Bit Interleaved Coded Multiple Beamforming

Enis Akay Student Member, IEEE, Ersin Sengul Student Member, IEEE,
Ender Ayanoglu Fellow, IEEE

Abstract

In this paper, we investigate the performance of bit interleaved coded multiple beamforming (BICMB). We provide interleaver design criteria such that BICMB achieves full spatial multiplexing of \( \min(N, M) \) and full spatial diversity of \( NM \) with \( N \) transmit and \( M \) receive antennas over quasi-static Rayleigh flat fading channels. If the channel is frequency selective, then BICMB is combined with OFDM (BICMB-OFDM) in order to combat intersymbol interference (ISI) caused by the frequency selective channels. BICMB-OFDM achieves full spatial multiplexing of \( \min(N, M) \), while maintaining full spatial and frequency diversity of \( NML \) for an \( N \times M \) system over \( L \)-tap frequency selective channels when an appropriate convolutional code is used. Both systems analyzed in this paper assume perfect channel state information both at the transmitter and the receiver. Simulation results show that when the perfect channel state information assumption is satisfied, BICMB and BICMB-OFDM provide substantial performance or complexity gains when compared to other spatial multiplexing and diversity systems.

Keywords: BICM, Beamforming, Bit Interleaved Coded Modulation, Diversity, Spatial Multiplexing, BICMB

Portions of this work were presented in IEEE WCNC 2006, Las Vegas, USA, April 2006.

E. Akay was and E. Sengul and E. Ayanoglu are with the Center for Pervasive Communications and Computing, Department of Electrical Engineering and Computer Science, The Henry Samueli School of Engineering, University of California, Irvine. E-mails: enisakay@ieee.org, esengul@uci.edu, ayanoglu@uci.edu.
Bit Interleaved Coded Multiple Beamforming

I. INTRODUCTION

Multi-input multi-output (MIMO) systems provide significant capacity and diversity advantages [1]. A basic challenge in a MIMO system design is to achieve a high diversity order as well as high throughput. In a MIMO system, it is possible to transmit multiple streams of data over multiple antennas. This technique is known as spatial multiplexing [2]. This is a good alternative solution to providing high data rates with restrictions on the constellation size and the available bandwidth. Spatial multiplexing can utilize different receivers some of which are: maximum likelihood (ML) receiver, successive cancelation (SUC) receiver, ordered SUC receiver, minimum mean squared error (MMSE) receiver, and zero forcing (ZF) receiver [2]. The ML receiver performs vector decoding and is the optimal receiver for systems that utilize the channel knowledge only at the receiver, although it has very high complexity. In fact, ML receiver achieves full receive diversity for uncoded systems regardless of the number of streams transmitted [2]. Recently the ML receiver has been simplified by a technique known as sphere decoding [3], [4]. The complexity of this technique is much less than the ML receiver but can still be significant [5]. On the other hand, MMSE and ZF receivers are easy to implement, but their performance are inferior to the performance of the ML receiver. None of these systems employs channel state information at the transmitter (CSIT).

Clearly, the presence of CSIT can improve overall performance significantly. A technique that provides high diversity and coding gain with the help of CSIT is known as beamforming [6]. The optimum such technique (in terms of the number of channels) is singular value decomposition (SVD). SVD separates the MIMO channel into parallel subchannels. Therefore, multiple streams of data can be transmitted easily. Single beamforming (i.e., sending one symbol at a time) was shown to achieve the maximum diversity in space with a substantial coding gain compared to space-time codes [7]. If more than one symbol at a time are transmitted, then the technique is called multiple beamforming. It can be expected that there would be a tradeoff between spatial multiplexing and diversity order in such systems. In fact, for uncoded multiple beamforming systems using uniform power allocation, while the data rate increases, one loses the diversity order with the increasing number of streams used over flat fading channels [8].

Bit interleaved coded modulation (BICM) was introduced as a way to increase the code diversity [9], [10]. BICM has been deployed with OFDM and MIMO-OFDM systems to achieve high diversity orders [11], [12], [13], [14]. In Section II, we analyze bit interleaved coded multiple beamforming (BICMB).
We show that with the inclusion of BICM to the system, one actually does not need to lose the diversity order with multiple beamforming even when all the subchannels are used. That is, in Section III we show that BICMB achieves full diversity order of \( NM \), and full spatial multiplexing order\(^1\) of \( \min(N,M) \) for a system with \( N \) transmit and \( M \) receive antennas over Rayleigh flat fading channels. We provide design criteria for the interleaver which guarantee full diversity and full spatial multiplexing.

If there is frequency selectivity in the channel, then BICMB is combined with OFDM in order to combat intersymbol interference (ISI). In Section V we show that BICMB-OFDM achieves full diversity order of \( NML \) and full spatial multiplexing order of \( \min(N,M) \) for a system with \( N \) transmit and \( M \) receive antennas over \( L \)-tap frequency selective channels, when an appropriate convolutional code is used.

We would like to reiterate that we assume perfect channel state information at the transmitter and at the receiver. As will be shown in Section VI, the systems investigated here provide substantial performance or complexity gains. It may be expected that there will be significant improvement even in the presence of channel estimation errors and limited feedback.

