations (Cholesky factorization). The product of an \(m \times n\) complex matrix \(A\) by an \(n \times p\) complex matrix \(B\) involves \(mnp\) floating point complex
multiply operations. The computation of the determinant of an \( n \times n \) Hermitian complex matrix requires \( n^3/3 \) floating point complex multiply operations (Cholesky factorization). Calculation of each \( C_k \) involves an \( N_R \times N_R \) matrix inversion of \( \Theta_{\Delta_h} \) given by (10) and a number of matrix multiplications which depends on index \( k \). The overall required complexity to compute capacities \( \{ C_k \}_{k=1}^{K} \) is a function of the partition \( \pi(\mathcal{A}|K) \) and of the decoding order \( \sigma \) and is expressed as

\[
 f_{\{ C_k \}_{k=1}^{K}} (\pi, \sigma) = \sum_{k=1}^{K} \left( |\Delta_k|^{3/3} + 2N^2_R|\Delta_k| + N^3_R/3 + \sum_{j=k+1}^{K} N_R|\Delta_j|^2 + N_R|\Delta_j| \right). \tag{44}
\]

In order to lessen the complexity, a fast matrix inversion update should rather be implemented. Indeed, noting that

\[
 \sum_{j \geq k}^{K} H_{\Delta_j} \Theta_{x_{\Delta_j}} H_{\Delta_j}^{\dagger} = \sum_{j \geq k+1}^{K} H_{\Delta_j} \Theta_{x_{\Delta_j}} H_{\Delta_j}^{\dagger} + H_{\Delta_k} \Theta_{x_{\Delta_k}} H_{\Delta_k}^{\dagger}
\]

the Woodbury identity can be applied to \( \Theta_{x_{\Delta_k}}^{-1} \) yielding the updating formula

\[
 \Theta_{x_{\Delta_k}}^{-1} = \Theta_{x_{\Delta_k}}^{-1} - \Theta_{x_{\Delta_k}}^{-1} H_{\Delta_k} A^{-1}_k H_{\Delta_k}^{\dagger} \Theta_{x_{\Delta_k}}^{-1}.
\]

This recursive formula is initialized with the identity matrix, i.e., \( \Theta_{x_{K}} = I_{N_R} \). Getting the successive values still involves a matrix inversion, but (in most cases) of reduced dimension \( |\Delta_k| \times |\Delta_k| \), since \( |\Delta_k| \leq N_R \). As a result, the complexity of \( \Theta_{x_{\Delta_k}}^{-1} \) is similarly dominated by the residual matrix inversion or by the matrix product, and is reduced to \( 2N^2_R|\Delta_{k+1}| + N_R|\Delta_{k+1}|^2 + |\Delta_{k+1}|^{3/3}/3 \). The overall required complexity to compute capacities \( \{ C_k \}_{k=1}^{K} \) is a function of the partition \( \pi(\mathcal{A}|K) \) and of the decoding order \( \sigma \) and is expressed as

\[
 g_{\{ C_k \}_{k=1}^{K}} (\pi, \sigma) = \sum_{k=1}^{K} \left( |\Delta_k|^{3/3} + 2N^2_R|\Delta_k| + 2N^2_R|\Delta_{k+1}| + N_R|\Delta_{k+1}|^2 + |\Delta_{k+1}|^{3/3}/3 \right) \tag{47}
\]

(with \( |\Delta_{K+1}| = 0 \). The complexity of the exhaustive search procedure requires summing the quantity \( g_{\{ C_k \}_{k=1}^{K}} \) over all possible ordered partitions, keeping in mind that many of them involve the same number of complex multiplications (this number is, in fact, primarily dictated by the distribution \( \{ |\Delta_k| \}_{k=1}^{K} \) for any fixed \( K \)). The complexity evaluation can be greatly simplified if we introduce the intermediate quantities

\[
 g^{(k)}_{\max} = \max_{(\pi,\sigma): \mathcal{A}(\pi) = K} g_{\{ C_k \}_{k=1}^{K}} (\pi, \sigma). \tag{48}
\]

If the decoding order is considered, the exhaustive search procedure has to run over all acceptable ordered partitions and its complexity is roughly upper bounded by

\[
 \sum_{k=1}^{N_T} k! S_{NT}^{(k; \leq N_R)} g^{(k)}_{\max}. \tag{49}
\]
If it is not, the exhaustive search procedure scans all acceptable non-ordered partitions and its complexity is roughly upper bounded by
\[
\sum_{k=1}^{N_T} S_{N_T}^{(k, \leq N_R)} g_{\text{max}}^{(k)}.
\]

