Trellis Precoding for MIMO Broadcast Signaling

Aaron Callard †, Amir Khandani †, Aladdin Saleh ‡
† University of Waterloo, Waterloo, ON, Canada, {aaron, khandani}@uwaterloo.ca
‡Bell Canada, Mississauga, ON, Canada, aladdin.saleh@bell.ca

***Abstract*— Channel Inversion, and its Minimum Mean Square Error (MMSE) variation, are low complexity methods for Space Division Multiple Access (SDMA) in Multiple Input Multiple Output Broadcast Channel (MIMO-BC). As the channel matrix deviates from orthogonal, these methods result in a waste of transmit power. This paper proposes a trellis precoding method (across time and space) to improve the power efficiency. Adopting a 4-state trellis shaping method from [1], the complexity of the proposed method, which is entirely at the transmitter side, is equivalent to the search in a trellis with 4^N states where N is the number of transmit antennas. Numerical results are presented showing that the achievable gains, which depend on the channel realization, can be significantly higher than the traditional shaping gain which is limited to 1.53dB.**

Keywords: MIMO, Trellis Shaping, Multi-User, MMSE, Broadcast Channel, Dirty Paper Coding, Vector Precoding

I. Introduction

Information theoretic results show that the capacity of a point-to-point wireless fading channel scales by $\min(N_t, N_r)$ where N_t and N_r are the number of transmit and receive antennas, respectively [2]. In practice, data is usually transmitted to several resource limited mobile units, each with a small number of antennas (MIMO-BC). Provided that the transmitter (with N_t antennas) knows the channel, through precoding one can still achieve a scaling of $\min(N_t, N_r)$ for the sum rate where N_r is the total number of receive antennas [3]. In this paper, we assume $N_t = N_r = N$ and consider only single antenna users; however, the methods presented can be easily generalized to include multi-antenna receivers.

The most widely known method for precoding over MIMO-BC is based on adding a vector modulo operation (precoding across space) at the transmitter/receiver sides. This results in many different symbols to represent the same message [4]. The transmitter finds a symbol among possible choices which has a small energy, but is still, after the modulo operation, equivalent to the original data. In the context of canceling known interference, it is known that a similar modulo operation can be performed with respect to the voronoi region of a large dimensional shaping lattice (precoding across time) [5] [6] [7]. By utilizing the trellis nature of the shaping code, this minimization can be performed using Viterbi algorithm. In view of these earlier works, the current article proposes a method for joint precoding across both space and time using a trellis structure. To do this, each user is equipped with trellis shaping across time, and the trellis representations of different users are combined into a super trellis[1] to facilitate spatial precoding. The branches in the super trellis correspond to vectors of spatial dimensions across different transmit antennas. To calculate the branch metrics, we examine two cases: For Zero Forcing (ZF), we allow no interference among users and branch metrics are the transmit energy. In the MMSE case, we allow for interference among users and the branch metrics reflect the MMSE criterion. The amount of gain depends on the channel realization, and, in the MMSE case, on the power level (the lower the power, the greater the gain).

Section II is devoted to problem statement. Section III reviews the concept of sign bit shaping, the simplest form of trellis shaping, and develops the super trellis to perform precoding across time and space. In Section IV, we combine the process of vector precoding with trellis shaping. In Section V, we examine the performance of the proposed method, and finally in Section VI, we provide our conclusions.

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. System Model and Problem Formulation

This paper deals with channels which can be represented by

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

where $\mathbf{y}$ is a column vector representing the received signals, $\mathbf{s}$ represents the transmitted signals subject to a power constraint of $\overline{P}$, $\mathbf{n}$ represents complex 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. Without loss of generality, (1) can be written as

$$\begin{aligned} \mathbf{y} &= \mathbf{HH}^{-1}\frac{\sqrt{P}}{\gamma}(\mathbf{u}+\boldsymbol{\epsilon}) + \mathbf{n} \text{ where } \gamma = \|\mathbf{H}^{-1}(\mathbf{u}+\boldsymbol{\epsilon})\| \\ &= \frac{\sqrt{P}}{\gamma}(\mathbf{u}+\boldsymbol{\epsilon}) + \mathbf{n} \\ s.t. &\quad E\{P\} \leq \overline{P}, \end{aligned}$$

where $\mathbf{u}$ and $\boldsymbol{\epsilon}$ represent message and interference terms, respectively. 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 that $c = \frac{\sqrt{P}}{\gamma}$ to remain