**Notation:** \( N \) is the number of transmit antennas, \( M \) is the number of receive antennas, \( K \) is the number of subcarriers within one OFDM symbol, and \( L \) is the number of taps in a frequency selective channel. The minimum Hamming distance of a convolutional code is denoted as \( d_{\text{free}} \). The symbol \( S \) denotes the total number of symbols transmitted at a time (spatial multiplexing order), in other words the total number of streams used. The minimum Euclidean distance between two constellation points is given by \( d_{\text{min}} \). The superscripts \( (\cdot)^H \), \( (\cdot)^T \), \( (\cdot)^* \), \( (\cdot) \), and the symbol \( \forall \) denote the Hermitian, transpose, complex conjugate, binary complement, and for-all respectively.

## II. BIT INTERLEAVED CODED MULTIPLE BEAMFORMING (BICMB): SYSTEM MODEL

BICMB is a combination of BICM and multiple beamforming. The output bits of a binary convolutional encoder are interleaved and then mapped over a signal set \( \mathcal{C} \) of size \( |\mathcal{C}| = 2^m \) with a binary labeling map \( \mu : \{0,1\}^m \rightarrow \mathcal{C} \). The \( d_{\text{free}} \) of the convolutional encoder should satisfy \( d_{\text{free}} \geq S \). The interleaver is designed such that the consecutive coded bits are

1) transmitted over different subchannels that are created by beamforming.
2) the code and the interleaver should be picked such that each subchannel created by SVD is utilized at least once within \( d_{\text{free}} \) distinct bits between different codewords.

\(^{1}\)In this paper we use the term “spatial multiplexing” or “spatial multiplexing order” to describe the number of spatial subchannels, as in [2]. It should be noted that this term is different than “spatial multiplexing gain” defined in [15].
The reasons for the interleaver design are given in Section III. Gray encoding is used to map the bits onto symbols. Since Gray encoding allows independent decoding of each bit [16], a Viterbi decoder is deployed at the receiver. During transmission, the code sequence $c$ is interleaved by $\pi$, and then mapped onto the signal sequence $\mathbf{x} \in \chi$.

Beamforming separates the MIMO channel into parallel subchannels. The beamforming vectors used at the transmitter and the receiver can be obtained by the singular value decomposition (SVD) [17] of the MIMO channel. Let $\mathbf{H}$ denote the quasi-static, Rayleigh flat fading $M \times N$ MIMO channel. Then the SVD of $\mathbf{H}$ can be written as

$$
\mathbf{H} = \mathbf{U} \Lambda \mathbf{V}^H = [u_1 \ u_2 \ldots u_M] \Lambda [v_1 \ v_2 \ldots v_N]^H
$$

where $\mathbf{U}$ and $\mathbf{V}$ are $M \times M$ and $N \times N$ unitary matrices, respectively, and $\Lambda$ is an $M \times N$ diagonal matrix with singular values of $\mathbf{H}$, $\lambda_i \in \mathbb{R}$, on the main diagonal with decreasing order. If $S$ symbols are transmitted at the same time, then the $S \times M [u_1 \ u_2 \ldots u_S]^H$ and the $N \times S [v_1 \ v_2 \ldots v_S]$ matrices are employed at the receiver and the transmitter, respectively. The system input-output relation at the $k^{th}$ time instant can be written as

$$
\mathbf{y}_k = [u_1 \ u_2 \ldots u_S]^H \mathbf{H} [v_1 \ v_2 \ldots v_S] \mathbf{x}_k + [u_1 \ u_2 \ldots u_S]^H \mathbf{n}_k
$$

where $\mathbf{x}_k$ is an $S \times 1$ vector of transmitted symbols, $\mathbf{y}_k$ is an $S \times 1$ vector of the received symbols, $\mathbf{n}_k$ is an $M \times 1$ additive white Gaussian noise with zero-mean and variance $N_0 = N/\text{SNR}$. Note that the total power transmitted is scaled as $N$. The channel elements $h_{nm}$ are modeled as zero-mean, unit-variance complex Gaussian random variables. Consequently, the received signal-to-noise ratio is $\text{SNR}$. Uniform power allocation is deployed for each subchannel. An adaptive modulation and coding scheme for BICMB was introduced in [18].

For an uncoded multiple beamforming system using uniform power allocation, if $S$ symbols are transmitted at a time, then the diversity order is equal to $(N - S + 1)(M - S + 1)$ [19].

