Trellis Precoding for the Multi-User Environment

Aaron Callard, Amir Khandani
E&CE Department, University of Waterloo
Email: aaron, khandani@shannon2.uwaterloo.ca

Aladdin Saleh
Wireless Technology, Bell Mobility
aomsaleh@Mobility.com

Abstract— When transmitting over the multi-user MIMO Broadcast Channel, the transmitter has both knowledge of, and control over, the interference between users. Methods based on interference cancellation focus only on the transmitter's knowledge of the interference. This paper proposes a method which takes advantage of the transmitter's control over the interference among users. This is done by considering all users simultaneously. First, redundancy is introduced into the transmitted data similarly to the trellis shaping method, then the Mean Square Error (MSE) is minimized by exploiting this redundancy. The proposed method is unique in that it considers all users simultaneously, and that the MSE is minimized rather than the transmit power. This method can achieve a BER of 10^{-5}, only 1 dB above the minimum Signal to Noise Ratio predicted by the capacity of the channel. Depending on the channel parameters, this demonstrates a 1-3 dB improvement over the best reported results. The improvement in the performance is achieved at the price of an increase in the complexity (equivalent to the Viterbi decoding of a Trellis with 2^{2N} states for N transmit antennas).

I. INTRODUCTION

The use of wireless technology is on the rise. As both the transmit power and the available bandwidth are severely limited, transmission schemes with higher spectral efficiency must be used.

One method for achieving this increase in spectral efficiency is to use multiple antennas at the transmitter and/or the receiver. Telatar [1] showed that the capacity of a channel can be increased by a factor of $\min(N_t, N_r)$, where N_t, and N_r are the number of transmit and receive antennas, respectively. Therefore, by adding additional antennas at both the transmitter and the receiver, the capacity of a channel can be increased dramatically.

However, in most real world applications, the majority of data transmission will be from a central base station to resource-limited mobile units. These mobile units are often quite restricted in both size and complexity, which can make it difficult for multiple receive antennas to be located on these units. Therefore, we only consider the case of single antenna users in this paper.

When restricted to single antenna users, we can still achieve the increased capacity promised by [1], if we communicate to several users at once. Provided that the transmitter knows the channel, the sum capacity increases with the minimum of N_t and N, where N_t is the number of transmit antennas and N is the number of single antenna users [2]. In this paper, we only consider the scenario where the number of transmit antennas equals to the number of single antenna users. The methods presented here can be easily extended to include the more general case.

The literature proposes two approaches to precoding over the MIMO broadcast channel. The first approach [3] [4] is based on the theoretical application of 'Dirty Paper Coding' (DPC) [5], and the second [6] [7] is a more pragmatic approach based on a generalization of the works of Tomlinson [8], and Harashima [9] [10].

In the DPC approach, the channel is first transformed into a series of parallel sub-channels with known interference. Then, using this knowledge of the interference, a signal is transmitted over these sub-channels.

The first step is performed using the LQ decomposition of the channel matrix. Because a sum power constraint is not affected by a unitary transformation, the channel matrix can be transformed into an equivalent lower triangular matrix. This means that the ith user will experience no interference from the signal sent to the jth user, $j < i$. By choosing the signals successively, the transmitter can decide which symbol to transmit without affecting the previous users received signals.

The theory behind DPC [5] states that a channel with an interference, which is known non-causally at the transmitter, has the same capacity as that same channel without the interference. Practical techniques for these interference channels are still not ideal and each provides some loss. When this loss is compounded over several users, it becomes significant. This causes these methods to operate well away from the capacity region.

Practical techniques for the interference channel are based on adding a modulo operation at the receiver. This modulo operation allows many different symbols to represent the same data. The transmitter can find a series of symbols which are close to the interference, and therefore require small transmit power, but are still, after the modulo operation, equivalent to the original data. This modulo operation is often with respect to large multi-dimensional regions.

Perhaps the best method for interference cancellation is Trellis Precoding. Trellis Precoding was first proposed for dealing with shaping loss [11], then modified for the ISI channel [12], and then finally for interference cancellation [3]. This method is based on a modulo operation with respect to the Voronoi Region of a shaping Trellis Code. This is equivalent to sending a certain amount of redundant data which is encoded using a simple convolutional or trellis code. However, instead of this data containing any information, it is chosen to minimize the transmit power. By utilizing the trellis

nature of the shaping code, this minimization can be done using the standard Viterbi algorithm.