In Table I, some numerical results are listed for typical values of \(N_T\) and \(N_R\). It is clear that the search space cardinality grows very fast with \(N_T\) when the decoding order is taken into account. This is due to the \(k!\) factors in (49). In particular, the last term \(N_T!S_{N_T}^{(N_T, \leq N_R)} = N_T!\) often dominates in (49) which suggests that the partition corresponding to PARC could be systematically discarded in order to reduce the complexity of the exhaustive search procedure.

No performance loss was observed with such an additional restriction. By way of example, we evaluate the complexity for a \(4 \times 4\) configuration. Without ordering, PARC needs 238 operations to compute the layered capacities whereas 3838 are required for LSTC-RPC. When the decoding order is activated, these values increase up to 5720 and 18778, respectively. In this case, 30% of the complexity of LSTC-RPC is due to the PARC partition.

VI. Numerical results

A. Simulation settings

We show average (ergodic) spectral efficiency, for Rayleigh flat-fading channels, to evaluate performance of previously presented algorithms. All simulations are averaged over 1000 channel realizations. Transmission occurs on a correlated quasi-static \(N_T \times N_R\) MIMO channel whose expression
\[
H = R_{N_R}^{1/2} H_w R_{N_T}^{1/2}
\]
captures reasonably the effects of spatial correlation. \(H_w\) is the spatially white MIMO flat-fading channel whose entries of \(H\) are circularly-symmetric complex zero-mean Gaussian random variables, i.e., \(h_{i,j} \sim \mathcal{CN}(0, 1)\). The \(N\)-dimensional correlation matrix follows the model in [19], i.e.,
\[
R_N = \begin{bmatrix}
1 & \alpha & \cdots & \alpha \\
\alpha & 1 & \ddots & \vdots \\
\vdots & \ddots & \ddots & \alpha \\
\alpha & \cdots & \alpha & 1
\end{bmatrix}
\]

where \(\alpha \in [0, 1]\) acts as the correlation parameter. This structure obviously represents a worst case, since in practice, correlation may decrease with antennas spacing. An equal power distribution is always assumed inside a block. For rate adaptation, transmission rates are chosen from
For conciseness, results are only presented for this rate set which is taken from [20] (OFDM case) since we believe it represents a reasonable assumption for future systems. However, we emphasize that the gains brought by the different investigated DoFs strongly depend on the available rates at transmitter. It seems natural that the LSTC-RPC encoding employs Space-Time Bit Interleaved Coded Modulation (STBICM) since proposed scalar MCSs in [20] are based on BICM. Furthermore, lots of contributions have already shown that this space time coding scheme is able to get very close to the outage probability under the assumption of Iterative Decoding (ID) (see, e.g., [21]). Thanks to (perfect) partial feedback, the channel is never in outage. It is thus reasonable to consider STBICM codebooks as capacity achieving. In this case, the finite discrete rate distribution $D_{|\Delta_k|}$ is easily obtained from the reference Table II by simply multiplying each final rate (column 4) by factor $|\Delta_k|$. Capacity results of LSTC-RPC are always for an optimized partition (adaptive mode), joint optimization with decoding order is not considered unless specified, e.g., LSTC-RPC with ordering.