The bit interleaver of BICMB can be modeled as $\pi : k' \rightarrow (k, s, i)$ where $k'$ denotes the original ordering of the coded bits $c_{k'}$, $k$ denotes the time ordering of the signals $x_{k,s}$ transmitted, $s$ denotes the subchannel used to transmit $x_{k,s}$, and $i$ indicates the position of the bit $c_{k'}$ on the symbol $x_{k,s}$. 
Let $\chi^i_b$ denote the subset of all signals $x \in \chi$ whose label has the value $b \in \{0, 1\}$ in position $i$. Then, the ML bit metrics can be given by using (3), [9], [10]

$$\gamma^i(y_{k,s}, c_{k'}) = \min_{x \in \chi^i_{k'}} |y_{k,s} - \lambda_s x|^2. \quad (4)$$

The ML decoder at the receiver can make decisions according to the rule

$$\hat{c} = \arg\min_{c' \in C} \sum_{k'} \gamma^i(y_{k,s}, c_{k'}). \quad (5)$$

III. BICM: Pairwise Error Probability (PEP) Analysis

In this section we show that by using BICM and the given interleaver design criteria, coded multiple beamforming can achieve full spatial diversity order of $NM$ while transmitting $S \leq \min(N, M)$ symbols at a time. Assume the code sequence $c$ is transmitted and $\hat{c}$ is detected. Then, using (4) and (5), the PEP of $c$ and $\hat{c}$ given CSI can be written as

$$P(c \rightarrow \hat{c}|H) = P\left(\sum_{k'} \min_{x \in \chi^i_{k'}} |y_{k,s} - \lambda_s x|^2 \geq \sum_{k'} \min_{x \in \chi^i_{k'}} |y_{k,s} - \lambda_s x|^2\right) \quad (6)$$

where $s \in \{1, 2, \ldots, S\}$.

For a convolutional code, the Hamming distance between $c$ and $\hat{c}$, $d(c - \hat{c})$, is at least $d_{\text{free}}$. For $c$ and $\hat{c}$ under consideration for PEP analysis, assume $d(c - \hat{c}) = d_{\text{free}}$. Then, $\chi^i_{c_{k'}}$ and $\chi^i_{\hat{c}_{k'}}$ are equal to one another for all $k'$ except for $d_{\text{free}}$ distinct values of $k'$. Therefore, the inequality on the right hand side of (21) shares the same terms on all but $d_{\text{free}}$ summation points. Hence, the summations can be simplified to only $d_{\text{free}}$ terms for PEP analysis.

$$P(c \rightarrow \hat{c}|H) = P\left(\sum_{k',d_{\text{free}}} \min_{x \in \chi^i_{k'}} |y_{k,s} - \lambda_s x|^2 \geq \sum_{k',d_{\text{free}}} \min_{x \in \chi^i_{k'}} |y_{k,s} - \lambda_s x|^2\right) \quad (7)$$

where $\sum_{k',d_{\text{free}}}$ denotes that the summation is taken with index $k'$ over $d_{\text{free}}$ different values of $k'$.

Note that for binary codes and for the $d_{\text{free}}$ points at hand, $\hat{c}_{k'} = \hat{c}_{k'}$. For the $d_{\text{free}}$ bits let’s denote

$$\tilde{x}_{k,s} = \arg\min_{x \in \chi^i_{c_{k'}}} |y_{k,s} - \lambda_s x|^2 \quad \tilde{x}_{k,s} = \arg\min_{x \in \chi^i_{\hat{c}_{k'}}} |y_{k,s} - \lambda_s x|^2. \quad (8)$$

It is easy to see that $\tilde{x}_{k,s} \neq \tilde{x}_{k,s}$ since $\tilde{x}_{k,s} \in \chi^i_{c_{k'}}$ and $\tilde{x}_{k,s} \in \chi^i_{\hat{c}_{k'}}$, where $\chi^i_{c_{k'}}$ and $\chi^i_{\hat{c}_{k'}}$ are complementary sets of constellation points within the signal constellation set $\chi$. Also, $|y_{k,s} - \lambda_s x_{k,s}|^2 \geq |y_{k,s} - \lambda_s \tilde{x}_{k,s}|^2$ and $x_{k,s} \in \chi^i_{c_{k'}}$. 

For convolutional codes, due to their trellis structure, \( d_{\text{free}} \) distinct bits between any two codewords occur in consecutive trellis branches. Let’s denote \( d \) such that \( d_{\text{free}} \) bits occur within \( d \) consecutive bits. The bit interleaver can be designed such that \( d \) consecutive coded bits are mapped onto distinct symbols (interleaver design criterion 1). This guarantees that there exist \( d_{\text{free}} \) distinct pairs of \((\tilde{x}_{k,s}, \hat{x}_{k,s})\), and \( d_{\text{free}} \) distinct pairs of \((x_{k,s}, \hat{x}_{k,s})\). The PEP can be rewritten as

\[
P(\mathbf{c} \rightarrow \hat{\mathbf{c}} | \mathbf{H}) = P \left( \sum_{k,d_{\text{free}}} |y_{k,s} - \lambda_s \tilde{x}_{k,s}|^2 - |y_{k,s} - \lambda_s \hat{x}_{k,s}|^2 \geq 0 \right)
\]

\[
\leq P \left( \sum_{k,d_{\text{free}}} |y_{k,s} - \lambda_s x_{k,s}|^2 \geq \sum_{k,d_{\text{free}}} |y_{k,s} - \lambda_s \hat{x}_{k,s}|^2 \right)
\]

\[
= P \left( \sum_{k,d_{\text{free}}} |n_{k,s}|^2 \geq \sum_{k,d_{\text{free}}} |\lambda_s(x_{k,s} - \hat{x}_{k,s}) + n_{k,s}|^2 \right)
\]

