Project	**IEEE 802.16 Broadband Wireless Access Working Group <http://ieee802.org/16>**
Title	Closed-Loop MIMO Precoding with Limited Feedback
Date Submitted	**2004-08-24**
Source(s)	Jianzhong (Charlie) Zhang, Hua Zhang, Nico Van Waes, Anthony Reid, Victor Stolpman Nokia 6000 Connection Drive Irving, TX 75039 Voice: 972-374-0958, 678-576-2462, 972-894-5669, 972-374-0644, 972-894-6872, Fax: 972-894-5937 charlie.zhang@nokia.com, huazhang@ece.gatech.edu, nico.vanwaes@nokia.com, tony.reid@nokia.com, victor.stolpman@nokia.com,
Re:	IEEE 802.16e D4 Draft
Abstract	Closed-loop MIMO Precoding to improve MIMO link performance with limited feedback
Purpose	To incorporate the changes here proposed into the 802.16e D4 Draft. ~~Crossed-out indicates deleted text,~~ underlined blue indicates new text change to the Standard, and underlined green indicates newly added text from the original contribution
Notice	
Release	
Patent Policy and Procedures	

Closed-Loop MIMO Precoding with Limited Feedback

Jianzhong (Charlie) Zhang, Hua Zhang, Nico Van Waes, Tony Reid, Victor Stolpman
Nokia

1. Background

In 8.4.8.3.6 of IEEE 802.16e/D4, a MIMO precoding format is proposed. In this proposal, the output of the space-time encoder is weighted by a pre-coding weight matrix W, before being transmitted from the actual transmit antennas. However, this approach (Feedback type 0001 in table 298a) requires periodic feedback of the actual complex elements of the weight matrix W, and can be very demanding in terms of the feedback bandwidth (resources) needed to ensure the performance of the closed-loop system.

Here we propose a structured closed-loop MIMO precoding method that does not require the actual feedback of the weight matrix W. Instead, for each transmit antenna size we construct a set of precoding matrices and let this set be known at both the BS and SS. Consequently, the SS only need to feedback to the BS the index to a precoding matrix within this set. The set of the matrices (or the codebook) can be constructed to achieve the desired performance and feedback bandwidth tradeoff. Once the codebook is fixed, the number of feedback bits needed does not grow with the size of the matrix W itself, unlike in the existing approaches. We show that with the proposed precoding method, near-optimal precoding MIMO performance can be achieved with reasonably low amount of feedback bits.

2. MIMO Precoding with Limited Feedback

2.1 Precoding for a particular subcarrier

Consider an N_t transmit antenna, N_r receive antenna MIMO system. Let M_t be the number of spatially multiplexed data streams to be transmitted, and let the $M_t \times 1$ vector x denotes the signals carried on these data streams, the precoding matrix is a $N_t \times M_t$ weight matrix that transform the x vector into a z vector, which is of size $N_t \times 1$:

$$z = Wx \qquad (1)$$

note that the z vector is the actual signal being transmitted on the transmit antennas. The signal received at the receive antennas are given as:

$$r = HWx + n \qquad (2)$$

where H is the channel matrix and n is the AWGN noise vector.

If we do not have constraints on the feedback bandwidth, the optimal choice of W is well-known to be the right singular vectors of H matrix. However, feeding back these singular vectors can be very expensive, especially when fast update is needed in a system. Here we propose a structured closed-loop MIMO precoding method that does not require the actual feedback of the weight matrix W. Instead, for each transmit antenna size we construct a set of precoding matrices and let this set be known at both the BS and SS. We call this set of matrices as the "codebook" and denote it $\boldsymbol{P} = \{P_1,..., P_L\}$. Here $L=2^q$ denotes the size of the codebook and q is the number of (feedback) bits needed to index the codebook. Note that each matrix in the codebook is a unitary matrix and the design of the codebook is shown to be a subspace-packing problem in a Grassmann manifold [1][2]. We propose to use the structured block-circulant codebook designed in [1], as it requires the least amount of storage at both the transmitter and receiver.

