

Trellis and Convolutional Precoding for Transmitter-Based Interference Pre-Subtraction

Wei Yu

Electrical and Computer Engineering Dept., University of Toronto
10 King's College Road, Toronto, Ontario M5S 3G4, Canada
email: weiyu@comm.utoronto.ca

David P. Varodayan and John M. Cioffi

Electrical Engineering Dept., Stanford University
Stanford, CA 94305 USA
email: varodayan@stanford.edu, cioffi@stanford.edu

Abstract

This paper studies the combination of practical trellis and convolution codes with Tomlinson-Harashima precoding (THP) for the pre-subtraction of multiuser interference that is known at the transmitter but not known at the receiver. It is well known that a straightforward application of Tomlinson-Harashima precoding suffers power, modulo and shaping losses. This paper proposes generalizations of Tomlinson-Harashima precoding that recover some of these losses. At a high SNR, the precoding loss is dominated by the shaping loss, which is about 1.53dB. To recover shaping loss, a trellis shaping technique is developed that takes into account the knowledge of the entire non-causal interfering sequence rather than the instantaneous interference. At rates of 2 and 3 bits per transmission, trellis shaping is shown to be able to recover almost all of 1.53dB shaping loss. At a low SNR, the precoding loss is dominated by power and modulo losses, which can be as large as 3-4dB. To recover these losses, a technique that incorporates partial interference pre-subtraction within convolutional decoding is developed. At rates of 0.5 and 0.25 bits per transmission, partial interference pre-subtraction is able to recover 1-1.5dB of the power loss. For intermediate SNR channels, a combination of the two schemes is shown to recover both power and shaping losses.

Keywords

Broadcast Channel, Dirty Paper Precoding, Tomlinson-Harashima Precoding, Convolutional Codes, Trellis Codes, Shaping Codes, Trellis Shaping, Channels with Side Information, Gel'fand-Pinsker Channel, Vectored Digital Subscriber Line.

Manuscript submitted to *IEEE Transactions on Communications* on March 25, 2004. This work was supported by a Stanford Graduate Fellowship, and by Natural Science and Engineering Research Council (NSERC), the Canada Research Chair program, Bell Canada University Laboratory and Communication and Information Technology Ontario (CITO). The conference version of this work has appeared in Globecom 2001. Correspondence should be addressed to Wei Yu.

I. INTRODUCTION

Consider a multiuser communications scenario in which a centralized transmitter wishes to transmit independent information to several remote users in geographically different locations at the same time. This downlink transmission environment is often modeled as a broadcast channel. Because of the mutual interference caused by multiple signals, the capacity region and the optimal transmission strategy for the broadcast channel have been long-standing open problems in multiuser information theory. Recently, however, a transmitter-based interference pre-subtraction scheme has gained much attention. The basic idea of pre-subtraction is as follows: Suppose that the transmitter wishes to send a signal x_1 to the first user and a signal x_2 to the second user. Since both signals are completely known prior to transmission, the transmitter may pre-subtract x_2 from x_1 so that the first receiver receives x_1 as if the interference does not exist. Moreover, it turns out that such a pre-subtraction scheme may be implemented without an increase in transmit power. Multiuser interference precoding scheme was first proposed in [1] and [2]. Later, several independent results [3] [4] [5] further suggested that multiuser interference pre-subtraction in fact achieves the sum capacity of the multi-antenna broadcast channel. Most recently, [6] finally fully established that the same is true for the entire capacity region.

The key theoretical insight for multiuser interference pre-subtraction is a result due to Costa [7] known as “writing on dirty paper”. Consider a Gaussian noise channel corrupted by an additive interference signal S that is known to the transmitter but not to the receiver:

$$Y = X + S + Z, \quad (1)$$

where X and Y are the transmitted and received signals respectively, S is the known interference, and Z is the unknown Gaussian noise. In a classical paper, Costa [7] showed that, surprisingly, the capacity of this channel under a transmit power constraint is the same as if S does not exist *provided that S is known non-causally at the transmitter*. For Costa’s result to hold, the transmitter has to know not only the present and past history of S , but also the future values of S . This setup in fact precisely models a multiuser downlink situation, where the interfering signal is the transmit signal to other users. Interference is known at the transmitter non-causally because the source information bits are typically stored in a buffer, and future values of the interference can be pre-constructed from the buffered bits. Costa’s paper is titled “Writing on Dirty Paper” because it models a transmitter which attempts to encode information on a piece of paper partially corrupted by dirt that is seen at the transmitter but not known at the receiver. Precoding methods for the dirty paper channel are sometimes referred to as “dirty-paper precoding”.

The main practical motivation for this paper is a vectored digital subscriber line application where crosstalk interference can be pre-subtracted at the transmitter. Digital Subscriber Line (DSL) is a high-speed data transmission standard for the conventional telephone twisted-pairs. Telephone wires

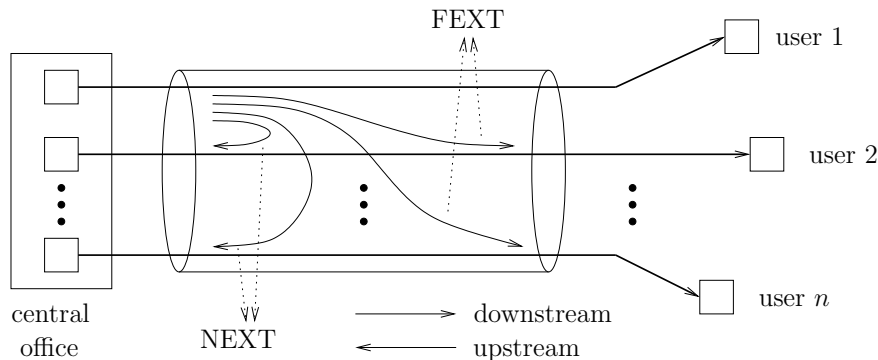


Fig. 1. Vectored Digital Subscriber Lines

are usually bundled together in a group of 50-100 lines at the central office. Because of their physical proximity, the crosstalk caused by the electromagnetic interference between the neighbouring lines is often the dominating noise source in the transmission environment. Fig. 1 illustrates the nature of the crosstalk. The near-end crosstalk (NEXT) refers to crosstalk emitted into the receivers located at the same side as the transmitter. The far-end crosstalk (FEXT) refers to crosstalk emitted into the receivers located at the opposite end. Both NEXT and FEXT can have detrimental effects on DSL transmission.

If each DSL line transmits and receives information independently, the transmission environment must be modelled as an interference channel, where interference is typically regarded as noise. However, if multiple transmitters at the central office coordinate, then array signal processing techniques may be used, and crosstalk cancellation becomes feasible. Because both NEXT and FEXT are completely known at the transmitter, the dirty-paper result implies that the capacity of the DSL channels can be as high as if the interference does not exist. This idea of interference pre-subtraction was first explored in [1] where a “Vectored DSL” system is proposed. Vectored DSL uses Tomlinson-Harashima precoding (THP) to pre-subtract crosstalk, and it has the potential to double the transmission rate for short-range DSL services.

This paper studies practical coding methods for interference pre-subtraction at the transmitter. We use Tomlinson-Harashima precoding as a starting point. Tomlinson-Harashima precoding [8] is a pre-equalization technique originally proposed for channels with intersymbol interference (ISI). It uses a modulo operation to pre-subtract interference with minimal power increase. However, compared to a channel with no interference, Tomlinson-Harashima precoding suffers from several sources of losses. As was pointed out in [9], precoding for the ISI channel incurs shaping loss in the high SNR regime and power and modulo losses at the low SNR regime. The objective of this paper is to characterize the nature of similar losses in Tomlinson-Harashima precoding for multiuser interference subtraction, and to propose practical ways to recover some of these losses for trellis and convolutional coded systems. As the main result of this paper shows, when combined with

Tomlinson-Harashima precoding, convolutional and trellis codes designed for the usual additive white Gaussian noise (AWGN) channel, $Y = X + Z$, can be adopted for the dirty-paper channel with only a small performance loss and with moderate amount of additional complexity.