\[
= P \left( \beta \geq \sum_{k,d_{\text{free}}} |\lambda_s(x_{k,s} - \hat{x}_{k,s})|^2 \right) \leq Q \left( \sqrt{\frac{d_{\text{min}}^2 \sum_{s=1}^S \alpha_s \lambda_s^2}{2N_0}} \right) \tag{9}
\]

where \( \beta = \sum_{k,d_{\text{free}}} \beta_{k,s} \), \( \beta_{k,s} = \lambda_s(\hat{x}_{k,s} - x_{k,s})^* n_{k,s} + \lambda_s(\hat{x}_{k,s} - x_{k,s}) n_{k,s}^* \), \( \alpha_s \) denotes how many times the \( s^{\text{th}} \) subchannel is used within the \( d_{\text{free}} \) bits under consideration, and \( \sum_{s=1}^S \alpha_s = d_{\text{free}} \). For given \( \mathbf{H} \), \( \beta_{k,s} \) are independent zero-mean Gaussian random variables with variance \( 2N_0|\lambda_s(\hat{x}_{k,s} - x_{k,s})|^2 \). Consequently, \( \beta \) is a Gaussian random variable with zero mean and variance \( 2N_0 \sum_{k,d_{\text{free}}} |\lambda_s(\hat{x}_{k,s} - x_{k,s})|^2 \).

If the interleaver is designed such that the consecutive coded bits are not spread over different subchannels created by beamforming, then the performance is dominated by the worst singular value. In other words, the error event on the trellis occurs on consecutive branches spanned by the worst subchannel, and \( \alpha_s = d_{\text{free}} \). This results in a diversity order of \( (N - S + 1)(M - S + 1) \) as in uncoded multiple beamforming. On the other hand, by spreading the consecutive coded bits over subchannels, bits that are transmitted over better subchannels can do better error correcting on nearby bits that are transmitted over worse subchannels (interleaver design criterion 1). Criteria 1 and 2 guarantee that \( \alpha_s \geq 1 \), for \( s = 1, 2, \ldots, S \).
Using an upper bound for the $Q$ function $Q(x) \leq (1/2)e^{-x^2/2}$, PEP can be upper bounded as

$$P(\xi \rightarrow \hat{\xi}) = E[P(\xi \rightarrow \hat{\xi}|H)] \leq E \left[ \frac{1}{2} \exp \left( -\frac{d_{\min}^2 \sum_{s=1}^{S} \alpha_s \lambda_s^2}{4N_0} \right) \right].$$

(10)

Let’s denote $\alpha_{\min} = \min\{\alpha_s : s = 1, 2, \ldots, S\}$. Then,

$$\left( \sum_{s=1}^{S} \alpha_s \lambda_s^2 \right) / S \geq \left( \alpha_{\min} \sum_{s=1}^{S} \lambda_s^2 \right) / S \geq \left( \alpha_{\min} \sum_{s=1}^{N} \lambda_s^2 \right) / N.$$

(11)

Note that,

$$\Theta \overset{\Delta}{=} \sum_{s=1}^{N} \lambda_s^2 = \|H\|^2_F = \sum_{n,m} |h_{n,m}|^2$$

(12)

is a chi-squared random variable with $2NM$ degrees of freedom (the elements of $H$, $h_{n,m}$, are complex Gaussian random variables). Using (10), (11), and (12) the PEP is upper bounded by

$$P(\xi \rightarrow \hat{\xi}) \leq E \left[ \frac{1}{2} \exp \left( -\frac{d_{\min}^2 \alpha_{\min} S}{4N_0N} \Theta \right) \right].$$

(13)

The expectation in (13) is evaluated with respect to $\Theta$ with pdf $f_\Theta(\theta) = \theta^{(NM-1)}e^{-\theta/2}/2^{NM}(NM-1)!$ [20]. Consequently,

$$P(\xi \rightarrow \hat{\xi}) = g(d, \alpha_{\min}, \chi) \leq \frac{1}{2^{NM+1}} \left( \frac{d_{\min}^2 \alpha_{\min} S}{4N_0N} + \frac{1}{2} \right)^{-NM}$$

$$\approx \frac{1}{2^{NM+1}} \left( \frac{d_{\min}^2 \alpha_{\min} S}{4N^2 SNR} \right)^{-NM}$$

(14)

(15)

for high SNR. The function $g(d, \alpha_{\min}, \chi)$ denotes the PEP of two codewords with $d(\xi - \hat{\xi}) = d$, with $\alpha_{\min}$ corresponding to $\xi$ and $\hat{\xi}$, and with constellation $\chi$. Note that the PEP function $g(\cdot)$ depends on $\alpha_{\min}$ as well as the distance $d$. That is, a set of codewords all of which are at Hamming distance $d$ from one another can have different PEPs depending on the corresponding $\alpha_{\min}$ between any two codewords.

