Multi-Mode Precoding Using Linear Receivers for Limited Feedback MIMO Systems

David J. Love† and Robert W. Heath Jr.
Dept. of Electrical and Computer Engineering
The University of Texas at Austin
Austin, TX 78712
{djlove, rheath}@ece.utexas.edu

Abstract—Multiple-input multiple-output (MIMO) wireless systems obtain large diversity and capacity gains by employing multi-element antenna arrays at both transmitter and receiver. The theoretical performance benefits, however, are irrelevant unless low error rate, high spectral efficiency spatio-temporal signaling techniques are found. Most work in space-time coding concentrates on either designing low error rate codes or high data-rate codes but not both simultaneously. This paper proposes a new method for designing high data-rate spatio-temporal signals with low error rates. The basic idea is to use transmitter channel information in the form of limited feedback to adaptively vary the transmission scheme for a fixed data-rate. This adaptation is done by varying the number of substreams and the rate of each substream in a precoded spatial multiplexing system. We show how this method can be implemented in a limited feedback scenario where only finite sets, or codebooks, of possible precoding configurations are known to both the transmitter and receiver. Monte Carlo simulations show substantial performance gains over beamforming and spatial multiplexing.

I. INTRODUCTION

Spatial multiplexing is a simple MIMO signaling approach that achieves large spectral efficiencies with only moderate transmitter complexity. Receivers for spatial multiplexing range from the high complexity, low error rate maximum likelihood decoding to the low complexity, moderate error rate linear receivers [1]. Unfortunately, rank deficiency of the channel means can cause all spatial multiplexing receivers to suffer increases in the probability of error [2]–[4].

Linear transmit precoding, where the transmitted data vector is premultiplied by a precoding matrix that is adapted to some form of channel information, adds resilience against channel ill-conditioning. Linear precoded spatial multiplexing has been proposed for transmitter’s with full channel knowledge, channel first-order statistics, channel second-order statistics, or limited feedback from the receiver. Optimization techniques for choosing the precoding matrix include minimizing the mean squared error [5], [6], maximizing the minimum distance [3], maximizing the minimum singular value [3], and maximizing the mutual information assuming Gaussian signaling [4], [7]. These precoding methods provide probability of error improvements compared to unprecoded spatial multiplexing, but linear precoding does not guarantee full diversity order.

This problem was addressed for the special case of antenna subset selection precoding in [8]–[11]. These papers studied systems where the size of the antenna subset, along with the spatial multiplexing constellation, could be varied in order to guarantee full diversity performance for a fixed data-rate. Various selection criteria were proposed for both dual-mode (i.e. selecting between spatial multiplexing and selection diversity) and multi-mode antenna selection. These methods provided substantial performance improvements compared with traditional spatial multiplexing. Antenna subset selection, however, is quite limited because of the restriction of the precoding matrices to columns of the identity matrix.

In this paper, we propose a modified version of linear precoding called multi-mode precoding. Multi-mode precoding varies the number of substreams contained in the precoded spatial multiplexing vector assuming a fixed data-rate. Because it is often impractical to assume perfect channel information at the transmitter, we employ a limited feedback similar to that developed in [12]–[15]. Limited feedback multi-mode precoders use precoder codebooks, finite sets of precoder matrices, for each of the supported substream configurations. The codebooks are designed offline and stored at both the transmitter and receiver. The chosen multi-mode precoder is then conveyed from the receiver to transmitter over a limited capacity feedback channel using a small number of bits. We present criteria for selecting the mode and precoder from the codebook. We also address codebook design for our limited feedback precoders.

This paper is organized as follows. Section II reviews the precoded spatial multiplexing system model. Criteria for choosing the optimal matrix from the codebook is presented in Section III. Design criteria for creation of the precoder codebook are derived in Section IV. Monte Carlo simulations results are presented in Section V. We conclude in Section VI.

II. SYSTEM OVERVIEW

Consider an $M_t$ transmit and $M_r$ receive antenna MIMO wireless system where $R$ bits (with $R$ fixed) are demultiplexed
into M different bit streams at each channel use. The bit streams are modulated using the same constellation, S, to produce a vector s_k at the kth channel use. The spatial multiplexing symbol vector s_k = [s_{k,1}, s_{k,2}, \ldots, s_{k,M}]^T is assumed to have power constraints so that E_{s_k} [s_k s_k^*] = \frac{E_s}{M} I_M. This constraint implies the average of the total transmitted power is independent of the number of substreams M.