This work is funded in part by the Bell University Research Laboratory (BUL) of Bell Canada, and in part by the Natural Sciences and Engineering Research Council of Canada. A preliminary version of this paper was presented in the Conference on Information Sciences and Systems 2006 (CISS2006)

[1] States in the super trellis correspond to all combinations of states in the individual users' trellis diagrams.

constant [4] which is the assumption hereafter. Assuming the power constraint is satisfied with equality, we obtain

$$\mathbf{y} = c(\mathbf{u} + \boldsymbol{\epsilon}) + \mathbf{n} \text{ where } c = \sqrt{\overline{P}/E\{\gamma^2\}}.$$

As we will see later, this model simplifies the derivation of the Minimum Mean Square Error (MMSE) solution.

III. Sign Bit Shaping for Multiple Users

A common method for shaping is based on using the voronoi region around the origin of a lattice constructed using a binary code $\mathbf{C}_{shaping}$ [8], and in specific, a convolutional code represented by a trellis diagram (Trellis Shaping) [1]. Sign bit shaping is the simplest form of trellis shaping in which only the first 2 bits of the transmitted constellation points (sign of the real and imaginary parts using Grey mapping) are modified by (modulo 2) addition of a codeword $\{\mathbf{c}_i\} \in \mathbf{C}_{shaping}$. Information can therefore be stored in the syndrome of the shaping code $\mathbf{C}_{shaping}$ and recovered using 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 [6], [1]. We make the observation that the loss can be avoided if the least significant bits are determined first, in which case the syndrome bits are better protected. Experiments show that using this strategy, the syndrome bits do not need to be channel coded at all.

Next we extend the concept of sign bit shaping to MIMO-BC. This involves two steps, first we design a trellis shaping code $\mathbf{C}_s$ for which users can decode their syndromes separately; and second, we derive an objective function to be used as trellis metric. The shaping code $\mathbf{C}_s$ is formed by combining N trellis shaping sub-codes $\mathbf{C}_{shaping}$. If $\mathbf{C}_{shaping}$ has generator matrix $\mathbf{G}$, then $\mathbf{C}_s$ has a generator matrix $\mathbf{G}_T$ which is block diagonal with $\mathbf{G}$ repeated N times along its diagonal. From $\mathbf{G}_T$, we can develop the trellis of $\mathbf{C}_s$ where the sign bits are calculated for each user and added to the part of the chosen shaping codeword $\{\mathbf{c}_i\}$ for that user. Next, we develop the branch metrics.

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 select $\{\mathbf{c}_i\}$ to maximize the scaling factor c.

$$\begin{aligned} \{\mathbf{c}_i\} &= \underset{\{\mathbf{c}_i\}\in\mathbf{C}_{shaping}}{\operatorname{argmax}} \frac{\overline{P}}{E\{\gamma^2\}} & (2)\\ &= \underset{\{\mathbf{c}_i\}\in\mathbf{C}_{shaping}}{\operatorname{argmin}} E\{\gamma^2\} & (3)\\ &= \underset{\{\mathbf{c}_i\}\in\mathbf{C}_{shaping}}{\operatorname{argmin}} E\{\|\mathbf{H}^{-1}\mathbf{u}\|^2\}. & (4)\end{aligned}$$

If we set the branch metric to be $\|\mathbf{H}^{-1}\mathbf{u}\|^2$, we can minimize the transmit energy (3) by finding the minimum metric path in the trellis of $\mathbf{C}_s$.

If interference is allowed among users, significant gains in the Signal to Interference plus Noise Ratio (SINR) can be achieved at low to medium power levels [10] [11][2]. Let us define our new objective function to be the MSE of the received signal, i.e.

$$\begin{aligned} MSE &= E\left\{\left\|\frac{\mathbf{y}}{c} - \mathbf{u}\right\|^2\right\} & (5)\\ \mathbf{y} &= c(\mathbf{u} + \boldsymbol{\epsilon}) + \mathbf{n}. & (6)\end{aligned}$$

This gives us two variables for minimizing MSE, namely $\boldsymbol{\epsilon}$ and $\mathbf{u}$. First, MSE is minimized with respect to $\boldsymbol{\epsilon}$ as shown in [10] [11] resulting in

$$\begin{aligned} \mathbf{y} &= c\mathbf{H}\mathbf{H}^H\left(\mathbf{H}\mathbf{H}^H + \alpha\mathbf{I}\right)^{-1}\mathbf{u} + \mathbf{n}\\ \alpha &= \frac{N\sigma^2}{\overline{P}}. & (7)\end{aligned}$$