The second class of approaches to precoding over the MIMO broadcast channel, Vector Precoding [6] and Shaping without Scrambling [7], use a much simpler modulo operation, but consider all users simultaneously. Rather than eliminating the interference for each user separately, these methods choose symbols to maximize the amount of constructive interference among users.

At the heart of Vector Precoding is the simple one dimensional modulo operation first introduced by Tomlinson and Harashima. In effect, this modulo operator extends the signal constellation periodically. This gives the transmitter an infinite number of possible symbols, all of which represent the same data. By properly choosing the transmitted symbols within the available selection, the transmitter can significantly reduce the transmit power. However, because of the simple modulo operation, this method suffers from shaping loss as every symbol, after the modulo operation, is equiprobable [13].

This paper proposes a method which combines conventional methods of interference cancellation with the benefits achieved by Vector Precoding. This is done by using the modulo operation designed for interference cancellation, but considering all users simultaneously, rather than individually.

To do this, we focus on the interference cancellation method of Trellis Precoding. As stated above, Trellis Precoding is equivalent to sending a certain amount of redundant data to each user. If we place the Trellis Shaping modulo operator at each of the receivers, then our problem is reduced to deciding what redundant data to send to each user.

Since the redundant data is encoded using convolutional codes, we would like to take advantage of the trellis structure to efficiently calculate which redundant data to transmit. To do this, we require two things: First, the convolutional codes for each of the users must be combined into one large trellis, and second, we need to calculate a branch metric to calculate the best path through the trellis.

The creation of a new trellis is relatively simple, and can be expressed as a function of our original single user trellis. To calculate the corresponding branch metric, we will examine two cases. In the first case, we allow no interference among users. This is often called Zero Forcing (ZF). In this case, the proper branch metric is the transmit power.

In the second case, we drop the ZF restriction and allow interference among users. To calculate how much interference to allow, we use the Mean Square Error (MSE) as our objective function. In the fast fading environment, this is equivalent to maximizing the Signal to Interference plus Noise Ratio (SINR). By setting the branch metric to be the MSE, we can efficiently calculate a path through the trellis which minimizes the MSE.

The amount of gain depends heavily on the channel realization, the number of users, and, in the MSE case, the power level. Therefore, depending on the circumstances, the amount of gain varies widely. However, some general trends can be observed: the larger the constellation size, the greater the gain,

in the MSE case, the lower the power level, the greater the gain, and finally, fading channels perform markedly better than their quasi-static counterpart.

In the MSE case gains of up to 2.5 dB can be achieved in a fast fading environment. This means that bit rates of 3 bits/dim/user can be transmitted with a Bit Error Rate (BER) of less than 10^{-5}, only 1 dB above the channel capacity. Gains are slightly less if we assume that the channel is quasi static, with average gains being near 1.5 dB. In the ZF case, we achieve gains of about 1 dB.

The paper is divided up as follows; in section II, we introduce our channel model and then develop some notations which allow for a simple distinction of useful signal from interference. We then go on in section III to review the concept of Sign Bit Shaping, the simplest form of trellis shaping. In section IV, we extend this single user method to the multi-user setting, developing the objective function and expanded trellis. We finish by examining the performance of this new method.

Notation: Bold uppercase (lowercase) represent matrices (column vectors); $E\{\cdot\}$, $(\cdot)^H$, represents the expectation, and Hermitian transpose respectively. $\mathbf{I}$ represents the identity matrix of size N.

II. THE MODEL

This paper deals with channels which can be represented by

$$\mathbf{y} = \mathbf{Hs} + \mathbf{n} \tag{1}$$
$$s.t. \quad E\{\|\mathbf{s}\|^2\} \leq \overline{P}$$

$\mathbf{y}$ is a column vector with elements representing the received signal at each user, $\mathbf{s}$ represents the transmitted signals and has an average power constraint $\overline{P}$, $\mathbf{n}$ represents additive white Gaussian noise (AWGN) with covariance matrix $\sigma^2\mathbf{I}$, and $\mathbf{H}$ represents the Channel matrix. We assume $\mathbf{H}$ to be square and invertible, although not necessarily well conditioned. This model can represent several practical systems such as Direct Sequence CDMA, the MIMO Broadcast Channel or DSL lines [14].