An M_k \times M linear precoding matrix F_M maps s_k to an M_k-dimensional spatio-temporal signal that is transmitted on M_t transmit antennas. The transmitted signal vector encounters an M_r \times M_t flat-fading matrix channel H before being added with an M_r-dimensional white Gaussian noise vector v_k. Assuming perfect pulse-shaping, sampling, and timing, this formulation yields an input/output relationship

\[ y = H F_M s + v \]

where the channel use index k has been suppressed because we are interested in vector-by-vector detection of s_k. We assume that H has independent and identically distributed (i.i.d.) entries with each distributed according to CN(0,1). As well, the noise vector v is assumed to have i.i.d. entries distributed according to CN(0, E_v).

The matrix HF_M can be thought of as an effective channel. The receiver decodes y using this effective channel and a linear receiver. It is assumed that the receiver has perfect knowledge of H. The receiver applies an M \times M_r matrix transformation G to y and then independently detects each entry of Gy. If a zero-forcing (ZF) decoder is used, G = (HF_M)^+\cdot. Minimum mean squared error (MMSE) decoding uses G = [F_M^* H^* H F_M + (N_0 M / E_v) I_M]^{-1} F_M^* H^*.

Note that the total transmitted power for this system is given by \|F_M^* F_M s\|^2. The precoder matrix F_M must therefore be constrained to limit the transmitted power. We restrict \|F_M^* F_M\| \leq 1 in order to limit the peak-to-average ratio. This means that E_{s_k} [s_k F_M F_M^* s_k] \leq E_s regardless of the modulation scheme or the value of M. It was shown in [6], [13], [14] that matrices of this form that optimize mean square error, capacity, and total channel power are all members of the set U(M_t, M). For this reason, we will further restrict that F_M \in U(M_t, M) for any chosen value of M.

The difference between multi-mode precoding and previously proposed linear precoders is that M is adapted using current channel conditions. We refer to the value of M as the mode of the precoder. Usually, only a subset of the M possible modes can be chosen. Examples of why only subsets of \{1, \ldots, M_t\} might be used include that i.) \frac{R}{M} is an integer for only a few of the integers between 1 and M or ii.) system architecture only supports a subset. We will denote the set of supported mode values as \mathcal{M}. For example, if R = 8 bits and M_t = 4 then only modes in the set \mathcal{M} = \{1, 2, 4\}

1We use T for transpose, \* for conjugate transpose, \# for the matrix pseudo-inverse, I_M for the M \times M identity matrix, \lambda_i(A) to denote the \ith largest singular value of the matrix A, E_v to denote expectation with respect to random variable s, and U(M, M) := \{ U \in \mathbb{C}^{M, M} \mid U^* U = I_M \}.

can be supported. Another example might be a dual-mode selection where \mathcal{M} = \{1, M_t\}. We assume that a selection function g : \mathbb{C}^{M_t \times M_t} \rightarrow \mathcal{M}. Thus, M = g(H).

The transmitter in this system model has no form of channel knowledge. This assumption is common for frequency division duplexing (FDD) systems where the separation between the forward and reverse link frequency bands is much larger than the coherence bandwidth. We overcome this lack of transmitter channel knowledge by using a low-rate feedback channel that can carry a limited number, denoted by B, of information bits from the receiver to the transmitter. The precoder F_M is chosen from a finite set, or codebook, of N_M different M_t \times M precoder matrices \mathcal{F}_M = \{F_{M,1}, F_{M,2}, \ldots, F_{M,N_M}\}. The precoder F_M is chosen from \mathcal{F}_M using a selection function f_M : \mathbb{C}^{M_t \times M_t} \rightarrow \mathcal{F}_M. Therefore, F_M = f_M(H).

Thus, we assume that there is a codebook for each supported mode value. Because there are a total of N_{total} = \sum_{m \in \mathcal{M}} N_m, a total feedback of B = \lceil \log_2(N_{total}) \rceil bits is required. Feedback can thus be kept to a reasonable amount by varying the size of N_M for each mode.