The dirty-paper channel has been the subject of many studies in the information theory literature. In [10], Erez, Shamai and Zamir considered the dirty-paper channel with causal side information and first pointed out the connection between Tomlinson-Harashima precoding and the dirty-paper result. They showed that at a high SNR, shaping loss is the only significant precoding loss, and at a low SNR a partial interference cancellation scheme can be implemented to recover some of the power and modulo losses. The focus of the present paper is on practical coding schemes for the dirty-paper channel. One of the main contributions of this paper is to illustrate that trellis precoding, previously developed for achieving a shaping gain in the intersymbol interference channel [11] [12], can be extended to the dirty-paper channel to recover most of the shaping loss in the high SNR regime. Empirically, trellis shaping is demonstrated to be within 0.2dB of the performance of the equivalent code on the AWGN channel.

In addition, this paper also investigates precoding methods for the low-SNR channel. It is previously known that in the low-SNR regime, partial interference pre-subtraction outperforms total interference pre-subtraction. However, the implementation of partial interference pre-subtraction has not been studied before. This paper proposes a decoding metric that combines partial interference pre-subtraction with convolutional codes to recover some of the power and modulo losses in the low SNR regime. The performance of the partial interference pre-subtraction is shown to be within 2.5dB of the equivalent codes over the AWGN channel at rates of 0.25 and 0.5 bits per dimension. These simulation results are consistent with similar results in the low-density parity-check (LDPC) context [13]. Finally, we also present a scheme for the intermediate SNR regime that combines shaping and partial interference pre-subtraction and achieves within 0.5dB as it would over the AWGN channel, thereby confirming Costa's capacity result practically.

To completely recover all of the power, modulo and shaping losses, a nested coding approach must be used. In this regard, Zamir, Shamai and Erez [14] showed that an inflated-lattice strategy can in theory achieve Costa's capacity with a nested lattice coding scheme. For practical coding systems, Erez and ten Brink [15] showed that nested LDPC code and convolutional code are effective. In a related field, Pradhan and Ramchandran [16] proposed a syndrome-based code construction. In general, however, true capacity-achieving practical codes are still not available for the dirty-paper channel.

The rest of the paper is organized as follows: Section II explains the connection between Tomlinson-Harashima precoding and the dirty-paper channel and illustrates the source of power, modulo and shaping losses. Section III presents the trellis shaping method and illustrates its performance on the high SNR channels. Section IV presents the partial interference pre-subtraction scheme and

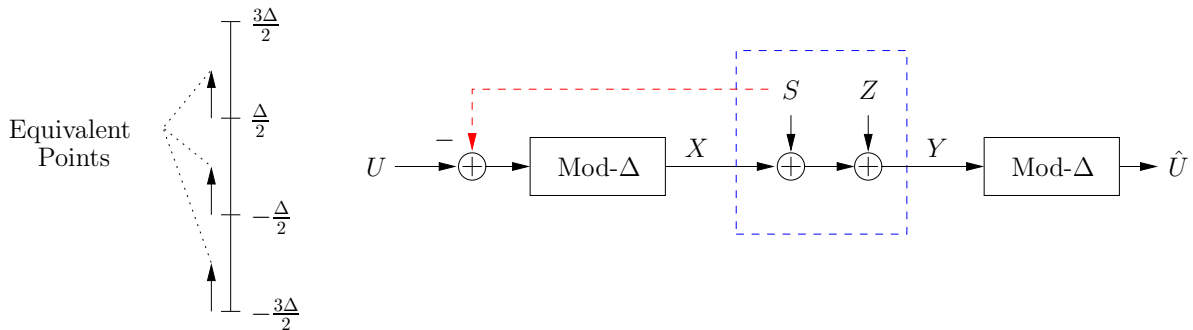


Fig. 2. Tomlinson-Harashima Precoding for the Dirty-Paper channel

illustrates its performance on the low and intermediate SNR regimes. Section V contains concluding remarks.

II. TOMLINSON-HARASHIMA PRECODING (THP)

A. Modulo Precoding

It is well-known that the capacity of an ordinary AWGN channel, $Y = X + Z$, with an input power constraint P_X and a noise power P_Z is

$$C = \frac{1}{2} \log_2 \left(1 + \frac{P_X}{P_Z} \right) \text{ bits/dim/transmission.} \quad (2)$$

Costa's result shows that the capacity does not change when the AWGN channel is corrupted by an interference signal with power P_S , if the interference is known non-causally at the transmitter. The realization of Costa's dirty-paper capacity involves a sophisticated binning strategy, which has to be implemented using nested codes [14]. However, at a high signal-to-noise ratio, dirty-paper precoding may be approximated by Tomlinson-Harashima precoding.

Tomlinson-Harashima precoding [8] [17] was originally designed for the intersymbol interference channel, and it can be readily extended for the pre-subtraction of multiuser interference at the transmitter. The main idea is as shown in Fig. 2. The intended message is denoted as U , and the interference signal is denoted as S . The goal is to convey U to the receiver in the presence of S . Without loss of generality, assume that P_S is large. (When P_S is small, pseudo-random dithering can always be added to both the transmitter and the receiver to make the effective P_S large.) Since S is known at the transmitter at every time instant, in order to convey an intended symbol U , a precoder may send $U' = U - S$ to compensate for the interference. However, when S is large, the power of U' may be exceedingly large. The main idea of Tomlinson-Harashima precoding is to constrain the intended symbol U to be within a finite interval $[-\Delta/2, \Delta/2)$. Instead of sending U' , the encoder sends U' modulo- Δ . In effect, all U' s differing by an integer multiple of Δ are regarded as the same symbol. (See Fig. 2.) In addition, the decoder recognizes the same equivalence relation

by implementing a modulo- Δ operation also. In the absence of noise, this equivalence relation allows U be completely recovered at the receiver.

Essentially, the modulo- Δ operation constrains the transmitted signal X to be within the finite interval $[\Delta/2, \Delta/2)$. When P_S is large, X is approximately uniformly distributed in the interval. In this case, the transmit power is reduced to approximately $\Delta^2/12$, which is independent of the interference power. This is all accomplished with the distinguishability U' maintained.

The encoding process in Tomlinson-Harashima precoding can also be thought of as the process of modifying S to one of the equivalent representatives of the symbol. This is a one-dimensional quantization process. The modulo- Δ geometry ensures that there is one representative within distance $\Delta/2$ from every possible S . As will be seen later, the Tomlinson-Harashima precoder can be interpreted as a one-dimensional implementation of “writing-on-dirty-paper”. One of the main contributions of this paper is a practical trellis precoding scheme that generalizes Tomlinson-Harashima precoding to high dimensions.

B. Precoding Losses

When compared to the performance of the standard PAM modulation on an AWGN channel, Tomlinson-Harashima precoding incurs some non-negligible performance losses. The precoding loss was first pointed out by Shamai and Laroia [9] for the intersymbol interference channel, and they characterized three sources of losses as shaping loss, power loss and modulo loss. This characterization is equally applicable to the multiuser interference pre-subtraction scenario being considered in this paper.

- *Shaping Loss*: First of all, because of the use of the modulo operation, the input signal U must be amplitude-limited. In particular, when S is large, X is uniformly distributed in $[-\Delta/2, \Delta/2)$. Such a uniform shape introduces the so-called shaping loss. Shaping loss is due to the fact that Shannon’s capacity formula for AWGN channel demands the capacity-achieving input distribution be a Gaussian distribution. Signaling using a uniform distribution introduces the finite shaping loss. The shaping loss due to M -PAM over AWGN channels is insignificant at low SNR, but tends to the ultimate shaping loss of 1.53dB for high SNR. In fact, as shown by Erez, Shamai and Zamir [10], shaping loss is intimately related to the fact that Tomlinson-Harashima precoding only takes the present value of the interference signal S into account, and does not take advantage of the knowledge of future interference. This observation is the motivation for the main result of this paper, which is a trellis shaping code that is capable of recovering shaping loss based on the entire interfering sequence.

