

# COSTA PRECODING IN ONE DIMENSION

Jinfeng Du, Erik G. Larsson and Mikael Skoglund

KTH/EE Communication Theory, Royal Institute of Technology  
Osqualdas väg 10, 100 44 Stockholm, Sweden. Email: erik.larsson@s3.kth.se.

## ABSTRACT

We design an optimum modulator for the Costa (dirty-paper) precoding problem under the constraint of a binary signaling alphabet, and assuming the interference symbols belong to a binary constellation. We evaluate the performance of our technique in terms of the mutual information between the channel input and output, and compare it to that of Tomlinson-Harashima precoding (THP) with optimized parameters. We show that our optimal modulator is always better than THP. In many relevant scenarios, the performance difference is significant.

## 1. INTRODUCTION

Costa showed in his 1983 paper [1] that the achievable rates of a communication channel remain unchanged if the receiver observes the transmitted signal in the presence of additive interference, *provided that the transmitter knows the interference non-causally*. More precisely, consider the setup in Figure 1, and suppose the interference  $z(t)$  and the noise  $n(t)$  are Gaussian. Then, if the transmitter has non-causal access to  $z(t)$ , the capacity of the channel from “TX” to “RX” is the same as it would be if  $z(t)$  were not present (under the same transmit power constraint).

The problem of designing a transmitter which achieves the channel capacity in the presence of  $z(t)$  is often called the “Costa (precoding) problem” or “dirty paper” coding problem (after the title of [1]). This problem is important because the known-interference scenario arises in a number of contexts, notably, when doing precoding for ISI channels and for the downlink multiuser MIMO channel [2, 3]. Consequently the problem has stimulated much research. Essentially, the strategy for achieving capacity is known (it is precisely the constructive proof in [1]; see also [4]): First quantize  $z(t)$  into a number of bins (this is essentially a source coding problem). Then, depending on what bin  $z(t)$  falls into, choose an appropriate code for the encoding of  $w(t)$ . The best Costa-precoding results known to us [5, 6] are based on this approach. For example, [5] uses a turbo-trellis code for the quantization of  $z(t)$ , and another turbo code for the encoding of  $w(t)$ .

References [5, 6] in fact, impressively, demonstrate the (near) achievability of Costa’s prediction. The downside of the approach therein, however, is complexity. In this light, it is natural to ask what one can do about the Costa problem when permitted to add no, or very little, extra complexity to the system compared to “classical” transmission. The goal of this paper is to shed some light on this question. More precisely, we consider the design of an optimal one-dimensional<sup>1</sup> scheme which maps a binary input bit  $i = \{0, 1\}$  and a binary interference symbol  $z = \pm\beta$ ,  $\beta \in \mathbb{R}$  (known to the

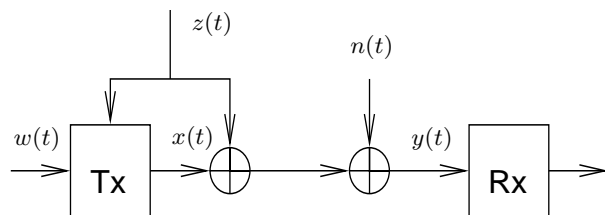


Fig. 1. System model.

transmitter but not to the receiver) onto an output symbol  $x \in \mathbb{R}$ . Thereby, strictly speaking, our focus is on modulation rather than on coding. The goal of our work is to obtain an understanding for what one can achieve in small (or a single) dimensions and at low complexity, rather than to achieve capacity. (Indeed achieving capacity is impossible with finite-dimensional precoding.)

We are not aware of any previous work that systematically treats the Costa problem in small (or a single) dimensions. We remark, however, that a special case of the one-dimensional precoding structure we propose here (and which we also take as a benchmark) is the Tomlinson-Harashima precoder (THP), originally proposed for ISI channels [7, 8]. THP takes  $x = (w - z) \bmod \Lambda$ , where  $w = w(i)$  is a function of  $i$  and  $\Lambda$  is a constant. Both  $w(i)$  and  $\Lambda$  can be optimized, see below. In this context also note [9] addressing a related problem, however without optimization over  $w(i)$  or  $\Lambda$ .