In the case of BICM with a rate $k_c/n_c$ binary convolutional code, the bit error rate (BER), $P_b$, can be bounded as

$$P_b \leq \frac{1}{k_c} \sum_{d=d_{\text{free}}}^{\infty} W_I(d) g_0(d, \mu, \chi)$$

(16)

where $W_I(d)$ denotes the total input weight of error events at Hamming distance $d$, and $g_0(\cdot)$ is the PEP
of two codewords with $d(\mathbf{c} - \mathbf{\hat{c}}) = d$, $\mu$ is a constellation labeling map, and $\chi$ is the constellation [10]. Since we have a fixed, Gray-encoded constellation labeling map, $\mu$ can be ignored. Needless to say, $g_0(\cdot)$ and $g(\cdot)$ are two different functions. In BICMB, $P_b$ can be calculated as

$$P_b \leq \frac{1}{k_c} \sum_{d=d_{free}}^{\infty} \sum_{i=1}^{W(d)} g(d, \alpha_{\min}(d, i), \chi).$$

(17)

For example, for the industry standard 1/2 rate, 64-state (133,171) convolutional code $W_I(d = d_{free} = 10) = 11$. Depending on the interleaver used, 11 codewords at a Hamming distance of 10 from the all-zero codeword may each have a different $\alpha_{\min}$. Therefore, we deviated from the usual notation for the union bound of convolutional codes of (16) to the one given in (17) with an extra summation inside specifically distinguishing the different $\alpha_{\min}$s for the codewords at a Hamming distance $d$. Note that the union bound in (17) provides a loose bound for quasi-static channels, and a limiting and averaging method should be used for a tighter bound [21], [22]. Nevertheless, the union bound is a very useful tool to provide an insight for the asymptotic behavior of the system, and therefore the diversity order. In this paper our goal is to provide the diversity order of the BICMB system when multiple streams of data are transmitted over multiple antennas, rather than providing tight bounds. Following (15) and (17)

$$P_b \leq \frac{1}{k_c} \sum_{d=d_{free}}^{\infty} \sum_{i=1}^{W(d)} \frac{1}{2^{NM+1}} \left( \frac{d_{\min}^2 \alpha_{\min}(d, i) S}{4N^2 SNR} \right)^{-NM}$$

(18)

As can be seen from (18), for all the summations, the $SNR$ component has a power of $-NM$. Consequently, BICMB achieves full diversity order of $NM$ independent of the number of spatial streams transmitted.

IV. BICMB-OFDM: SYSTEM MODEL

In order to combat the ISI in frequency selective channels, we combined BICMB with OFDM (BICMB-OFDM). The system model is similar to BICMB with few minor differences as given in this section. The multiplication with beamforming vectors are carried at each subcarrier before IFFT at the transmitter and after FFT at the receiver. The interleaver is designed such that the consecutive coded bits are

1) interleaved within one MIMO-OFDM symbol to avoid extra delay requirement to start decoding at the receiver,

2) transmitted over different subcarriers of an OFDM symbol,

3) transmitted over different subchannels that are created by beamforming.
By adding cyclic prefix (CP), OFDM converts the frequency selective channel into parallel flat fading channels for each subcarrier. Let $H(k)$ denote the quasi-static, flat fading $M \times N$ MIMO channel observed at the $k^{th}$ subcarrier, and $h_{mn} = [h_{mn}(0) \ h_{mn}(1) \ \cdots \ h_{mn}(L-1)]^T$ represent the $L$-tap frequency selective channel from the transmit antenna $n$ to the receive antenna $m$. Each tap is assumed to be statistically independent and modeled as zero mean complex Gaussian random variable with variance $1/L$. The SVD is formed for each $H(k)$ in order to calculate the beamforming matrices for each subcarrier. If $S$ symbols are transmitted on the same subcarrier over $N$ transmit antennas, then the system input-output relation at the $k^{th}$ subcarrier can be written as

$$
\mathbf{y}(k) = [\mathbf{u}_1(k) \ \mathbf{u}_2(k) \ \cdots \ \mathbf{u}_S(k)]^H \mathbf{H}(k)[\mathbf{v}_1(k) \ \mathbf{v}_2(k) \ \cdots \ \mathbf{v}_S(k)]\mathbf{x}(k) + [\mathbf{u}_1(k) \ \mathbf{u}_2(k) \ \cdots \ \mathbf{u}_S(k)]^H \mathbf{n}(k)
$$

(19)

$$
y_s(k) = \lambda_s(k)x_s(k) + n_s(k)
$$

(20)

for $s = 1, 2, \ldots, S$, and $k = 1, 2, \ldots, K$, where $\lambda_s(k)$ is the $s^{th}$ largest singular value of $H(k)$ and $n(k)$ is the additive white complex Gaussian noise with zero mean and variance $N/\text{SNR}$. The average total power transmitted over all the antennas at each subcarrier is scaled as $N$ such that the received signal-to-noise ratio over all the receive antennas is $\text{SNR}$. Note that the received signal-to-noise ratio at each subchannel for each subcarrier is directly proportional to the corresponding channel gain $\lambda_s(k)^2$.

