Project	IEEE 802.16 Broadband Wireless Access Working Group <http://ieee802.org/16 >
Title	MU-MIMO: Demodulation at the Mobile Station
Date Submitted	2009-01-05
Source(s)	M. A. Khojastepour, NEC, amir@nec-labs.com Ron Porat, Nextwave, rporat@nextwave.com
Re:	IEEE C802.16m-08/052 Call for SDD comments and contributions on 802.16m SDD (802.16m-08/003r6)
Abstract	This document proposes that each MU-MIMO user may be informed about some scheduling parameters of other co-scheduled MU-MIMO users, such as their beamforming vectors, power levels and modulations.
Purpose	Discussion and Decision
Notice	
Release	
Patent Policy	

MU-MIMO: Demodulation at the Mobile Station

N. Prasad, M. A. Khojastepour, M. Jiang, X. Wang and S. Rangarajan

I. INRODUCTION

In this document, we consider MU-MIMO where the base station (BS) may schedule several mobile stations on the same resource unit. Each scheduled mobile station (MS) is served only a single stream via beamforming. We examine the various demodulator options that are available to a scheduled MS (or user) depending on the information it has about the beam vectors, power levels and modulations used to serve the co-scheduled users. We assume that each scheduled user is informed about the beam vector, power level as well as the MCS assigned to it by the BS.

Consider the received signal model on any subcarrier at a user terminal k, which is equipped with N receive antennas and where the BS has M transmit antennas,

$$\boldsymbol{y}_k = \boldsymbol{H}_k \boldsymbol{x} + \boldsymbol{v}_k, \tag{1}$$

where $\boldsymbol{H}_k \in \mathbb{C}^{N\times M}$ is the channel matrix and $\boldsymbol{v}_k \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I})$ is the additive noise. The signal vector $\boldsymbol{x}$ transmitted by the BS can be expanded as

$$\boldsymbol{x} = \sum_{k\in\mathcal{S}} \boldsymbol{g}_k s_k \tag{2}$$

where $\mathcal{S}$ is the set of users that are scheduled on the same resource unit, $\boldsymbol{g}_k$ is the beamforming vector and s_k is the data symbol corresponding to user $k \in \mathcal{S}$. The beamforming vectors are selected from a codebook $\mathcal{C}$ of unit-norm vectors. The sum power constraint is given by $E[\boldsymbol{x}^\dagger \boldsymbol{x}] = \sum_{k=1}^{|\mathcal{S}|} E[|s_k|^2] \leq \rho$.

II. Linear Combiner

A. Match Filter Linear Combiner

A popular linear combiner is the match filter (a.k.a. maximum ratio combiner (MRC)) which is optimal in the absence of interference from the signals intended for the other co-scheduled users. In order to employ this linear combiner, the user of interest (say user 1) does not need to know any information about the co-scheduled users and it can demodulate its own data by completely neglecting the co-channel interference resulting from the transmission intended for the other co-scheduled users. In particular, user 1 employs the unit norm combiner given by,

$$\boldsymbol{u} = \frac{\boldsymbol{H}_1\boldsymbol{g}_1}{||\boldsymbol{H}_1\boldsymbol{g}_1||}. \tag{3}$$

The received signal at user 1 post-combining can now be written as

$$\boldsymbol{u}_1^\dagger\boldsymbol{y}_1 = \boldsymbol{u}_1^\dagger\boldsymbol{H}_1\boldsymbol{g}_1 s_1 + \sum_{k\in\mathcal{S},k\neq 1} \boldsymbol{u}_1^\dagger\boldsymbol{H}_1\boldsymbol{g}_k s_k + \tilde{v}_1 \tag{4}$$

and note that due to our normalization $E[|\tilde{v}_1|^2] = 1$. In order to compute the log-likelihood ratios (LLRs) for the coded bits corresponding to the symbol s_1, the user can completely neglect the interference term $\sum_{k\in\mathcal{S},k\neq 1} \boldsymbol{u}_1^\dagger\boldsymbol{H}_1\boldsymbol{g}_k s_k$ and assume the SNR to be $\rho_1||\boldsymbol{H}_1\boldsymbol{g}_1||^2$, where $\rho_1 = E[|s_1|^2]$. Alternatively, an estimate of the true SINR can instead determined. However, the accuracy of such an estimate depends on the number of sample observations available.

