This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2011 proceedings

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

Manar Mohaisen[†], Abedelaziz Mohaisen[‡], Yanhua Li[‡], and Pengkui Luo[‡]
[†] Korea University of Technology and Education, Cheonan 330-708, Korea
[‡] University of Minnesota, Minneapolis, Minnesota 55455, USA
E-mail: manar.subhi@kut.ac.kr, mohaisen@cs.umn.edu, yanhua2008@gmail.com, pluo@cs.umn.edu

***Abstract*—In this paper, we propose a parallel QRDM encoder (PQRDME) for multi-user MIMO (MU-MIMO) systems. The proposed algorithm transforms the full tree-search problem of the conventional QRDME algorithm into parallel partial trees that are processed in parallel, leading to a tremendous increase in the encoding throughput. The proposed algorithm outperforms the fixed-complexity sphere encoder (FSE) and performs close to the optimum performance for several scenarios. For instance, in a 4×4 MUM-MIMO system, the proposed PQRDME lags the optimum performance by 0.3 dB while outperforming the FSE by 2.3dB at a target BER of 10^{-4}. In this case, the proposed algorithm also doubles the encoding throughput of the conventional QRDME algorithm.**

I. Introduction

To achieve the maximum sum capacity at the downlink of multi-user multiple-input multiple-output (MU-MIMO) systems, several approaches relying on the information-theoretic principle of dirty-paper coding (DPC) were proposed. DPC was initially proposed by Costa, where he showed that the capacity of an interference channel with known interference is exactly the same as the interference-free channel [1]. In communication systems, the transmitted signal to one user can be seen as an interference by other users. Because this *interference* is known to the BS and the channel can be fed back by the users to the BS, inter-user interference (IUI) can be canceled, or highly reduced, by means of MU-MIMO precoding.

Linear zero-forcing (LZF) and minimum-mean square error (LMMSE) are the simplest precoding techniques [2]. When the channel matrix is ill-conditioned, channel inversion precoding inverses the small singular values so that a high transmission power is required, leading to degradation in system performance. Regularized inversion precoding alleviates the noise enhancement problem by regularizing the channel matrix such that the the conditionality of the precoding matrix is improved. Despite of the improvement due to the regularization, LMMSE precoder still has a mediocre performance.

Tomlinson-Harashima precoding (THP) limits the transmit power by introducing the non-linear modulo operation [3], [4]. As a consequence, out of constellation points at the output of the precoder are rounded off to a pre-defined range. A linearized version of the THP that consists of vector perturbation stage and IUI cancellation stage was presented in [5]. The vector perturbation stage perturb the data vector such that the transmit power is reduced. Then, the transmitted signal can be recovered at the receivers by the same modulo operation. The IUI cancellation stage can be either done successively or using any of the aforementioned linear precoders.

Transmit power can be further reduced by perturbing the transmitted vector as in [6], where the optimum perturbation vector is found using the sphere encoder (SE). Although SE has a small average computational complexity, its worst-case complexity is high. Besides, SE is sequential in the tree-search phase, which limits the potential for efficient hardware implementation. To overcome the random computational complexity of the SE, the QR-decomposition with M-algorithm encoder (QRDME) was proposed in [7]. The main idea of the QRDM-E is to retain a fixed number of candidates at each encoding level. Although the conventional QRDME achieves the same performance of the SE, its tree search strategy limits the possibilities for the efficient hardware implementation using pipelining.

In [8], a fixed-complexity sphere encoder (FSE) for single-stream transmission has been proposed to achieve high encoding throughput by pipelining the tree-search stage. An extension of the FSE for multi-stream transmission has been introduced in [9], where a mobile station can be equipped with one or more receive antennas. In this extension, the streams of each user are precoded independently from other users' streams. This is done by applying block diagonalization of the channel matrix prior to the vector perturbation stage, where the IUI is perfectly canceled out. Although FSE achieves both reduction in the computational complexity and increase in the encoding throughput, it lags the performance of the optimum encoder.

In this paper, we propose a parallel QRDM encoder (PQRDME) based on the set grouping of the candidates for the perturbing vector. As a result of the set grouping, the search tree of the conventional QRDME is divided into *partial tree searches* that can be processed in parallel to increase the encoding throughput. Therefore, a tradeoff between the encoding throughput and the performance of the proposed algorithm is achieved.