Example:

Consider a 4 Tx, 2 Rx MIMO system. To feedback the W matrix directly, we would need q = 4*2*qc bits, where qc is the number of bits needed to represented a complex number. With a typical precision of 8 bits for every real number, we have qc =2*8 =16 and the total number of feedback bits q= 4*2*16 = 144 bits. In contrast, we state that a codebook of size L=64 is enough to retain most of the performance gain from precoding in this 4 by 2 MIMO system, meaning that we need only

$q=\log_2 64 = 6$ feedback bits instead of 144 bits required by the direct feedback method. This results in tremendous reduction in the number of feedback bits needed.

Once the codebook is specified for a MIMO system, the receiver observes a channel realization, selects the best precoding matrix (codeword) to be used at the moment, and feedback the index of the codeword to the transmitter. The basic idea of the limited feedback precoding MIMO system is illustrated in Figure 1 below. The performance of the precoded MIMO system is illustrated in Figure 2 for a 2 Tx, 1 Rx narrowband system.

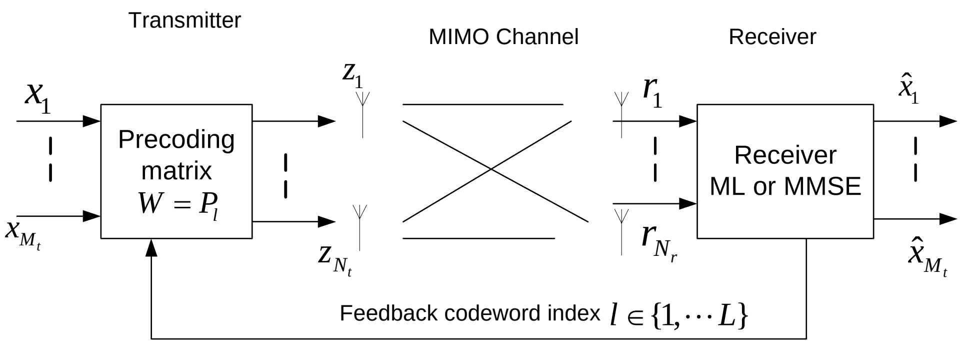

Fig. 1. Illustration of the Nt by Nr MIMO percoding, Mt data streams.

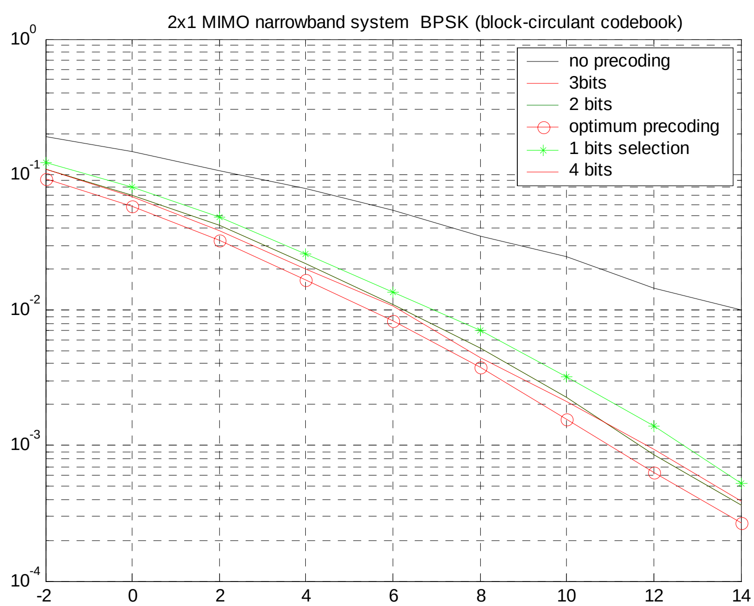

Fig. 2. Performance of the MIMO precoding method. With 3 bits of feedback, it is within 0.2 db of the optimal precoding solution, which assumes full precision channel feedback. Menatime, it is more than 6 dBs better than a baseline 1 by 1 open loop system

.

