# Parallel QRD-M encoder for multi-user MIMO systems

Manar Mohaisen, Abedelaziz Mohaisen, Merouane Debbah

**Telecommunication Systems manuscript No.**
(will be inserted by the editor)

# Parallel QRD-M Encoder for Multi-user MIMO Systems

**Manar Mohaisen · Abedelaziz Mohaisen · Merouane Debbah**

Received: date / Accepted: date

**Abstract** In the context of multi-user precoding, the idea behind vector perturbation (VP) lies in adding an integer vector to the data vector such that the overall transmit power is reduced, where the performance at the users end is consequently improved. In the literature, several techniques have been proposed to find a quasi-optimum perturbing vector, where this process was represented as an integer lattice search problem. In this paper, we propose a parallel QRD-M encoder (PQRDME) that, besides attaining a quasi-optimum diversity order, leads to tremendous reduction in the latency of the vector perturbation stage. Based on the set grouping, the proposed encoder transforms the full tree-search of the conventional QRDME into partial trees that can be pipelined to increase the encoding throughput. We evaluate the proposed algorithm under several scenarios with both perfect channel state information (PCSI) and imperfect CSI (ICSI) at the transmitter side, where simulation results show robust performance when compared to the optimum encoder.

**Keywords** MU-MIMO precoding · vector perturbation · QRD-M encoder · sphere encoder

Part of this paper has been presented at IEEE International Conference on Communications (ICC'11), Kyoto, Japan [16].

This work is supported by research subsidy for newly-appointed professor at Korea University of Technology and Education (KoreaTech) for the period 2012- 2013.

M. Mohaisen
Korea University of Technology and Education (KoreaTech)
Department of EEC Engineering
303-708 Cheonan, R. of Korea
E-mail: manar.subhi@kut.ac.kr

A. Mohaisen
University of Minnesota
Minneapolis, Minnesota 55455, USA
E-mail: mohaisen@cs.umn.edu

M. Debbah
SupElec, Paris, France
E-mail: merouane.debbah@supelec.fr

## 1 Introduction

Compared with the single-user multiple-input multiple-output (SU-MIMO) systems [18], multi-user MIMO (MU-MIMO) systems can achieve tremendous capacity gains without requiring additional spatial or spectral resources [22]. To this end, downlink precoding techniques based on the information-theoretic concept of dirty paper coding (DPC), first proposed by Costa in [5], can be used. Considering the knowledge of the data vector and the channel state information (CSI) at the base station (BS) side, inter-user interference (IUI) can be canceled so that each user only receives its designated data.

Several MU-MIMO precoding techniques were proposed in the literature in order to achieve the near-capacity. Among these techniques the linear precoding which consists of pre-processing the data vector using a criterion-based filtering matrix, which is constructed using either the zero-forcing (ZF) or the minimum-mean square error (MMSE) criterion [8] (a.k.a. channel inversion and regularized inversion, respectively.) Although the MMSE precoding alleviates the noise amplification problem appearing in the case of ZF precoding—especially when the channel matrix is ill-conditioned, it still performs much worse than the optimum brute-force encoder in terms of bit-error rate (BER) performance and diversity order.

To alleviate the noise amplification, lattice-base reduction techniques iteratively find a better-conditioned

base, i.e., a channel matrix with shorter and more orthogonal columns. Therefore, linear precoding is sufficient to achieve quasi-optimum performance due to the good conditionality of the reduced basis of the channel matrix. Lenstra-Lenstra-Lovasz (LLL) algorithm [11] and Seysen's algorithm (SA) [19] are extensively used to obtain better bases, where LLL algorithm is shown to require fewer computations than SA for an almost negligible degradation in the BER performance [1]. However, drawbacks of the lattice-reduction-based precoding techniques are in their iterative nature that slows down the encoding process, and their high worst-case computational complexity, which is exponential in the case of ill-conditioned channel matrix. Theoretically, Jalden *et al.* have shown in [9] that the number of iterations of the LLL algorithm is infinite when the channel matrix has a very high condition number.

To further improve the performance of MU-MIMO precoding without requiring lattice-basis reduction, the Tomlinson-Harashima precoding (THP) reduces the required transmit power via employing a non-linear modulus operation [6,21]. A linear version of this technique consists of two stages [12]: (i) a vector perturbation (VP) stage, where the data vector is perturbed by an integer-valued vector such that the required transmit power is reduced, and (ii) a precoding stage where the perturbed vector is linearly precoded using any of the aforementioned linear techniques. Despite of the apparent improvement by the THP, its vector perturbation stage is equivalent to the successive interference cancellation used in the MIMO detection literature, which is not optimum. As such, further reduction in the required transmit power can be achieved once the VP stage is optimized.

The VP problem was solved in [3,7] using the sphere encoder (SE) which suffers from sequential search nature and high worst-case complexity. In [23], a QRDM encoder (QRDME) was proposed to fix the computational complexity of the SE with quasi-optimum performance. Recently, we proposed a fixed-complexity SE (FSE) that overcomes both shortcomings of the SE, while lagging the performance of the optimum encoder [15].

**Contributions:** In this paper, we propose a parallel QRDME (PQRDME) that achieves a tradeoff between the attained performance and the precoding throughput. Based on the set grouping, the full vector perturbation (i.e., tree-search phase) is divided into partial vector perturbation (PVP) problems that can be processed in parallel, leading to a substantial speed-up of the precoding stage. A trade-off between precoding throughput and system performance is achieved by setting the number of PVPs; a high number of PVPs indicates higher gain in the precoding throughput but also leads to sacrificed performance, and vice-versa. In [16], the proposed encoder is evaluated under the assumption of single-stream transmission where the channel state is considered to be perfectly known at the transmitter. In this paper, we further consider the case of multi-stream communications, where users can be equipped with more than a single receive antenna. To this end, block diagonalization (BD) [4] is used as a pre-processing stage that eliminates the IUI. Furthermore, we consider imperfect knowledge of the CSI at the transmitter side [2]. Therefore, we employ two types of channel quantization: (i) the nonuniform Lloyd-Max quantizer which is used to independently quantize the real and imaginary parts of each channel coefficient, and (ii) the phase of the channel is quantized using a uniform quantizer, where the channel amplitude is considered to be known at the transmitter. This assumption is motivated by the fact that in future-generation communication systems, several uplink control signals, such as channel quality indicator (CQI) [10], indicating the amplitude or reliability of the channel are already assumed to be fed back to the BS from users. Finally, we evaluate the performance of the proposed encoder under several scenarios and compare it with the conventional MU-MIMO precoding techniques. Finally, we bring more insight into the selection of candidates set size from which the elements of the perturbing vector are drawn.