The MRC based demodulator can result in substantial performance degradation when the interference power is not negligible compared to that of the additive noise. This is particularly true at high SNR. In order design a better linear combiner for such scenarios, a quantization error minimization approach is employed in [1]. In particular, assuming $N < M$, the linear combiner used by the MS is now given by

$$\boldsymbol{u}_1 = \frac{(\boldsymbol{g}_1^\dagger\boldsymbol{H}_1^+)^\dagger}{||\boldsymbol{g}_1^\dagger\boldsymbol{H}_1^+||} \tag{5}$$

where $\boldsymbol{H}_1^+ = \boldsymbol{H}_1^\dagger(\boldsymbol{H}_1\boldsymbol{H}_1^\dagger)^{-1}$ denotes the pseudo-inverse. In general, the user can use any linear combiner $\boldsymbol{u}_1$ (which could have been pre-computed based only on the channel estimate $\boldsymbol{H}_1$).

Upon using such a combiner, the resulting model is given by (4) and the LLRs can either be computed by completely neglecting the interference and using the SNR $\rho_1||\boldsymbol{u}_1^\dagger \boldsymbol{H}_1 \boldsymbol{g}_1||^2$ or by using an estimated SINR.

B. MMSE Linear Combiner

In some scenarios, each scheduled user can also deduce the beamforming vectors and power levels used to serve the other co-scheduled users. In this case, assuming that the transmit power is equally split among all scheduled users, user 1 can determine the optimal minimum mean squared error (MMSE) linear combiner as

$$\boldsymbol{u}_1 = \left(\boldsymbol{I} + \tilde{\rho} \sum_{k \in \mathcal{S}, k \neq 1} \boldsymbol{H}_1 \boldsymbol{g}_k \boldsymbol{g}_k^\dagger \boldsymbol{H}_1^\dagger\right)^{-1} \boldsymbol{H}_1 \boldsymbol{g}_1, \tag{6}$$

where $\tilde{\rho} = \frac{\rho}{|\mathcal{S}|}$. The resulting SINR is determined to be

$$\tilde{\rho} \boldsymbol{g}_1^\dagger \boldsymbol{H}_1^\dagger \left(\boldsymbol{I} + \tilde{\rho} \sum_{k \in \mathcal{S}, k \neq 1} \boldsymbol{H}_1 \boldsymbol{g}_k \boldsymbol{g}_k^\dagger \boldsymbol{H}_1^\dagger\right)^{-1} \boldsymbol{H}_1 \boldsymbol{g}_1. \tag{7}$$

III. Non-linear Demodulators

In order to further improve the performance, the user may employ non-linear demodulators. Suppose that the user knows the beamforming vectors and power levels used to serve the other co-scheduled users. Moreover, the user also knows the modulations used to serve some or all of the other co-scheduled users. In particular, let $\mathcal{J} \subseteq \mathcal{S}$ denote the set of users whose corresponding modulations are known to user 1 and clearly $1 \in \mathcal{J}$. Then, user 1 can first design a filter to suppress the interference from the signals intended for the other co-scheduled users not in $\mathcal{J}$ to obtain the model

$$\boldsymbol{z} \triangleq \left(\boldsymbol{I} + \tilde{\rho} \boldsymbol{H}_1 \sum_{k \notin \mathcal{J}} \boldsymbol{g}_k \boldsymbol{g}_k^\dagger \boldsymbol{H}_1^\dagger\right)^{-1/2} \boldsymbol{y}_1 = \left(\boldsymbol{I} + \tilde{\rho} \boldsymbol{H}_1 \sum_{k \notin \mathcal{J}} \boldsymbol{g}_k \boldsymbol{g}_k^\dagger \boldsymbol{H}_1^\dagger\right)^{-1/2} \boldsymbol{H}_1 \sum_{k \in \mathcal{J}} \boldsymbol{g}_k s_k + \boldsymbol{\eta}, \tag{8}$$