- *Power Loss*: Second, the transmitted signal X in THP can have more power than the intended signal U . This difference is referred to as the power loss. Pre-subtracting Gaussian-distributed S from U and applying a modulo operation spreads X continuously over the modulo region. Assume

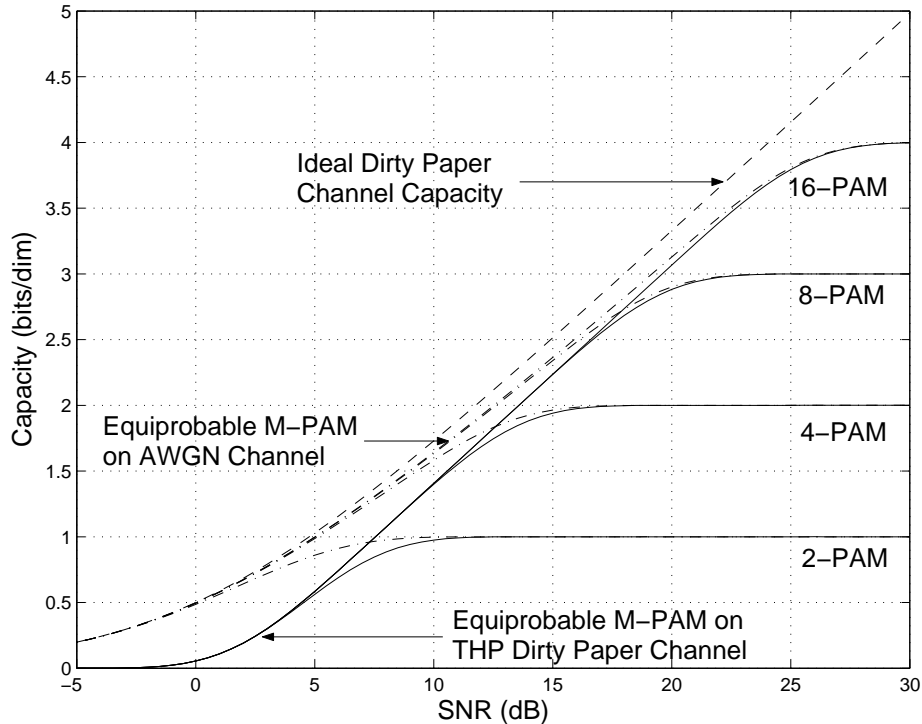


Fig. 3. Capacity of ideal dirty-paper channel compared with M -PAM capacities of AWGN and THP dirty-paper channels

that U is an equiprobable M -PAM with constellation points in $\{-M+1, -M+3, \dots, M-1\}$. When S has a sufficiently large power, X would also be uniform over the modulo region $[-M, M]$. Such a X has more energy than U , so THP introduces a power loss at the transmitter equal to P_X/P_U . A simple calculation shows that for the M -PAM constellation, the power loss equals $M^2/(M^2 - 1)$. Power loss is significant only for low SNR channels where small constellations are used.

- *Modulo Loss*: The modulo operation at the THP receiver introduces a different loss, called modulo loss. Observe that the receiver collapses all received symbols that differ by Δ to the same value, so they can be viewed as multiple representations of the same symbol. With THP, received symbols at the boundary of a constellation may now be mistaken for symbols at the opposite boundary of the constellation, thus incurring potential errors. Modulo loss is related to the proportion of constellation points at the boundaries, and is more pronounced for small constellations (on low SNR channels). Notice also that channel coding schemes for the THP dirty-paper channel must take this wrapping effect into account. For convolutional and trellis coded systems, this alters the calculation of branch costs in the Viterbi algorithm, which increases the number of nearest-neighbor error patterns.

Fig. 3 compares the capacity of the dirty-paper channel given in (2) with the equiprobable M -PAM capacities of the AWGN and THP dirty-paper channels. The capacity of the THP dirty-paper

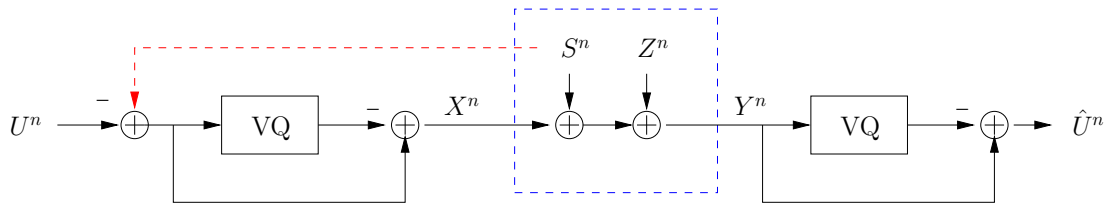


Fig. 4. Precoding via vector quantization

channel is computed as in [18]:

$$I(\hat{U}; U) = h(\hat{U}) - h(\hat{U}|U) \quad (3)$$

$$= h(Y \bmod \Delta) - h(Z \bmod \Delta). \quad (4)$$

Notice that the performance loss of a M -PAM coded system on the AWGN channel is purely shaping loss, while the THP dirty-paper channel suffers from additional power and modulo losses, which can be large in the low SNR regime. The goal of the rest of the paper is to present various approaches to recover these precoding losses in various SNR regimes.

III. TRELLIS CODING FOR THE DIRTY-PAPER CHANNEL

A. Trellis Precoding

At high SNR, Tomlinson-Harashima precoding on the dirty-paper channel incurs a negligible loss compared to M -PAM constellation on the AWGN channels. However, it suffers a shaping loss compared to the ideal dirty-paper capacity. The main objective of this section is to present a trellis precoding scheme that is capable of recovering most of the shaping loss in the high SNR regime. The key insight, first pointed out in [10], is that Tomlinson precoder pre-subtracts only the current S , and does not consider future values of S . With only causal side information, the modulo operation has to be done on a symbol-by-symbol basis, producing an output that is uniformly distributed between $-\Delta/2$ and $\Delta/2$. This uniform distribution corresponds to a high-dimensional cubic shape, thus incurring a shaping loss when compared to the spherical shape of an optimal Gaussian code. To recover the shaping loss, non-causal side information must be used to perform the modulo operation on high-dimensional spheres. The modulo operation in the Tomlinson-Harashima precoding can be thought of as a one-dimensional quantization process where the output is the quantization noise. The generalization of the high-dimensional modulo operation is therefore a vector quantizer which outputs the vector quantization noise.

A conceptual model for a vector quantization-based precoder is shown in Fig. 4. It works as follows. First, a codeword sequence U^n is generated by the encoder of an error-correcting code. The additive interference sequence S^n is pre-subtracted from the codeword, and the difference sequence is then quantized by a vector quantizer. The quantization noise is sent as the input to the channel. The