2.1.1 Codebook Design

We adopt the design proposed in [1] where the cross-correlations of the codewords follow a block-circulant structure. In this design, a codebook is fully specified once the first codeword P_1 and a diagonal rotation matrix
Q is provided. The other codewords in the codebook are given by:

$$\mathbf{P}_l = \mathbf{Q}^l \mathbf{P}_1, \quad \text{for } l = 2, \cdots L ,$$

where Q is a diagonal matrix fully parameterized by an integer vector $\mathbf{u} \triangleq [u_1, \cdots, u_{N_t}]$:

$$\mathbf{Q} = \begin{bmatrix} e^{j\frac{2\pi}{L}u_1} & & 0 \\ & \ddots & \\ 0 & & e^{j\frac{2\pi}{L}u_{N_t}} \end{bmatrix}$$

Furthermore, in this design, the first codeword P_1 is chosen to be a $N_t \times M_t$ submatrix of the $N_t \times N_t$ DFT matrix $\mathbf{D}_{N_t}$ whose (m,n) element is specified as $(\mathbf{D}_{N_t})_{m,n} = e^{j\frac{2\pi}{N_t}(m-1)(n-1)}$ where $1 \le m, n \le N_t$. Denoting $\mathbf{d}_c$ as the c^{th} column of the matrix $\mathbf{D}_{N_t}$, the first codeword is the collection of M_t columns parameterized by the set of column indices $\mathbf{c} \triangleq [c_1, \cdots, c_{M_t}]$, i.e,

$$\mathbf{P}_1 = [\mathbf{d}_{c_1}, \cdots, \mathbf{d}_{c_{N_t}}].$$

In Table 1, we tabulate the choices of $\mathbf{u} \triangleq [u_1, \cdots, u_{N_t}]$ and $\mathbf{c} \triangleq [c_1, \cdots, c_{M_t}]$ for different transmitter antenna numb and N_t, and spatially multiplexed data stream number M_t. Note that the choice of L is the result of trading off performance with number of feedback bits. For example, we picked L =8 for the N_t =2, M_t = 1 case since it is clear from Figure 2 that the performance gain is diminishing with more than 3 feedback bits.

N_t (# of Tx Antennas)	M_t (# of data streams)	L /(q=$\log_2$L) Codebook size/ (feedback bits)	$\mathbf{c} \triangleq [c_1, \cdots, c_{M_t}]$ Column indices	$\mathbf{u} \triangleq [u_1, \cdots, u_{N_t}]$ Rotation Vector
2	1	8 / (3)	[1]	[1, 0]
3	1	32/ (5)	[1]	[1, 26, 28]
3	2	32/ (5)	[1, 2]	[1, 26, 28]
4	1	64/ (6)	[1]	[1, 8, 61, 45]
4	2	64/ (6)	[0, 1]	[1, 7, 52, 56]
4	3	64/ (6)	[0, 2, 3]	[1, 8, 61, 45]

2.1.2 Codeword Selection at the receiver

After a codebook is chosen, the receiver observes a channel realization and makes a decision on the optimal codeword (precoding matrix) to be used at the transmitter. The index of the optimal codeword is then sent back through the designated feedback channel to the transmitter. We note that several receiver structures can be used in this MIMO system, including maximum likelihood (ML) and linear Minimum Mean Square Error (LMMSE) receivers. For the MMSE receiver, the MSE at the output of the receiver is a function of the precoding matrix $\mathbf{W} = \mathbf{P}_l$ used at the transmitter:

$$MSE(\mathbf{P}_l) = \frac{E_s}{N_o} tr\left\{\left(\mathbf{I}_{M_t} + \frac{E_s}{N_r N_o}\mathbf{P}_l^H \mathbf{H}^H \mathbf{H}\mathbf{P}_l\right)^{-1}\right\},$$

and the receiver does the following simple optimization to select the index of the precoding matrix to be conveyed to the transmitter:

$$l^{opt} = \arg\min_{l\in\{1,2,\cdots,L\}} MSE(\mathbf{P}_l).$$

2.2 Multi-carrier OFDMA MIMO system

For a single carrier system, we have shown that by utilizing a codebook of unitary matrices, the proposed limited feedback MIMO method achieves near-optimal beamforming performance with very few feedback bits. The extension of this method to a MIMO OFDMA system with N subcarriers in a subchannel is straightforward, once we decide that the same codebook $P = \{P_1,..., P_L\}$ can be used for all N subcarriers. In a direct extension of the precoding to OFDMA, the receiver selects the optimal codeword for each subcarrier in the subchannel of a particular SS (subscriber station), and use $q = \log_2 L$ bits to feedback the optimal codeword for that subcarrier. We denote this scheme (subcarrier) *independent precoding scheme*, in order to differentiate from a so-called *subspace-tracking precoding scheme* we introduce later. In the independent predcoing scheme, we need a total of Nq bits to feedback the optimal codeword choices for the SS with N subcarrier.

> Example:
> Consider the same 4 Tx, 2 Rx MIMO system, but with N = 108 out of 128 subcarriers assigned. We again use a codebook of size L=32, meaning that we need $q=\log_2 65 =6$ feedback bits for each subcarrier and a total of 108*6 = 648 bits for the whole system. The amount of feedback bits required becomes large when N increases.

To further reduce the number of feedback bits required for an SS with a large number of subcarriers, we proposed a *subspace tracking precoding scheme* where the choices of precoding matrices are dependent across the subcarriers. The proposed approach exploits the statistical correlation of the neighbouring subcarrier channels in an OFDMA system. The idea originates from the fact that due to the statistical correlation between two neighbouring subcarriers, it is highly likely that the two desired precoding matrices reside within a small neighbourhood in the high-dimensional Grassmann manifold. Consequently, we devise a mechanism for recursive selection of precoding matrices, which we term subcarrier-tracking algorithm here. In this tracking algorithm, we start with the first subcarrier and use the full precision ($q =\log_2 L$ bits) to select one of the best precoding matrix, W_1, out of 2^q possibilities.

Observing that the best precoding matrix for the second subcarrier, W_2, lives in the small neighbourhood of P_1, we are able to narrow our search. Assuming that the number of matrices in this small neighbourhood to be $2^{q'}$ (q'<q), we effectively reduce the number of feedback bits needed for the second subcarrier to q'. Recursively repeating this process to cover all N subcarriers involved, and we end up with a total requirement of q+(N-1)q' feedback bits, which is much less than the N*q bits necessary for the non-tracking approach. The search for W_2 in the neighbourhood of W_1 is illustrated in Figure 3.

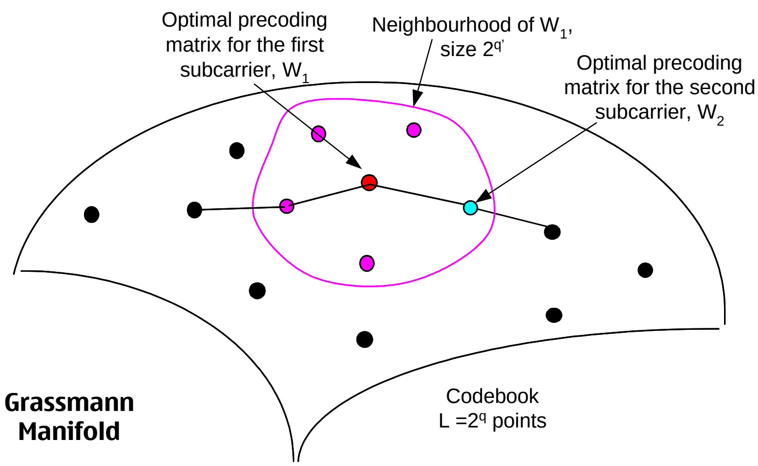

Fig. 3. Illustration of subspace tracking in Grassmann Manifold.

We summarize the *subspace tracking precoding scheme* as follows. Note here we assume that the codebook is the same across all the subcarriers. The codebook is P={P_1,…, P_L}, where L = 2^q is the codebook size and q is the feedback bits needed for the codeword selection from the whole codebook. Meanwhile, we used an additional parameter called step size for flexibility.