**Organization:** This paper is organized as follows. In Section 2 we introduce the system model and statement of problem. In Section 3 we briefly describe the conventional VP techniques and outline our motivations. In Section 4, we introduce the proposed PQRDME techniques and outline its merits. In Section 5, we show simulation results and finally draw conclusions in Section 6.

## 2 System Model and Statement of Problem

In this paper, we consider a downlink system in which a single BS equipped with $n_T$ transmit antennas simultaneously communicates with $n_U$ users each equipped with $n_R$ receive antennas. In this paper, we consider the minimum number of required transmit antennas, that is, $n_T = (n_U \times n_R)$. Also, let the system be transformed into the $K$-dimensional real Euclidean space with $N = (2 \times n_T)$, then under the assumption of narrow-band flat-fading, the downlink MU-MIMO channel is given by:

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_1^T & \mathbf{H}_2^T & \cdots & \mathbf{H}_{n_U}^T \end{bmatrix}^T, \tag{1}$$

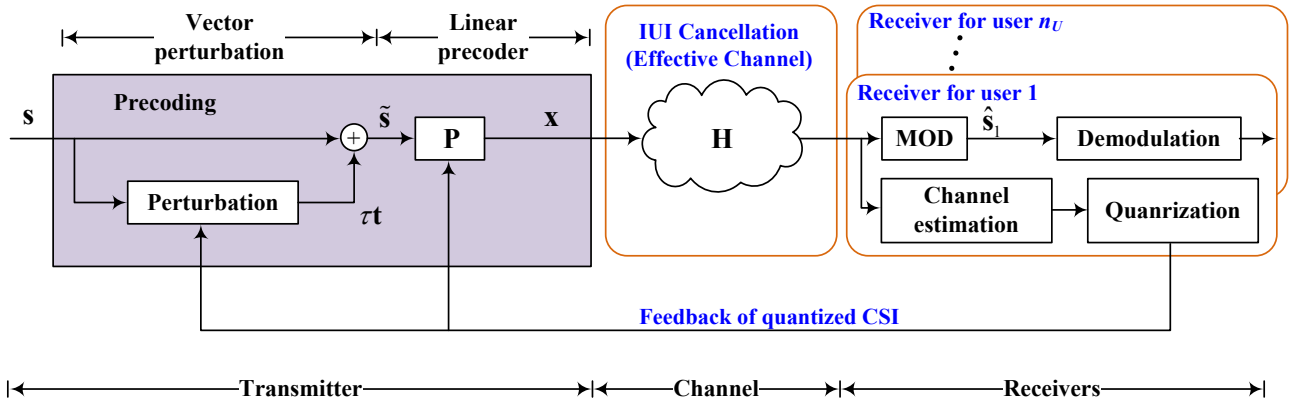


**Fig. 1** Structure of the MU-MIMO precoding system with vector perturbation and imperfect CSI at the transmitter.

where $\mathbf{H}_i \in \mathbb{R}^{(2\times n_R)\times N}$ is the channel coupling the $n_T$ transmit antennas to the $n_R$ receive antennas of user $i$, and $(\cdot)^T$ denotes the matrix transpose.

Having both the data vector $\mathbf{s} \in \mathbb{R}^N$ and the channel matrix $\mathbf{H}$ at the BS, linear precoding techniques treat the data vector using a filtering matrix to obtain the following precoded vector:

$$\mathbf{x} = \mathbf{P}\frac{\mathbf{s}}{\sqrt{\gamma}} = (\mathbf{H}^T\mathbf{H} + \alpha\mathbf{I}_N)^{-1}\mathbf{H}^T\frac{\mathbf{s}}{\sqrt{\gamma}}, \tag{2}$$

where $\alpha$, the regularization coefficient, is set to zero in the case of ZF precoding and to a value proportional to the noise variance in the case of MMSE precoding. Also, the scaling factor $\gamma$ is present to fix the expected transmit power to a predefined value, which is set to 1. Therefore, the receive signal-to-noise ratio (SNR) at each receive antenna is given by $E(s^*s)/\gamma\sigma_n^2$. Since $\gamma$ is inversely proportional to the conditionality of channel matrix, the receive SNR is degraded when the channel matrix is ill-conditioned.

As an alternative to the linear precoding techniques, THP reduces the required transmit power using a non-linear modulo operation that limits the power of the precoded symbols. Conventionally, symbols are precoded successively, where each precoded symbol, e.g., $s_i$, is rounded into a predefined range. In [12], a linearized version of the THP was introduced where the algorithm is split into two successive stages; namely, the VP stage and the linear precoding stage as shown in Fig. 1. The VP stage of the THP algorithm is given as:

$$\tilde{\mathbf{s}} = \mathbf{s} + \tau\mathbf{t}, \tag{3}$$

where $\mathbf{t}$ is an $N$-dimensional vector with integer elements. In [7], $\tau$ is defined as an integer given by:

$$\tau = 2\left(|c_{max}| + \Delta/2\right), \tag{4}$$

where $|c_{max}|$ is the absolute value of the symbol with the largest magnitude, and $\Delta$ is the spacing between any two neighbor symbols. Now, let the precoder fully equalizes for the channel effect, then the data symbols can be recovered at the receivers using a simple modulo operation that reduces the range of the received signal $\mathbf{y}$ to $[-K, K)$ with $K$ as the square root of the cardinality of the modulation set.