While this model is easy to build, it can be cumbersome to work with. By some slight manipulations, the transmitted signal can be divided into useful information ($\mathbf{u}$) and interference (ϵ). By defining $\mathbf{s}$ to be $\mathbf{H}^{-1}(\mathbf{u}+\epsilon)$ as well as including the scaling factor γ, and energy variable P, the channel can be rewritten without loss of generality as

$$\mathbf{y} = \mathbf{HH}^{-1}\frac{\sqrt{P}}{\gamma}(\mathbf{u}+\epsilon) + \mathbf{n}$$
$$= \frac{\sqrt{P}}{\gamma}(\mathbf{u}+\epsilon) + \mathbf{n}$$
$$\gamma = \|\mathbf{H}^{-1}(\mathbf{u}+\epsilon)\|$$
$$s.t. \quad E\{P\} \leq \overline{P}$$

For the users to decode their received signals, they must have knowledge of the scaling factor $\frac{\sqrt{P}}{\gamma}$. The simplest way for this to occur is for $\frac{\sqrt{P}}{\gamma}$ to remain constant for any value

of $\overline{P}$ and $\mathbf{H}$. We have therefore assumed that this value $\frac{\sqrt{P}}{\gamma}$ defined as c henceforth remains constant. It was shown in [6] that this assumption leads to enhanced performance over the fast fading channel.

Furthermore, if we make the reasonable assumption that our transmission scheme performs better when more power is transmitted, then we may remove the inequality and define c to be the largest value which satisfies the power constraint. Thus, the channel can be written as

$$
\begin{aligned}
\mathbf{y} &= c(\mathbf{u} + \epsilon) + \mathbf{n} &\quad (2) \\
c &= \sqrt{\overline{P}/E\{\gamma^2\}} \\
\gamma &= \|\mathbf{H}^{-1}(\mathbf{u} + \epsilon)\|
\end{aligned}
$$

This channel model is easy to work with as it allows a clear distinction between the useful and detrimental portions of the signal. It also allows the solutions to the MMSE equations to be written quite simply and without any linearity assumptions.

For simulation purposes, we have assumed that the elements of $\mathbf{H}$ are independent unit mean complex Gaussian variables. In the fast fading case, we assume that for every transmission there is a new independent sample of $\mathbf{H}$. In the quasi-static case, we assume that $\mathbf{H}$ remains constant throughout a block of transmissions and is then replaced by a new independent realization.

III. SIGN BIT SHAPING

In this section, we review the application of Sign Bit Shaping to the AWGN channel. Over the AWGN channel, the introduction of a modulo operation was first proposed to realize some shaping gain [11] [15] [16] [17]. The idea is to give the transmitter many different options when determining which symbol to transmit. By choosing correctly, the transmitter can transmit the same data, yet at a lower transmit power.

This is very similar to the method proposed in [3] [4] for dealing with the interference envisioned in DPC. The only difference is that the transmitter chooses the symbol which is closest to the interference, rather than the symbol with the smallest magnitude. As this effectively minimizes the transmit power, it is easy to visualize signal shaping as the special case of interference cancellation, where the interference is zero.

Perhaps the best and simplest modulo operator uses the Voronoi region of a trellis code $\mathbf{C}_{shaping}$. Effectively, this means transmitting a certain amount of redundant data which, while decodable at the receiver, is of no direct use. As this data is encoded using a Convolutional, or Trellis code, the underlying method is termed Trellis Shaping.

Sign Bit Shaping is the simplest form of Trellis Shaping. Only the first 2 bits of the transmitted constellation are modified by the shaping code, while the rest remain unaffected.

Consider an $M \times M$ square constellation. This constellation is capable of transmitting $R = \log_2 M^2$ bits per two dimensions. If this constellation is labeled using standard Grey mapping, then the first two bits are in charge of the sign of the real and imaginary parts of the symbol, and are thus called

the sign bits. We will denote these two bits by the sequence $\{\mathbf{z}_j\} = \{(z_{1,j}, z_{2,j})\}$.