and note that $E[\boldsymbol{\eta}\boldsymbol{\eta}^\dagger] = \boldsymbol{I}$. Letting $\boldsymbol{B} = \left(\boldsymbol{I} + \tilde{\rho} \boldsymbol{H}_1 \sum_{k \notin \mathcal{J}} \boldsymbol{g}_k \boldsymbol{g}_k^\dagger \boldsymbol{H}_1^\dagger\right)^{-1/2} \boldsymbol{H}_1 \boldsymbol{G}_{\mathcal{J}}$ with $\boldsymbol{G}_{\mathcal{J}} = [\boldsymbol{g}_k]_{k \in \mathcal{J}}$ denoting the $N \times |\mathcal{J}|$ matrix formed by the beam vectors employed to serve users in $\mathcal{J}$ and $\boldsymbol{s}_{\mathcal{J}}$ denoting the $|\mathcal{J}| \times 1$ vector formed by the corresponding symbols transmitted to those users, the model in (8) can be re-written as

$$\boldsymbol{z} = \boldsymbol{B}\boldsymbol{s}_{\mathcal{J}} + \boldsymbol{\eta}. \tag{9}$$

Now we can employ several suitable non-linear demodulators, such as the soft-output sphere decoder [2] and its recent variants over the model in (9). *The key point to be noted is that the LLRs need to be generated only for the coded bits in the symbol s_1. Note that in order to generate these LLRs, we do not need any information about the coding rates employed to serve any user in $\mathcal{J}$. This is particularly useful since there are only three distinct modulations that can be used to serve each scheduled user. Also, no attempt is made to decode the codeword of any other user (apart from user 1) in $\mathcal{J}$.*

To illustrate one such non-linear demodulator, henceforth referred to as the partial-MLD, suppose that $\mathcal{J} = \{1, 2\}$ and that user-1 has two receive antennas so that $\boldsymbol{B} = [\boldsymbol{b}_1, \boldsymbol{b}_2]$ is a 2×2 matrix. Denote $\widetilde{\boldsymbol{b}}_i \triangleq \boldsymbol{b}_i/\|\boldsymbol{b}_i\|, i = 1, 2,$ where $\|\boldsymbol{b}_i\| = \sqrt{\boldsymbol{b}_i^\dagger \boldsymbol{b}_i}$. Let $\boldsymbol{B} = \|\boldsymbol{b}_2\|\boldsymbol{U}\boldsymbol{L}$ be the modified QR decomposition of $\boldsymbol{B}$ with $\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2]$ being a semi-unitary matrix such that $\boldsymbol{U}^\dagger \boldsymbol{U} = \boldsymbol{I}$ and $\boldsymbol{L}$ being lower triangular with positive diagonal elements. In particular, we have

$$\boldsymbol{u}_1 = \frac{\widetilde{\boldsymbol{b}}_1 - \rho\widetilde{\boldsymbol{b}}_2}{\sqrt{1 - |\rho|^2}}, \quad \boldsymbol{u}_2 = \widetilde{\boldsymbol{b}}_2, \quad \text{with} \;\; \rho \triangleq \widetilde{\boldsymbol{b}}_2^\dagger \widetilde{\boldsymbol{b}}_1, \tag{10}$$

$$\text{and} \qquad \boldsymbol{L} = \begin{bmatrix} l_{11} & 0 \\ l_{21} & 1 \end{bmatrix}, \quad \text{with} \;\; l_{11} = \frac{\|\boldsymbol{b}_1\|}{\|\boldsymbol{b}_2\|}\sqrt{1 - |\rho|^2}, \quad l_{21} = \frac{\|\boldsymbol{b}_1\|}{\|\boldsymbol{b}_2\|}\rho. \tag{11}$$

We then obtain

$$\boldsymbol{w} \triangleq \boldsymbol{U}^\dagger \boldsymbol{z}/\|\boldsymbol{b}_2\| = \boldsymbol{L}\boldsymbol{s}_{\mathcal{J}} + \widehat{\boldsymbol{n}}, \tag{12}$$

where $\boldsymbol{w} = [w_1, w_2]^T, \boldsymbol{s}_{\mathcal{J}} = [s_1, s_2]^T$ and $E[\widehat{\boldsymbol{n}}\widehat{\boldsymbol{n}}^\dagger] = \|\boldsymbol{b}_2\|^{-2}\boldsymbol{I}$. Let s_i^{R} and s_i^{I} denote the real and imaginary parts of $s_i, i = 1, 2$, respectively. For the j-th possibility of symbol s_1, denoted as $s_{1,j}$, we define the metric

$$Q(s_{1,j}) \triangleq |w_1 - l_{11}s_{1,j}|^2 + \min_{s_2 \in \mathcal{S}_2} |w_2 - l_{21}s_{1,j} - s_2|^2, \tag{13}$$