Despite the improvement achieved by the THP compared to the linear precoders, its VP stage is equivalent to the successive interference cancellation, where only a single $t$ candidate is retained at each VP level. This technique is not optimum, which means that further reduction in the required transmit power, i.e., smaller $\gamma$, can be achieved once the VP stage is optimized.

To this end, the VP is represented as an integer lattice-search as following:

$$\begin{aligned}\mathbf{t} &= \arg\min_{\mathbf{t}\in\mathbb{Z}^N}\left\{(\mathbf{s}+\tau\mathbf{t})^T\mathbf{P}^T\mathbf{P}(\mathbf{s}+\tau\mathbf{t})\right\},\\ &= \arg\min_{\mathbf{t}\in\mathbb{Z}^N}\left\|\mathbf{P}(\mathbf{s}+\tau\mathbf{t})\right\|^2.\end{aligned} \tag{5}$$

Now, let the transpose of the matrix $\mathbf{H}$ be factorized into the product of a unitary matrix $\mathbf{Q}$ and an upper triangular matrix $\mathbf{R}$, thus, the search problem in (5) based on the zero-forcing criterion is simplified into:

$$\begin{aligned}\mathbf{t} &= \arg\min_{\mathbf{t}\in\mathbb{Z}^K}\left\|\mathbf{L}(\mathbf{s}+\tau\mathbf{t})\right\|^2,\\ &= \arg\min_{\mathbf{t}\in\mathbb{Z}^K}\sum_{i=1}^{K}\left\|L_{i,i}(s_i+\tau t_k) + \sum_{j=1}^{i-1}L_{i,j}(s_j+\tau\hat{t}_j)\right\|^2,\end{aligned} \tag{6}$$

where the lower-triangular matrix $\mathbf{L} = (\mathbf{R}^{-1})^T$. As such, the VP problem is defined as finding the integer vector $\mathbf{t}$ that minimizes $\gamma$. In the sequel, $t$ will be drawn from the integer set $\mathcal{A}$ of size $T$. In the following section, we introduce several conventional VP techniques and the motivations for the proposed PQRDME.

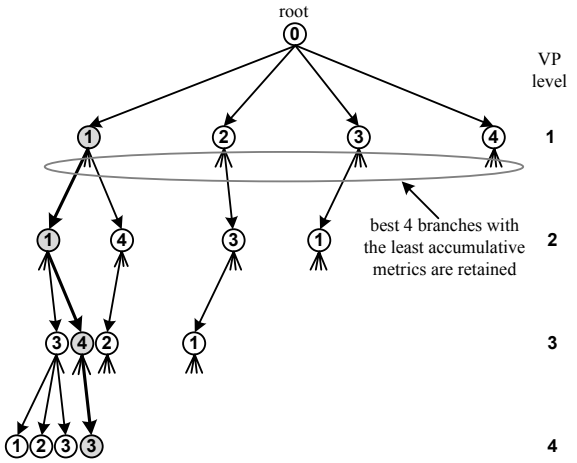


**Fig. 2** An example of the conventional QRDM encoder for $M = T = K = 4$.

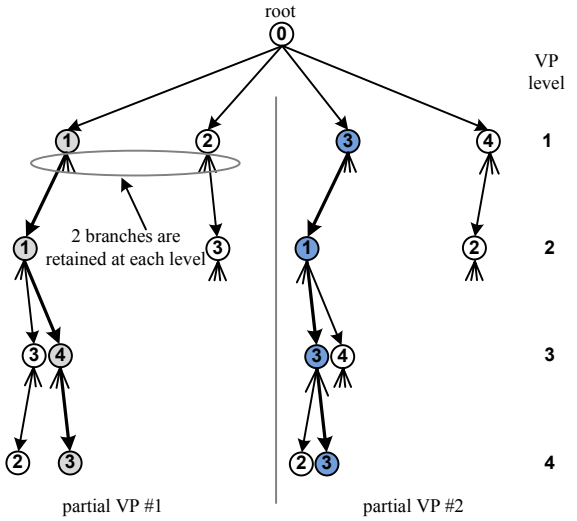


**Fig. 3** An example of the proposed parallel QRDM encoder for $p = 1$, $G = M_G = 2$, and $T = K = 4$.

## 3 Conventional VP techniques and Motivations

In this section, we briefly introduce three conventional MU-MIMO precoding techniques, namely, SE, QRDME, and FSE.

The main idea of the SE is to restrict the search for the optimum perturbing vector inside a hypersphere of a predefined radius. That is,

$$\mathbf{t} = \underset{\mathbf{t}\in\mathbb{Z}^N}{\arg\min} \left( \|\mathbf{P}(\mathbf{s} + \tau\mathbf{t})\|^2 \leq d^2 \right), \tag{7}$$

where $d$ is the search radius. SE is shown to achieve a quasi-optimum performance with low average computational complexity [7]. However, SE has two drawbacks: (i) its worst-case complexity is exponential, i.e., comparable to that of the brute-force search, and (ii) it has an iterative tree-search phase that limits the efficient implementation using pipelining.

To overcome the random computational complexity of the SE, the QRDME has been proposed in [23]. In the QRDME, the best $M$ candidates with the least accumulative metrics are retained at each encoding level. Fig. 2 depicts an example of the QRDME $M = T = K = 4$. At the first encoding level, the best 4 candidates are retained for the second encoding level. At the last encoding level, the perturbed vector with the least accumulative metric is precoded and transmitted. Although QRDME achieves a quasi-optimum performance, its tree-search phase cannot be efficiently implemented using pipelining, where consequently the encoding throughput is reduced.

In [15], we proposed the FSE, which overcomes both drawbacks of the SE. The employed tree search strategy in the FSE achieves a full parallelization of the VP stage, leading to tremendous gains in the precoding throughput. The only drawback of the FSE is that it lags the optimum BER performance.