B. Discussion

1) Optimized partition in LSTC-RPC: In this subsection, we consider a $4 \times 4$ spatially uncorrelated antenna configuration and equal power distribution. To this end, we perform an exhaustive search throughout all possible partitions. We observe from Fig. 2 that antenna partitioning DoF provides a great improvement over conventional PARC, at 10 bits/c.u. the gap is about 1.4 dB. We point out that the gap becomes larger at high SNR, e.g., 2.5 dB for a 16 bits/c.u. transmission. Furthermore, due to the upper rate limit (e.g., 64QAM with a rate 3/4 in our simulations), both systems will reach the same performance at a sufficient high SNR. This asymptotic limit clearly depends on the maximal discrete rate available. Decoding order when optimized (optionally combined with the other DoFs) actually provides substantial capacity gains at all SNRs, thus corroborating initial results in [14]. We note that the decoding order DoF is more efficient in PARC than LSTC-RPC. Indeed, PARC with optimized decoding order can approach the performance provided by LSTC-RPC without optimized decoding order. This comment actually holds for all tested antenna configurations, i.e., $N_T \times N_R$ with $N_T \leq 6$ and $N_R \leq 6$. However, the number $N_T!$ of possible decoding orders (assuming PARC) is greater than the number of possible partitions (without ordering) as shown in Table I. Although, decoding order does not
require extra feedback, this quantitative observation may speak in favor of this adaptive partition process in terms of receiver complexity. The feedback load of LSTC-RPC is only 1 bit larger than the one of PARC, i.e., 15 bits compared to 14 bits for PARC. Furthermore, if LSTC-RPC does not consider the PARC partition, the feedback decreases to 13 bits while the complexity decreases by 6% and the performance remains the same. To conclude, LSTC-RPC without PARC partition is better than PARC with ordering in terms of feedback, complexity and performance for the simulation setting of Fig. 2.

2) Power quantization in LSTC-RPC: In this subsection, we investigate the performance enhancement of power control algorithms. For 32 power levels (resp. 4) in a $4 \times 2$ spatially uncorrelated MIMO channel, LSTC-RPC and PARC have the same feedback load with 34 (resp. 22) bits. We plot an upper bound which is the 4-user vector MAC capacity with sum power constraint, i.e., PARC with optimal power allocation and continuous rate. Fig. 3 shows that the ORPQ algorithm is particularly efficient at low to moderate SNR for 32 power levels. Its interest is thus two-fold, since it provides similar performance as the vectorial SRPQ2 (version of the vectorial SRPQ with optimized decoding order as suggested by [14]) and avoids the full search over decoding orders. At a transmission of 4 bits/c.u., the ORPQ algorithm outperforms SRPQ1 and LSTC-RPC without power optimization by, respectively, 1 dB and 1.75 dB. The amount of power quantization is also addressed in Fig. 3. With only 4 power levels, power control algorithms perform poorly. Indeed, a coarse power quantization induces a huge capacity degradation that challenges the interest of power control. Consequently, power control needs a lot of feedback information to be attractive.

3) Antenna selection in LSTC-RPC: Comparison between power quantization and antenna selection is not pursued since these strategies require very different feedback loads. As a result, in this subsection, we compare the proposed joint iterative antenna selection algorithm (with either forward or reverse ordering) to the exhaustive joint selection, the serial selection\(^1\) of [15] and the disjoint selection. The latter consists of applying the quantization rate once the selected set has been chosen. We plot an upper bound which is in fact antenna selection with the assumption of a continuous rate set. It is worth pointing out that the ordering rule in [15] is based on MMSE-SIC SINRs. Not only does this initialization not yield a significant increase of performance but it requires additional complexity. In Fig. 4, comparison arises in a $4 \times 4$ highly

\(^1\)In [15], equation (14) has a wrong dimension, it should be entries of a $n_T - i + 1$ square matrix. Our simulations employ the corrected formula.
correlated channel with $\alpha = 0.6$ for both $R_{NR}$ and $R_{NT}$. It illustrates the effectiveness of our approach. As expected, the proposed joint iterative selection procedure performs close to the exhaustive search. Hence, near-optimal performance is reached while keeping complexity low. For the forward ordering $\sigma_{fwd}$, an enhancement of 1 dB and 2.7 dB at, respectively, transmission rate of 8 bits/c.u. and 14 bits/c.u. is obtained compared to the serial selection algorithm. We notice that the forward decoding order performs always better than the reverse one in PARC systems. The same trends, but more pronounced, can be observed for a highly correlated $4 \times 2$ channel with $\alpha = 0.6$ for both $R_{NR}$ and $R_{NT}$, see Fig. 5. For this antenna configuration, LSTC-RPC requires 15 bits compared to 14 bits for PARC. The substantial degradation observed at high SNR for the disjoint antenna selection is due to the large quantization noise it undergoes. As a result, the disjoint antenna selection chooses the best subchannels regardless of the largest MCS at disposal. It was shown in [26] that the percentage of active antennas on average is larger in the case of the proposed joint iterative antenna selection than in the case of the disjoint antenna selection.