Parameter	Assumption
Bandwidth	10.0 MHz
FFT size	1024
Resource block size	24 subcarriers and 6 OFDM symbols
Sub-carrier spacing	15.0 kHz
Number of data sub-carriers per OFDM symbol	900
Channel encoder	Convolutional Code(CC) Rate =1/2; 2/3; 3/4
Number of information bits per block	768 (R=1/2); 512(R=2/3); 576(R=3/4)
Beamforming codebook	3-bit codebook V(4,1,3) defined in 802.16e
Number of antennas at Base Station (BS)	4
Number of antennas at Mobile Station (MS)	2
Channel model	WINNER C2 (urban Macro-cell)

TABLE I
SIMULATION PARAMETERS

where $\mathcal{S}_2$ denotes the constellation (modulation-type) of user-2. Suppose the modulation $\mathcal{S}_1$ assigned to user-1 is M-QAM, Using (13) we determine the M metrics $\{Q(s_{1,j})\}$.

The QAM symbol s_1, corresponds to $\log_2 M$ bits, i.e., is represented by a $\log_2 M$ length bit-vector. The M metrics $\{Q(s_{1,j})\}$ defined above are sufficient to determine the max-log LLR for each bit associated with s_1. To see this, consider the $\log_2 M$ bits associated with s_1. Then letting $\lambda_{1,\ell}$ denote the max-log LLR of the ℓ-th bit $b_{1,\ell}$ associated with s_1 and assuming equal *a priori* bit probabilities, we have

$$\lambda_{1,\ell} = \|\boldsymbol{b}_2\|^2 \min_{s_{1,j}\in\mathcal{S}_1:b_{1,\ell}=0} Q(s_{1,j}) - \|\boldsymbol{b}_2\|^2 \min_{s_{1,j}\in\mathcal{S}_1:b_{1,\ell}=1} Q(s_{1,j}), \quad 1 \leq \ell \leq \log_2 M. \tag{14}$$

IV. SIMULATION RESULTS

The block error rate (BLER) performance of the proposed scheme is investigated via simulations. We assume that the base station (BS) has four transmit antennas whereas the user (or MS) of interest has two receive antennas. Each MS reports its preferred PMI (per resource block) to the BS. The BS then pairs the user of interest with another user such that their reported PMIs are (almost) mutually orthogonal. The BS uses the user reported PMIs as beamforming vectors and divides the power equally among scheduled users. The performance obtained using the match filter (MF) based demodulator, the LMMSE based demodulator and the partial-MLD

Table 400—V(4,1,3)

Vector index	1	2	3	4	5	6	7	8
v1	1	0.3780	0.3780	0.3780	0.3780	0.3780	0.3780	0.3780
v2	0	–0.2698 – *j*0.5668	–0.7103 + *j*0.1326	0.2830 – *j*0.0940	–0.0841 + *j*0.6478	0.5247 + *j*0.3532	0.2058 – *j*0.1369	0.0618 – *j*0.3332
v3	0	0.5957 + *j*0.1578	–0.2350 – *j*0.1467	0.0702 – *j*0.8261	0.0184 + *j*0.0490	0.4115 + *j*0.1825	–0.5211 + *j*0.0833	–0.3456 + *j*0.5029
v4	0	0.1587 – *j*0.2411	0.1371 + *j*0.4893	–0.2801 + *j*0.0491	–0.3272 – *j*0.5662	0.2639 + *j*0.4299	0.6136 – *j*0.3755	–0.5704 + *j*0.2113