In this paper, we propose a PQRDME that achieves a tradeoff between the achieved performance and the precoding throughput by selecting the appropriate number of PVPs.

## 4 Proposed Parallel QRDM Encoder

The goal of the PQRDME is to achieve a tradeoff between encoding throughput, i.e., speed, and system performance. To do that, the set of candidates for the elements of $\mathbf{t}$, i.e., $\mathcal{A}$, is divided into non-overlapping subsets of equal cardinalities. That is,

$$\mathcal{A} = \bigcup_{i=1}^{G} D_i, \tag{8}$$

where $G$ is the number of subsets and the $i$-th subset is designated $D_i$. The tree-search of the conventional QRDME is then divided into independent PVPs that are processed in parallel, where the candidates for $t_1$ at the $j$-th PVP is drawn from the subset $D_j$. Moreover, all the candidates for $t$ are retained at the first $p$ levels. When $p = 1$, each PVP represents a smaller tree-search equivalent to that of the conventional QRDME with the difference that the candidates for $t_1$ are drawn from the partial subset $D$. However, when $p = 2$, all the resulting branches at the first and second encoding levels are retained. In the following, PQRDME for $p = 1$ and $p = 2$ are referred to as PQRDME-$p1$ and PQRDME-$p2$, respectively. Note that the candidates for $t_{p+1}$ to $t_K$ are drawn from the full set $\mathcal{A}$.

Figure 3 depicts an example of the proposed PQRDME for $p = 1$, $G = M_G = 2$, and $T = K = 4$, where $M_G$ is the number of candidates retained at each encoding level per PVP. Note that the number inside the circle representing each node indicates the index of the hypothesis for $t$. At the first encoding level, the set of candidates $\mathcal{A}$ is divided into two subsets of equal sizes. As a result, the full tree is divided into two parallel and independent PVP problems. In the $j$-th PVP, the root node is extended to all possible candidates of $(s_1 + \tau t)$, where $t \in D_j$. All obtained branches are retained and their metrics are computed based on (5). At the second encoding level, all the retained candidates from the first level are extended to all possible candidates of $(s_2 + \tau t)$, where $t \in \mathcal{A}$, the accumulative metric of all the resulting branches are computed using (5) and the best two (i.e., $M_G$) candidates with the least accumulative metrics are retained to the following encoding levels. This process is repeated up to the last encoding level. For each PVP, the best perturbed vector, with the least accumulative metric, is retained. These best candidates per each PVP are compared and the best among them is announced as the result of the proposed PQRDME-$p1$ algorithm. The perturbed vector is then precoded and transmitted via the $n_T$ transmit antennas.

Figure 4 depicts an example of the proposed PQRDME for $p = 2$, $G = M_G = T = 2$, and $K = 4$. At each PVP, the root is extended to all possible candidates of $(s_1 + \tau t)$ with $t \in D_j$, where all candidates are retained (two candidate per PVP in this example). The retained candidates at the first encoding level are extended to all possible candidates $(s_2 + \tau t)$, where $t \in \mathcal{A}$. Also, all the resulting candidates are retained. In the third encoding level, the extended candidates are sorted based on their accumulative metrics and the two (i.e., $M_G$) with the least accumulative metrics are retained for the next encoding levels. The process is repeated as with PQRDME-$p1$, where at the last encoding level the best candidate that requires the least transmit power is precoded and transmitted.

Based on the above description, it stems out that the proposed PQRDME achieves a tradeoff between the parallelization capabilities, by controlling the number of PVPs ($G$), and the performance of the MU-MIMO system. That is, the proposed encoder is at least $G$ times faster than the conventional QRDME for a tolerable degradation in the BER performance. Also, compared to the conventional QRDME, the proposed PQRDME requires a lower number of comparisons to select the retained hypotheses at each encoding level. The QRDME retains all candidates at the first encoding level, therefore no comparisons are required. The $T = M$ candidates retained at the first encoding level are expanded

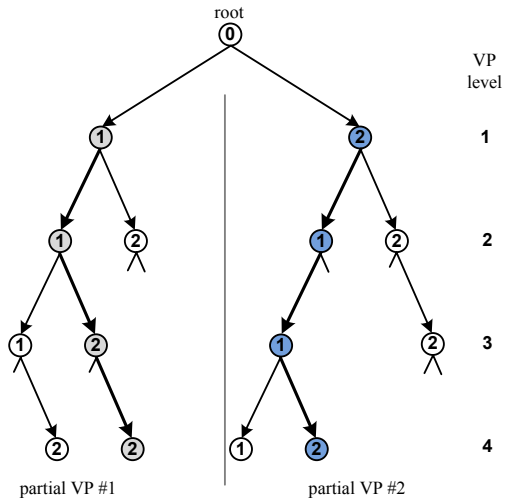


**Fig. 4** An example of the proposed parallel QRDM encoder for $p = 2$, $G = M_G = T = 2$, and $K = 4$.

to all possible hypotheses, leading to $MT$ branches. To order these branches and select the best $M$ candidates, we require $((M \times T)^2 - 1)$ comparisons. The same applies to all the encoding levels except the first, leading to an overall number of comparisons of $((M \times T)^2 - 1)(K - 1)$, which is equal to 765 for the parameters given in Fig. 2. In the case of PQRDME-$p1$, the first encoding level does not require any comparison. In the following levels, $((M_G \times T)^2 - 1)$ comparisons are required per PVP to sort the extended branches and select the best $M_G$ candidates. This leads to a total of $G((M_G \times T)^2 - 1)(K - 1)$ comparisons, which is equal to 378 for the parameters given in Fig. 3. Finally, PQRDME-$p2$ does not require comparisons in both the first and second encoding levels, therefore it requires $G((M_G \times T)^2 - 1)(K - 2) = 60$ comparisons, for the parameters given in Fig. 4. Thus, the proposed PQRDME requires a fraction of the number of comparisons performed by the conventional QRDME. Since comparisons are done serially, which add further delay to the VP stage, the proposed encoder achieves additional gains in the encoding throughput besides reducing the computational power required to perform the comparisons. In terms of the number of visited nodes and based on the parameters of Figs. 2-4, the conventional QRDME visits $T$ nodes at the first encoding level and $(M \times T)$ nodes at each of the remaining encoding levels leading to a total of $(T + (M \times T)(K - 1)) = 52$ nodes, for the parameters given in Fig. 2. The proposed PQRDME-$p1$ visits $T$ nodes at the first encoding level, and $M_G T$ nodes per level per PVP at the remaining encoding levels, leading to a total of $(T + (GM_G T)(K - 1)) = 52$ nodes, for the parameters given in Fig. 3. Finally, the proposed PQRDME-$p2$ visits $T$ nodes at the first encoding level,

