Physical-Layer MIMO Relaying

Anatoly Khina
Dept. of EE-Systems, TAU
Tel Aviv, Israel
Email: anatolyk@eng.tau.ac.il

Yuval Kochman
EECS Dept., MIT
Cambridge, MA 02139, USA
Email: yuvalko@mit.edu

Uri Erez*
Dept. of EE-Systems, TAU
Tel Aviv, Israel
Email: uri@eng.tau.ac.il

Abstract—The physical-layer network coding (PNC) approach provides improved performance in many scenarios over "traditional" relaying techniques or network coding. This work addresses the generalization of PNC to wireless scenarios where network nodes have multiple antennas. We use a recent matrix decomposition, which allows, by linear pre- and post-processing, to simultaneously transform both channel matrices to triangular forms, where the diagonal entries, corresponding to both channels, are equal. This decomposition, in conjunction with precoding, allows to convert any two-input multiple-access channel (MAC) into parallel MACs, over which single-antenna PNC may be used. The technique is demonstrated using the two-way relay channel with multiple antennas. For this case it is shown that, in the high signal-to-noise regime, the scheme approaches the cut-set bound, thus establishing the asymptotic network capacity.

Index Terms—network modulation, physical-layer network coding, network capacity, multiple access channel, structured codes, nested lattices, MIMO channels, two-way relay channel

I. INTRODUCTION

The capacity region of networks is a long-standing problem in Information Theory. Wireless networks are of special practical interest. Traditionally, such networks were treated at two different levels: A physical-layer local code which translates the wireless channels into "bit-pipes", and a network code over the bit network. This separation is in general sub-optimal, as was demonstrated in recent years by *physical-layer network coding* (PNC) approaches.

In the most basic variant of PNC (for single-antenna nodes), relay nodes simply forward their inputs using power adjustment only ("amplify-and-forward"). This analog-PNC approach helps in opening network bottlenecks, by allowing a node to assist in transmission even if it cannot decode the message, and indeed it is optimal in some cases, e.g., some limits of the parallel relay network [1], [2]. However, analog-PNC suffers from noise accumulation: Without decoding, the relays also forward noise. Structured-PNC is an alternative approach which solves both the bottleneck and noise accumulation effects at the same time. It uses structured/linear codes, building on the property that an integer linear combination of codewords is a codeword as well. A relay node may be able to decode the combination codeword, even if it cannot decode the individual messages. It was first presented by Wilson et

al. [3] for the simple two-way relay channel. Nam et al. [4] have shown that such a scheme achieves a rate, within half a bit of the cut-set bound, for any channel coefficients. A more general network is treated by Nazer and Gastpar's "compute-and-forward" strategy [5].

Wireless communication widely uses multiple-input-multiple-output (MIMO) techniques to obtain degrees of freedom. It is therefore natural to ask, what can the combination of PNC and MIMO achieve over a network where terminals have multiple antennas. While some works considered MIMO versions of analog-PNC (see, e.g., [6]–[8]), little was done in the context of structured-PNC. In [9], Zhan et al. demonstrate that in the two-way relay channel, structured-PNC can gain from multiple relay antennas – but the work is restricted to single-antenna terminal nodes, not allowing to fully enjoy MIMO gains.

In this work we introduce a way to combine structured-PNC and MIMO transmission, where a relay node receives the linear combination of two transmit nodes via MIMO channels. The approach we propose builds on a recently introduced matrix decomposition [10], that allows to simultaneously transform two channel matrices into *triangular* form with equal diagonals, by applying unitary operations at both sides. The transformation is applied to the channel matrices in conjunction with dirty-paper precoding at the transmitters, which yields parallel channels.

An essential property of the matrix decomposition, which allows to independently apply structured-PNC to each sub-channel without loss, is that the ratio between the gains of both transmitters is fixed for all the parallel channels. We note that the two-way MIMO relay channel problem was recently also considered in [11], where it was proposed to use the generalized singular value decomposition (GSVD) [12] to transform the matrices into triangular form. However the GSVD fails to achieve a fixed ratio. Thus, in general it yields sub-optimal results.

In this work we concentrate on the case of two users. Nonetheless, extensions to more than two users can be constructed, as explained in [13].

The rest of the paper is organized as follows. In Section II we present the problem and the main result. In Section III we provide a constructive proof using the aforementioned decomposition. We conclude in Section IV by presenting performance comparisons and discussing some extensions.