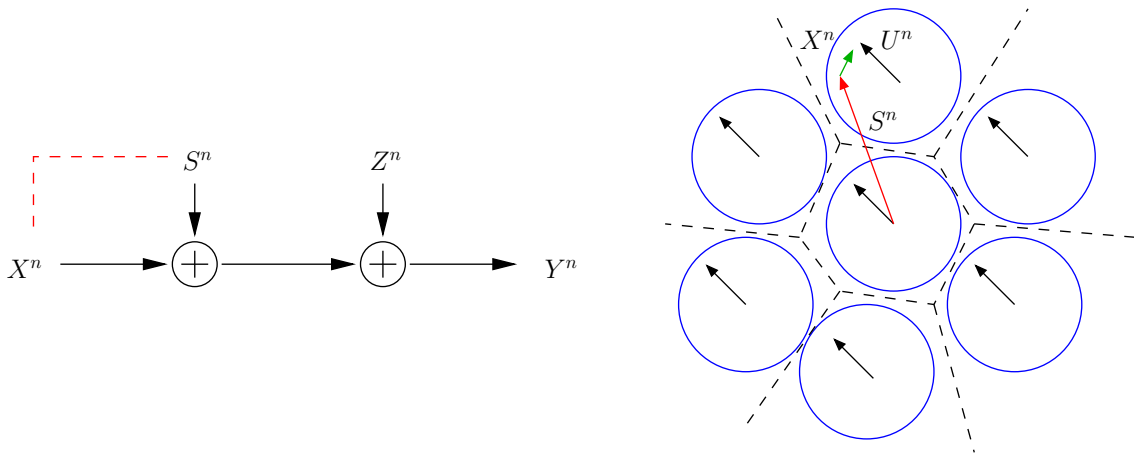


Fig. 5. Precoding with respect to a Voronoi region

channel adds the interference and noise. At the decoder, the received sequence is first quantized by the same vector quantizer. The quantization noise is sent to the decoder of the error-correcting code to recover the message. Fig. 5 illustrates the relation between U^n , S^n and X^n . The spheres denote the Voronoi regions associated with the quantizer outputs. The codeword sequence U^n is designed to be confined within the Voronoi region. Each codeword is given multiple equivalent representatives corresponding to multiple quantizer outputs. The equivalent representatives are illustrated by the arrows in the figure. Since the interference sequence S^n is known non-causally, the precoder can construct an input sequence X^n to steer S^n to the closest representative of U^n . It is easy to see that as long as the codeword is confined to the Voronoi region, perfect reconstruction is possible in the absence of noise. This vector quantization approach can be viewed as a generalization of Tomlinson-Harashima precoding. The one-dimensional modulo- Δ operation is replaced by a vector quantizer which performs a modulo operation with respect to a Voronoi region. After the modulo operation, the precoder outputs are approximately uniformly distributed in the Voronoi region. The Voronoi region of an ideal vector quantizer is a high dimensional sphere, thus achieving a shaping gain. In other words, the transmit power may be reduced by up to 1.53dB when compared to a cubic shape.

The idea of using Voronoi region for shaping was proposed by Forney in [11] for the additive white Gaussian noise (AWGN) channel, where a trellis code is used as a vector quantizer to achieve a shaping gain. For an AWGN channel, the capacity achieving distribution is a Gaussian distribution which is equivalent to a uniform distribution over an n -dimensional sphere over a block length of n . However, traditional rectangular constellation-based trellis codes cover the n -cube uniformly, thus suffering from a shaping loss up to 1.53dB. To recover the shaping loss, [11] proposed first to expand the constellation size slightly, then to modulo the expanded sequence with respect to the Voronoi region of a trellis shaping code. It is clear that shaping for AWGN channel is very similar to shaping

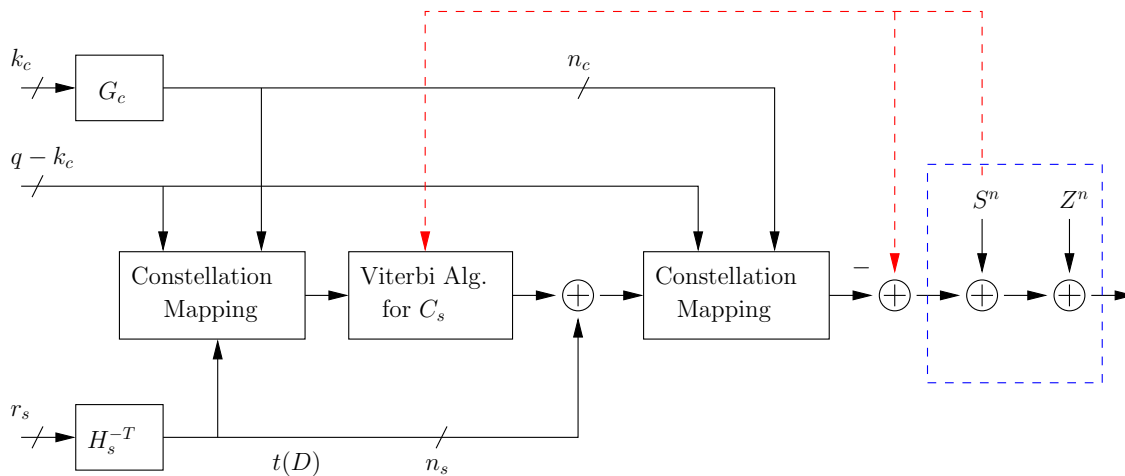


Fig. 6. Implementation of Voronoi precoder

for channel with side-information. In fact, Eyuboglu and Forney further extended trellis shaping by combining shaping with one-dimensional Tomlinson precoding and trellis coding for the ISI channel [12]. This combined structure can be easily modified for channels with non-causal side information.

We now describe a trellis precoding scheme for channels with side information. The operation of the precoder is shown in Fig. 6. A trellis code can be viewed as a collection of paths through the constellation. These paths have a good minimum distance property, and they can be represented by a finite state machine, thus allowing efficient decoding. Such a collection of trellis paths can also be used as reconstruction values in a vector quantizer.

Two codes are working independently. C_c is a trellis channel code, with G_c as its k_c/n_c convolutional coset encoder. C_s is a k_s/n_s convolutional shaping code, whose Voronoi region will be used as the basis of modulo operation. The input bits are divided into two groups. First q bits are the inputs to the trellis channel code, where k_c of which are encoded by the convolutional code G_c , and the rest as uncoded bits selecting constellation points within each coset. The output of the trellis code is a constellation of size $2^{q-k_c+n_c}$. This constellation is repeated 2^{n_s} times in two dimensions resulting in 2^{n_s} non-overlapping regions. At any given time instance, one of these regions is selected by the second group of r_s inputs and the shaping convolutional code C_s . Therefore, the entire codeword can be thought of as a combination of 2^q trellis paths within each region, and the sequence of regions. Two sequences of regions are deemed equivalent if their difference is a valid codeword for C_s . The initial sequence is determined by r_s input bits and an inverse syndrome decoder H_s^{-T} for C_s . In a convolutional code, sequences differing by a valid codeword all have the same syndrome. Thus, using r_s input bits as the syndrome uniquely determines an equivalent class of n_s -bit sequences $t(D)$. Every sequence in this equivalent class can be represented by $t(D)$ plus some valid codeword in C_s . So, the shaping encoder only needs to select the appropriate codeword to be added. The appropriate codeword is selected by the Viterbi algorithm for C_s . The criterion for such selection can in

principle be anything as far as decoding is concerned. To minimize transmit power for channels with side information, the criterion here is the minimization of the square difference between the output sequence and the side information sequence. Thus, the Viterbi algorithm for C_s compares the side information sequence with the constellation sequence determined jointly by the trellis encoder C_c and the syndrome sequence for C_s , and outputs a codeword for C_s which modifies the sequence of region selects so that the resulting codeword is as close to the side information as possible. Since the side information will be added by the channel later, the encoder sends out the difference between the two to the channel. On the decoder side, the bits are recovered with a usual trellis decoder for C_c and a syndrome mapper for C_s . The decoder is identical to the one in [12].

There are two key differences between this trellis precoder for a channel with side-information and a trellis shaping precoder for an ISI channel as described in [12]. In [12], a shaping code selects the region sequence so that the output codeword has the minimum energy, while here a precoder selects the region sequence so that it takes the minimum energy to steer the side information to a correct codeword. The second difference is more subtle. In a shaping code for an ISI channel, a small amount of constellation expansion suffices, thus a rate 1/2 shaping code is sufficient. For precoding for a channel with side information, it is desirable to have a code rate such as 3/4 or 5/6 to force greater constellation expansion. The reason for expansion is to ensure that the side information sequence lies entirely within the expanded constellation. In practice where the magnitude of the side information sequence is not known in advance, it is necessary to add a modulo- M operation outside of the expanded constellation. If the expansion is insufficient with respect to S , the output distribution becomes uniform inside a cubic shape and shaping loss may not be recovered. Note that the actual transmit symbol has a much smaller amplitude compared to the expanded constellation. Since only the difference between the intended signal and the side information sequence is transmitted, the expanded constellation does not affect the dynamic range requirement of a practical transmitter.