III. MULTI-MODE PRECODER SELECTION

The selection of the mode and precoder matrix will determine the performance of the entire system. As stated earlier, we will consider decoders using ZF and MMSE linear receivers. We will present probability of symbol vector error (i.e. the probability that at least one symbol is in error) and capacity selection. The probability of error selection is based on the previous work in [3], [9]–[11] while the capacity selection is similar to work presented in [3], [7].

**Probability of Error Selection**

Assuming a probability of error selection criterion, the optimal selection criterion would obviously be to choose the mode and precoder that provide the lowest probability of symbol vector error. Selection using this criterion, however, is unrealistic because closed-form expressions for the probability of symbol vector error conditioned on a channel realization are not available to the authors’ knowledge. The Nearest Neighbor Union Bound (NNUB) can be successfully employed in place of this bound for asymptotically tight selection. The NNUB for an M_r \times M effective channel is given by

\[ P_e^{\text{NNUB}}(HF_M, M) = 1 - \left( 1 - N_c(M, R) Q \left( \sqrt{SNR_{\text{min}}(F_M)} \frac{d_{\text{min}}(M, R)}{2} \right) \right)^M \]

where SNR_{\text{min}}(F_M) is the minimum receive substream SNR after linear equalization (expressions can be found in [3]) and d_{\text{min}}(M, R) is the minimum distance of the normalized rate R/M constellation. Using the NNUB result, the following selection criterion is obtained.
NNUB Selection Criterion: Choose $M$ and $F_M$ such that

$$f_M(H) = \arg \min_{F' \in F_M} P_{e, \text{NNUB}}(HF', M)$$

$$g(H) = \arg \min_{M' \in M} P_{e, \text{NNUB}}(Hf_{M'}(H), M').$$

(2)

The cost function can be implemented using a brute force search because there are only a finite number of matrices per codebook. The values of $N_e(M, R)$ and $d_{\text{min}}(M, R)$ can be computed offline and stored in the receiver’s memory.

Unfortunately, the Q-function is often problematic to implement. This issue can be overcome by further simplifying the cost function evaluated in (2). The Q-function in (2) numerically dominates the constant terms because it decreases exponentially. For this reason, it is often beneficial to use the following approximation selection criterion.

SNR Selection Criterion: Choose $M$ and $F_M$ such that

$$f_M(H) = \arg \max_{F' \in F_M} SNR_{\text{min}}(F')$$

$$g(H) = \arg \max_{M' \in M} SNR_{\text{min}}(f_{M'}(H))d_{\text{min}}^2(M', R).$$

(3)

The SNR criterion represents a large complexity cost savings compared to the NNUB criterion because of the removal of the Q-function. This criterion can be easily implemented as long as the receiver and SNR are known. Situations could arise where a selection criterion that is independent of the kind of linear receiver and the SNR is preferable.

Note that $HF_M$ is at most a rank $M$ matrix. Thus for each value of $M$

$$\max_{F_M \in F_M} SNR_{\text{min}}(F') \leq \frac{E_s}{MN_0} \max_{F_M \in F_M} \lambda^2_M\{HF'_M\}$$

(4)

by the Poincare Separation Theorem [16, pp. 190]. A selection criterion based on selecting a mode only if the best achievable SNR provides approximately minimizes the NNUB. This corresponds to the following selection criterion.

Singular Value (SV) Selection Criterion: Choose $M$ and $F'_M$ such that

$$f_M(H) = \arg \max_{F' \in F_M} \lambda^2_M\{HF'_M\}$$

$$g(H) = \arg \max_{M' \in M} \lambda^2_{M'}\{H\}d_{\text{min}}^2(M', R).$$

(5)

Despite the fact that this bound is approximate, the probability of error performance was shown in [9] to be quite close to both the SNR selection and NNUB selection. The important advantage of this criterion is that the selection of $M$ and $F_M$ are decoupled. The computation of the smallest singular value of $HF'_M$ for all codebook matrices $F'_M$ must only be completed for one value of $M$ rather than for every value in $M$.