## 2. SYSTEM MODEL

Consider Figure 1. From now on, we consider a discrete, one-dimensional, channel so all quantities are real-valued and scalar. (We omit the time index  $t$  for simplicity of notation.) The modulator maps an information symbol index  $i \in \mathbb{Z}$  and an interference symbol  $z \in \mathbb{R}$  onto a modulated symbol  $x \in \mathbb{R}$ . Expressing the (nonlinear) modulator mapping as  $x = x(i, z)$  we can write

$$y = x(i, z) + z + n \quad (1)$$

where  $z$  is the interference symbol (known to transmitter), and  $n$  is noise. The receiver does not know  $z$ , however, we shall assume that it knows the probability distribution of  $z$ , say  $p_z(u)$ . (This assumption is weak if  $z$  is drawn from a stationary and ergodic process.)

We assume that the noise is Gaussian:  $n \sim N(0, \sigma^2)$  where  $\sigma^2$  is known. Also, we shall treat only here the special case when  $i$  and  $z$  are discrete, binary random variables (over  $\mathbb{Z}$  and  $\mathbb{R}$ , respectively) as follows:

$$P(i = 0) = P(i = 1) = 1/2 \quad (2)$$

$$P(z = -\beta) = P(z = \beta) = 1/2 \quad (3)$$

This work was supported in part by the Swedish Research Council (VR), VINNOVA, and Wireless@KTH.

<sup>1</sup>Extension to inphase/quadrature (narrowband) modulation, or to other

orthogonal multiplexing formats is immediate by treating each dimension independently. See also Section 6.

That is, the input alphabet is binary ( $i = 0, 1$ ) and the interference comes from a scaled BPSK constellation  $z = \pm\beta$ . Also, all combinations of  $i, z$  are equally likely. Further, we assume that the available transmit power is  $P$ , i.e., the modulator operates under the constraint  $E[x^2] \leq P$ .

### 3. STATE-OF-THE-ART (IN ONE DIMENSION)

We first present some baseline strategies for the problem in Figure 1.

**No interference.** If there is no interference ( $z = \beta = 0$ ) then taking  $x = \pm\sqrt{P}$  (say,  $x(i) = (2i - 1)\sqrt{P}$ ) is the best we can do, with a binary alphabet and subject to the power constraint. The optimal receiver (in the minimum error-probability sense) is the one that maximizes the a posteriori probability of  $i$  when  $y$  is received:

$$\hat{i}_{\text{MAP}} = \operatorname{argmax}_i P(i|y) = \operatorname{argmin}_i |y - (2i - 1)\sqrt{P}|$$

**No interference cancellation.** If the transmitter does not know the interference, but the receiver knows  $p_z(u)$  (an assumption we do make throughout the paper) then we may take, say,  $x(i) = (2i - 1)\alpha$  for some constant  $\alpha$ . Note that,  $\alpha = \sqrt{P}$  “works” (in the sense that the power constraint is satisfied). However, this choice of  $\alpha$  is not necessarily optimal. In our comparisons, we choose the value of  $\alpha$  (subject to  $\alpha \leq \sqrt{P}$ ) which maximizes performance. The optimal receiver is

$$\hat{i}_{\text{MAP}} = \operatorname{argmax}_i p_y(y|i) = \operatorname{argmax}_i p_{(z+n)}(y - (2i - 1)\alpha)$$

where  $p_{z+n}(u)$  is the convolution of  $p_z(u)$  and  $p_n(u)$ .

**Interference subtraction.** Arguably the transmitter could cancel  $z$  by taking  $x = (2i - 1)\alpha - z$ . However, since we must have  $E[x^2] = \alpha^2 + \beta^2 \leq P$ , doing so would work only if  $\beta^2 < P$ . Also, even under this rather strong condition, i.e., weak interference, it is not optimal. This technique therefore is not a meaningful baseline for comparison.

**Tomlinson-Harashima Precoding [7, 8].** This fits into our framework by setting  $w(i) = (2i - 1)\alpha$  for some constant  $\alpha$  and then taking

$$x = (w - z) \bmod \Lambda \quad (4)$$

so that,

$$y = ((w - z) \bmod \Lambda) + z + n = w + k\Lambda + n = w + e$$

where  $k$  is an integer which depends on  $i$  and where we defined  $e = k\Lambda + n$  ( $e$  also depends on  $i$ ).<sup>2</sup>

For us, the purpose of introducing THP is only to have a good baseline for comparison. (A more specific motivation is that THP has been proposed for the downlink MIMO problem [10, 11].) However, as a byproduct of our work we also obtained the optimal receiver for THP (an explicit derivation of which we were unable to pinpoint in the literature). The optimal receiver (see the next paragraph) differs from the heuristic (and suboptimal) detector

$$\hat{i}_{\text{subopt}} = \operatorname{argmin}_i |(y \bmod \Lambda) - w|$$

which is usually used in papers dealing with THP. The difference in performance between the two receivers, however, is usually not large except for “unlucky” choices of  $\alpha, \Lambda$ .