V. BICMB-OFDM: PEP Analysis

Assume the code sequence $\mathbf{c}$ is transmitted and $\hat{\mathbf{c}}$ is detected. Then, using (20), the PEP of $\mathbf{c}$ and $\hat{\mathbf{c}}$ given CSI can be written as

$$
P(\mathbf{c} \rightarrow \hat{\mathbf{c}}|H(k), \forall k) = P\left(\sum_{k'} \min_{x \in \chi_{x,k'}} |y_s(k) - \lambda_s(k)x|^2 \geq \sum_{k'} \min_{x \in \chi_{x,k'}} |y_s(k) - \lambda_s(k)x|^2\right)
$$

(21)

where $s \in \{1, 2, \ldots, S\}$.

Similar to Section III, for $d_{\text{free}}$ bits under consideration for the PEP analysis, let’s denote

$$
\tilde{x}_s(k) = \arg \min_{x \in \chi_{x,k}} |y_s(k) - \lambda_s(k)x|^2 \quad \quad \hat{x}_s(k) = \arg \min_{x \in \chi_{x,k}} |y_s(k) - \lambda_s(k)x|^2.
$$

(22)

It is easy to see that $\tilde{x}_s(k) \neq \hat{x}_s(k)$ since $\tilde{x}_s(k) \in \chi_{\tilde{x}_s,k}$ and $\hat{x}_s(k) \in \chi_{\hat{x}_s,k}$, where $\chi_{\tilde{x}_s,k}$ and $\chi_{\hat{x}_s,k}$ are complementary sets of constellation points within the signal constellation set $\chi$. Also, $|y_s(k) - \lambda_s(k)x_s(k)|^2 \geq |y_s(k) - \lambda_s(k)\tilde{x}_s(k)|^2$ and $x_s(k) \in \chi_{x,k}$.

Interleaver design criteria 2 and 3 suggest that the bit interleaver should be designed such that $d$
consecutive coded bits are mapped onto distinct symbols and onto distinct subcarriers. This guarantees that there exist \(d_{\text{free}}\) distinct pairs of \((\hat{x}_s(k), \hat{x}_s(k))\), and \(d_{\text{free}}\) distinct pairs of \((x_s(k), \hat{x}_s(k))\) with \(d_{\text{free}}\) distinct values of \(k\). The PEP can be rewritten as

\[
P(\mathfrak{c} \rightarrow \mathfrak{c} | \mathbf{H}(k), \forall k) = P \left( \sum_{k, d_{\text{free}}} |y_s(k) - \lambda_s(k) \hat{x}_s(k)|^2 - |y_s(k) - \lambda_s(k) \hat{x}_s(k)|^2 \geq 0 \right)
\]

for some known \(s\) at each subcarrier \(k\), where \(\beta = \sum_{k, d_{\text{free}}} \lambda_s(k) (\hat{x}_s(k) - x_s(k))^* n_s(k) + \lambda_s(k) (\hat{x}_s(k) - x_s(k)) n_s^*(k)\). For given \(\mathbf{H}(k), \forall k\), \(\beta\) is a Gaussian random variable with zero mean and variance \(2N_0 \sum_{k, d_{\text{free}}} |\lambda_s(k)(\hat{x}_s(k) - x_s(k))|^2\).

The interleaver can be designed such that the consecutive coded bits are transmitted on different subchannels (interleaver design criterion 3). This way, on the trellis, within the \(d_{\text{free}}\) bits under consideration, coded bits that are transmitted on better subchannels can provide better error correcting on the neighboring bits that are transmitted on worse subchannels. Using an upper bound for the \(Q\) function \(Q(x) \leq (1/2)e^{-x^2/2}\), PEP can be upper bounded as

\[
P(\mathfrak{c} \rightarrow \mathfrak{c}) = E[P(\mathfrak{c} \rightarrow \mathfrak{c} | \mathbf{H}(k), \forall k) \leq E \left[ \frac{1}{2} \exp \left( - \frac{\sum_{k, d_{\text{free}}} \lambda_s(k)^2}{4N_0} \right) \right].
\]

Assuming high frequency selectivity in the channel, \(\lambda_s(k)s\) are independent for different \(k\), and identically distributed for the same \(s\). Let’s denote \(\mu_s(k) = \lambda_s(k)^2\), the marginal pdfs of each \(\mu_s(k)\) as \(f(\mu_s(k))\), and \(\alpha_s\) as the number of times the \(s^{th}\) channel is used within \(d_{\text{free}}\) bits under consideration such that \(\sum_{s=1}^{S} \alpha_s = d_{\text{free}}\). Note that, criterion 3 guarantees \(\alpha_s \geq 1, \forall s\). The expectation in (25) can be evaluated using the marginal pdfs as

\[
P(\mathfrak{c} \rightarrow \mathfrak{c}) \leq \frac{1}{2} \prod_{s=1}^{S} \left[ \int_{0}^{\infty} \exp \left( - \frac{\mu_s(k)}{4N_0} \right) f(\mu_s(k)) d\mu_s(k) \right]^{\alpha_s}.
\]