Capacity Selection

While capacity selection is not optimal from a probability of error point-of-view, it can provide insight into the attainable spectral efficiencies given the multi-mode precoding system model. The uniformed transmitter capacity (the maximum mutual information given uncorrelated, Gaussian signaling) is given by

$$C_{UT}(F_M) = \log_2 \det \left( I_M + \frac{E_s}{MN_0} F^*_M H^* HF_M \right).$$

The following criterion can be succinctly stated based on maximizing the effective channel capacity.

Capacity Selection Criterion: Choose $M$ and $F'_M$ such that

$$f_M(H) = \arg \max_{F' \in F_M} C_{UT}(F_M)$$

$$g(H) = \arg \max_{M' \in M} C_{UT}(f_{M'}(H)).$$

(7)

IV. MULTI-MODE CODEBOOK DESIGN

The next major issue that must be addressed is the distribution of $N_{\text{total}}$ matrices among $N_1, \ldots, N_M$, and the design of the codebooks $F_1, \ldots, F_M$. Because the selection chooses $M$ and then chooses $F_M$, we can decouple the codebook design into the design of $\text{card}(M)$ independent codebooks.

A. Distribution of Bits

The feedback amount $B$ is often specified offline by general system design constraints. For example, only $B$ bits of control overhead might be available in the reverse-link frames. For this reason, we will assume that $B$ is a fixed system parameter. Thus we wish to understand how to distribute $B$ bits (or equivalently $N_{\text{tot}} = 2^B$ matrices) among the $\text{card}(M)$ modes.

Before proceeding, note that the codebook $F_{M_t}$ corresponds to choosing traditional spatial multiplexing. Thus $I_{M_t} \in F_{M_t}$. The identity matrix, however, is the only matrix in $F_{M_t}$. The reason for this is that if there existed $F' \in F_{M_t}$ with $F' \neq I_{M_t}$, then the performance of precoded spatial multiplexing using $F'$ over Rayleigh fading would exactly match the performance of traditional spatial multiplexing. This follows easily from the fact that $HF'$ has the same distribution as $H$ when $H$ is an uncorrelated, Rayleigh fading matrix [17]. Therefore $F_{M_t} = \{I_{M_t}\}$. Thus there is no reason to optimize the bit disbursement while including $F_{M_t}$. Therefore, the problem simplifies to figuring out how to assign $N_{\text{tot}} - 1$ matrices over $\text{card}(M) - 1$ modes.

The first step in assigning these bits is the determination of a distortion function $D(\cdot)$. Because the NNUB, SNR, and SV selection all relate directly or indirectly to maximizing $\lambda^2_M\{HF_M\}$, we will define the distortion conditioned to be

$$D(F_M, M) = d_{\text{min}}^2(M, R) \left| \lambda^2_M\{HF_{\text{sv, opt}}\} - \lambda^2_M\{HF_M\} \right|^2$$

where $F_{\text{sv, opt}}$ maximizes the minimum singular value of the effective channel over $d(M, M)$.

The capacity selection will use the conditional distortion function given by

$$D(F_M, M) = \left| \log_2 \det \left( I_M + \frac{E_s}{MN_0} F^*_{c,\text{opt}} H^* HF_{c,\text{opt}} \right) - \log_2 \det \left( I_M + \frac{E_s}{MN_0} F^*_M H^* HF_M \right) \right|^2$$

$$\left| I_M + \frac{E_s}{MN_0} F^*_M H^* HF_M \right|$$

Where $F_{c,\text{opt}}$ maximizes the minimum singular value of the effective channel over $d(M, M)$.

The feedback amount $B$ is often specified offline by general system design constraints. For example, only $B$ bits of control overhead might be available in the reverse-link frames. For this reason, we will assume that $B$ is a fixed system parameter. Thus we wish to understand how to distribute $B$ bits (or equivalently $N_{\text{tot}} = 2^B$ matrices) among the $\text{card}(M)$ modes.

Before proceeding, note that the codebook $F_{M_t}$ corresponds to choosing traditional spatial multiplexing. Thus $I_{M_t} \in F_{M_t}$. The identity matrix, however, is the only matrix in $F_{M_t}$. The reason for this is that if there existed $F' \in F_{M_t}$ with $F' \neq I_{M_t}$, then the performance of precoded spatial multiplexing using $F'$ over Rayleigh fading would exactly match the performance of traditional spatial multiplexing. This follows easily from the fact that $HF'$ has the same distribution as $H$ when $H$ is an uncorrelated, Rayleigh fading matrix [17]. Therefore $F_{M_t} = \{I_{M_t}\}$. Thus there is no reason to optimize the bit disbursement while including $F_{M_t}$. Therefore, the problem simplifies to figuring out how to assign $N_{\text{tot}} - 1$ matrices over $\text{card}(M) - 1$ modes.