<sup>2</sup>Conditioning on  $i$  is equivalent to conditioning on  $w(i)$ , a fact we will use repeatedly.

To find the optimal receiver for THP, first note that  $k$  has the conditional distribution in (5), at the top of the next page. In (5),  $F_z(t) = P(z \leq t)$  is the cumulative distribution function of  $z$ .<sup>3</sup> Thus  $p_{y|i}(y) = \sum_{\kappa=-\infty}^{\infty} P(k = \kappa|w) p_n(y - w(i) - \kappa\Lambda)$ . The optimal receiver is

$$\hat{i}_{\text{MAP}} = \operatorname{argmax}_i \sum_{\kappa=-\infty}^{\infty} P(k = \kappa|w) \exp\left(-\frac{1}{2\sigma^2}(y - w(i) - \kappa\Lambda)^2\right) \quad (6)$$

In practice the sum in (6) can be truncated to a few terms.

The parameters  $\alpha$  and  $\Lambda$  in THP can be optimized, subject to the power constraint  $E[x^2] \leq P$ . We do not dwell into this optimization, as it is not the focus of the paper. In fact, this can be done as a special instance of our optimal modulator (enforcing an additional constraint in the optimization), which we present next.

### 4. DESIGN OF AN OPTIMUM MODULATOR FOR INTERFERENCE AVOIDANCE

As criterion for optimization of the modulator mapping  $x = x(i, z)$ , we will use the mutual information  $I(y; i)$  between  $i$  and  $y$ , under the constraints presented in Section 2. This quantity is relevant at least if Figure 1 is thought of as an inner “code” and additional coding is used outside. (This would be the case in most real systems, anyway.)

The mutual information  $I(y; i)$  can be written as in (7), see the top of the next page. In (7), we used that  $\int_{-\infty}^{\infty} p_y(y|i) dy = 1, \forall i$ . Also, in (7),

$$\begin{aligned} p_y(y|i) &= p_y(y|i, z = -\beta)P(z = -\beta) + p_y(y|i, z = \beta)P(z = \beta) \\ &= \frac{1}{2}(p_y(y|i, z = -\beta) + p_y(y|i, z = \beta)) \end{aligned} \quad (8)$$

where

$$p_y(y|i, z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-z-x(i,z))^2/(2\sigma^2)}$$

In practice,  $I(y; i)$  can easily be computed by Monte-Carlo integration. Naturally  $p_y(y|i)$  (and  $I(y; i)$ ) depend on the specific modulator mapping  $x(i, z)$  used. We shall select the mapping  $x = x(i, z)$  which maximizes  $I(y; i)$ .

Under the assumptions of Section 2, there are four combinations of  $i$  and  $z$ , so we can write

$$\begin{aligned} x(i = 0, z = -\beta) &\triangleq a_0 \\ x(i = 0, z = \beta) &\triangleq a_1 \\ x(i = 1, z = -\beta) &\triangleq a_2 \\ x(i = 1, z = \beta) &\triangleq a_3 \end{aligned} \quad (9)$$

By symmetry ( $z$  and  $n$  have symmetric densities), we must have  $x \in \{-a, -b, b, a\}$  for some positive constants  $a, b$ . The problem is then to find  $a, b$  and to map  $a_0, \dots, a_3$  onto the set  $\{-a, -b, b, a\}$ . With no constraint on the ordering of  $a$  and  $b$ , there are  $4! = 24$  possibilities, of which 12 are redundant (because  $a$  and  $b$  are not ordered). The set of possible mappings to be considered therefore is

$$\begin{aligned} a_0 = -a, a_1 = -b, a_2 = a, a_3 = b \\ a_0 = -b, a_1 = -a, a_2 = a, a_3 = b \\ a_0 = -a, a_1 = b, a_2 = -b, a_3 = a \\ a_0 = -b, a_1 = b, a_2 = -a, a_3 = a \\ a_0 = a, a_1 = -a, a_2 = -b, a_3 = b \end{aligned} \quad (10)$$