Since the instantaneous received signal to noise ratio for each \(s\) and each \(k\) depends directly on \(\mu_s(k)\), the diversity and coding gains for average bit error rate at high \(SNR\) depend only on the behavior of \(f(\mu_s(k))\) around the origin \(\mu_s(k) = 0\) [23], [24]. Using a Taylor series expansion around 0, the first order
approximation of the marginal pdf of $\mu_s(k)$ is given by [8], [19], [24], [25]

$$f(\mu_s(k)) = \kappa_s \mu_s(k)^{(N-s+1)(M-s+1)-1}$$  \hspace{1cm} (27)

where $\kappa_s$ is a constant [24]. Consequently, (26) can be calculated as

$$P(\mathbf{c} \rightarrow \mathbf{\hat{c}}) = g(d, \alpha, \chi) \leq \frac{1}{2} \prod_{s=1}^{S} \gamma_s^{\alpha_s} \left( \frac{d_{\min}^2}{4NSNR} \right)^{-\alpha_s(N-s+1)(M-s+1)}$$  \hspace{1cm} (28)

where $\gamma_s$ is a constant, which depends on $\kappa_s$ [24]. The function $g(d, \alpha, \chi)$ denotes the PEP of two codewords that are at a Hamming distance of $d$ from one another with the corresponding vector $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_S]$. The coefficients, $\alpha_s$ for $s = 1, \ldots, S$, are calculated depending on the codewords $\mathbf{c}$ and $\mathbf{\hat{c}}$, $d(\mathbf{c} - \mathbf{\hat{c}})$, and the interleaver used. Note that the function $g(\cdot)$ changes for different $\alpha$. Therefore, the PEPs of a set of codewords (all of which are at a distance $d$ from one another) can be different depending on the corresponding $\alpha$. In a similar fashion to Section III, we use the usual union bound to illustrate the diversity order of the system. The bit error rate $P_b$ can be calculated as

$$P_b \leq \frac{1}{k_{c}} \sum_{d=d_{\text{free}}}^{\infty} \sum_{i=1}^{W_I(d)} g(d, \alpha(d, i), \chi).$$  \hspace{1cm} (29)

For a set of codewords at a Hamming distance $d$, there may be a different vector $\alpha$. $\alpha(d, i) = [\alpha_1(d, i), \alpha_2(d, i), \ldots, \alpha_S(d, i)]$ denotes the vector $\alpha$ for a codeword at a distance $d$ from the all-zero codeword with the given interleaver. For the same Hamming distance of $d$ from the all-zero codeword, there are a total of $W_I(d)$ different codewords, and therefore there may be $W_I(d)$ different $\alpha(d, i)$ vectors (some of which could be the same). Let’s define

$$\Delta(\alpha(d, i)) = \sum_{s=1}^{S} \alpha_s(d, i)(N - s + 1)(M - s + 1)$$  \hspace{1cm} (30)

$$\alpha(d_{\text{free}}, j) = \arg \min_{i=1, \ldots, W_I(d_{\text{free}})} \Delta(\alpha(d_{\text{free}}, i)).$$  \hspace{1cm} (31)

Note that $\Delta(\alpha(d_{\text{free}}, j))$ is the minimum for all $d \geq d_{\text{free}}$, since convolutional codes are trellis-based and
for any \( d > d_{\text{free}} \), \( \Delta(\alpha(d, i)) \geq \Delta(\alpha(d_{\text{free}}, j)) \). Combining (28) - (31), \( P_b \) can be calculated as

\[
P_b \leq \frac{1}{2k_c} \prod_{s=1}^{S} \gamma_s(d_{\text{free}}, j) \left( \frac{d_{\text{free}}^2}{4N} \right)^{-\alpha_s(d_{\text{free}}, j)(N-s+1)(M-s+1)} + \frac{1}{k_c} \sum_{d=d_{\text{free}}}^{\infty} \sum_{i=1}^{N} \sum_{j=1}^{M} g(d, \alpha(d, i), \chi) \tag{32}
\]

For asymptotically high \( SNR \), the diversity order of a system is determined by the smallest power of \( SNR \), since the higher order terms yield to zero faster with increasing \( SNR \). Consequently, BICMB-OFDM provides a diversity order of \( \Delta(\alpha(d_{\text{free}}, j)) \) for a spatial multiplexing order of \( S \). Note that, if the interleaver design criterion 3 is not met, then the maximum diversity order reduces to \( (N-S+1)(M-S+1)d_{\text{free}} \) for spatial multiplexing order of \( S \) with \( \alpha_S(d_{\text{free}}, j) = d_{\text{free}} \). It is known that the maximum diversity order of MIMO systems over \( L \)-tap frequency selective channels is \( NML \) [26], [27]. As will be shown in Section VI, BICMB-OFDM achieves full diversity order of \( NML \) when \( NML \leq \Delta(\alpha(d_{\text{free}}, j)) \) for spatial multiplexing order of \( S \).

A very low complexity decoder for BICM-OFDM can be implemented as in [28], [29]. The same decoder can be used for BICMB and BICMB-OFDM as well: Instead of using the single-input single-output (SISO) channel value of BICM-OFDM for the decoder ([28] and [29]), one should use \( \lambda_s \) for BICMB, and \( \lambda_s(k) \) for BICMB-OFDM. Hence, BICMB and BICMB-OFDM provide full spatial multiplexing, full diversity, and easy-to-decode systems.