The first step in assigning these bits is the determination of a distortion function $D(\cdot)$. Because the NNUB, SNR, and SV selection all relate directly or indirectly to maximizing $\lambda^2_M\{HF_M\}$, we will define the distortion conditioned to be

$$D(F_M, M) = d_{\text{min}}^2(M, R) \left| \lambda^2_M\{HF_{\text{sv, opt}}\} - \lambda^2_M\{HF_M\} \right|^2$$

where $F_{\text{sv, opt}}$ maximizes the minimum singular value of the effective channel over $d(M, M)$.

The capacity selection will use the conditional distortion function given by

$$D(F_M, M) = \left| \log_2 \det \left( I_M + \frac{E_s}{MN_0} F^*_{c,\text{opt}} H^* HF_{c,\text{opt}} \right) - \log_2 \det \left( I_M + \frac{E_s}{MN_0} F^*_M H^* HF_M \right) \right|^2$$

Where $F_{c,\text{opt}}$ maximizes the minimum singular value of the effective channel over $d(M, M)$.

Where $F_{c,\text{opt}}$ maximizes the minimum singular value of the effective channel over $d(M, M)$. For this reason, we will assume that $B$ is a fixed system parameter. Thus we wish to understand how to distribute $B$ bits (or equivalently $N_{\text{tot}} = 2^B$ matrices) among the $\text{card}(M)$ modes.
where $\mathbf{F}_{opt}$ maximizes the capacity of the effective channel over $\mathcal{U}(M_1, M)$. We wish to minimize the average distortion

$$D = E_{\mathbf{M}} \left[ E_{\mathbf{H}} \left[ D(f_M(\mathbf{H}), \mathbf{M}) \mid \mathbf{M} \right] \right]$$

(8)

using our chosen distortion. The random variable $\mathbf{M}$ has a distribution obtained from the conditional probabilities given that the mode can take only the spatial multiplex mode using the approximate SV selection criterion (i.e. $Pr^{(SV)}(M = j) = Pr^{(SV)}(M = j \mid M \neq M_i)$). Because the SV selection criterion is decoupled from the codebook design, it can be easily employed numerically to evaluate these conditional probabilities.

We will use a uniform, high-resolution approximation for $E_{\mathbf{H}} \left[ D(f_M(\mathbf{H}), \mathbf{M}) \mid \mathbf{M} \right]$ [18, p. 163]. This assumes that

$$E_{\mathbf{H}} \left[ D(f_M(\mathbf{H}), \mathbf{M}) \mid \mathbf{M} \right] \approx K/N_M^2$$

(9)

where $K$ is a positive real constant. While this approximation is very rough, it is sufficient enough for us to roughly allocate the bits among the $\text{card}(\mathbf{M}) = \text{card}(\mathbf{M}) - 1$ supported levels other than pure spatial multiplexing and removes the dependence of $D$ on the form of the selection function $f_M(\cdot)$.

One solution to this codebook allocation problem, as presented in [18, p. 231], can then be formulated as

$$b_j = B - \left[ \log_2(\text{card}(\mathbf{M})) \right] + \frac{1}{2} \log_2 \frac{Pr^{(SV)}(M = j)}{P}$$

where $P$ is the geometric mean of $\{Pr^{(SV)}(M = j)\}_{j \in \mathbf{M}}$. Given this allocation scheme, $N_j$ can then be computed as

$$N_j = \text{round}(2^{b_j})$$

This constraint will solve the bit allocation problem but returns real numbers [18]. Since we are only concerned with the total bit allocation $B$ rather than the individual codebook sizes, we will directly compute $N_j$ using the computed $b_j$ for each $j \in \mathbf{M}$.