$GT$ nodes at the second encoding level, and $M_G T$ nodes per level per PVP at the remaining $(K-2)$ levels, leading to a total of $(T(G+1)+(GM_GT)(K-2))$ nodes, which is equal to 22 for the parameters given in Fig. 4. This emphasizes on the effectiveness of the proposed PQRDME-$p2$ in terms of both number of comparisons and number of visited nodes, as compared to the conventional QRDME.

## 5 Simulation Results, Discussions, and Analyses

### 5.1 Simulation environment

In the following, we use the linear MMSE criterion in the precoding stage due to its superior performance, as compared to the ZF criterion. Bits are mapped to symbols using quadrature-phase shift keying (QPSK) modulation. Also, we consider that the channel is either perfectly or imperfectly known at the transmitter, where the imperfection in the channel knowledge is due to quantization error. Users are considered to be decentralized, hence they are non-cooperative. The BD of the channel matrix is achieved via the singular-value decomposition block diagonalization.

### 5.2 Setting the size of $\mathcal{A}$

Unlike most works in the MIMO detection literature, where the candidates for the transmitted data belong to a predefined modulation set, the size of the set $\mathcal{A}$ is not theoretically defined, but rather depends on the system configuration and the employed VP technique. Therefore, in this paper we use intensive simulations to identify the size of the set $\mathcal{A}$ for the conventional and the proposed encoders.

Table 1 summarizes the probability of $t$ for $t \in \mathcal{A}$ in the perturbing vector. We remarked that prob$(t = +a)$ = prob$(t = -a)$ for $a \in \mathcal{A}$, that is why we present the values as prob$(t = \pm a)$ which is equal to $2\times$prob$(t = +a)$. We remark that the size of the set $\mathcal{A}$ for the less optimized algorithms, such as the THP and FSE-$p1$, is smaller compared to that of the more optimized QRDME. Also, we remarked that the probability of $(|t| > 1)$ is very small but it has an impact on the BER performance, especially at high value of signal-to-noise ratio (see [15], Fig. 3.) Besides, we remarked that these values are not affected by the size of the system (i.e., $n_U$ and $n_T$). Finally, since the proposed PQRDME-$p2$ algorithm achieves a quasi-optimal performance with $T = 3$ (see Fig. 7), where a marginal gain in the BER can be achieved by increasing $T$, the size of the set $\mathcal{A}$ is therefore restricted to 3.

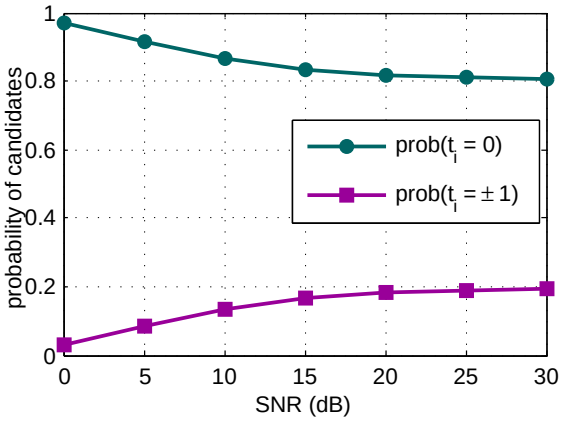


**Fig. 5** Probability of the candidates for $t$ using the proposed PQRDME for $p = 2$, $G = M_G = T = 3$, and $K = 8$.

Fig. 5 depicts the probability of candidates for $t$ using the proposed PQRDME for $p = 2$, $G = M_G = T = 3$, and $K = 8$ for several SNR values. At low SNR values, prob$(t = 0)$ is close to 1, which indicates that the VP stage is not efficient and the resulting transmitter is equivalent to the MMSE precoder. However, as the SNR increases, prob$(t = \pm 1)$ increases, where the required transmit power is consequently reduced. This can be seen from the BER curves where the performance of the linear equalization and the proposed encoder coincide at low SNR values while the proposed PQRDME achieves better performance at high SNR close to the optimum encoder.

### 5.3 Single-stream transmission with perfect CSI at the transmitter

Fig. 6 shows the BER performance of the proposed PQRDME for $p = 1$ and $M = (G \times M_G) = T = 8$ and those of several conventional precoding techniques. Note that we set $T = 5$ and $T = 7$ in the case of THP and FSE-$p1$, respectively, since no further improvement is achieved by increasing $T$. For $G = 2$, PQRDME-$p1$ performs close to the QRDME algorithm, while increasing the encoding throughput by a factor of two. When $G$ is increased to 4, the encoding throughput is increased by a factor of four, while the performance is degraded due to the increased parallelization. However, the proposed PQRDME-$p1$ still outperforms the FSE algorithm. Also, as shown in Figure 6, THP and LMMSE have lower performance and diversity order compared to the proposed PQRDME-$p1$ algorithm.