This work is supported by research subsidy for newly-appointed professor at Korea University of Technology and Education (KUT) for the period 2010-2011.

978-1-61284-233-2/11/$26.00 ©2011 IEEE

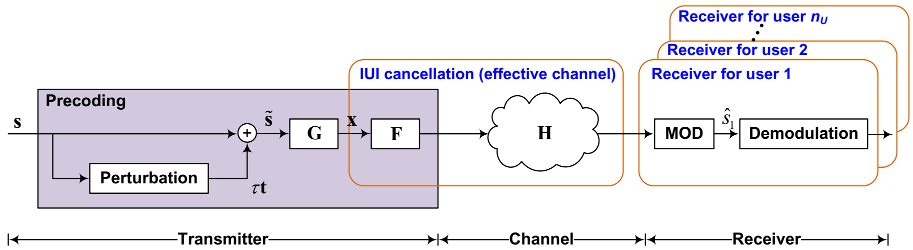

Fig. 1. Structure of the end-to-end broadcast MIMO system with vector perturbation and signal precoding.

The rest of this paper is organized as follows. In Section II and III, we introduce the system model and review several conventional precoding techniques. In Section IV we introduce the proposed PQRDME algorithm. In Section V we show simulation results and finally draw conclusions in Section VI.

II. System Model

In this paper, we consider a downlink MU-MIMO system (i.e., BS to users) in which a single BS equipped with n_T transmit antennas and communicates simultaneously with n_U single-antenna users. Without any loss of generality, we let $n = n_T = n_U$. Also, we consider a flat-fading and slowly time-varying channel, whose state information is perfectly known at the transmitter, unless otherwise stated. The MU-MIMO system can be then modeled as following:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{y} \in \mathbb{C}^{n_U}$ is the vector whose element y_k is the received signal at the k-th user, and $\mathbf{x} \in \mathbb{C}^{n_T}$ is the precoded transmitted vector. $\mathbf{n} \in \mathbb{C}^{n_U}$ is the additive white Gaussian noise whose mean and variance are zero and σ_n^2, respectively. Finally, $\mathbf{H}$ is the channel matrix whose element $h_{k,i}$ is the transfer function between the i-th transmit antenna and the single antenna of the k-th user. The elements of $\mathbf{H}$ are independent and follow an unbiased complex Gaussian distributions. Note that the channel matrix $\mathbf{H}$ can be rewritten as $\mathbf{H} = [\mathbf{h}_1^T \ \ \mathbf{h}_2^T \ \ \cdots \ \ \mathbf{h}_{n_U}^T]^T$, where the row-vector $\mathbf{h}_k$ contains the channel coefficients coupling the single antenna of the k-th user to the n_T transmit antennas. The system is then converted to the K-dimensional real Euclidean space, where $K = 2n$ and all scalars, vectors, and matrices are considered to be real-valued.

III. Review of Precoding Techniques and Problem Statement

A. Linear Precoding Techniques

Linear precoding techniques are the simplest in terms of computational complexity. In the case of linear zero-forcing precoding (LZF) the effect of the channel is canceled by precoding the transmitted data vector using the pseudo-inverse of the channel matrix.

$$\mathbf{x}_{\text{zf}} = \frac{1}{\sqrt{\gamma}}\mathbf{H}^{\dagger}\mathbf{s}, \tag{2}$$

where $\mathbf{s}$ is the vector of transmitted signals and the scaling factor γ is present to fix the *expected* total transmit power to (P_T); that is, $\gamma = \frac{1}{\text{P}_T}\text{Tr}\left\{\left(\mathbf{H}\mathbf{H}^H\right)^{-1}\right\}$, where $\text{Tr}(\cdot)$ refers to the trace operation. If the channel matrix is ill-conditioned, γ becomes large and consequently the post-processing signal to noise ratio (SNR) is decreased. To partially overcome this drawback, linear minimum mean-square error (LMMSE) precoding regularizes the channel matrix. The precoded signal using LMMSE is given by:

$$\mathbf{x}_{\text{mmse}} = \mathbf{H}^H\left(\mathbf{H}\mathbf{H}^H + \alpha\mathbf{I}_K\right)^{-1}\mathbf{s}, \tag{3}$$

where $\alpha = K\sigma_n^2/P_T$ is the regularization factor. Although the LMMSE precoder reduces the required transmit power, i.e., reduces γ, its performance is still mediocre and further improvement can be achieved.