II. The Two-Way MIMO Relay Channel

The two-way relay channel consists of two terminals and a relay. We define the channel model as follows. Transmission takes place in two phases, each one, w.l.o.g., lasting N channel uses. At each time instance in the first stage, terminal i ($i = 1, 2$) transmits a signal $X_{i,n}$ and the relay receives Y_n according to some memoryless multiple-access channel (MAC) $W_{\text{MAC}}(Y|X_1, X_2)$. At each time instance in the second phase, the relay transmits a signal X_n and the terminals receive $Y_{i,n}$ according to some memoryless broadcast (BC) channel $W_{\text{BC}}(Y_1, Y_2|X)$. Before transmission begins, terminal i ($i = 1, 2$) possesses an independent message of rate R_i, unknown to the other nodes; at the end of the two transmission phases, each terminal should be able to decode, with arbitrarily low error probability, the message of the other terminal. The closure of all achievable pairs (R_1, R_2) is the capacity region of the network.

We consider a Gaussian MIMO setting, where terminal i ($i = 1, 2$) has $N_{t;i}$ transmit antennas and the relay has N_r receive antennas, during the MAC phase, which is depicted in Figure 1a. Denoting by bold letters the transmit and receive vectors, we have the MAC channel:

$$\mathbf{Y} = H_1\mathbf{X}_1 + H_2\mathbf{X}_2 + \mathbf{Z},$$

where H_i are $N_r \times N_{t;i}$ matrices, $\mathbf{Z}$ is circularly-symmetric white Gaussian noise with unit variance and the inputs are subject to the same total power constraint:

$$E\left[\mathbf{X}_i^\dagger \mathbf{X}_i\right] \le P, \qquad i = 1, 2.$$

We assume that the number of transmit antennas at each node $N_{t;i}$ is at least as large as the number of receive antennas N_r, and that the matrices H_1 and H_2 are full-rank, i.e., have rank N_r.[1] We further assume that the messages have the same rate $R_1 = R_2 = R$ and that the products of the non-zero singular values, of each of the channel matrices, are equal to 1, or equivalently:

$$\det\left(H_i H_i^\dagger\right) = 1, \quad i = 1, 2.\tag{1}$$

The exact nature of the BC channel is not material in the context of this work. We characterize it using its common message capacity C_{common}.

First consider the single-antenna case, $N_r = N_{t;1} = N_{t;2} = 1$. It is shown in [3] that one can achieve a rate of: [2]

$$R_{\text{PNC}} = \min\left\{\log\left(\frac{1}{2} + P\right), C_{\text{common}}\right\}.\tag{2}$$

By the min-cut theorem, one cannot achieve a rate greater than the point-to-point capacity of the MAC links or the common-message capacity of the BC channel:
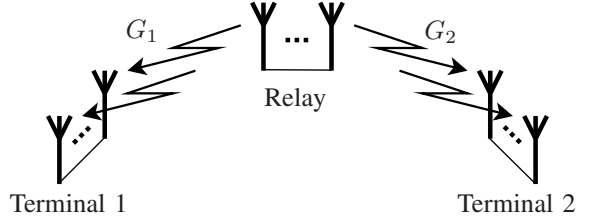
$$R_{\text{CS}} = \min\left\{\log\left(1 + P\right), C_{\text{common}}\right\}.\tag{3}$$

[1]For the cases in which the matrices are not full-rank or have more receive antennas see Section IV.
[2]Wilson et al. [3] considered a real-valued network, thus an additional $1/2$ pre-log factor was present in their original expression.
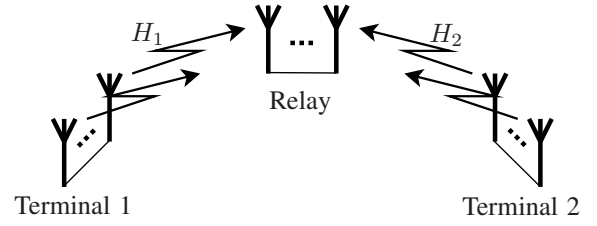
(a) MIMO MAC phase.

(b) BC phase for the special case where it is a MIMO BC channel.

Fig. 1: MIMO two-way relay channel. The second (BC) phase may be different in general.