We define the addition of a codeword $\mathbf{c}_i$ to be the addition of $\{\mathbf{c}_i\}$ to $\{\mathbf{z}_j\}$, with modulo 2 arithmetic. As stated previously, the receivers are equipped with a modulo operation with respect to the code $\mathbf{C}_{shaping}$. This means that any codeword $\{\mathbf{c}_i\} \in \mathbf{C}_{shaping}$ can be added to our original sign bits $\{\mathbf{z}\}$ and the information will not be changed.

Therefore, after adding a code word $\{\mathbf{c}_i\}$, the sign bits are $\{\mathbf{u}\} = \{\mathbf{z} \oplus \mathbf{c}_i\}$. As the codeword $\{\mathbf{c}_i\}$ does not affect the information transmitted, it can be chosen to minimize the transmit power. This minimization can be done using the Viterbi Algorithm to search through the trellis representation of $\mathbf{C}_{shaping}$, using the transmit power as the branch metric.

We have now outlined how a codeword is added to our original sign bits in order to minimize energy. Now, we must consider how to perform this modulo operation and recover information in the original sign bits $\mathbf{z}$. The answer lies in the syndrome of the shaping code $\mathbf{C}_{shaping}$.

For any rate k/n binary linear convolutional code $\mathbf{C}$ with generator matrix $\mathbf{G}$, there exists an $n \times (k - n)$ syndrome former matrix $\mathbf{F}^T$ such that

$$
\mathbf{GF}^T = \mathbf{0}
$$

Therefore, for any $\{\mathbf{c}\} \in \mathbf{C}$, $\mathbf{cF}^T = \mathbf{0}$. Furthermore, by simple arithmetic, we can see that addition of a codeword will not affect the syndrome of a code [11].

$$
\begin{aligned}
\mathbf{uF}^T &= \mathbf{z} \oplus \mathbf{c}_i\mathbf{F}^T &\quad (3) \\
&= \mathbf{zF}^T \oplus \mathbf{c}_i\mathbf{F}^T &\quad (4) \\
&= \mathbf{zF}^T &\quad (5)
\end{aligned}
$$

Information can therefore be stored in the syndrome of the shaping code $\mathbf{C}_{shaping}$. For a rate k/n code, $n - k$ bits can be stored in the syndrome. Thus, the syndrome in Sign Bit Shaping can hold one bit of information per transmission. Given data bits $\mathbf{b}$, the sign bits can be generated using an inverse syndrome former $\mathbf{F}^{-T}$.

$$
\mathbf{z} = \mathbf{bF}^{-T}
$$

Note that $\mathbf{F}^{-T}$ is not, in general, unique and is any left inverse of $\mathbf{F}^T$.

Recovering this information requires the use of a syndrome former. As the syndrome former depends on multiple received signals, these bits suffer from noise amplification. This is mentioned as a source of loss in several papers on Trellis Shaping [3], [11]. There is a simple method to avoid this loss. If the least significant bits are decoded first and then pre-subtracted in a DFE-like manner, then the syndrome bits are much better protected than the less significant bits. Experiments show that for power levels near capacity, the syndrome bits are so well protected that they do not need to be encoded at all.

IV. SIGN BIT SHAPING FOR MULTIPLE USERS

We now extend the concept of Sign Bit Shaping to the multi-user setting. This involves two steps, first we must design
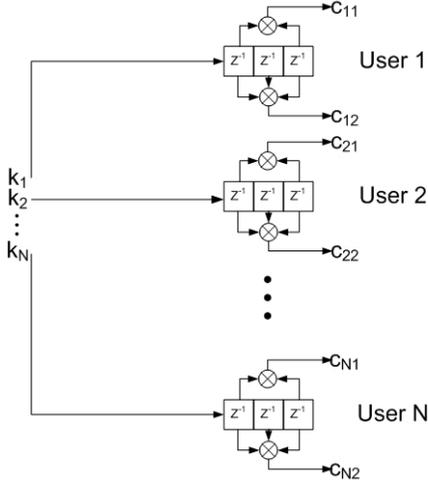
Fig. 1. An example of a convolutional code which is made up of smaller independent codes

a shaping code $\mathbf{C}_s$, with syndromes that can be decoded separately for each user, but still provide the shaping we require; and second, we must design an objective function to use as a branch metric when determining the individual codeword used from our shaping code.