B. THP and Vector Perturbation

The main idea of the THP is to limit the precoded signal to a pre-defined range such that the transmit power is reduced. This is done using a non-linear modulo (MOD) operation, which is defined as follows.

$$M(a) = a - \left\lfloor \frac{a}{\tau} + \frac{1}{2} \right\rfloor \tau, \tag{4}$$

where τ is an integer that depends on the used modulation scheme, and $\boldsymbol{t}$ is an N-dimensional integer vector. In [6], $\tau = 2\left(|c_{max}| + \Delta/2\right)$, where $|c_{max}|$ is the absolute value of the symbol with the largest magnitude, and Δ is the spacing between any two adjacent symbols. THP algorithm can be linearized, where the linear version consists of two stages; namely, a vector perturbation stage and a linear precoding stage [9]. The vector perturbation stage is employed to reduce the required transmission power. The perturbed vector $\tilde{\boldsymbol{s}}_i$ is then derived from the THP technique as follows:

$$\tilde{\boldsymbol{s}}_i = \boldsymbol{s}_i + \tau\boldsymbol{t}. \tag{5}$$

Since in the ideal case the precoding equalizes for the channel effect, the i-th received element of $\mathbf{s}$ is given by:

$$y_i = \tilde{s}_i + n_i, \tag{6}$$

where n_i is the additive-white Gaussian noise vector. At the receiver, the original data vector s_i is recovered, without knowledge of the vector t, using the non-linear modulo operation as follows.

$$\hat{s}_i = \text{mod}(y_\text{i}), \tag{7}$$

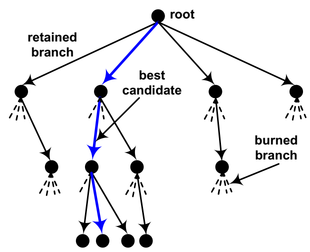

Fig. 2. An example of the conventional QRDME for $T = M = 4$, and $K = 3$.

where $\mathrm{mod}(\cdot)$ is the modulo operation that reduces the range of the received signal to the interval $[-U, U)$, where U depends on the used modulation scheme [10]. Specifically, $U = \sqrt{|\Omega|}$, where $|\Omega|$ is the cardinality of the modulation set Ω. For instance, U = 2 and 4 for QPSK and 16-QAM modulation schemes, respectively.

THP is shown to have superior performance to those of the linear precoders. However, further improvement in the performance can be achieved if the perturbing vector ($\mathbf{t}$) is optimally obtained. In the following, we introduce several powerful vector perturbation techniques.

C. Review of vector Perturbation techniques

The end-to-end MU-MIMO communication system employing vector perturbation and linear precoding is depicted in Figure 1. $\mathbf{F}$ in Figure 1 can be seen as a power allocation matrix. In this paper, we consider that users' data is drawn from the same modulation set and we set $\mathbf{F}$ to the identity matrix, and the $\mathbf{G}$ matrix is the linear precoding matrix.

The vector perturbation can be represented as an integer-lattice search, where at the transmitter, $\mathbf{t}$ is chosen such that γ is minimized; that is,

$$\begin{aligned}\mathbf{t} &= \underset{\mathbf{t}\in\mathbb{Z}^K}{\arg\min} \left\{(\mathbf{s}+\tau\mathbf{t})^T\mathbf{G}^T\mathbf{G}(\mathbf{s}+\tau\mathbf{t})\right\}, \\ &= \underset{\mathbf{t}\in\mathbb{Z}^K}{\arg\min} \left\|\mathbf{G}(\mathbf{s}+\tau\mathbf{t})\right\|^2. \end{aligned} \tag{8}$$