It follows that PNC is optimal in the high-SNR limit. It is interesting to compare these to common relaying approaches. Using decode-and-forward (D&F) relaying, the relay must decode both messages with sum-rate $2R$. Instead of forwarding both message it can use a network-coding approach and XOR them, then each terminal can XOR out its own message to obtain the desired one. The resulting rate is given by:

$$R_{\text{DF}} = \min\left\{\frac{1}{2}\log\left(1 + 2P\right), C_{\text{common}}\right\}.\tag{4}$$

In the high-SNR limit, when the MAC stage is the bottleneck, the D&F rate is about half the cut-set bound. In this scenario, an approach that we denote "amplify-and-forward" (A&F) can be used, where the relay forwards the noisy sum of the terminal transmissions.[3] This approach achieves:

$$R_{\text{AF}} = \log(1 + \alpha P),\tag{5}$$

where $\alpha = \alpha(P, C_{\text{common}})$ is a noise amplification factor which is strictly smaller than one for channels of finite capacity. It follows that the A&F rate has a finite difference from the cut-set bound in the high-SNR limit.

Moving back to the MIMO case, the cut-set bound is given by:

$$R_{\text{CS}} = \min\left\{C_1, C_2, C_{\text{common}}\right\},\tag{6}$$

where

$$C_i \triangleq \max_{\text{trace}(C_\mathbf{x}) \le P} \log\det(I + H_i C_\mathbf{x} H_i^\dagger), \ i = 1, 2.$$

In the high-SNR limit, under the normalization (1), we have:

$$\lim_{P \to \infty} C_i - N_r \log\frac{P}{N_r} = 0.\tag{7}$$

[3]Note that it does not necessarily involve analog amplification, if the BC links are not AWGN channels. Instead in that case we compress the sum of the codewords, and treat the unbiased quantization noise as channel noise.

The next theorem, which is proved in Section III, shows that the PNC rate (2) generalizes to the MIMO case as follows.

Theorem 1: The capacity of the Gaussian MIMO two-way relay network with full-rank channel matrices, with $N_t^{(1)}, N_t^{(2)} \geq N_r$ satisfying (1), is lower-bounded by

$$R_{\text{PNC}} = \min \left\{ N_r \log \frac{P}{N_r}, C_{\text{common}} \right\} . \qquad (8)$$

On account of (7), the rate R_{PNC} approaches the cut-set bound for high SNR as in the single-antenna case. High-SNR conditions hold when

$$\frac{(\lambda_{i;j})^2 P}{N_r} \gg 1, \qquad \begin{matrix} i = 1, 2 \\ j = 1, ..., N_r \end{matrix}, \qquad (9)$$

where $\{\lambda_{i;j}\}_{j=1}^{N_r}$ are the singular values of H_i.

Theorem 1 assumes the use of white input. At any finite SNR, it can be improved by optimizing over the input covariance matrices. however, this rate is already better than the rate achievable by the D&F approach, except for very low SNR.

III. STRUCTURED MIMO-PNC SCHEME

In this section we provide a constructive proof for the main result, stated in Theorem 1. The key ingredient is using the matrix decomposition of [10, Sec. IV] to obtain equivalent parallel single-antenna networks, over which we can use the structured-PNC strategy of [3]. The decomposition is given in the following theorem of [10], transposed in order to accommodate for the MAC setting.

Theorem 2: Let A_1 and A_2 be complex-valued full-rank matrices, of dimensions $m \times n_1$ and $m \times n_2$, respectively, such that $n_1, n_2 \geq m$ (meaning A_i are of rank m). If the products of their singular values are equal, then A_1 and A_2 can be jointly decomposed as

$$A_1 = U T_1 V_1^\dagger$$
$$A_2 = U T_2 V_2^\dagger , \qquad (10)$$

where V_1, V_2 and U are unitary matrices of dimensions $n_1 \times n_1$, $n_2 \times n_2$ and $m \times m$, respectively; and T_1 and T_2 are generalized lower-triangular matrices (matrices with zero entries above the main diagonal, i.e., $T_{1;ij} = 0$ and $T_{2;ij} = 0$ for $i < j$, where $T_{i;k,j}$ denotes the (k, j) entry of T_i) with positive *equal* diagonal elements.

We apply the decomposition of Theorem 2 to the channel matrices H_1 and H_2. and denote the diagonal entries of T_1 (which are equal to those of T_2 by assumption (1)) by $t_1, \ldots, t_{N_r}$.