To assure that each user can decode their syndromes individually, we design our shaping code $\mathbf{C}_s$ by combining N shaping sub-codes $\mathbf{C}_{shaping}$. Figure 1 is an example of one such shaping code.

The generator matrix of this larger code may be readily written as a function of the constituent codes $\mathbf{C}_{shaping}$. If $\mathbf{C}_{shaping}$ has generator matrix $\mathbf{G}$, then the new code has a generator matrix $\mathbf{G}_T$ which is block diagonal with the $\mathbf{G}$'s repeated N times along its diagonal.

$$\mathbf{G}_T = \begin{bmatrix} \mathbf{G} & 0 & \cdots & 0 \\ 0 & \mathbf{G} & \cdots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & & \mathbf{G} \end{bmatrix}$$

From the generator matrix $\mathbf{G}_T$, it is simple to develop the overall trellis.

Because our shaping code is made up of independent shaping sub-codes, we can use the same method as Sign Bit Shaping to encode our data. That is, using a syndrome former, the sign bits are calculated for each user and then added in modulo 2 arithmetic to the part of our chosen shaping codeword $\{\mathbf{c}_i\}$ for that user. If we define the output symbols for a given codeword to be $\mathbf{u}$, we need some method to select the best codeword.

Next, we develop the objective function or branch metric. We first consider the relatively simple case in which no interference is allowed among users, we then remove this restriction to consider the more general MSE case.

A. Trellis Shaping with Zero Forcing

If we allow no interference among users (Zero Forcing), then the received signal for a given $\mathbf{u}$ is simply $c\mathbf{u}$. Therefore, we wish to determine the codeword $\{\mathbf{c}_i\}$ which will maximize the scaling factor c.

$$\begin{align} \{\mathbf{c}_i\} &= \underset{\{\mathbf{c}_i\}\in\mathbf{C}_{shaping}}{\operatorname{argmax}} \frac{\overline{P}}{E\{\gamma^2\}} \tag{6} \\ &= \underset{\{\mathbf{c}_i\}\in\mathbf{C}_{shaping}}{\operatorname{argmin}} E\{\gamma^2\} \tag{7} \\ &= \underset{\{\mathbf{c}_i\}\in\mathbf{C}_{shaping}}{\operatorname{argmin}} E\{\|\mathbf{H}^{-1}\mathbf{u}\|^2\} \tag{8} \end{align}$$

This problem can be solved by exploiting the trellis properties of our shaping code. If we set the branch metric to be $\|\mathbf{H}^{-1}\mathbf{u}\|^2$, we can minimize the transmit power (7), by using the Viterbi algorithm on the trellis describing the shaping code $\mathbf{C}_s$.

B. Trellis Shaping with Mean Square Error

If interference is allowed among users, significant gains in the SINR can be achieved at low to medium power levels. Let us then remove the Zero Forcing restriction, and define our new objective function to be the MSE of the received signal. For any given codeword, the MSE can be written as

$$\begin{align} MSE &= E\left\{\left\|\frac{\mathbf{y}}{c} - \mathbf{u}\right\|^2\right\} \tag{9} \\ \mathbf{y} &= c(\mathbf{u}+\epsilon) + \mathbf{n} \end{align}$$

This gives us two variables to consider when minimizing the MSE; $\mathbf{u}$ which is modified by the choice of codeword, and ϵ which can be chosen arbitrarily. Using simple vector calculus, the MSE can be minimized with respect to ϵ as shown in [18]. This results in

$$\begin{align} \mathbf{y} &= c\mathbf{H}\mathbf{H}^H\left(\mathbf{H}\mathbf{H}^H + \alpha\mathbf{I}\right)^{-1}\mathbf{u} + \mathbf{n} \\ \alpha &= \frac{2N\sigma^2}{\overline{P}} \end{align}$$

We next find a codeword $\{\mathbf{c}_i\}$ which will minimize the MSE. Note that this is a key difference between the proposed method and other methods reported in the literature. Both Vector Precoding and Trellis Precoding focus on minimizing the transmit power, and thus, do not perform well when interference among users is allowed. To minimize the MSE, the branch metric could be set to (9), however there does exist a simplification of the equation [18]. For a given $\mathbf{u}$, the MSE can be written as

$$MSE = \alpha\left\|\sqrt{\left(\mathbf{H}\mathbf{H}^H + \alpha\mathbf{I}\right)^{-H}}\mathbf{u}\right\|^2 \tag{10}$$

By setting the branch metric to (10), this function can be solved by exploiting the trellis structure of the underlying shaping code.