It is probable that $\sum_{j \in \mathbf{M}} N_j$ may not equal $N_{tot} - 1$. If $\sum_{j \in \mathbf{M}} N_j > N_{tot} - 1$, an ad hoc approach is to remove the extra matrices from the least likely modes. More than likely, the algorithm will return $\sum_{j \in \mathbf{M}} N_j > N_{tot} - 1$. The remaining codebook allocation slots can be distributed to the least most likely modes.

B. Codebook Criterion Given the Number of Substreams

Now that we have determined an algorithm that gives an ad hoc allocation of the possible codebook matrices among the modes, it is now imperative to present the designs of $\mathcal{F}_M$ for each mode.

There are two special cases of the multi-mode precoder codebooks. The column vectors in $\mathcal{F}_1$ correspond to beamforming vectors [12]. The design of limited feedback beamforming was explored in [12], [19]. In particular, it was shown in [12], [19] that the set of vectors should be designed by thinking of the vectors as representing lines in $\mathbb{C}^M$. The lines can then be optimally spaced by maximizing the minimum angular separation between any two lines. The set $\mathcal{F}_M$ is trivially designed because we will require that $\mathcal{F}_M = \{I_M\}$. This precoder matrix corresponds to sending a standard spatial multiplexing vector.

Codebook design for limited feedback precoding assuming a fixed number of substreams was studied in [13]–[15], [20] for a variety of selection functions. The SNR selection criterion, which uses the cost function $SNR_{min}(\mathbf{F}_M)$, can be approximately maximized by maximizing $\lambda_M(\mathbf{H}_M)$. It was shown in [20] that $\mathcal{F}_M$ should be designed by maximizing the minimum projection two-norm subspace distance between any two matrices in $\mathcal{F}_M$. The capacity selection criterion, however, motivates maximizing the minimum Fubini-Study subspace distance between any two codebook subspaces. These codebooks can easily be designed using the matrix codebooks designed for non-coherent constellations in [21] with modified distance functions.

V. Simulations

Limited feedback multi-mode precoding was simulated to exhibit the available minimization in the probability of error and the increase in capacity. The capacity results are compared with the results in [22].

Experiment 1: This experiment addresses the probability of symbol vector error of $4 \times 4$ multi-mode precoding with a ZF receiver. The results are shown in Fig. 1. The rate is fixed at $R = 8$ bits per channel use with QAM constellations. The set of supported modes is $\mathcal{M} = \{1, 4\}$. Four bits of feedback is assumed to be available. Limited feedback beamforming using four bits (see [12], [19]) and spatial multiplexing are simulated for comparison. Multi-mode precoding with the SV selection criterion provides approximately a $0.25$ dB performance improvement over limited feedback beamforming. The SNR and NNUB selection criteria perform approximately the same, $0.55$ dB better than the SV selection criterion.
Experiment 2: The capacity gains available with the capacity selection criterion are illustrated in Fig. 2 for a $2 \times 2$ MIMO system. The plot shows the ratio of the computed capacity with the capacity of a transmitter with perfect channel knowledge. The capacity of an uninformed transmitter (UT) and the limited feedback covariance optimization capacity results published in [22] are shown for comparison. Note that limited feedback precoding outperforms limited feedback covariance optimization for both two and three bits of feedback. This result is striking because, unlike covariance optimization, multi-mode precoding does not require any form of water-filling. Thus our scheme, on average, always transmits with the same power on each antenna. The high-rate feedback performance difference between limited feedback covariance optimization and multi-mode precoding can be most likely contributed to this power-pouring.

![Graph showing capacity comparison](image)

Fig. 2. Capacity comparison of multi-mode precoding, limited feedback covariance optimization [22], and the uninformed transmitter.

VI. SUMMARY AND CONCLUSIONS

We proposed limited feedback multi-mode precoding for multi-antenna wireless systems. Both the number of substreams and the linear precoding matrix were optimally chosen based on current channel conditions from a set of possible precoder configurations and conveyed from receiver to transmitter via a low-rate feedback link. We used codebooks of precoding matrices that allow the multi-mode precoder to be selected from a finite set of possible precoder matrices.

It is of interest to find efficient selection functions for multi-mode precoding that efficiently minimize the probability of error exactly. This would require finding closed-form probability of error expressions for spatial multiplexing conditioned on current channel conditions. It is also of interest to characterize the diversity-multiplexing tradeoff curve as in [2] of multi-mode precoding.

REFERENCES