<sup>3</sup>Note that (5) does not require  $i$  and  $z$  to be binary. Therefore this equation is valid also for a more general scenario than that defined in Section 2.

$$\begin{aligned}
P(k = \kappa|i) &= P(k|w) = P((w - z) \bmod \Lambda - (w - z) = \kappa\Lambda |w) \\
&= P(w - z \in [-(\kappa + 1/2)\Lambda, -(\kappa - 1/2)\Lambda] |w) \\
&= P(w + (\kappa - 1/2)\Lambda \leq z \leq w + (\kappa + 1/2)\Lambda |w) \\
&= F_z(w + (\kappa + 1/2)\Lambda) - F_z(w + (\kappa - 1/2)\Lambda)
\end{aligned} \tag{5}$$

$$\begin{aligned}
I(y; i) &= H(i) - H(i|y) = \sum_{i=0}^1 \int_{-\infty}^{\infty} P(y, i) \log P(i|y) dy - \sum_{i=0}^1 P(i) \log P(i) \\
&= \sum_{i=0}^1 \left[ \int_{-\infty}^{\infty} p_y(y|i) P(i) \log \frac{p_y(y|i) P(i)}{p_y(y)} dy - P(i) \log P(i) \right] \\
&= \sum_{i=0}^1 P(i) \int_{-\infty}^{\infty} p_y(y|i) \log \frac{p_y(y|i)}{\sum_{j=0}^1 p_y(y|j) P(j)} dy
\end{aligned} \tag{7}$$

$$a_0 = a, a_1 = -b, a_2 = -a, a_3 = b$$

$$a_0 = a, a_1 = b, a_2 = -a, a_3 = -b$$

$$a_0 = a, a_1 = b, a_2 = -b, a_3 = -a$$

$$a_0 = b, a_1 = -a, a_2 = a, a_3 = -b$$

$$a_0 = b, a_1 = -b, a_2 = a, a_3 = -a$$

$$a_0 = -b, a_1 = b, a_2 = a, a_3 = -a$$

$$a_0 = -a, a_1 = b, a_2 = a, a_3 = -b$$

(Possibly this set can be reduced further.)

By symmetry,  $x \in \{-a, -b, b, a\}$  are equally likely so the power constraint translates into  $E[x^2] = (a^2 + b^2)/2 \leq P$ . We then search over a grid which contains all  $a, b$  that satisfy this constraint, and for each combination of  $a, b$  we examine the 12 combinations in (10). The optimization is computationally rather burdensome. However, it can be accomplished within a few hours on a standard desktop PC. Note that the optimization of THP (with respect to  $\alpha, \Lambda$ ) can be accomplished via the same procedure by restricting the search to those  $a_0, \dots, a_3$  which satisfy  $x = ((2i - 1)\alpha - z) \bmod \Lambda$  for some  $\alpha, \Lambda$ .

The optimal receiver has a simple form, simpler than the optimal receiver for THP indeed. To find its explicit form, note from (8) that

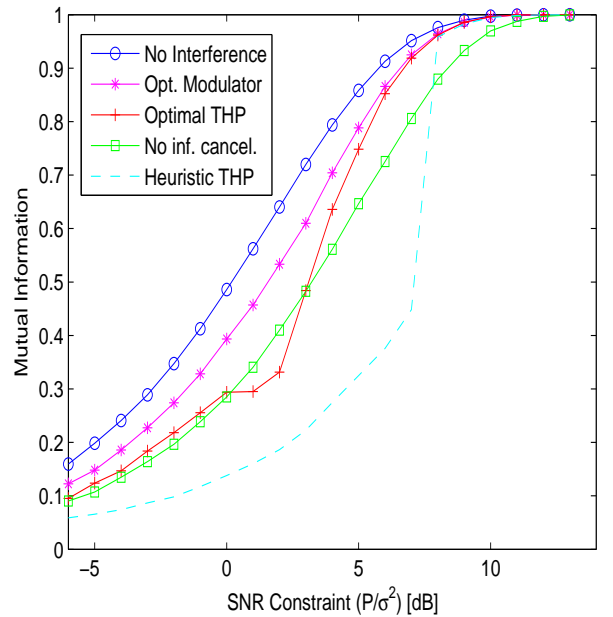
$$\hat{i}_{\text{MAP}} = \underset{i}{\operatorname{argmax}} \left( e^{-(y+\beta-x(i,-\beta))^2/(2\sigma^2)} + e^{-(y-\beta-x(i,\beta))^2/(2\sigma^2)} \right)$$

## 5. NUMERICAL RESULTS

Mutual information between the received signal  $y$  and the information bit  $i$  was used as performance measure. The input constellation is binary, so assuming an outer code with rate  $r$  the interesting region would be  $I(y; i) > r$ . Monte-Carlo simulation was used to obtain the results.