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 (8) based on the zero-forcing criterion is simplified to:

$$\begin{aligned}\mathbf{t} &= \underset{\mathbf{t}\in\mathbb{Z}^K}{\arg\min} \left\|\mathbf{L}(\mathbf{s}+\tau\mathbf{t})\right\|^2, \\ &= \underset{\mathbf{t}\in\mathbb{Z}^K}{\arg\min} \sum_{i=1}^{K} \left\| L_{i,i}(s_i+\tau t_k) + \sum_{j=1}^{i-1} L_{i,j}(s_j+\tau\hat{t}_j)\right\|^2, \end{aligned} \tag{9}$$

where the lower triangular matrix $\mathbf{L}$ equals $\left(\mathbf{R}^{-1}\right)^T$. When the MMSE criterion is used, the extended matrix $\tilde{\mathbf{H}} = [\mathbf{H}^T \ \sqrt{\alpha}\mathbf{I}]^T$ is factorized into the $\mathbf{Q}$ and $\mathbf{R}$ matrices, where $\mathbf{L}$ also equals $\left(\mathbf{R}^{-1}\right)^T$. The elements of $\mathbf{t}$ are drawn from a predefined integer set A of length T and centered at the origin. T is chosen in such a way a tradeoff between complexity and performance is achieved.

The search problem in (9) can be solved as in [6], [11], [12] using the SE. However, SE suffers from a high worst-case complexity and low capabilities for parallel processing due to its sequential nature. In [7], a QRDM encoder (QRDME) was proposed to solve the integer lattice problem in (9) and to overcome the high worst-case complexity of the SE. The main idea of the QRDME is to retain a fixed number of candidates for t at each tree-search level. Figure 2 depicts an example of the conventional QRDME for $T = 4$ and $K = 3$. At each encoding level, all the retained branches are extended to all possible candidates, their accumulative metrics are calculated and sorted. The M branches with the least accumulative metrics are retained for the next encoding levels. Although QRDME performs close to the optimum encoder, its tree-search strategy lacks the capability of fast processing via pipelining.

In [8], [9], we proposed an FSE that leads to full parallelization of the tree-search, where consequently the encoding throughput is increased. However, FSE lacks the optimum bit-error-rate (BER) performance.

IV. Proposed Parallel QRDME

The goal of the proposed parallel QRDME (PQRDME) is to achieve a tradeoff between encoding throughput, i.e., speed, and system performance. In reference to the conventional QRDME, the set of candidates for the elements of $\mathbf{t}$, i.e., A, is divided into non-overlapping subsets of equal number of elements. That is,

$$A = \bigcup_{i=1}^{G} D_i, \tag{10}$$

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 partial tree-search processes. In the sequel, these independent tree-searches are referred to as *partial vector perturbation processes* (PVP), 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, to achieve better performance as in the case of the FSE [8], [9]. 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 A.

Figure 3 depicts an example of the proposed PQRDME for $p = 1$, $M = 4$, $T = 4$, $G = 2$, and $K = 3$. Note that M herein is the total number of retained candidates by all the PVPs. At the first encoding level, the set of candidates 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 (9). 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 $t \in A$, the accumulative metric of all the resulting branches are computed using (9) and the best two (i.e., M/G)

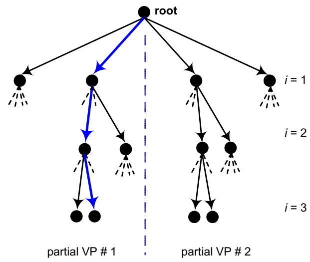

Fig. 3. An example of the proposed PQRDME for $p = 1$, $T = 4$, $G = 2$, and $K = 3$.

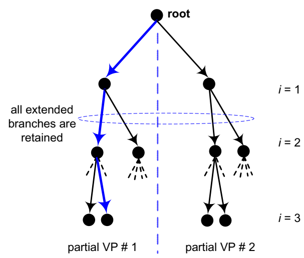

Fig. 4. An example of the proposed PQRDME for $p = T = G = 2$, and $K = 3$.

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 PVP are compared and the best among them is announced as the result of the proposed PQRDME-$p1$ algorithm.

Figure 4 depicts an example of the proposed PQRDME for $p = 2$, $M = 4$, $T = 2$, $G = 2$, and $K = 3$. 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 (one 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 $t \in A$. Also, all the resulting candidates are retained since $p = 2$. 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 level. The process is repeated as in the case of PQRDME-$p1$, where at the last encoding level the best candidate that requires the least transmit power is encoded and transmitted.

Based on the above description, the proposed PQRDME achieves a tradeoff between the parallelization capabilities, by controlling the number PVPs (G), and the performance of the MU-MIMO system.

V. Simulation Results and Discussion

In this section, we evaluate the performance of the proposed algorithm in a 4×4 MU-MIMO system. The channel state information is assumed to be perfectly know at the transmitter and transmitted symbols are QPSK modulated.