We use dirty-paper precoding at each of the terminals to cancel the off-diagonal elements of T_1 and T_2. This results in N_r parallel channels, with gains given by the (equal for both terminals) diagonal elements. Denote these elements of T_1 (which are also equal to those of T_2) by $t_1, \ldots, t_{N_r}$. If we use each such channel as part of an independent single-antenna two-way relay network with input power P/N_r, we can achieve a total rate of:

$$R_1 = R_2 = \sum_{k=1}^{N_r} r_k \overset{(a)}{\geq} \sum_{k=1}^{N_r} \log \left(\frac{1}{2} + \frac{t_k^2 P}{N_r} \right)$$

$$\geq N_r \log \frac{P}{N_r} + \sum_{k=1}^{N_r} \log t_k^2 = N_r \log \frac{P}{N_r} , \qquad (11)$$

where r_k is the rate conveyed over sub-channel k, and (a) holds true due to (2).

We now describe in detail a scheme which allows to achieve this rate.

Codebooks generation: Apply the decomposition of Theorem 2 (where we set $n_i = N_t^{(i)}$ and $m = N_r$) to the channel matrices H_1 and H_2 to obtain matrices U, V_1, V_2, T_1 and T_2. At each terminal, divide the message into N_r sub-messages of rate $r_k = R_{\text{PNC}}/N_r + \log t_k^2$.[4] Now for each k ($k = 1, \ldots, N_r$), both terminals use the *same* nested-lattice code [14] with nesting ratio r_k. Denote the coarse lattice (common for all sub-messages) by Λ, and the fine lattices by $\{\Lambda_k\}$. The second moment of Λ is P/N_r. Λ is assumed to be Rogers-good, while $\{\Lambda_k\}$ are good for AWGN coding.

MAC encoding: Terminal i chooses a lattice point $\ell_{i,k}$ for the k-th sub-message,[5] then sequentially computes:

$$\tilde{X}_{i,k} = \left[\ell_{i,k} - \frac{1}{t_k} \sum_{j=1}^{k-1} T_{i;k,j} \tilde{X}_{i,j} \right] \mod \Lambda . \qquad (12)$$

Finally, the transmit signal is

$$\mathbf{X}_i = V_i \tilde{\mathbf{X}}_i .$$

MAC decoding: The relay computes $\tilde{\mathbf{Y}} = U^\dagger \mathbf{Y}$ and then uses lattice decoding to find $\hat{\ell}_k$, the closest point of Λ_k (modulo Λ), and recovers messages $\{m_k\}$ according to the codebook mapping.

BC stage: The relay conveys the messages $\{m_k\}$ to both terminals, using any capacity-approaching common-message BC scheme.

Final decoding: The terminals recover $\{\ell_k\}$, then compute for each sub-message:

$$\hat{\ell}_{\bar{i},k} = [\hat{\ell}_k - \ell_{i,k}] \mod \Lambda , \qquad (13)$$

where

$$\bar{i} \triangleq \begin{cases} 2 & i = 1 \\ 1 & i = 2 \end{cases} .$$

This estimated codeword is mapped back to a message.

Proof of Theorem 1: Since Λ is Rogers good with second moment P/N_r and V_i are orthogonal, the transmission satisfies the power constraint. Now,

$$\begin{aligned} \tilde{\mathbf{Y}} &= U^\dagger (H_1 V_1 \tilde{\mathbf{X}}_1 + H_2 V_2 \mathbf{X}_2 + \mathbf{Z}) \\ &= T_1 \tilde{\mathbf{X}}_1 + T_2 \tilde{\mathbf{X}}_2 + \tilde{\mathbf{Z}} , \end{aligned}$$

---

[4]If this quantity is negative for some k, set $r_k = 0$; since this only improves the achievable rate, we can assume for the sake of analysis that it is always positive.

[5]$\ell_{i,k}$ and the signals that follow are n-dimensional vectors; we still reserve boldface to denote spatial (antenna) dimension $1, \ldots, N_r$.
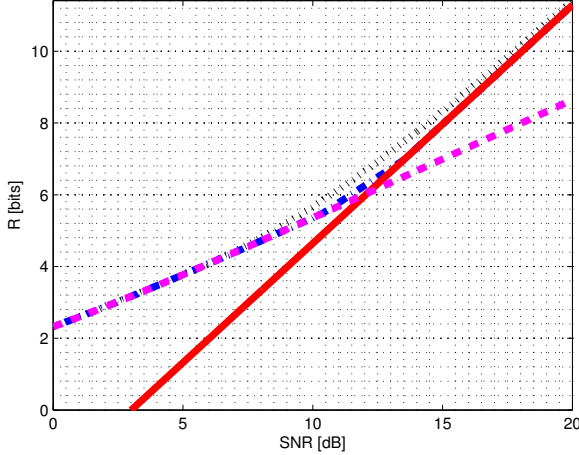
Fig. 2: Rates as a function of the MAC SNR P in Example 1 where the MAC section is the bottleneck. Dotted black line – R_{CS} (cut-set bound); red continuous line – R_{PNC}; dashed magenta line – R_{DF}; dashed-dotted blue line – time-sharing between PNC and D&F.