VII. CONCLUSION AND OPEN RESEARCH TOPICS

In this paper, a new transceiver architecture has been developed which combines space-time coding and spatial multiplexing for transmission over wireless MIMO channels with partial feedback. The core idea is to exploit an additional DoF, namely the partitioning of the transmit antennas into groups. It has been proved that this design theoretically achieves the MIMO MAC capacity under a sum power constraint. Combinations of this third DoF with the other DoFs at disposal, i.e., power control and decoding order have been proposed and a variety of algorithms described aiming at achieving a good trade-off between complexity, feedback rate and performance. Both the usefulness of the concept and the efficiency of the algorithms are supported with convincing simulations. Indeed, in all investigated scenarios, fair improvements in terms of sum discrete-rate are observed compared to conventional PARC. An extension of this work to narrowband MIMO Rayleigh fading multipath channel model is straightforward [22, Section 4]. Further research topics include:

- Power quantization: it could be interesting to investigate the effect of a non-uniform power quantization in our optimization algorithms (as opposed to the uniform quantization policy considered here in each antenna group).
- Feedback load: we could formalize and conduct optimization under a sum feedback rate
constraint drawing our inspiration from [27].

- Error propagation in MMSE-SIC: capacity analysis assumes error-free decoding, but of course, with actual codes, errors are made. There is no doubt that the error propagation phenomenon increases with the number of antennas. Thus, we may expect a performance improvement of LSTC-RPC over conventional PARC due to the reduced number of decoding stages (at the price of an increased decoding complexity for each individual stage, except if iterative sub-optimal MMSE-PIC is performed).

- Imperfect channel estimation: an imperfect estimate of the MIMO wireless channel does not only produce residual cancellation errors in MMSE-SIC but it leads to possibly inappropriate feedback CQIs too. This detrimental effect should be studied carefully.

All those topics will be addressed in future contributions.

REFERENCES


Algorithm 1 Vectorial SRPQ

\[
P_{\text{remaining}} = P;
\]

for \( k = K \) to 1 do

Step 1 Equidistribute the remaining power: \( P_{\sigma(k)} = \frac{P_{\text{remaining}}}{k} \)

Step 2 Compute \( C_{\sigma(k)} \) from (9), and choose the closest discrete rate \( R_{\sigma(k)} \), see (22)

Step 3 Update \( P_{\sigma(k)} \) solution of (23) and quantize it \( \hat{P}_{\sigma(k)} \)

Step 4 Check the power budget violation:

if \( (P_{\text{remaining}} - \hat{P}_{\sigma(k)} < 0) \) then

Step 4.1 Take the immediate inferior discrete rate \( R_{\sigma(k)} \)

Step 4.2 Update \( P_{\sigma(k)} \) and quantize it \( \hat{P}_{\sigma(k)} \)

end if

\[
P_{\text{remaining}} = P_{\text{remaining}} - \hat{P}_{\sigma(k)};
\]

end for
Algorithm 2 ORPQ

\(P_{\text{remaining}} = P;\)

Init Find \(\sigma\) according to (25) (26)

for \(k = K\) to 1 do

Step 1 Assign to \(P_{\sigma(k)}\) the remaining power: \(P_{\sigma(k)} = P_{\text{remaining}}\)

Step 2 Compute \(C_{\sigma(k)}\) from (9), and choose the closest but inferior discrete rate \(R_{\sigma(k)}\)

Step 3 Update \(P_{\sigma(k)}\) solution of (23) and quantize it \(\hat{P}_{\sigma(k)}\)

Step 4 Check the power budget violation:

if \((P_{\text{remaining}} - \hat{P}_{\sigma(k)} < 0)\) then

Step 4.1 Take the immediate inferior discrete rate \(R_{\sigma(k)}\)

Step 4.2 Update \(P_{\sigma(k)}\) and quantize it \(\hat{P}_{\sigma(k)}\)

end if

\(P_{\text{remaining}} = P_{\text{remaining}} - \hat{P}_{\sigma(k)};\)

end for
Algorithm 3 Proposed joint selection

Init Order $S$ according to (31)
Initialize $S_{l=1}$ to $S$ and $R_{\text{sum}}^{(0)} = 0$

while $l \neq N_T$ do

Step 1 Let us evenly balance the power budget over antennas of $S_l$

Step 2 Scan all partitions:
for all $\pi_j$ do

Step 2.1 Initialize $\sigma$ according to (33)

Step 2.2 Calculate the sum discrete-rate $R_{\text{sum}}^{(l)}(S_l, \pi_j, \sigma)$ and stack the set of rates $\{R_{\phi}\}_{k=1}^{K}$
end for