Figure 5 shows the BER performance of the proposed PQRDME for $p = 1$ and $M = T = 8$ and those of several

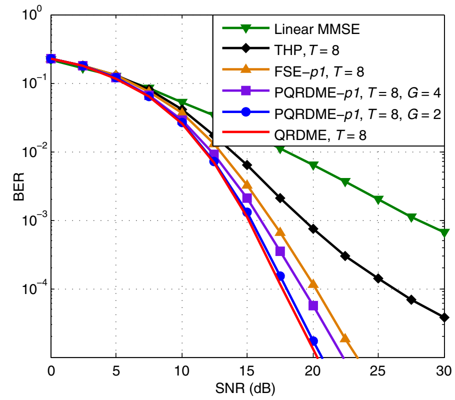

Fig. 5. BER performance of the proposed PQRDME for $p = 1$ and $M = T = 8$, using QPSK modulation.

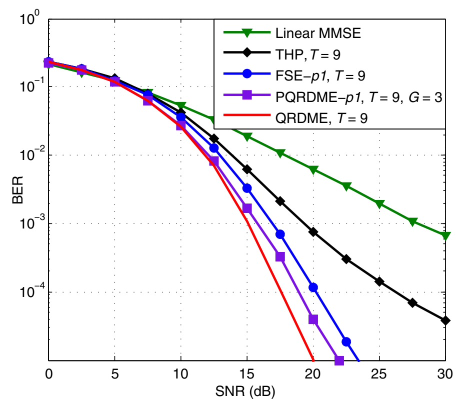

Fig. 6. BER performance of the proposed PQRDME for $p = 1$ and $M = T = 9$, using QPSK modulation.

conventional precoding techniques. 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 5, THP and LMMSE have lower performance and diversity order compared to the proposed algorithm.

Figure 6 shows the BER performance of the proposed PQRDME for $p = 1$ and $M = T = 9$ and $G = 3$. The proposed PQRDME-$p1$ performs in the middle between the conventional QRDME and FSE for a gain in the encoding throughput by a factor of three compared to the QRDME. Figure 7 depicts the performance of the proposed PQRDME-$p2$. Note that for $p = 2$, the performance of the FSE and the the proposed PQRDME algorithms are improved compared to the case $p = 1$. Yet, the proposed PQRDME-$p2$ outperforms the FSE-$p2$ by about 1dB at a target BER of 10^{-4}.

Table I depicts the mean and standard deviation of the accumulative metrics corresponding to the retained vector candidates at the last encoding level. It is evident that PQRDM-$p2$ ($G = 3$)and EFSE-$p2$ lead to both low mean and standard de-

TABLE I

MEAN AND STANDARD DEVIATION OF THE METRICS CORRESPONDING TO THE RETAINED CANDIDATES AT THE LAST ENCODING LEVEL AVERAGED OVER 100,000 INDEPENDENT CHANNEL REALIZATIONS AT SNR = 15DB.

System	QRDME ($T = 9$)		PQRDME-$p1$ ($T = 9$)		PQRDME-$p2$ ($T = 3$)		FSE-$p1$ ($T = 9$)		FSE-$p2$ ($T = 3$)	
	mean	std	mean	std	mean	std	mean	std	mean	std
4×4	9.28	2.73	21.40	11.31	11.77	5.05	40.84	32.40	20.77	13.84
8×8	7.36	1.56	13.05	5.20	8.68	2.41	22.17	14.27	13.09	4.87

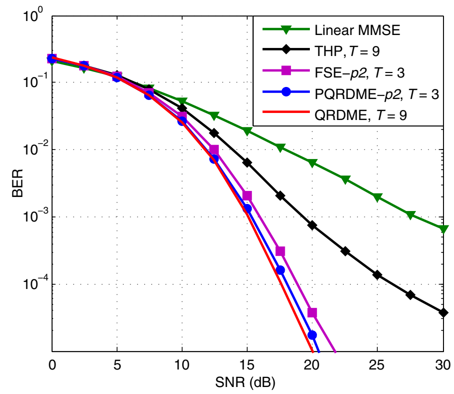

Fig. 7. BER performance of the proposed PQRDME for $p = 2$, $M = 9$, and $T = G = 3$, using QPSK modulation..