1. For the first subcarrier, use the full q bits to select the precoding matrix W_1 out of the L codewords.
2. Defining a step size K such that K|N (K is a factor of N), we will skip the subcarriers 2,…,K and move to subcarrier K+1. The search for W_{K+1} will be limited in the neighbourhood of W1 defined by the set $\mathbf{P}_{S_1} = \{\mathbf{P}_i, \text{ s.t. } d(\mathbf{P}_i, \mathbf{W}_1) \le \delta_1\}$, where $d(\mathbf{P}_i, \mathbf{W}_1) \triangleq M_t - \left\|\mathbf{P}_i^H \mathbf{W}_1\right\|_F^2$ is the chordal distance between $\mathbf{P}_i$ and $\mathbf{W}_1$ in the Grassmann manifold, and $\left\|\bullet\right\|_F$ denotes Frobenius norm. The parameter δ_1 is selected to chosen such that the size of the set $\left|\mathbf{P}_{S_1}\right| \le 2^{q'}$, where q' denotes the number of feedback bits needed for the K+1 th subcarrier.
3. Repeat step 2 for subcarriers 2K+1, 3K+1,… (N/K-1)K+1.

According to the above subspace tracking precoding scheme, the total number of bits required for all N subcarriers is $q + \frac{N}{K} q'$ bits. The parameters K and q' are selected to achieve the best performance/feedback bandwidth tradeoff. In the Figure 4 below, we demonstrate the efficacy of the proposed algorithms using a 128 subcarrier, 32 guard carriers, 4 Tx, 2Rx MIMO OFDMA system.

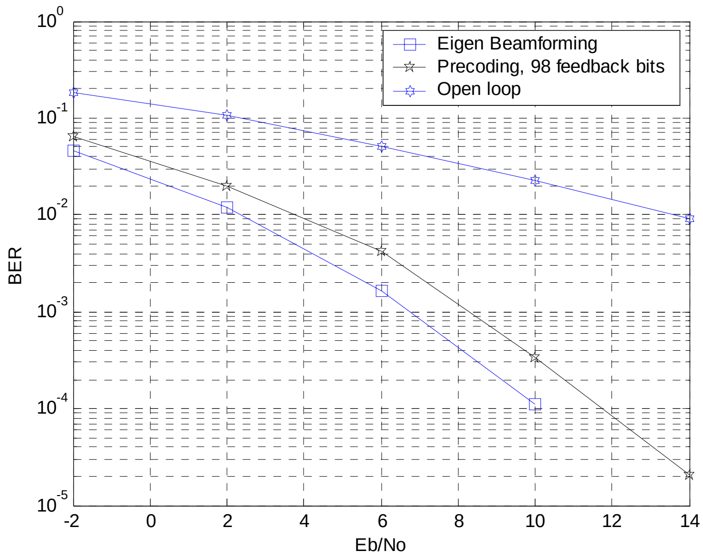

Fig 4. Performance of MIMO precoding. 4 Tx, 2 Rx, 96 subcarriers assigned. Meanwhile, q =6, q' =4 and K=4. With only about 1 bit per subcarrier, the precoding method is within 1dB of the optimal eigen-beamforming solution. Meanwhile, the precoding solution is 10 dBs better than the open loop non-STBC solution.

3. Specific Text Changes

[Modify the following Table 298a in section 8.4.5.3.12.1]

Table 298a. CQICH Enhanced allocation IE format

Syntax	Size (bits)	Notes
CQICH_Enhanced_Alloc_IE() {		
Extended DIUC	4	0x09
Length	4	Length in bytes of following fields
CQICH_ID	variable	Index to uniquely identify the CQICH resource assigned to the MSS
Period (=p)	2	A CQI feedback is transmitted on the CQICH every 2^p frames
Frame offset	3	The MSS starts reporting at the frame of which the number has the same 3 LSB as the specified frame offset. If the current frame is specified, the MSS should start reporting in 8 frames
Duration (=d)	3	A CQI feedback is transmitted on the CQI channels indexed by the CQICH_ID for 10 x 2^d frames. If d== 0, the CQICH is de-allocated. If d == 111, the MSS should report until the BS command for the MSS to stop.