Step 3 Find $J$ such that $J = \arg \max \ R_{\text{sum}}^{(l)}(S_l, \pi_j, \sigma)$

Step 4 Compare maximum values:
if $R_{\text{sum}}^{(l)}(S_l, \pi_J, \sigma) > R_{\text{sum}}^{(l-1)}(A^*, \pi^*, \sigma^*)$ then

Step 4.1 Update optimal values: $A^* = S_l$, $\pi^* = \pi_J$ and $\sigma^* = \sigma$

Step 4.2 Take off worst element of the set: $S_l = S_l \setminus \{S_l(\text{last})\}$ and set $l$ to $l + 1$
else

Step 4.1 Exit the algorithm
end if
end while
<table>
<thead>
<tr>
<th>Function</th>
<th>$N_T$</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_{N_T}^{( \leq N_R) }$</td>
<td>$N_T \geq N_T$</td>
<td>2</td>
<td>15</td>
<td>203</td>
<td>4140</td>
<td>115975</td>
</tr>
<tr>
<td></td>
<td>$N_R = 4$</td>
<td>2</td>
<td>15</td>
<td>196</td>
<td>3795</td>
<td>99146</td>
</tr>
<tr>
<td></td>
<td>$N_R = 2$</td>
<td>2</td>
<td>10</td>
<td>76</td>
<td>764</td>
<td>9496</td>
</tr>
<tr>
<td>$\sum_{k=1}^{N_T} k! S_{N_T}^{(k \leq N_R) }$</td>
<td>$N_T \geq N_T$</td>
<td>3</td>
<td>75</td>
<td>4683</td>
<td>545835</td>
<td>102247563</td>
</tr>
<tr>
<td></td>
<td>$N_R = 4$</td>
<td>3</td>
<td>75</td>
<td>4670</td>
<td>543130</td>
<td>101519250</td>
</tr>
<tr>
<td></td>
<td>$N_R = 2$</td>
<td>3</td>
<td>66</td>
<td>3690</td>
<td>385560</td>
<td>64751400</td>
</tr>
<tr>
<td>$N_T!$</td>
<td></td>
<td>2</td>
<td>24</td>
<td>720</td>
<td>40320</td>
<td>3628800</td>
</tr>
</tbody>
</table>

**TABLE I**

Search space cardinality as a function of $N_T$ and $N_R$
**TABLE II**  
**AVAILABLE MCSs**

<table>
<thead>
<tr>
<th>Modulation format</th>
<th>Outer coding rate</th>
<th>Repetition factor</th>
<th>Final discrete rate (bits/c.u.)</th>
<th>MCS index</th>
</tr>
</thead>
<tbody>
<tr>
<td>QPSK</td>
<td>1/4</td>
<td>19</td>
<td>0.026</td>
<td>1</td>
</tr>
<tr>
<td>QPSK</td>
<td>1/4</td>
<td>9</td>
<td>0.055</td>
<td>2</td>
</tr>
<tr>
<td>QPSK</td>
<td>1/4</td>
<td>2</td>
<td>0.25</td>
<td>3</td>
</tr>
<tr>
<td>QPSK</td>
<td>1/4</td>
<td>1</td>
<td>0.5</td>
<td>4</td>
</tr>
<tr>
<td>QPSK</td>
<td>1/2</td>
<td>1</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>QPSK</td>
<td>3/4</td>
<td>1</td>
<td>1.5</td>
<td>6</td>
</tr>
<tr>
<td>16QAM</td>
<td>1/2</td>
<td>1</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>16QAM</td>
<td>3/4</td>
<td>1</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>64QAM</td>
<td>3/4</td>
<td>1</td>
<td>4.5</td>
<td>9</td>
</tr>
</tbody>
</table>
Fig. 1. LSTC-RPC transceiver scheme
Fig. 2. $4 \times 4$ Rayleigh channel, $\alpha = 0$, effect on partitioning and decoding order DoFs
Fig. 3. $4 \times 2$ Rayleigh channel, $\alpha = 0$, effect of quantized power control
Fig. 4. $4 \times 4$ Rayleigh channel, $\alpha = 0.6$, comparison of antenna selection procedures
Fig. 5. $4 \times 2$ Rayleigh channel, $\alpha = 0.6$, comparison of antenna selection procedures