Next, to select $\mathbf{u}$, the branch metrics are computed using (6) and (7), however there exists a simplification [11] as:

$$MSE = \alpha\left\|\sqrt{\left(\mathbf{H}\mathbf{H}^H + \alpha\mathbf{I}\right)^{-H}}\mathbf{u}\right\|^2. \quad (8)$$

IV. Combining Vector Precoding and Trellis Precoding

One benefit of considering all users simultaneously is that an integer vector $\mathbf{p}$ multiplied by some constant τ can be added to the transmitted signal without affecting the data [4] [10] [11]. This is compensated by applying another vector modulo operation at the receiver side. Given data vector $\mathbf{u}$, we have

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

where $\mathbf{p}$ is selected to minimize MSE, i.e.

$$\hat{\mathbf{p}} = \underset{\mathbf{p}}{\operatorname{argmin}}\left\|\sqrt{\left(\mathbf{H}\mathbf{H}^H + \alpha\mathbf{I}\right)^{-H}}(\mathbf{u} + \tau\mathbf{p})\right\|^2. \quad (9)$$

This minimization can be performed, for example, using a sphere encoder. The final branch metric is

$$MSE = \alpha\left\|\sqrt{\left(\mathbf{H}\mathbf{H}^H + \alpha\mathbf{I}\right)^{-H}}(\mathbf{u} + \tau\hat{\mathbf{p}})\right\|^2. \quad (10)$$

V. Performance Results

In this section, we compare the proposed method with the MMSE vector precoding method of [10] [11]. The elements of $\mathbf{H}$ are simulated as independent unit mean complex Gaussian variables and $\mathbf{H}$ remains constant throughout a block of length 100 to permit accurate energy calculation (reducing the influence of trellis boundary/termination). Figure 1 shows the SINR comparisons for N=4, M=64. For a fair comparison, vector precoding is given a 3 dB energy offset to account for the extra bit it transmits. Note that such conversion between rate and energy is commonly used in shaping literature (see [1] and references therein).

To compare with capacity bounds, following [4], we have simulated the method for $N = 4$, $M = 64$ over a fast fading channel using Turbo-codes[3]. Less significant bits are

[2]We became aware of [10] after [11] was published.

[3]Note that although assumption of fast fading is unrealistic for MIMO-BC, a similar setup is used in [4] to make it possible to provide comparisons with the Ergodic capacity limits.

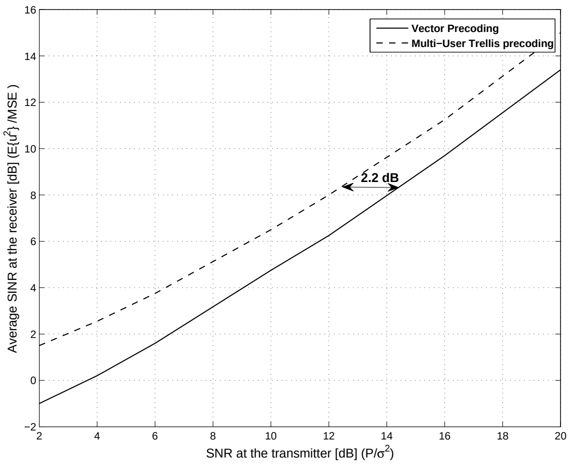

Fig. 1. SINR for Multi-User Trellis Precoding and Vector Precoding (N=4, M=64).

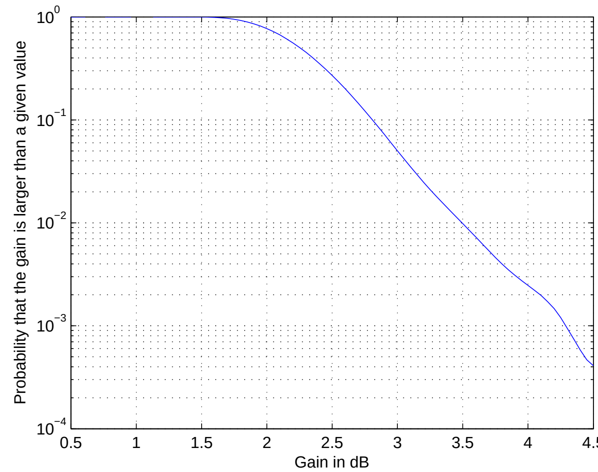

Fig. 2. Probability that the gain due to the application of the proposed method vs. conventional MMSE vector precoding is larger than a given value for N=4, M=64 (effective rate is 3 bits/user).

encoded using a rate 1/2 Turbo-code (resulting in a data rate of 3 bits/user) with feedback and feed-forward polynomials $1 + D^2 + D^3$ and $1 + D + D^3$, respectively, over a block-length of 4000 where 10000 such blocks are simulated. This method achieves a BER of 10^{-5} at only 1 dB away from the Ergodic capacity limit. A Turbo-code setup with similar parameters is used in [4] showing significantly larger gap to the Ergodic capacity.