N_T actual BS antennas	3	001 = Reserved 010 = 2 actual antennas 011 = 3 actual antennas 100 = 4 actual antennas 101 = 5 actual antennas 110 = 6 actual antennas 111 = 7 actual antennas 000 = 8 actual antennas
Feedback_type	4	0000 = Open loop precoding. Pilots in burst to be precoded with *W*. SS to rely only on pilots in burst for channel estimation. 0001 = Complex weight of specific element of *W*- 0010 = Fast DL measurement 0011 = Layer specific channel strengths 0100 = MIMO mode and permutation zone feedback 0101 = Feedback of subset of antennas to use.- 0110 = Feedback of the precoding matrix index. Where the set of precoding matrices is a pre-defined matrix codebook known at both the transmitter and receiver. 0111 ~ 1111 reserved
CQICH_Num	4	Number of CQICHs assigned to this CQICH_ID is (CQICH_Num +1)
for (i=0;i<CQICH_Num;i++) {		
Allocation index	6	Index to the fast feedback channel region marked by UIUC=0
}		
if (Feedback_type != 0100) { MIMO_permutation_feedback cycle }	2	00 = No MIMO and permutation mode feedback 01 = the MIMO and permutation mode indication shall be transmitted on the CQICH indexed by the CQICH_ID every 4 frames. The first indication is sent on the 8th CQICH frame. 10 = the MIMO mode and permutation mode indication shall be transmitted on the CQICH indexed by the CQICH_ID every 8 frames. The first indication is sent on the 8th CQICH frame. 11 = the MIMO mode and permutation mode indication shall be transmitted on the CQICH indexed by the CQICH_ID every 16 frames. The first indication is sent on the 16th CQICH frame.
if (Feedback_type != 0110) { Feedback cycle for index of the precoding matrix }	2	00 = Feed back one precoding matrix index every frame. 00 = Feed back one precoding matrix index every 4th frame. 00 = Feed back one precoding matrix index every 8th frame. 00 = Feed back one precoding matrix index every 16th frame.
Padding	*variable*	The padding bits are used to ensure the IE size is integer number of bytes.

[Add the following text into section 8.4.8.3.7]

The space time coding output can be weighted by a matrix before mapping onto transmit antennas:

$$z = Wx$$

where *x* is a vector with the output from the space-time coding (per-subcarrier), M_t is the number of antennas at the output of the space-time coding scheme. The matrix W is an weighting matrix where the quantity N_t is the number of actual transmit antennas. The vector z contains the signals after weighting for the different actual antennas. The labeling of the elements in the weighting matrix W is performed in accordance with the example of W given below for the case of 4 actual antennas and 2 space-time coding output antennas:

$$W = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \\ W_{31} & W_{32} \\ W_{41} & W_{42} \end{bmatrix}$$

The space-time weighting matrix W belongs to a codebook $\mathbf{P} = \{\mathbf{P}_1, \cdots, \mathbf{P}_L\}$, the codebook is fully specified once the first codeword P_1 and a diagonal rotation matrix Q is provided. The other codewords in the codebook are given by:

$$\mathbf{P}_l = \mathbf{Q}^l \mathbf{P}_1, \quad \text{for } l = 2, \cdots L,$$

where Q is a diagonal matrix fully parameterized by an integer vector $\mathbf{u} \triangleq [u_1, \cdots, u_{N_t}]$:

$$\mathbf{Q} = \begin{bmatrix} e^{j\frac{2\pi}{L}u_1} & & 0 \\ & \ddots & \\ 0 & & e^{j\frac{2\pi}{L}u_{N_t}} \end{bmatrix}.$$

Furthermore, in this design, the first codeword P_1 is chosen to be a $N_t \times M_t$ submatrix of the $N_t \times N_t$ DFT matrix $\mathbf{D}_{N_t}$ whose (m,n) element is specified as $(\mathbf{D}_{N_t})_{m,n} = e^{j\frac{2\pi}{N_t}(m-1)(n-1)}$ where $1 \le m, n \le N_t$. Denoting $\mathbf{d}_c$ as the c[th] column of the matrix $\mathbf{D}_{N_t}$, the first codeword is the collection of M_t columns parameterized by the set of column indices $\mathbf{c} \triangleq [c_1, \cdots, c_{M_t}]$, i.e,

$$\mathbf{P}_1 = [\mathbf{d}_{c_1}, \cdots, \mathbf{d}_{c_{N_t}}].$$

In Table eee, the choices of $\mathbf{u} \triangleq [u_1, \cdots, u_{N_t}]$ and $\mathbf{c} \triangleq [c_1, \cdots, c_{M_t}]$ are tabulated for different transmitter antenna numb and N_t, and spatially multiplexed data stream number M_t. Note that the choice of L is the result of trading off performance with number of feedback bits.