where $\tilde{\mathbf{Z}} = U$. By the precoder operation (12) we have:

$$\left[\frac{\tilde{Y}_k}{t_k} \right] \bmod \Lambda = \left[\ell_k + \frac{\tilde{Z}_k}{t_k} \right] \bmod \Lambda \,,$$

where $\ell_k = [\ell_{1,k} + \ell_{2,k}] \bmod \Lambda$, and $\tilde{\mathbf{Z}}$ has the same distribution as $\mathbf{Z}$. Since ℓ_k is a valid codeword of Λ_k/Λ, this codeword effectively passed through a mod-Λ channel with SNR $P_k = t_k^2 P/N_r$. Moreover, since Λ_k is good for AWGN coding at rate $\log P_k$, we have that $\hat{\ell}_k = \ell_k$ with arbitrarily low error probability (as the dimension $n \to \infty$). Thus, the relay may decode the messages $\{m_k\}$, and the total rate of the latter is

$$R - i = N_r \log \frac{P}{N_r} \,, \qquad i = 1, 2 \,.$$

Of course, if C_{common} is lower than that, we can achieve the lower rate. Now, these messages can be conveyed to the terminals with arbitrarily low error probability using rate C_{common}. If there was no error in any of the stages, then (13) yields $\hat{T}_{\bar{i},k} = T_{\bar{i},k}$; by our choice of codebooks, this corresponds to a transmission rate R_{PNC}. ∎

Example 1: Figure 2 depicts the achievable rate obtained when the channel matrices are:

$$H_1 = \begin{pmatrix} 1/2 & 0 \\ 0 & 2 \end{pmatrix}, \quad H_2 = \begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix},$$

and the MAC phase is assumed to be the bottleneck, compared to the cut-set bound and to the D&F rate. As expected, the proposed scheme approach the upper bound as the MAC SNR increases.

At this point, one can understand why R_{PNC} of Theorem 1 is higher than that of Yang et al. [11]. The GSVD-based

scheme suggested there induces equivalent triangular channels with *different* diagonals, thus suffering substantial loss when applying the scalar PNC scheme. In fact, as the GSVD provides the "most spread diagonal ratios" (see [10]), a scheme that does not apply any precoding matrices U_i should have better performance than the GSVD one.

IV. EXTENSIONS

Time-sharing. As is true also for the SISO case, D&F is better than PNC at low rates; for example, it yields a positive rate for any positive P_k, while PNC fails for $P_k < 1$ due to the loss of the "1" (see expressions (2) and (11)). As noted for the single-antenna case in [3], one may use time-sharing in the intermediate region, to improve over both. Figure 2 compares the different bounds, as a function of the MAC SNR which is assumed to be the bottleneck in this case.

Non full-rank matrices. In the case when the channel matrices H_1 and H_2 are not full-rank or when there are more receive antennas than transmit ones, other strategies needed to accompany the approach proposed in this work. Consider, for instance the following two channel matrices:

$$H_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad H_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

In this case the relay sees two parallel channels, from each of the two terminals. Note that in this case a good strategy would be to sum the two channel outputs at the relay, to decode the sum-codeword and send it to the two terminals during the BC phase, or alternately to decode each of the messages, sum them (modulo lattice) and send the result over the BC channel (it is more efficient to send a common message than two private messages). For more general non full rank matrices, such strategies can be combined with the strategy suggested in Section III as well applying the joint triangularization of Theorem 2 to augmented versions of the channel matrices as is done in [10].

Use of non zero-forcing elements. We have used the lattice codes in a "naïve" manner, achieving $R_k = \log P_k$. That is, we lost the "entire" 1. Indeed, using dithering and MMSE estimation as in [3] one can achieve the higher rate $R_{PNC,k}$ (2). This will improve performance at low signal to noise ratios (though not extremely low, as $R_{PNC,k}$ is only positive for $P_k > 1/2$). Furthermore, instead of the "zero-forcing" decomposition proposed in the current work, an MMSE version could be used, which decomposes augmented versions of the channel matrices along with using non-white covariance matrix, as is done in the common message Gaussian MIMO BC case, see [10].