The shaping gain for the precoder depends on the shape of the Voronoi region of the shaping code. Therefore, the trellis shaping codes in [11] can be used directly for multiuser precoding with exactly the same shaping gain. In particular, a simple 4-state trellis shaping code already achieves almost 1dB shaping gain.

As far as the transmit power is concerned, trellis shaping is capable of recovering all of 1.53dB of the shaping loss [11]. However, as noted by Fischer et al [19] [20], a trellis shaping receiver can suffer from error propagation since errors in decoding the channel coded stream affect the uncoded shaping stream as well. The alternative method suggested in [19] selects the initial shaping sequence using an intersymbol interference precoder. However, as the dirty-paper channel has no intersymbol interference, this remedy is not directly applicable here. Fortunately, as seen in the simulation results, the effect of error propagation is not significant for the dirty-paper channel.

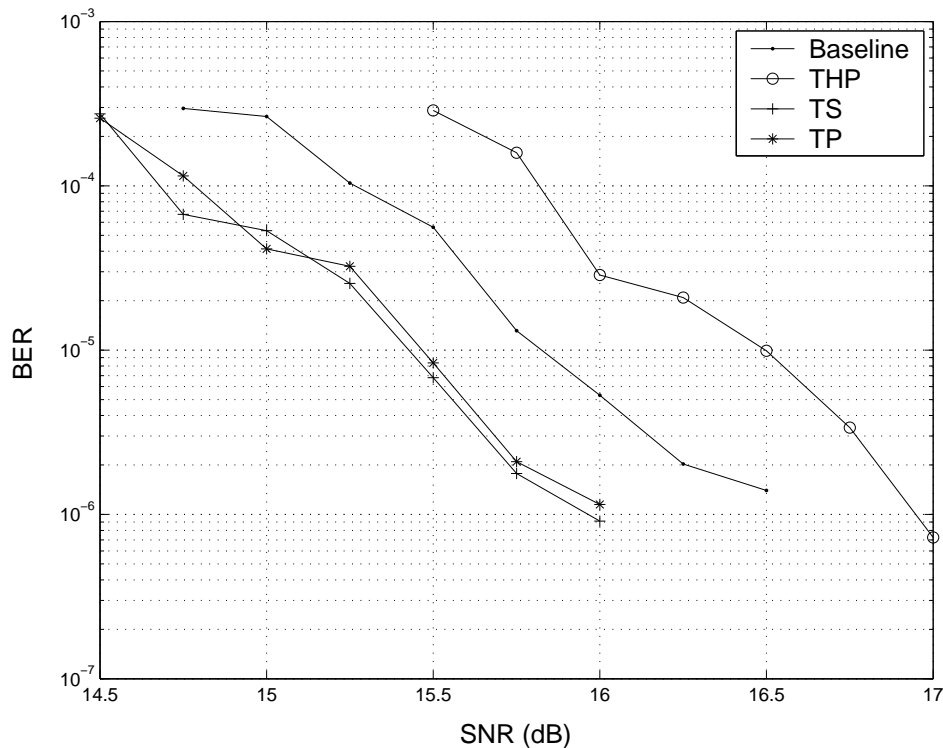


Fig. 7. BER vs SNR for schemes at 2 bits/dim/transmission, using 64-state channel codes and 8-state shaping codes.

B. High SNR Simulation Results

The performance of trellis precoding is evaluated on the dirty-paper channel. Simulation results for four different schemes are presented for the purpose of comparison: the baseline trellis coded modulation on AWGN channels, the trellis shaping (TS) scheme over the AWGN channel, the Tomlinson-Harashima precoding (THP) scheme over the dirty paper channel, and the trellis precoding (TP) scheme with trellis shaping over the dirty-paper channel. Figs. 7 and 8 depict the performance of implementations operating at 2 and 3 bits/dimension/transmission, respectively. All the channel trellis codes in these two figures incorporate identical 64-state, rate 2/3 systematic convolutional encoders, specified by feedforward polynomials $h_2(D) = 036$ and $h_1(D) = 052$ and feedback $h_0(D) = 115$ in octal notation. (In octal notation, 115 represents the transfer function $D^6 + D^3 + D^2 + 1$.) In Fig. 7, the baseline and THP systems use 2 additional signal select bits, while the trellis shaping and trellis precoding schemes use 1 signal select bit and 1 shaping bit. The respective schemes in Fig. 8 each use 1 further signal select bit. The shaping codes for all the trellis shaping and trellis precoding schemes in both figures correspond to 8-state, rate 2/3 systematic convolutional encoders, specified by feedforward polynomials $h_2(D) = 04$ and $h_1(D) = 02$ and feedback $h_0(D) = 11$ and with 3 signal select bits each.

The key observation is that trellis precoding scheme on the dirty-paper channel performs almost

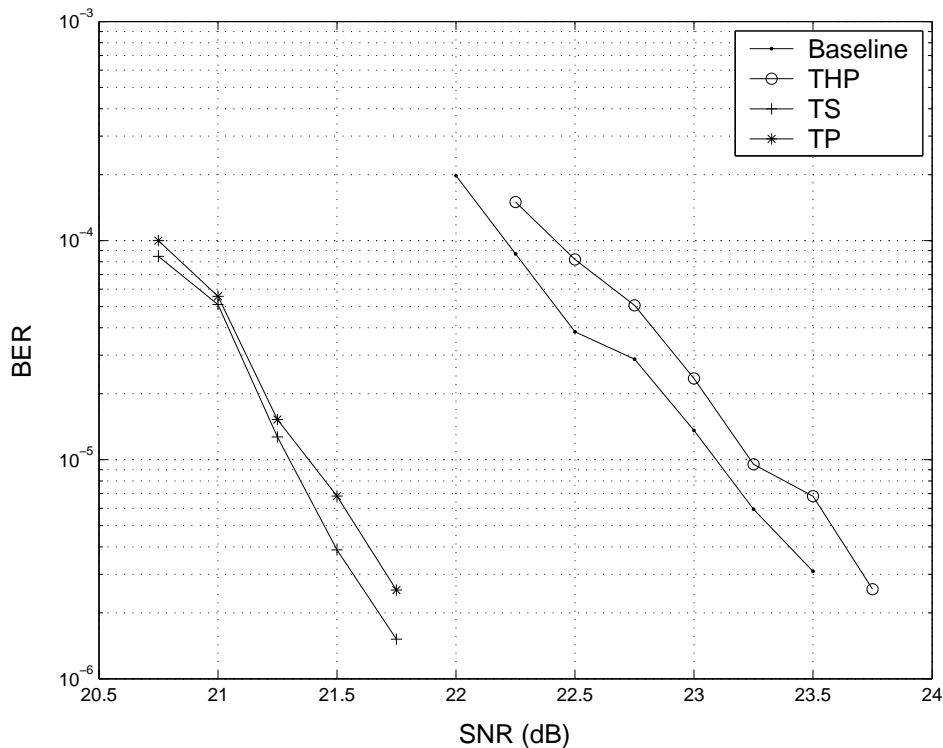


Fig. 8. BER vs SNR for schemes at 3 bits/dim/transmission, using 64-state channel codes and 8-state shaping codes

as well as the usual trellis shaping scheme over the AWGN channel. The performance discrepancy of less than 0.2dB is a small power loss due to the fact that the transmitted signal for trellis shaping is discrete but for trellis precoding it is continuous. With respect to capacity, performance using 64-state channel codes is approximately 3.75dB away at $\text{BER} = 10^{-5}$ for the transmission rate of 2 bits/dim; it is 3.25dB away at the same BER for 3 bits/dim. Note that although the channel codes used in all four schemes are similar, they differ slightly in the number of signal select bits. The constellation expansion needed in trellis shaping scheme over the AWGN channel and the trellis precoding scheme requires more bits to be used as signal selects. Note also that trellis precoding over the dirty-paper channel has the same implementation complexity as the trellis shaping scheme over the AWGN channel.