V. Combining Vector Precoding and Trellis Precoding

One benefit to considering all users simultaneously is that a perturbation vector **p**, like that introduced in [6] [18], can be added to the transmitted signal. We can add another modulo operator identical to the one used for THP or Vector Precoding to this design. Then, any integer multiple of some constant τ can be added to the received signal without affecting the data to be transmitted. Thus, given a vector of original data symbols $\mathbf{u}_0$, any signal of the form

$$\mathbf{s} = c\mathbf{H}^H \left(\mathbf{H}\mathbf{H}^H + \alpha\mathbf{I}\right)^{-1} (\mathbf{u}_0 + \tau\mathbf{p})$$
$$\mathbf{p} \in \{\mathbf{a} + i\mathbf{b} | \mathbf{a} \in \mathbf{Z}^N, \mathbf{b} \in \mathbf{Z}^N\}$$

transmits the same data.

Originally, this vector **p** was chosen to minimize the transmit power γ^2 [6]. An improved version, presented in [18], proposes choosing **p** to minimize the MSE directly. We will use this new choice of vector **p** shown below.

$$\mathbf{p} = \underset{\mathbf{p}}{\operatorname{argmin}} \left\| \left(\mathbf{H}\mathbf{H}^H + \alpha\mathbf{I}\right)^{-H/2} (\mathbf{u} + \tau\mathbf{p}) \right\|^2 \quad (11)$$

By using a sphere encoder, this minimization can be done efficiently for the small number of users we are considering. Thus, the new branch metric would be

$$MSE = \alpha \left\| \sqrt{\left(\mathbf{H}\mathbf{H}^H + \alpha\mathbf{I}\right)^{-H}} (\mathbf{u} + \mathbf{p}) \right\|^2 \quad (12)$$

VI. Performance Results

The performance of this new method depends on several variables. In the following, we compare the proposed method of Multi-User Trellis Precoding to the improved version of Vector Precoding proposed in [18]. We have selected [18] for comparison as it is the best form of precoding reported in the literature.

To evaluate the gains achieved by the proposed method, we consider the MSE as our objective function. First, the MSE using Multi-User Trellis Precoding is calculated. Then, the power required to achieve this same MSE using Vector Precoding is computed. The gains presented are the ratio of these two power levels, with Vector Precoding given a 3 dB advantage because of the extra bit it transmits.

In the MSE case, the amount of gain depends on the power level. To determine which power level to consider, we have calculated the minimum power required for the channel capacity region to contain the transmitted rate vector **R**. We have assumed that the unaffected bits are encoded using a rate $1/2$ code, and that the syndrome bits are uncoded. Thus, for an $M \times M$ constellation $R_i = \log_2 M$ bits are transmitted. To determine the minimum energy required, we have simply solved for the minimum power for the DPC region to contain **R** as shown in [19].

For the quasi static case, the gains presented in table II - III were simulated over many blocks of length 100 [1]. In the fast

[1]The number of simulated blocks is large enough to guarantee a variance less than 0.05 in the simulated average values.

TABLE I

GAIN DUE TO TRELLIS PRECODING WITH FAST FADING AND MMSE

N by M	4	16	36	64
1	-3.01	3.8	3.4	4.2
2	0.22	3.06	2.69	2.44
3	0.93	2.70	2.35	2.14
4	1.47	2.48	2.3	2.2

TABLE II

GAIN DUE TO TRELLIS PRECODING WITH QUASI-STATIC FADING AND MMSE

N by M	4	16	36	64
1	-3.01	1.25	1.32	1.44
2	-0.28	1.61	1.66	1.67
3	0.40	1.68	1.65	1.71
4	1.41	2.3	1.88	1.69

fading case, the gains presented in table I,IV, were simulated over block lengths of 10000.