Figures 2–3 show  $I(y; i)$  for the five different transmitter structures/scenarios (i) no interference, (ii) interference but no cancellation, (iii) THP with the heuristic parameter choice  $\Lambda = 3\alpha$ , used in most papers we found, (iv) THP with optimized parameters  $\alpha, \Lambda$ , and (v) our proposed optimal modulator. Results are displayed as a function of the maximum allowed transmit power ( $P$ ), keeping the noise power and the interference power ( $\beta^2$ ) fixed. Figure 2 shows performance with  $\beta^2 = 4, \sigma^2 = 4$ . In this case, the interference-to-noise ratio, INR, is equal to 0 dB. The signal-to-noise ratio (SNR) equals the signal-to-interference (SIR).<sup>4</sup> Figure 3

<sup>4</sup>Strictly speaking, for a given ratio  $P/\sigma^2$ , the actual SNR may be less than  $P/\sigma^2$ , because the optimal modulator does not necessarily use all avail-



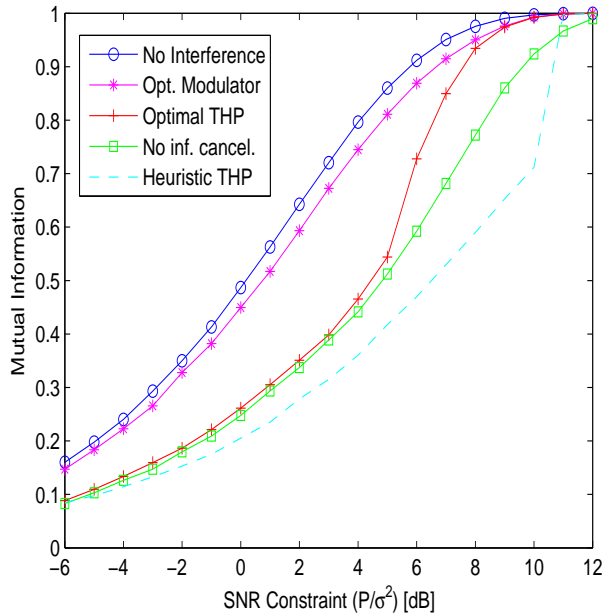
**Fig. 2.** Mutual information for  $\beta^2 = 4, \sigma^2 = 4$ , INR = 0 dB, SNR = SIR.

shows the performance at  $\beta^2 = 10, \sigma^2 = 5$  (here INR = 3 dB and SNR = SIR + 3 dB).

Our optimal modulator is always the best performing one (this is no surprise since all variants of THP are a special case of the mapping that we optimize). Interestingly, there are values of  $P$  for which one generally does better without THP than with it.

In Figure 4, we fix  $\beta^2 = 4$  and vary  $P, \sigma^2$ . We show performance at an SNR of 1, 3 and 6 dB, as a function of the SIR. We also show the result for SIR =  $\infty$  (dotted, horizontal lines), for some different SNR values to get a sense for how much performance loss (in terms of equivalent SNR loss) the binary interference gives rise to. The conclusion from this plot is that the equivalent SNR loss induced by the binary interference is at most 1.5 dB, irrespectively of the SIR (at least for the SNR values considered in the plot). The loss without interference precoding, however, is much larger. (The THP curves are omitted to keep the plot readable.)

able power. Yet we refer to  $P/\sigma^2$  as SNR because this facilitates a well-defined comparison with the no-interference case.