IV. CONVOLUTIONAL CODING FOR THE DIRTY-PAPER CHANNEL

A. Partial Interference Pre-Subtraction

We have so far dealt with the high SNR channel where shaping loss is the main concern. For the low SNR dirty-paper channels, the power and/or modulo loss of Tomlinson-Harashima precoder must also be taken into account. In this section, we use a partial interference pre-subtraction (PIP) scheme as an effective way to partially recover the power loss at a low SNR. This scheme draws from

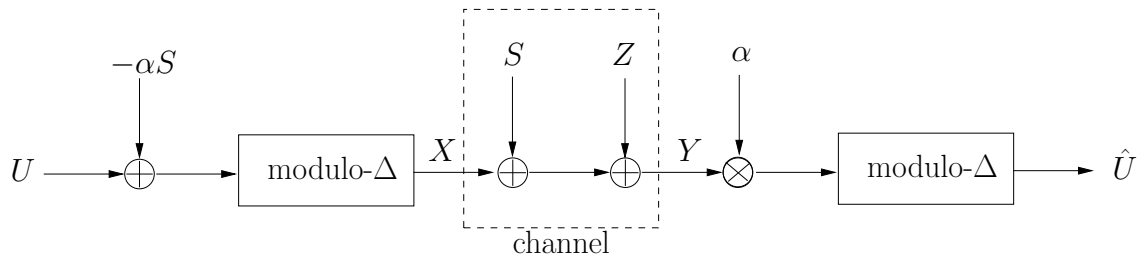


Fig. 9. Partial Interference Pre-subtraction THP

Costa's information theoretical proof of the dirty-paper channel capacity, and is a one-dimensional implementation of Erez, Shamai and Zamir's re-interpretation in the context of inflated lattice strategy [10]. The main contribution of this section is in showing how partial pre-subtraction may be incorporated in the Viterbi decoding of convolutional codes.

At a low signal-to-noise ratio, U is a binary signal. The Tomlinson-Harashima precoding loss for a binary signal comes from two sources. First, precoding creates a uniformly distributed signal, which has a significantly higher power than a binary signal. In addition, compared to the usual uncoded BPSK signal on an AWGN channel where each constellation point has exactly one nearest neighbour, with Tomlinson-Harashima precoding, the number of nearest neighbors increases to two. This is also significant and it leads to a modulo loss. The partial interference pre-subtraction scheme, shown in Fig. 9, is designed to recover some of these losses.

The main idea is due to [10]. Instead of subtracting S , the transmitter pre-subtracts αS , where $\alpha \in (0, 1]$. At the other end, the receiver multiplies the received signal by α before applying the modulo operation. Note that:

$$\hat{U} = (\alpha X + \alpha S + \alpha Z) \bmod \Delta \quad (5)$$

$$= (U - (1 - \alpha)X + \alpha Z) \bmod \Delta, \quad (6)$$

(since $\alpha S = (U - X) \bmod \Delta$.) The effective noise is thus the modulo of the weighted sum of two independent random variables, the uniform X and the Gaussian Z . By choosing $\alpha < 1$, the combined noise can be made smaller than the Gaussian noise Z alone. In fact, the optimal α is the value that maximizes $I(\hat{U}; U)$. However, in practice, an α that minimizes the power of $(1 - \alpha)X + \alpha Z$ is also found to work well.

The choice of α that minimizes the total noise power turns out to coincide with Costa's choice based on information theoretical considerations: $\alpha = P_X / (P_X + P_Z)$. Fig. 10 quantifies the gain due to partial pre-subtraction, by comparing the capacity of the dirty-paper channel given in (2) with the equiprobable M -PAM capacities of the THP dirty-paper channel and the partial interference pre-subtraction THP dirty-paper channel. At 0.5 bits per transmission or below, the gain is over 2dB, which is significant.

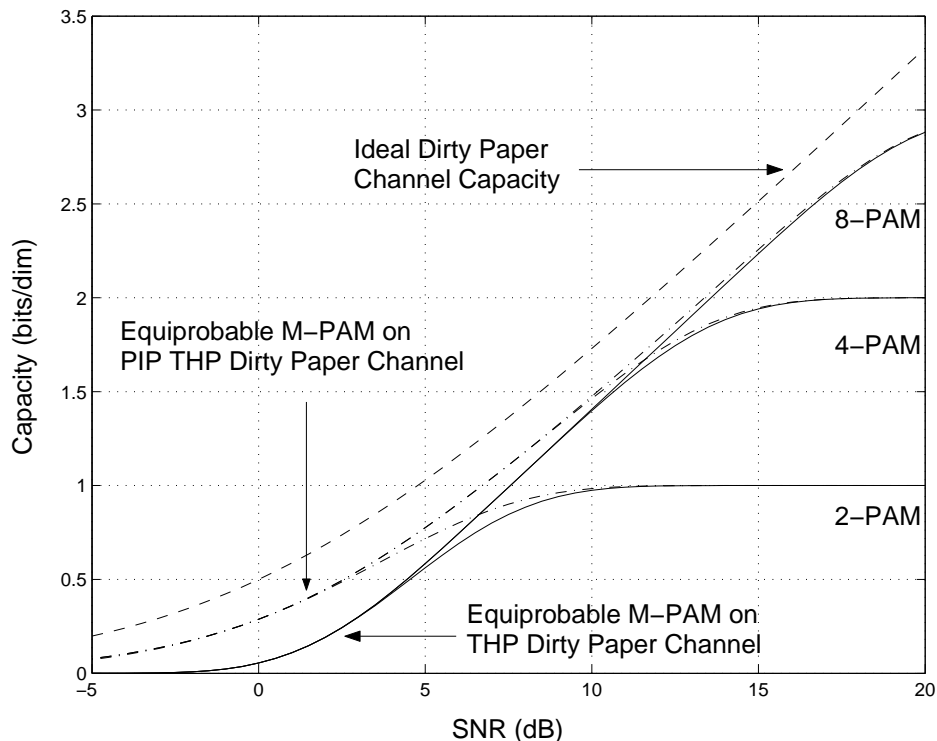


Fig. 10. Capacity of ideal dirty-paper channel compared with equiprobable M -PAM capacities of THP and Partial Interference Pre-subtraction (PIP) THP dirty-paper channels

The main point of this section is to illustrate how partial interference pre-subtraction may be combined with convolutional codes. The decoding process is now slightly different when $\alpha < 1$ is chosen. The usual Viterbi decoding of the convolutional code over the AWGN channel uses a Euclidean distance metric. However, under partial interference pre-subtraction, the effective noise is a combination of Gaussian and uniform components as per (6). Decreasing α increases the power of the uniform part and flattens the overall noise distribution. Therefore, at $\alpha < 1$, the usual Euclidean distance metric in the Viterbi algorithm gives too much preference to small magnitude noise over large magnitude noise. Instead, the following new metric should be used:

$$d(x, y) = -\log f(x - y), \quad (7)$$

where $f(x)$ is the pdf of the noise $(-(1 - \alpha)X + \alpha Z) \bmod \Delta$. This metric captures exactly the self-information of the noise and the Viterbi algorithm minimizes it over the whole received signal. In terms of complexity, the new metric is no worse than the Euclidean one since both are ultimately implemented as table look-ups.