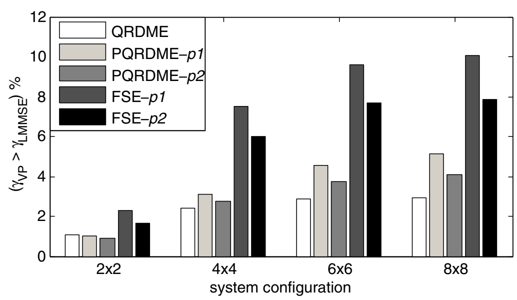

Fig. 8. Percentage that the MU-MIMO system with vector perturbation requires more power than the conventional linear precoding.

viation of those metrics as compared to the PQRDME-$p1$ and FSE-$p1$, respectively. Moreover, the proposed PQRDME has a lower mean which means that it leads to further optimization in the perturbation stage. Note that when the variance in the accumulative metrics of the retained candidates is large, this means that there are retained candidates with high metrics and can be canceled out at early encoding stages without affecting the performance. The proposed algorithm has a lower standard deviation compared to the FSE algorithm, which outlines the reason behind the out superiority of the proposed algorithm in terms of system performance.

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 precoder as a reference. Figure 8 depicts the percentage of achieving *an increase* in the transmit power by several vector perturbation techniques. That is, the probability (given in %) that the system employing vector perturbation 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 vector perturbation 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.

VI. 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. Simulation results show that the proposed PQRDME performs close to the optimum encoder and attains the optimum diversity order, while achieving higher encoding throughput. Moreover, the proposed PQRDME outperforms the FSE for all the simulated scenarios.

REFERENCES

[1] M. Costa, “Writing on dirty paper,” *IEEE Transactions on Information Theory,* vol. IT-29, pp. 439-441, May 1983.

[2] C. Peel, B. Hochwald, and L. Swindlehurst, “A vector-perturbation technique for near-capacity multiantenna multiuser communication - Part I: Channel inversion and regularization,” *IEEE Transactions on Communications,* vol. 53, no. 1, pp. 195-202, Jan. 2005.

[3] M. Tomlinson, “New automatic equalizer employing modulo arithmetic,”*Electronics Letters,* vol. 7, pp. 138-139, Mar. 1971.

[4] H. Harashima and H. Miyakawa, “Matched-transmission technique for channels with intersymbol interference,” *IEEE Transactions on Communications,* vol. 20, no. 4, pp. 774-780, Aug. 1972.

[5] J. Liu, and W. Krzymien, “Improved Tomlinson-Harashima precoding for the downlink for multi-user MIMO systems,” *Canadian Journal of Electrical and Computer Engineering,* vol. 32, no. 3, pp. 133-144, Summer 2007.

[6] B. Hochwald, C. Peel, and L. Swindlehurst, “A vector-perturbation technique for near-capacity multiantenna multiuser communication - Part II: Perturbation,” *IEEE Transactions on Communications,* vol. 53, no. 3, pp. 537-544, Mar. 2005.

[7] J.Z. Zhang and K.J. Kim, “Near-capacity MIMO multiuser precoding with QRD-M algorithm,” in *Proc. of IEEE ACSSC,* Nov. 2005, pp. 1498-1502.

[8] M. Mohaisen and K.H. Chang, “Fixed-complexity sphere encoder for multi-user MIMo systems,” *Journal of Communication and Networks,* under second round review.

[9] M. Mohaisen, H. Bing, K.H. Chang, S.H. Ji, and J.S. Joung, “Fixed-complexity vector perturbation with block diagonalization for MU-MIMO systems,” in *Proceedings IEEE Malaysia International Conference on Communications,* Dec. 2009, pp. 238 - 243.

[10] R. Fischer, *Precoding and Signal Shaping for Digital Transmission.* NY: Wiley, 2002.

[11] S. Shim, C-B. Chae, and R. Heath, Jr., “A lattice-based MIMO broadcast precoder with block diagonalization for multi-stream transmission,” in *Proc. IEEE Global Telecommunications Conf.,* Nov. 2006, pp. 1-5.

[12] C-B. Chae, S. Shim, and R. Heath, Jr., “Block diagonalization vector perturbation for multiuser MIMO systems,” *IEEE Transactions on Wireless Communications,* vol. 7, no. 11, pp. 4051-4057, Nov. 2008.