Fig. 7 depicts the performance of the proposed PQRDME-$p2$. Note that for $p = 2$, the performance of the FSE

**Table 1** Probability of $t$ in the perturbation vector $\tau\mathbf{t}$ in a MU-MIMO system for $K = 8$, $n_U = 4$, and SNR = 20dB, averaged over 100,000 independent trials.

| Algorithm | prob($t = 0$) | prob($t = \pm1$) | prob($t = \pm2$) | prob($t = \pm3$) | prob($t = \pm4$) |
|---|---|---|---|---|---|
| THP ($T = 9$) | 0.8760 | 0.1231 | 0.0009 | 0 | 0 |
| FSE-$p1$ ($T = 9$) | 0.8228 | 0.1737 | 0.0034 | 0.0002 | 0 |
| QRDME ($T = 9$) | 0.8133 | 0.1862 | 0.0005 | 0.0002 | 0.0001 |
| PQRDME-$p1$ ($T = 9$, $G = 3$) | 0.8228 | 0.1737 | 0.0034 | 0.0002 | 0.0001 |
| PQRDME-$p1$ ($T = 8$, $G = 4$) | 0.8278 | 0.1703 | 0.0019 | 0.0002 | 0 |
| PQRDME-$p1$ ($T = 8$, $G = 2$) | 0.8194 | 0.1800 | 0.0006 | 0.0002 | 0 |
| FSE-$p2$ ($T = 3$, $G = 3$) | 0.8096 | 0.1904 | - | - | - |
| PQRDME-$p2$ ($T = 3$, $G = 3$) | 0.8160 | 0.1840 | - | - | - |

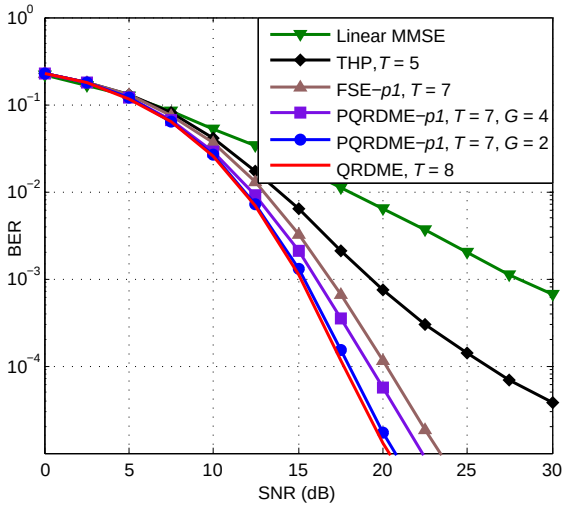


**Fig. 6** BER performance of the proposed PQRDME for $p = 1$, $M = T$, and $K = 8$.

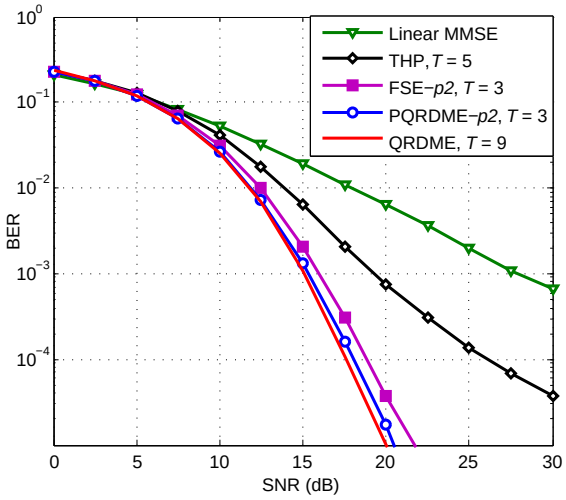


**Fig. 7** BER performance of the proposed PQRDME for $p = 2$, $M_G = T$, and $K = 8$.

and that of the proposed PQRDME algorithm are improved compared to the case of $p = 1$. Yet, the proposed PQRDME-$p2$ outperforms the FSE-$p2$ by about 1dB at a target BER of $10^{-4}$. Also, the proposed PQRDME-$p2$, at least, triples the speed of the VP stage. This happens because the proposed encoder not only parallelizes the VP stage but also reduces the number of visited nodes and the performed comparisons to select the best candidates at each encoding level.

### 5.4 Multi-stream transmission with perfect CSI at the transmitter

When the number of users increases and each user is equipped with more than one antenna, the encoder becomes extremely slow due to the high dimension of the tree to be searched. To overcome this problem, the IUI can be cancelled out at the transmitter side by means of block diagonalization (BD). Therefore, the goal of the BD is to transform the MU-MIMO channel into several parallel SU-MIMO channels with zero inter-channel interference [20]. That is, BD find the matrix $\mathbf{B}$ such that

$$\mathbf{H}_{\text{eff}} = \mathbf{HB} = \begin{bmatrix} \mathbf{H}_{\text{eff},1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_{\text{eff},2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{H}_{\text{eff},n_U} \end{bmatrix}. \tag{9}$$

In this scenario, $\mathbf{B}$ can be seen as a beamforming matrix, with $\mathbf{H}_{\text{eff},i}$ as the effective channel matrix for user $i$. As a result of the BD, users' data can be perturbed and encoded independently, leading to further gains in the encoding throughput.

In this paper, we consider a $(n_T, n_U, n_R)$ MU-MIMO system with $n_T = (n_U \times n_R)$. Fig. 8 depicts a BER comparison between the proposed PQRDME-$p2$ and FSE-$p2$ in $(8, 4, 2)$ and $(8, 2, 4)$ systems. The proposed PQRDME-$p2$ outperform the conventional encoder by 0.8dB and 1.5dB in the $(8, 2, 4)$ and $(8, 4, 2)$ systems, respectively.