The need for the modified metric highlights the difficulty in truly achieving the capacity of the dirty-paper channel in the low SNR regime using newer iterative decoding methods such as the low-density parity-check (LDPC) codes. As SNR goes to zero, the optimal α also goes to zero. This

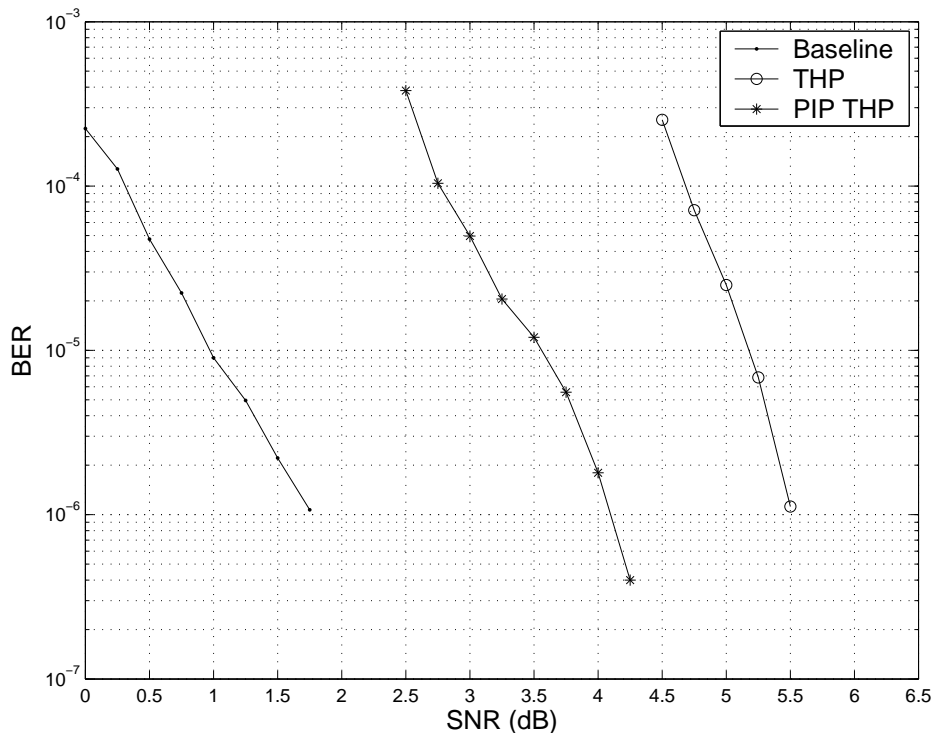


Fig. 11. BER vs SNR for schemes at 0.25 bits/dim/transmission, using 64-state convolutional codes.

means that the noise in the channel is essentially uniform. The LDPC codes with iterative decoding are effective in the Gaussian noise channels, but usually perform poorly in uniform noise channels. The decoding problem with uniform noise is equivalent to the quantization problem, which is known to be difficult task for iterative decoding methods. Such a difficulty does not exist for Viterbi-based decoders.

B. Low SNR Simulation Results

Figs. 11 and 12 show the performance of baseline convolutional codes over AWGN channels and the same codes with THP and partial interference pre-subtraction (PIP) THP over dirty-paper channels, for transmission rates of 0.25 bits/dimension and 0.5 bits/dimension, respectively. The three schemes in Fig. 11 use identical 64-state convolutional codes, specified in octal notation by generators $g_{11}(D) = 163$, $g_{12}(D) = 147$, $g_{13}(D) = 135$ and $g_{14}(D) = 135$. The schemes in Fig. 12 also use identical 64-state convolutional codes, specified by $g_{11}(D) = 155$ and $g_{12}(D) = 117$. The plots indicate that partial interference pre-subtraction recovers a sizeable portion of the total THP loss for low SNR channels. The remaining 2 to 3dB loss closely matches the modulo loss in Fig. 10. We also note that the performance gap to capacity of these partial interference pre-subtraction THP schemes is approximately 7dB at $\text{BER} = 10^{-5}$ for a transmission rate of 0.25 bits/dim, and 6dB at the same BER for 0.5 bits/dim. In effect, about 2 to 2.5dB of the gap to capacity is due

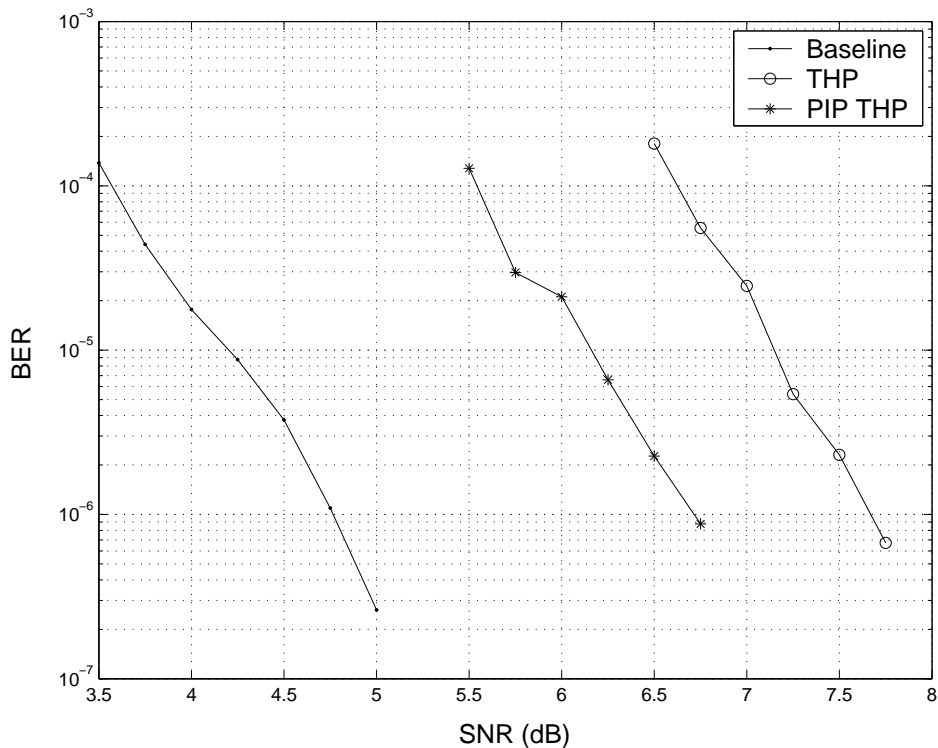


Fig. 12. BER vs SNR for schemes at 0.5 bits/dim/transmission, using 64-state convolutional codes.

to modulo loss and 4dB is due to the sub-optimality of the 64-state channel codes. In other words, partial interference pre-subtraction is within 2.5dB of an equivalent code on AWGN channels. This confirms Costa's capacity result for in the low SNR dirty-paper channel for practical codes.

C. Intermediate SNR Simulation Results

At intermediate SNR, both power loss and shaping loss are present, as shown in Fig. 3. To recover as much of the loss as possible, we combine trellis precoding and partial interference pre-subtraction. This scheme differs very slightly from the trellis precoding described in the previous section. At the transmitter, the Viterbi algorithm for the shaping code selects a representative that lies close to αS , from which αS is pre-subtracted. At the receiver, the received signal is multiplied by α before further processing. Notice that it is not necessary to alter the cost metric in the Viterbi algorithm for the channel code, because the transmitted signal is Gaussian. So the effective noise is also Gaussian and the Euclidean metric suffices.

Fig. 13 shows the performance of baseline trellis coding and trellis shaping (TS) over the AWGN channel and THP and partial interference pre-subtraction trellis precoding (PIP TP) over the dirty-paper channel, at a transmission rate of 1 bit/dimension. All channel trellis codes consist of identical 64-state, rate 1/2 systematic convolutional encoders, specified by feedforward polynomial $h_1(D) = 054$ and feedback $h_0(D) = 161$. The baseline and THP schemes use 1 signal select bit, while the