N_t (# of Tx Antennas)	M_t (# of data streams)	L /(q=$\log_2$L) Codebook size/ (feedback bits)	$\mathbf{c} \triangleq [c_1, \cdots, c_{M_t}]$ Column indices	$\mathbf{u} \triangleq [u_1, \cdots, u_{N_t}]$ Rotation Vector
2	1	8 / (3)	[1]	[1, 0]
3	1	32/ (5)	[1]	[1, 26, 28]
3	2	32/ (5)	[1, 2]	[1, 26, 28]
4	1	64/ (6)	[1]	[1, 8, 61, 45]
4	2	64/ (6)	[0, 1]	[1, 7, 52, 56]
4	3	64/ (6)	[0, 2, 3]	[1, 8, 61, 45]

After a codebook is chosen, the receiver observes a channel realization and makes a decision on the optimal codeword (precoding matrix) to be used at the transmitter. The index of the optimal codeword is then sent back through the designated feedback channel to the transmitter. We note that several receiver structures can be used in this MIMO system, including maximum likelihood (ML) and linear Minimum Mean Square Error (LMMSE) receivers. For the MMSE receiver, the MSE at the output of the receiver is a function of the precoding matrix $\mathbf{W} = \mathbf{P}_l$ used at the transmitter:

$$MSE(\mathbf{P}_l) = \frac{E_s}{N_o} tr\left\{\left(\mathbf{I}_{M_t} + \frac{E_s}{N_r N_o}\mathbf{P}_l^H \mathbf{H}^H \mathbf{H}\mathbf{P}_l\right)^{-1}\right\},$$

and the receiver does the following simple optimization to select the index of the precoding matrix to be conveyed to the transmitter:

$$l^{opt} = \arg \min_{l \in \{1,2,\cdots,L\}} MSE(\mathbf{P}_l).$$

If more than one subcarriers are assigned to one SS, the feedback bits of each subcarriers are obtained by:

1. For the first subcarrier, use the full q bits to select the precoding matrix W_1 out of the L codewords.
2. Defining a step size K such that K|N (K is a factor of N), we will skip the subcarriers 2,…,K and move to subcarrier K+1. The search for W_{K+1} will be limited in the neighbourhood of W1 defined by the set $\mathbf{P}_{S_1} = \{\mathbf{P}_i, \text{ s.t. } d(\mathbf{P}_i, \mathbf{W}_1) \le \delta_1\}$, where $d(\mathbf{P}_i, \mathbf{W}_1) \triangleq M_t - \left\|\mathbf{P}_i^H \mathbf{W}_1\right\|_F^2$ is the chordal distance between $\mathbf{P}_i$ and $\mathbf{W}_1$ in the Grassmann manifold, and $\left\|\bullet\right\|_F$ denotes Frobenius norm. The parameter δ_1 is selected to chosen such that the size of the set $\left|\mathbf{P}_{S_1}\right| \le 2^{q'}$, where q' denotes the number of feedback bits needed for the K+1 th subcarrier.
3. Repeat step 2 for subcarriers 2K+1, 3K+1,… (N/K-1)K+1.

4 References

[1] B. M. Hochwald, T. L. Marzetta, T. J. Richardson, W. Sweldens, R. Urbanke, " Systematic design of unitary space-time constellation", IEEE Transactions on Information Theory, vol. 46, pp. 1962-1973, Sep 2000.

[2]. D. Love and R. W. Heath, " Limited feedback precoding for spatial multiplexing systems," GLOBECOM 2003, vol. 4, Dec. 2003, pp. 1857-1861.