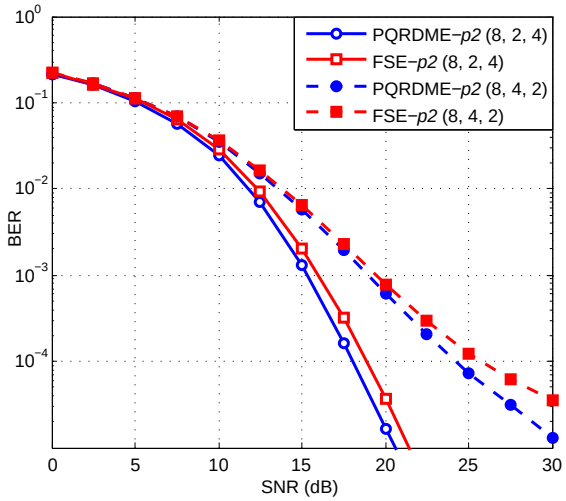


**Fig. 8** BER performance of the proposed PQRDME-$p2$ for $M_G = T = 3$, and a $(n_T, n_U, n_R)$ MU-MIMO system.

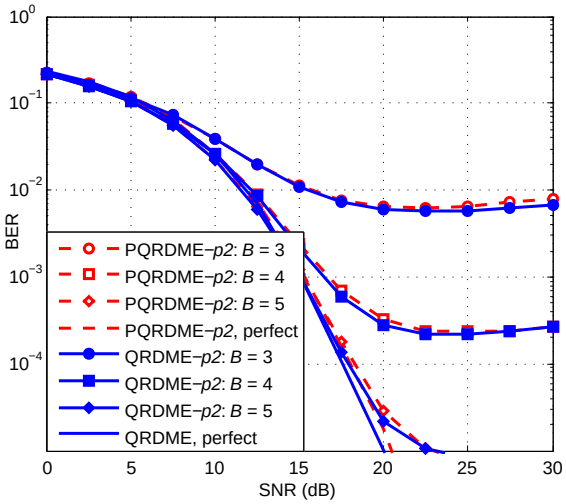


**Fig. 9** BER performance of the proposed PQRDME-$p2$ for $M_G = T = 3$ and $K = 8$ with imperfect CSI at the transmitter. The real and imaginary parts are quantized using Lloyd-Max non-uniform quantizer, where each part is represented by $B$ bits.

### 5.5 MU-MIMO system with imperfect CSI at the transmitter

In this section, we consider the channel coefficients to be quantized at the mobile stations. Then, the quantized version of each channel coefficient is fed back to the BS over a fast allocated control channel. In this paper, we consider the following two quantization schemes:

1. **Lloyd-Max non-uniform quantizer:** Uniform quantization divides the range to which the variable belongs into equal intervals. If the variable to be quantized belongs to a certain interval, then the centroid of the interval is considered as the quantized version. This quantizer is suitable for uniformly distributed variables, which is not the case of the MU-MIMO system. In this paper, we consider the real and imaginary parts of the channel to be i.i.d. centered Gaussian variables. Therefore, the non-uniform Lloyd-Max quantizer, which takes the probability density function (pdf) of the variables to be quantized into consideration, is more suitable [13,14]. The Lloyd-Max quantizer iteratively finds the intervals' endpoints so that the mean square error (MSE) between each channel coefficient and its quantized version is minimized. This can be achieved by allocating shorter intervals when the pdf has high values (i.e., the variable is most probable) and longer intervals when the pdf has low values. Fig. 9 depicts the performance of the proposed PQRDME-$p2$ for several values of $B$; the number of quantization bits for each part. For a low $B$, the BER increases, where when $B = 5$ the degradation due to the quantization error is tolerable.
2. **Uniform phase quantization:** In future-generation communication systems, such as LTE/LTE-A [10], the channel power is quantized and fed back to the BS via a control signal to enable adaptive modulation and coding (AMC), scheduling, etc. Therefore, it is of interest to consider that the channel amplitude is known at the transmitter and only quantize the phase of each channel coefficient. Note that the amplitude of the channel coefficient follows a Rayleigh distribution which is a positive distribution with much more concentration than the Gaussian one. Therefore, a lesser number of bits is required to quantize the amplitude as compared to quantizing the real and imaginary parts which follow a double-sided distribution with higher variance. Fig. 10 shows the BER performance of the proposed PQRDME-$p2$ for several values of $B$; the number of quantization bits for each phase value. $B = 6$ is suitable for quantizing the phase of the channel coefficients since the degradation due to the quantization error is tolerable.

### 5.6 A remark on the optimality of the proposed PQRDM encoder

The proposed encoder is quasi-optimum, which means that the linear MMSE encoder might have a lower $\gamma$ if the VP stage deviates from the optimum solution. Therefore, the efficiency of the vector perturbation can be measured by the amount of reduction it can achieve in the required transmit power, taking the linear pre-

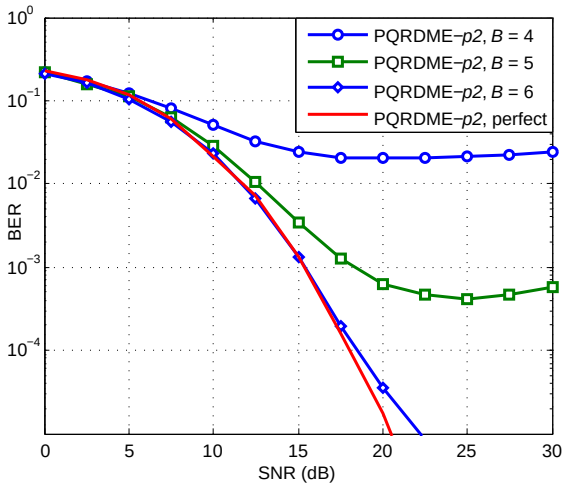


**Fig. 10** BER performance of the proposed PQRDME-$p2$ for $M_G = T = 3$ and $K = 8$ with imperfect CSI at the transmitter. The phase of each channel coefficient is uniformly quantized and represented by $B$ bits.