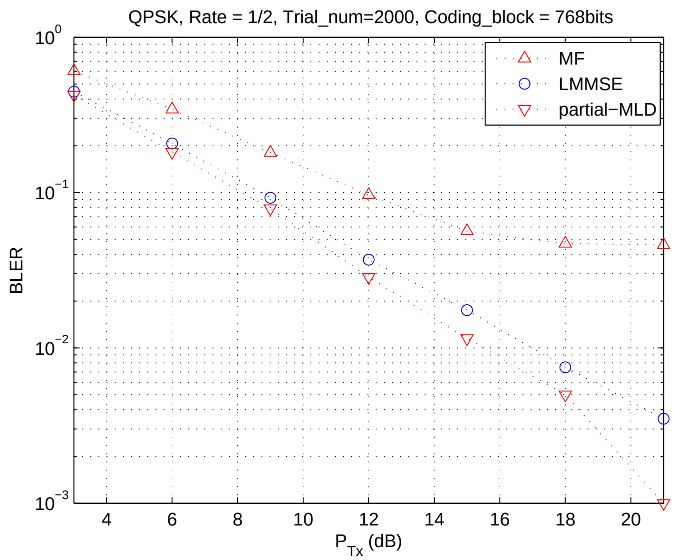
Fig. 1. Beamforming codebook: 3-bit 802.16e V(4,1,3)

based demodulator is shown in the following figures. Other specific simulation parameters are listed in Table I.

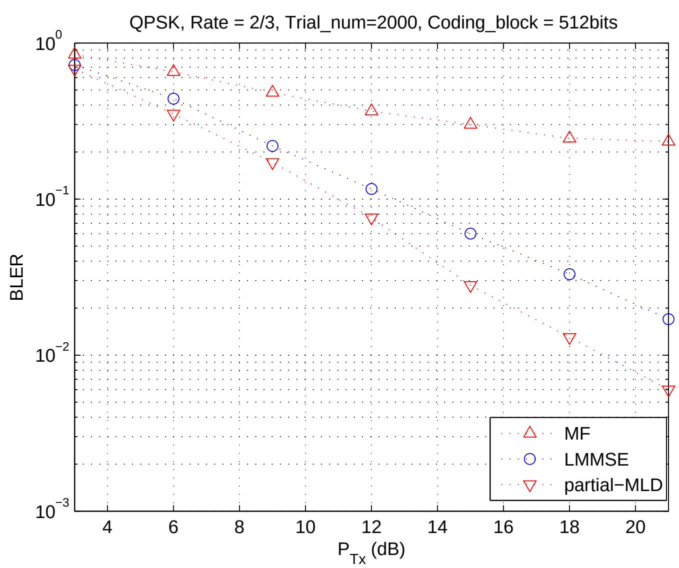
In Fig. 2(a) and (b), we compare the performance obtained using the linear match filter, the linear MMSE and the non-linear partial-MLD demodulators, respectively, for QPSK modulation. The beamforming index/precoding matrix index (PMI) feedback is reported for each resource block. It is seen that the partial-MLD demodulator has much better performance than both MF and LMMSE demodulators. For example, at a BLER of 10^{-2}, the partial-MLD has 1.8dB and 4.0dB gains, respectively, over the LMMSE for R= $1/2$ and R= $2/3$.

Fig. 3 illustrates the BLER performance of the linear MMSE and the partial-MLD demodulators for 16-QAM modulation. As seen from the figure, the partial-MLD demodulator significantly outperforms the linear MMSE one.

In a practical scenario, each MS will be allowed to report only one (or a few) PMIs. We anticipate more gains from the partial-MLD demodulator over the linear demodulators in this case since more severe multi-user interference is expected, which is particularly detrimental to the performance of linear demodulators.



(a) Block error rate performance for QPSK, R=1/2.



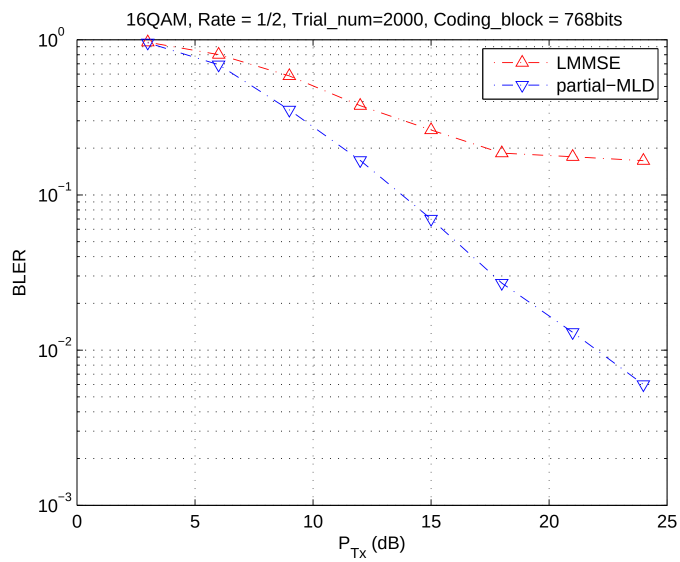
(b) Block error rate performance for QPSK, R=2/3.