Non-symmetric setting. The approach can be extended beyond the symmetric setting, to find achievable rate pairs (R_1, R_2). Suppose that the ratio between the singular-values-product of H_1 and H_2 is ρ, then we can work at each of the parallel networks with ratio ρ^{1/N_r}. If this quantity is not integer, then there is a tradeoff between rounding and noise

amplification [5]. However the loss can be bounded as in [4]. Furthermore, we may shape the ratios between the diagonals, such that all the ratios are integers, except for maybe one, or alternately, if the decomposed matrices satisfy a certain majorization condition, it is possible to take all the ratios to be equal to 1 except for one, see [10] for details.

Application: Colored two-way relay channel. In the special case where the channel matrices are diagonal, the problem is equivalent to a single-antenna colored channel with piecewise-constant spectrum; by increasing the matrix dimension, arbitrary spectra can be accommodated for. Interestingly, the precoding operation at the terminals (12) is in the *frequency domain* in this case, unlike traditional time-domain Tomlinson-Harashima precoding [15], [16].

Coding for the broadcast section. The only assumption we used regarding the BC section, is that it has a common-message capacity high enough such that it does not limit performance. However, it seems likely that the links to the terminal will be wireless MIMO ones as well. In that case, depicted also in Figure 1b, the complexity may be considerably reduced by using the scheme which is based upon the decomposition of Theorem 2, for that section as well; see [10].

REFERENCES

[1] B. Schein and R. G. Gallager. The Gaussian parallel relay channel. In *Proc. Int. Symp. Inform. Theory (ISIT), Sorrento, Italy*, page 22, June 2000.

[2] M. Gastpar and M. Vetterli. On the capacity of large Gaussian relay networks. *IEEE Trans. Info. Theory*, IT-51:765–779, March 2005.

[3] M. P. Wilson, K. Narayanan, H. Pfister, and A. Sprintson. Joint physical layer coding and network coding for bidirectional relaying. *IEEE Trans. Info. Theory*, 56:5641–5654, Nov. 2010.

[4] W. Nam, S.-Y. Chung, and Y. H. Lee. Capacity of the Gaussian two-way relay channel to within 1/2 bit. *IEEE Trans. Info. Theory*, 56:5488–5494, Nov. 2010.

[5] B. Nazer and M. Gastpar. Compute-and-forward: Harnessing interference through structured codes. *IEEE Trans. Info. Theory*, submitted. Available: arXiv:cs/0908.2119.

[6] R. Zhang, Y.-C. Liang, C.-C. Chai, and S. Cui. Optimal beamforming for two-way multi-antenna relay channel with analogue network coding, June 2009.

[7] H. Shi, T. Abe, and H. Yoshino. Relaying schemes using matrix triangularization for MIMO wireless networks. *IEEE Trans. Comm.*, 55:1683–1688, Sep. 2007.

[8] C. Li, L. Yang, and W.-P. Zhu. Two-way MIMO relay precoder design with channel state information. *IEEE Trans. Comm.*, 58:3358–3363, Dec. 2010.

[9] J. Zhan, B. Nazer, M. Gastpar, and U. Erez. MIMO compute-and-forward. In *Proc. of the ISIT-2009, Seoul, Korea*.

[10] A. Khina, Y. Kochman, and U. Erez. Joint unitary triangularization for MIMO networks. *IEEE Trans. Signal Processing*, Submitted, Dec. 2010. Availabel: arXiv:cs/1012.4715.

[11] H. J. Yang, C. Joohwan, and A. Paulraj. Asymptotic capacity of the separated MIMO two-way relay channel with linear precoding. In *48th Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL*, Sept. 29–Oct. 1, 2010.

[12] C. F. Van Loan. Generalizing the singular value decomposition. *SIAM J. Numer.*, Anal. 13:76–83, 1976.

[13] A. Khina, A. Hitron, and U. Erez. Modulation for MIMO networks with several users. In *Proc. Int. Symp. Inform. Theory (ISIT), St. Petersburg, Russia*, July 31–Aug. 5 2011.

[14] R. Zamir, S. Shamai (Shitz), and U. Erez. Nested linear/lattice codes for structured multiterminal binning. *IEEE Trans. Info. Theory*, 48:1250–1276, June 2002.

[15] M. Tomlinson. New automatic equalizer employing modulo arithmetic. *Electronic Lett.*, vol. 7:138–139, Mar. 1971.

[16] H. Harashima and H. Miyakawa. Matched-transmission technique for channels with intersymbol interference. *IEEE Trans. Comm.*, COM-20:774–780, Aug. 1972.