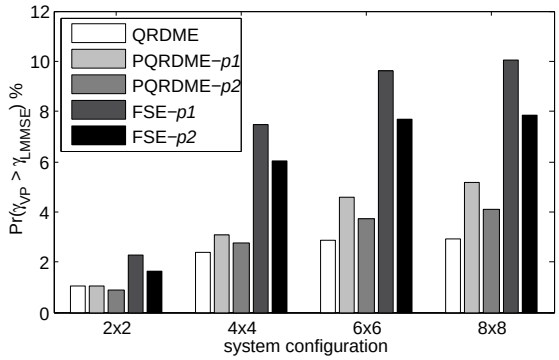


**Fig. 11** Percentage that the MU-MIMO system with vector perturbation requires more power than the conventional linear precoding.

coder as a reference. Fig. 11 depicts the percentage of achieving *an increase* in the transmit power by several VP techniques. That is, the probability (given in percent) that the system employing VP requires more transmit power than the simple linear MMSE encoder. Intuitively, this percentage increases proportionally with the number of transmit antennas. This is because for higher system dimensions, the probability that the VP technique deviates from the optimum solution increases. QRDME achieves the best performance followed by the proposed algorithms, where at low system dimensions the proposed PQRDME-$p2$ performs close to QRDME.

## 6 Conclusions

In this paper, we proposed a parallel QRDM encoder (PQRDME) for MU-MIMO systems. Unlike the conventional QRDME scheme, which has a limited capability for parallel implementations, the proposed PQRDME has a parallel tree-search structure, leading to higher efficiency for hardware implementation via pipelining. Besides, the proposed PQRDME-$p2$ achieves a tremendous reduction in the computational complexity due to its improved VP stage. In this paper, we considered realistic scenarios, where users are equipped with more than a single receive antenna in a multi-stream transmission. Furthermore, we consider the channel coefficients to be quantized and fed back to the BS rather than being perfectly known at the transmitter. To demonstrate the efficiency of our proposed encoder in such scenario, two quantization schemes have been addressed. In all scenarios, simulation and analytical results show that the proposed PQRDME performs close to the optimum encoder and attains the optimum diversity order, while achieving higher encoding throughput and much lower computational complexity. Moreover, the proposed PQRDME outperforms the FSE and THP for all the simulated scenarios.

## References

1. An, H.-S., Mohaisen, H., & and Chang, K.-H. (2009). Lattice reduction aided precoding for multiuser MIMO using Seysen's algorithm. In proceedings of PIMRC, 2479-2483.
2. Aruna, T. & Suganthi, M. (2011). Variable power adaptive MIMO OFDM system under imperfect CSI for mobile ad hoc networks. Telecommunication Systems. doi:10.1007/s11235-010-9387-y.
3. Chae, C-B., Shim, S., & Heath, Jr., R. (2008). Block diagonalization vector perturbation for multiuser MIMO systems. IEEE Transactions on Wireless Communications, 7(11), 4051-4057.
4. Choi, L-U., & Murch, R. (2004). A transmit preprocessing technique for multiuser MIMO systems using a decomposition approach. IEEE Transaction on Wireless Communications, 3(1), 20-24.
5. Costa M. (1983). Writing on dirty paper. IEEE Transactions on Information Theory, 29, 439-441.
6. Harashima, H. & Miyakawa, H. (1972). Matched-transmission technique for channels with intersymbol interference. IEEE Transactions on Communications, 20(4), 774-780.
7. Hochwald, B., Peel, C., & Swindlehurst, L. (2005). A vector-perturbation technique for near-capacity multiantenna multiuser communication - Part II: Perturbation. IEEE Transactions on Communications,53(3), 537-544.
8. Holakouei, R., Silva, A., & Gameiro, A. (2011). Multiuser precoding techniques for a distributed broadband wireless system. Telecommunication Systems. doi:10.1007/s11235-011-9496-2.
9. Jalden, J., Seethaler, D., Matz, G., (2008). Worst- and average-case complexity of LLL lattice reduction in MIMO wireless systems. In proceedings of ICASSP, 2685-2688.
10. Khan, F. (2009). LTE for 4G broadband: Air interface technologies and performance. Cambridge: Cambridge University Press.
11. Lenstra, A. K., Lenstra, H. W., & Lovasz L. (1982). Factoring polynomials with rational coefficients. Math. Annalen, 261, 515-534.

12. Liu, J., & Krzymien, J. (2007). Improved Tomlinson-Harashima precoding for the downlink for multi-user MIMO systems. Canadian Journal of Electrical and Computer Engineering, 32(3), 133-144.
13. Lloyd, S. P. (1982). Least squares quantization in PCM. IEEE Transactions on Information Theory, 28(2), 129-137.
14. Max, J. (1960). Quantizing for minimum distortion. IEEE Transactions on Information Theory, 6(l), 7-12.
15. Mohaisen, M., & Chang, K.-H. (2011). Fixed-complexity sphere encoder for multi-user MIMO systems. Journal of Communications and Networks (JCN), 13(1), 63-69.
16. Mohaisen M. *et al.* (2011). Parallel QRD-M encoder for decentralized multi-user MIMO systems. in Proceedings of IEEE ICC, 1-5.
17. Peel, C., Hochwald, B., & Swindlehurst, L. (2005). A vector-perturbation technique for near-capacity multi-antenna multiuser communication - Part I: Channel inversion and regularization. IEEE Transactions on Communications, 53(1), 195-202.
18. Pollock, T., Abhayapala T., & Kennedy, R. (2003). Introducing Space into MIMO Capacity Calculations. Telecommunication Systems, 24(2-4), 415-436.
19. Seysen, M. (1993). Simultaneous reduction of a lattice basis and its reciprocal basis, Combinatorica, 13, 363-376.
20. Stankovic, V., Haardt, M., & Fuchs, M. (2004). Combination of block diagonalization and THP transmit filtering for downlink beamforming in multi-user MIMO systems. In Proceedings of European Conf. on Wireless Technology, 145-148.
21. Tomlinson, M. (1971). New automatic equalizer employing modulo arithmetic. Electronics Letters, 7, 138-139.
22. Weingarten, H., Steinberg, Y., & Shamai, S. (2006). The capacity region of the Gaussian multiple-input multiple-output broadcast channel. IEEE Transactions on Information Theory, 52(9), 3936-3964.
23. Zhang, J. Z., & Kim, K.J. (2005). Near-capacity MIMO multiuser precodingwith QRD-M algorithm. In Proceedings of IEEE ACSSC, 1498-1502.