Fig. 2. BLER performance with independent PMI per RB.

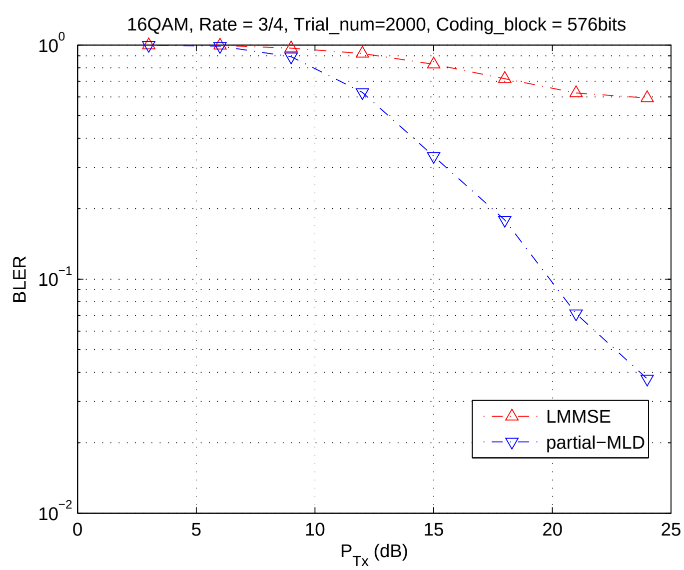
V. Discussion

It is evident that the linear combiner which requires no information about other co-scheduled users results in the least amount of (feedforward) signaling overhead. Such combiners rely on a large extent on the precoder employed by the transmitter (BS) to mitigate or remove the interference seen by them. If the BS has perfect knowledge of the channel matrix [1] of each user

[1]As suggested in [1], each user can use a linear combiner to convert its channel matrix into an effective channel vector and report the latter to the BS.



(a) R=1/2.



(b) R=3/4.

Fig. 3. BLER performance with independent PMI per RB for 16QAM.

on each subcarrier, it can employ a zero-forcing based precoder to ensure that each scheduled user sees no interference from transmission to other co-scheduled users. In such a scenario the MRC combiner is optimal.

However, in practical FDD systems with limited feedback, providing such perfect channel knowledge about each user to the BS is not possible, particularly in the wideband OFDMA based downlink, where each user's channel response matrix varies across subcarriers. Consequently, the users may need to employ more sophisticated demodulators to combat residual interference and help the system realize the benefit of multi-user MIMO. These sophisticated demodulators need the knowledge about some scheduling parameters of other co-scheduled users such as their beam vectors, power levels and modulations. Fortunately, such parameters for each scheduled user are anyway transmitted in the user-specific part of the unicast service control channel. The signaling should be designed in an appropriate manner so that an MU-MIMO user can deduce such parameters of other co-scheduled MU-MIMO users.

References

[1] N.Jindal,"Antenna Combining for the MIMO Downlink Channel,", *IEEE Trans. Wireless Communications*, Vol. 7, No. 10, pp. 3834-3844, Oct. 2008.

[2] B. M. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, pp. 389-399, Mar. 2003.

PROPOSED TEXT

[Modify the text in section 11.8.1.6.2 “Signaling Support for MU-MIMO”]

In the downlink MU-MIMO, the precoding matrix shall be signaled via explicit signaling if common demodulation pilots are used, or via dedicated pilots. Each scheduled MU-MIMO user may be informed about some scheduling parameters of other co-scheduled MU-MIMO users, such as their beam vectors, power levels and modulations. The exact choice of these scheduling parameters is FFS.