**Fig. 3.** Mutual information for  $\beta^2 = 10$ ,  $\sigma^2 = 5$ , INR = 3 dB, SNR = SIR + 3 dB.

## 6. CONCLUSIONS

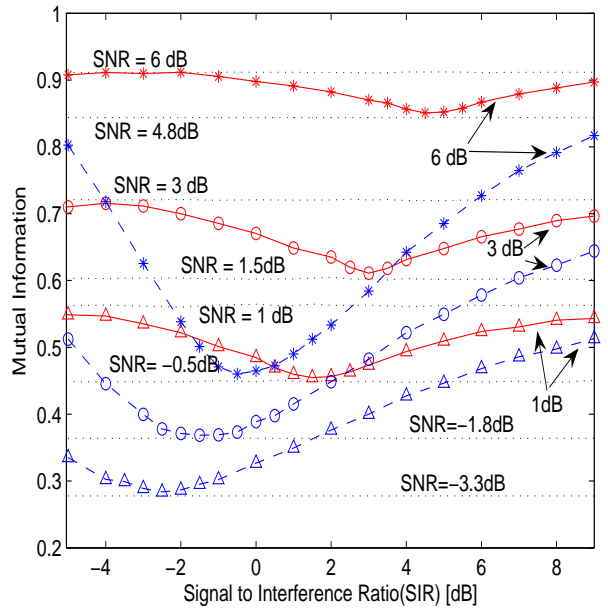
The goal of this paper was to study in some depth the simplest possible instance of the dirty-paper problem, namely, in one dimension and with binary signals and interference. We obtained the optimum precoder (rather, modulator) for this case and demonstrated that it typically outperforms Tomlinson-Harashima precoding, even when the parameters of the latter are optimally chosen. A more specific conclusion was that provided the optimal modulator is used, binary interference—of arbitrary power—can never hurt the performance more than what a 1.5 dB decrease in SNR would do (at least for the SNR values considered in Figure 4).

All conclusions we have drawn under the assumption of binary constellations do not necessarily translate (at least not quantitatively) to the case of larger constellations. However, our study does indicate that rather impressive interference suppression (rather, avoidance) performance can be achieved in a single dimension. This result serves as motivation to continue study low-complexity approaches to the Costa problem.

The work can be extended in several directions. First, the constraint of binary signal constellations may be relaxed. In this case, it is not clear how the resulting optimization problem can be solved: an exhaustive search over the mapping  $x(i, z)$  does not seem feasible. However, preliminary experiments not showed here have indicated that optimization of a subclass of the mapping  $x(i, z)$  (such as THP) is possible. Second, one may attempt to extend our strategy to a higher (but small) dimension; that is, let  $x, i, z, n, y$  be vectors and work with a multivariate mapping  $x(i, z)$ . An implementation of Costa precoding in practice will likely rely on operations in a space of small dimension, so the problems outlined here would be of much interest.

## 7. REFERENCES

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



**Fig. 4.** Mutual information at  $\beta^2 = 4$ , as a function of the SIR for different SNR values ("\*" = 6 dB SNR, "o" = 3 dB SNR, "Δ" = 1 dB SNR). The figure shows the performance of no interference (dotted, horizontal lines), the optimal modulator (solid lines), and no interference cancellation (dashed lines).

[2] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*, Cambridge University Press, 2005.

[3] G. Caire and S. Shamai, "On the Achievable Throughput of a Multiantenna Gaussian Broadcast Channel," *IEEE Transactions on Information Theory*, vol. 49, pp. 1691-1706, July 2003.

[4] R. Zamir, S. Shamai and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Transactions on Information Theory*, vol. 48, pp. 1250-1276, June 2002.

[5] A. Bennatan, D. Burshtein, G. Caire and S. Shamai, "Superposition Coding for Side-Information Channels." Submitted to *IEEE Transactions on Information Theory*, April 2004.

[6] U. Erez and S. ten Brink, "Approaching the dirty paper limit for cancelling known interference," in Proc. of *Allerton Conference on Communications, Control and Computing*, Oct. 2003.

[7] M. Tomlinson, "New automatic equalizer employing modulo arithmetic," *Electronics Letters*, pp. 138-139, March 1971.

[8] H. Harashima, H. Miyakawa, "Matched-transmission technique for channels with intersymbol interference," *IEEE Trans. on Communications*, vol. COM-20 No. 4, pp. 774-780, Aug. 1972.

[9] R. D. Wesel, J. M. Cioffi, "Achievable rates for Tomlinson-Harashima precoding," *IEEE Transactions on Information Theory*, vol. 44 No. 2, pp. 824-831, March 1998.

[10] M. Airy, A. Forenza, R. W. Heath Jr., and S. Shakkottai, "Practical Costa Precoding for the Multiple Antenna Broadcast Channel," in Proc. of *GLOBECOM*, Dec. 2004.

[11] C. Windpassinger, R. F. H. Fischer, T. Vencel, and J. B. Huber, "Precoding in multiantenna and multiuser communications," *IEEE Transactions on Wireless Communications*, vol. 3, pp. 1305-1316, July 2004.