In a MIMO-BC with fixed rates, the amount of transmit energy for a given message depends on the channel realization and can have significant fluctuations around its average value. This complicates the design of the power amplifiers. To study this effect, we have plotted the probability that the gains due to the proposed method is larger than a given value. As the amount of gain depends on the transmit energy, first a suitable power level must be found. To do this, we calculate the minimum power required for the channel capacity region [12] (corresponding to any given channel realization) to contain the corresponding rate vector. Using $M = 64$, we have assumed that the shaping bits are uncoded and the rest of bits are encoded using a rate $1/2$ Turbo-code, resulting in an effective rate of 3 bits/user. Using this power level for the proposed method, we compute the resulting SINR and perform an exhaustive search over the power level required in MMSE vector precoding [10] [11] such that the difference in the SINR values of the two methods are negligible. The result of such computation is shown in figure 2.

VI. Conclusion

It is shown that by applying the precoding across both time (trellis shaping [1], or more generally voronoi constellation [9]) and space (vector precoding), one can improve the MIMO-BC system in terms of both average transmit energy and its fluctuations. The improvement in the performance is achieved at the cost of an increase in the complexity. As the shaping code $\mathbf{C}_s$ is composed of N shaping sub-codes, the number of states in its trellis grows exponentially with N. For the example of sign bit shaping based on 4-states sub-codes, the complexity is equivalent to Viterbi decoding of a trellis with 2^{2N} states. In practical systems, the number of transmit antennas, N, is limited, resulting in an overall manageable complexity for the proposed method. There also exist many sup-optimal decoding techniques which could be applied [13] if the trellis becomes too large. Note that unlike the case of channel decoding where selecting a sub-optimum trellis path with a slightly higher metric can result in a significant increase in the error probability, in precoding, the loss in performance (increase in transmit energy) due to such slight sub-optimality is usually negligible.

References

[1] G. Forney, "'Trellis shaping,'" *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 281-300, March 1992.

[2] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommunications*, vol. 10, pp. 585-595, Nov/Dec 1999.

[3] 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.

[4] B.M. Hochwald, C.B. Peel and A.L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part II: perturbation," *IEEE Transactions on Communications*, vol. 53, no. 3, pp. 537-544, March 2005.

[5] R. Zamir, S. Shamai, U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1250-1276, June 2002.

[6] W. Yu, D. P. Varodayan, J. M. Cioffi "Trellis and Convolutional Precoding for Transmitter-Based Interference Pre-Subtraction", *IEEE Transactions on Communications*, vol. 53, no. 7, pp. 1220-1230, July 2005.

[7] U. Erez, S. Shamai, R. Zamir, "Capacity and lattice strategies for canceling known interference," *IEEE Transactions on Information Theory*, vol. 51, no. 11, pp.3820-3833, Nov. 2005.

[8] G. D. Forney, Jr., "Multidimensional constellations, Part II: voronoi constellations," *IEEE Journal on Selected Areas in Communications*, vol. 7, pp. 941-958, Aug. 1989.

[9] G. D. Forney Jr. and L.-F. Wei, "Multidimensional constellations, Part I: Introduction, figures of merit, and generalized cross constellations *IEEE Journal on Selected Areas in Communications*, vol. 7, no. 6, pp. 877-892, Aug. 1989.

[10] D. Schmidt, M. Joham, and W. Utschick, "Minimum Mean Square Error Vector Precoding," Proc. *16th IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC'05)*, Berlin, Germany, September 11th-14th 2005

[11] 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

[12] S. Vishwanath, N. Jindal and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels" *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2658-2668, Oct. 2003.

[13] G.J. Pottie and D.P. Taylor, "A comparison of reduced complexity decoding algorithms for trellis codes," *IEEE Journal on Selected Areas in Communications*, vol. 7, no. 9, pp. 1369-1380, Dec. 1989.