There are three possible sources of improvement for the proposed method.

1) A shaping gain for each user as it would apply in the single user case. This gain is limited by 1.53 dB [20].
2) The optimization for the Vector Precoding vector **p** is now over a new more flexible region.
3) In the fast fading environment, the time dependent nature of the trellis can take advantage of fluctuations in the channel by transmitting smaller signals when users are in deep fade.

The amount of gain depends heavily on the channel realization **H**, the number of users N, the size of the original constellation M, and in the MSE case, the power level considered. However, there are some general trends which can be seen by examining the tables.

If **H** is held constant over the entire transmission block, then the proposed method only benefits from the standard shaping gain (this is equal to that shown for the single user case), and the increase in flexibility due to the larger optimization region (1 and 2 above). We can see that the benefit of this larger region increases with N and is more prominent for small constellation sizes.

To demonstrate that this gain can be realized as a true reduction of the BER, we have simulated this method over the fast fading channel. The less significant bits are encoded using rate $1/2$ Turbo Codes with feedback polynomial $(1+D^2+D^3)$ and feedforward polynomial $(1+D+D^3)$ over blocklengths of 4000. The channel was simulated as a four antenna fast fading MIMO broadcast channel. We have $N = 4$ and $M = 64$, we would therefore expect a gain of 2.2 dB. This is consistent with the gain observed in Figure VI. As we are transmitting 3 bits/dim/user, the minimum energy required is 11.2 dB, as one can see a BER of 10^{-5} is achieved at 12.2 dB, only 1 dB away from the capacity limit.

TABLE III

GAIN DUE TO TRELLIS PRECODING WITH QUASI-STATIC FADING AND

ZERO FORCING

N by M	4	16	36	64
1	-3.01	0.78	0.74	0.89
2	-0.44	0.99	1.02	1.03
3	-0.10	1.02	1.01	1.02
4	0.14	0.99	0.96	0.96

TABLE IV

GAIN DUE TO TRELLIS PRECODING WITH FAST FADING AND ZERO

FORCING

N by M	4	16	36	64
1	-3.01	3.97	2.09	1.41
2	-0.90	2.03	1.69	1.96
3	0.32	1.48	1.54	1.47
4	0.47	1.26	1.23	1.24

VII. CONCLUSION

Multi-User Trellis Precoding outperforms all known methods for transmitting over the MIMO Broadcast Channel. As it considers all users simultaneously, it results in an optimal, in the MSE sense, amount of interference. This makes it superior to all forms of precoding with the exception of Vector Precoding. As this method suffers much less shaping loss, it also outperforms Vector Precoding.

The improvement in the performance is achieved at the price of an increase in the complexity. As the shaping code $\mathbf{C}_s$ is composed of N shaping sub-codes, the number of states of the trellis grows exponentially with N. For the Sign Bit Shaping example above, the complexity is equivalent to Viterbi decoding of a Trellis with 2^{2N} states, which is unmanageable for a large number of users. However, in practical systems, the number of transmit antennas is quite limited, resulting in an overall manageable complexity for the proposed method.
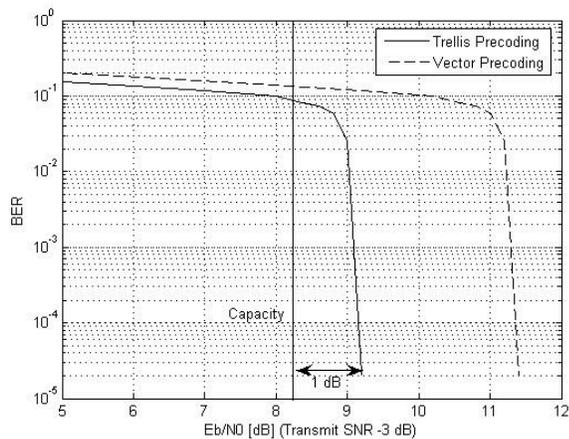