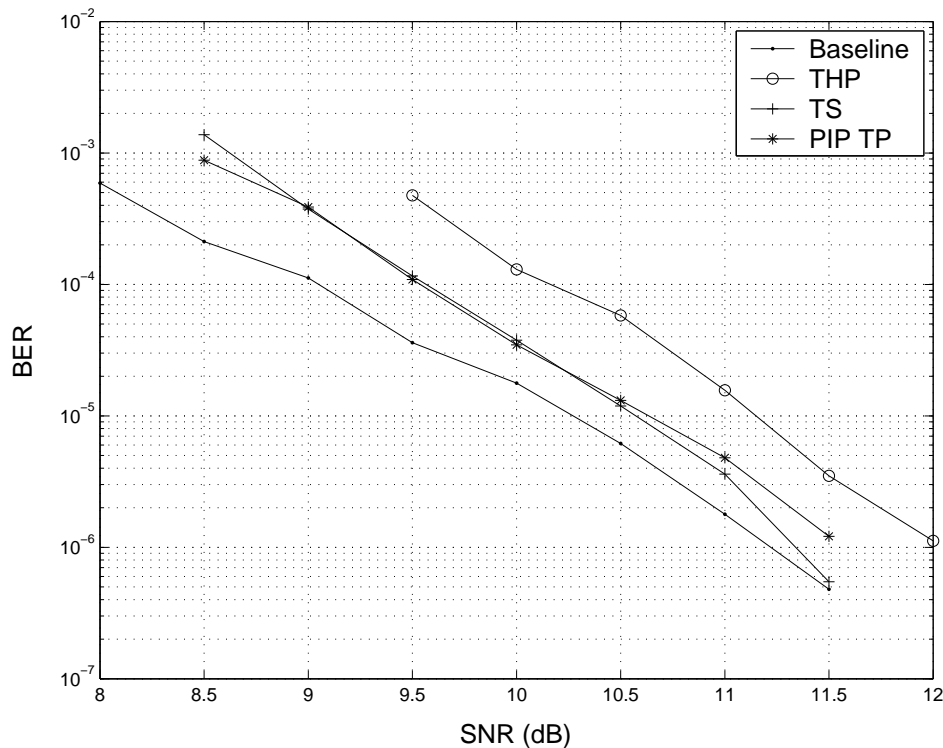


Fig. 13. BER vs SNR for schemes at 1 bit/dim/transmission, using 64-state channel codes and 8-state shaping codes.

trellis shaping and trellis precoding schemes use 1 shaping bit instead. These last two schemes have shaping codes that correspond to 8-state, rate 2/3 systematic convolutional encoders, specified by feedforward polynomials $h_2(D) = 04$ and $h_1(D) = 02$ and feedback $h_0(D) = 11$ and with 3 signal select bits each.

It can be seen from the figure that partial interference pre-subtraction trellis precoding performs within 0.5dB from the baseline case. At $\text{BER} = 10^{-5}$, this 64-state trellis method has a gap to capacity of 5.75dB. Using the theoretical plots in Fig. 10, we estimate that 2dB of the gap is due to modulo loss, leaving 3.75dB resulting from the sub-optimality of the 64-state channel code.

V. CONCLUSIONS AND DISCUSSIONS

This paper studies the combination of practical trellis and convolutional coding with Tomlinson-Harashima precoding for the pre-subtraction of multi-user interference at the transmitter. Tomlinson-Harashima precoding (previously designed to the intersymbol interference channel) incurs significant precoding losses compared to the dirty-paper channel capacity. The precoding loss can be characterized as shaping loss, power loss and modulo loss. Shaping loss dominates in the high SNR regime. Power and modulo losses dominate in the low SNR regime.

The main contribution of this paper is a trellis precoding method that is capable of almost completely recovering shaping loss for the high SNR channels with known interference. Trellis precoding

is based on the combination of Tomlinson-Harashima precoding and trellis shaping. The modulo operation in Tomlinson-Harashima precoding is generalized to a vector quantization process. By pre-subtracting the entire sample path of the interfering signal, rather than the symbol-by-symbol subtraction, the transmit signal can be shaped to approximate a Gaussian distribution, thus recovering the shaping loss. Simulation results show that at rates 1, 2 and 3 bits/dimension, trellis precoding is within 0.5dB of the performance of an equivalent code for AWGN channels.

This paper also studies the low SNR dirty-paper channels and proposes a technique to combine partial interference pre-subtraction with convolutional coding to recover the power loss. The main idea is to subtract interference partially, and to code for an effective noise that is a combination of self-noise and Gaussian noise, but has a smaller overall variance. The implementation of partial interference pre-subtraction requires a modification of the Viterbi decoding in convolutional codes. Simulation results show that it is within 2.5dB of the performance of an equivalent code for AWGN channels at rates 0.25 and 0.5 bits/dim.

The basis of comparison in this paper is between equivalent codes on channels with or without known interference. Although more advanced coding strategies such as turbo codes and LDPC codes are expected to outperform the trellis codes and convolutional codes presented here, the comparison concerning the effectiveness of trellis precoding and partial interference pre-subtraction is expected to be valid even when more advanced error correcting codes are used. In particular, we expect the trellis shaping scheme to continue to recover almost all of 1.53dB shaping gain when a turbo code is used in the underlying trellis code structure.

The coding strategies developed in this paper may find applications in many multiuser system designs where interference pre-subtraction is a key strategy in achieving the capacity. These applications are expected to include not only the multiuser cancellation of crosstalk in digital subscriber lines mentioned earlier, but also multi-user multi-antenna wireless communication systems and high-speed serial link designs.

REFERENCES

- [1] G. Ginis and J. M. Cioffi, "Vectored transmission for digital subscriber line systems," *IEEE J. Selected Areas Comm.*, June 2002.
- [2] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *IEEE Trans. Inform. Theory*, June 2003.
- [3] W. Yu and J. M. Cioffi, "Sum capacity of Gaussian vector broadcast channels," *accepted in IEEE Trans. Inform. Theory*, 2004.
- [4] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates and sum-rate capacity of Gaussian MIMO broadcast channels," *accepted in IEEE Trans. Info. Theory*, 2003.
- [5] P. Viswanath and D. Tse, "Sum capacity of the multiple antenna Gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. Info. Theory*, July 2003.
- [6] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian MIMO broadcast channels," in *Conference on Information Systems and Sciences (CISS)*, March 2004.
- [7] M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

- [8] M. Tomlinson, "New automatic equalizer employing modulo arithmetic," *Electr. Lett.*, vol. 7, pp. 138–139, Mar. 1971.
- [9] S. Shamai and R. Laroia, "The intersymbol interference channel: lower bounds on capacity and channel precoding loss," *IEEE Trans. Info. Theory*, Sep. 1996.
- [10] U. Erez, S. Shamai, and R. Zamir, "Capacity and lattice strategies for cancelling known interference," in *Inter. Symp. Inform. Theory and Its Appl. (ISITA)*, Nov. 2000.
- [11] G. D. Forney Jr., "Trellis shaping," *IEEE Trans. Inform. Theory*, vol. 38(II), pp. 281–300, March 1992.
- [12] M. V. Eyuboglu and G. D. Forney Jr., "Trellis precoding: combined coding, precoding and shaping for intersymbol interference channels," *IEEE Trans. Inform. Theory*, vol. 38(II), no. 2, pp. 301–314, March 1992.
- [13] G. Caire and S. Shamai, "Writing on dirty tape with LDPC codes," in *DIMACS Workshop on Signal Processing for Wireless Transmission*, Oct. 2002.
- [14] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1250–1276, June 2002.
- [15] U. Erez and S. ten Brink, "Approaching the dirty-paper limit for cancelling known interference," in *Allerton Conf. Comm. Cont. Comp.*, 2003.
- [16] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS)," *IEEE Trans. Info. Theory*, vol. 49, no. 3, Mar. 2003.
- [17] M. Miyakawa and H. Harashima, "A method of code conversion for a digital communication channel with intersymbol interference," *Trans. Inst. Elec. Comm. Eng. Japan*, vol. 52-A, pp. 272–273, June 1969.
- [18] R. D. Wesel and J. M. Cioffi, "Achievable rates for Tomlinson-Harashima precoding," *IEEE Trans. Inform. Theory*, vol. 44, no. 2, pp. 824–831, March 1998.
- [19] R. F. H. Fischer, W. H. Gerstacker, and J. B. Huber, "Dynamics limited precoding, shaping, and blind equalization for fast digital transmission over twisted pair lines," *IEEE J. Selected Areas Comm.*, vol. 13, pp. 1622–1633, Dec. 1995.
- [20] R. F. H. Fischer, C. Stierstorfer, and C. Windpassinger, "Precoding and signal shaping in space-time transmission," *preprint*, 2003.