Fig. 2.   BER vs transmit SNR for Trellis Precoding and Vector Precoding

[1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommunications*, vol. 10, pp. 585-595, Nov/Dec 1999.
[2] W. Yu and J. Cioffi, "Sum capacity of a Gaussian vector broadcast channel," in *Proceedings IEEE International Symposium on Information Theory*, p. 498, July 2002.
[3] Wei Yu, David Varodayan, John Cioffi: "Trellis and Convolutional Precoding for Transmitter-Based Interference Pre-Subtraction"',accepted in *IEEE Transactions on Communications*, 2005.
[4] U. Erez, S. Brink. "A Close-to-Capacity Dirty Paper Coding Scheme" *IEEE Transactions on information theory*, Vol. 51, No. 10, October 2005.
[5] M. H. M. Costa. "Writing on dirty paper". *IEEE Transactions on Information* Theory, 29(3):439-441, May 1983.
[6] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A Vector-Perturbation Technique for Near-Capacity Multi-Antenna Multi-User Communication-Part 1" *IEEE Transactions on Communications*, VOL. 53, NO. 1, January 2005 195
[7] 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 JSAC*, pp.1622-1633, Dec. 1995.
[8] M. Tomlinson, "New automatic equalizer employing modulo arithmetic," *Electron. Lett.*, vol. 7, pp. 138-139, Mar. 1971.
[9] H. Miyakawa and H. Harashima, "Information transmission rate in matched transmission systems with peak transmitting power limitation," in *Nat. Conf. Rec., Inst. Electron., Inform., Commun. Eng. of Japan*, Aug. 1969, no. 1269, p. 1268.
[10] H. Harashima and H. Miyakawa, "Matched-transmission technique for channels with intersymbol interference," *IEEE Trans. Commun., vol. COM-20*, pp. 774-780, Aug. 1972.
[11] G. Forney, "'Trellis shaping,'" *IEEE Trans. on Information Theory*, vol. IT-38, no. 2, pp. 281-300, Mar. 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), pp. 301-314, Mar. 1992.
[13] R. D. Wesel and J. M. Cioffi, "Precoding and the MMSE-DFE," in *Proceedings of the 28th Asilomar Conference on Signals, Systems and Computers*, Monterey (CA), pp. 1144–1148, Nov. 1994.
[14] C. Windpassinger, "Detection and Precoding for Multiple Input Multiple Output Channels Ph.D Dissertation, Friedrich Alexander University, Erlangen-Nrnberg, Germany, 2004.
[15] J. H. Conway and N. J. A. Sloane, "A fast encoding method for lattice codes and quantizers, *IEEE Trans. Inform. Theory*, Vol 29, pp. 820–824, Nov. 1983.
[16] G. D. Forney, Jr., "Multidimensional constellations— Part II:Voronoi constellations," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 941-958, Aug. 1989.
[17] A.R. Calderbank and L.H. Ozarow "Nonequiprobable signaling on the Gaussian channel'" *IEEE Transaction on Information Theory*, Vol. 36, No. 4, pp. 726-740, July 1990.
[18] A. Callard, A. Khandani, A. Saleh "Vector Precoding with MMSE for the Fast Fading and Quasi-Static Multi-User Broadcast Channel" *Annual Conference on Information Sciences and Systems*, Princeton, NJ, March 2006
[19] Vishwanath, S.; Jindal, N.; Goldsmith, A. "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels" *IEEE Transactions on Information Theory*, Volume 49, Issue 10, Oct. 2003 Page(s):2658 - 2668
[20] G. D. Forney, R. G. Gallagher, G. R. Lang, F. M. Longstaff, and S. U. Qureshi, "Efficient modulation for bandlimited channels," *IEEE J. Select. Areas Commun.*, vol. 2, no. 5, pp. 632–646, Sept. 1984.
[21] L. Rde, B. Westergren, *Mathematics Handbookfor Science and Engineering*. Berlin: Springer, 1999
[22] M. Joham, K. Kusume, M. H. Gzara, W. Utschick, and J. A.Nossek, "Transmit Wiener Filter for the Downlink of TDD DS-CDMA Systems," in *Proc. ISSSTA 2002*, vol. 1, pp. 9-13,September 2002.
[23] U. Madhow and M. Honig, "MMSE interference suppression for direct sequence spread spectrum CDMA", *IEEE Trans. Commun.*, vol. 42, pp. 3178-3188, Dec. 1994.