VI. SIMULATION RESULTS

In the simulations below, the industry standard 64-state 1/2-rate (133,171) \( d_{\text{free}} = 10 \) convolutional code is used. For BICMB, coded bits are separated into different streams of data and a random interleaver is used to interleave the bits in each substream. BICMB-OFDM deploys the interleaver given in [30]. The interleavers satisfy the design criteria of Sections II and IV. Each packet has 1000 bytes of information bits, and the channel is changed independently from packet to packet. Each OFDM symbol has 64 subcarriers, and has 4 \( \mu s \) duration, of which 0.8 \( \mu s \) is CP. All the comparisons below are carried at \( 10^{-5} \) BER.

A. BICMB

Figure 1 illustrates the results for BICMB with QPSK and with 4 transmit and 4 receive antennas. Note that when \( S = 4 \), and with 1/2-rate convolutional code, BICMB transmits 4 bits/s/Hz. Also note that all the curves have the same slope for high SNR. One can verify by using simulation results with
e.g., a 4 × 4 1/2-rate complex orthogonal STBC that our system achieves full spatial diversity order of 16 regardless of the number of streams transmitted simultaneously. A comparison with the systems in [31], [32] shows the same results for diversity. The systems in [31], [32] have comparable performance without CSIT, employing sphere decoding, which has significantly higher complexity than our system. The prospect of a high-performing, full diversity and maximum spatial multiplexing system without CSIT is very appealing. But, when used in an $N \times N$ MIMO system, sphere decoding results in a complexity of $O(A\mu N^2)$ where $A$ is the constellation size and $\mu$ is a number between 0 and 1, close to 1 for low SNR [5]. As a result, the complexity of such a system, although much lower than ML, is still high and is dependent on SNR.

A comparison of the 2 × 2, 3 × 3, and 4 × 4 cases, with full spatial multiplexing in each case, is given in Figure 2. In all the cases 16 QAM is deployed for transmission. Even though the 4 × 4 system transmits twice the data rate of 2x2 system, the performance of the 4 × 4 system is significantly better than the 2 × 2 system. This is due to the fact that the 4 × 4 system achieves a diversity order of 16 where the 2 × 2 system has a diversity order of 4. Consequently, BICMB provides both advantages of MIMO systems: It provides full diversity and full spatial multiplexing.

Figure 3 illustrates the importance of the interleaver design. We simulated a random interleaver such that consecutive coded bits are transmitted over the same subchannel. In other words, on a trellis path, consecutive bits of length $1/S$th of the coded packet size are transmitted over the same subchannel. Consequently, an error on the trellis occurs over paths that are spanned by the worst channel and the diversity order of coded multiple beamforming approaches to that of uncoded multiple beamforming with uniform power allocation. It is our experience that a straightforward use of the interleaver employed in the 802.11a standard [33] can result in this behavior, especially for $S = 2$ and 4.

Figure 4 shows the simulation results of BICMB when compared to a spatial multiplexing system using BICM at the transmitter and ML, MMSE, and ZF receivers. All the receivers deploy a soft Viterbi decoder. In this paper, for ZF and MMSE, the bit metrics in [34] and [35], respectively, are employed. All the systems have spatial multiplexing order of 2. While ML receiver achieves a high diversity order with substantial complexity, ZF achieves a diversity order of $M - N + 1$ [2], [36]. ML receiver is known as the optimal receiver for a spatial multiplexing system. Using BICM at the transmitter with an interleaver spreading the consecutive bits over the transmit antennas, and deploying ML receiver can be considered as the Vertical Encoding (VE) in [2]. Such a system is capable of providing a high diversity
order. However, as discussed previously, ML receiver has prohibitive complexity while its simplified form sphere decoder has still substantial complexity in real MIMO applications. Therefore, sub-optimal (therefore poorer performance) but easy-to-implement receivers are designed such as MMSE, ZF, SUC, and ordered SUC [2]. As illustrated for the 2x2 case, BICMB outperforms ML receiver by 4.5 dB, while the performance gain compared to MMSE and ZF receivers is more than 25 dB. It is possible that the base station (or the access point) has more antennas than the receiver. BICMB with 4 transmit and 2 receive antennas with spatial multiplexing of 2 outperforms ML receiver by 15.5 dB. Note that, for BICMB, the performance of the $4 \times 2$ and the $2 \times 4$ cases are identical. Therefore, the same high performance is available for both the downlink and the uplink.

We state once again that CSIT is absent in the systems we compare BICMB with whereas BICM employs perfect CSIT. However, the large performance or complexity gains achieved by BICMB leave room for more modest performance gains with channel estimation errors and limited feedback, and may be indicative of practical systems with good performance. Our goal in this paper is to merely quantify absolute performance bounds.

B. BICMB-OFDM

Figure 5 illustrates the results for BICMB-OFDM for different rms delay spread values, when 2 streams of data are transmitted at the same time. The maximum delay spread of the channel is assumed to be ten times the rms delay spread. The channel is modeled as in Section IV, where each tap is assumed to have equal power. The spectrum of (133,171) shows that there are 11 codewords with a Hamming distance of $d_{\text{free}}$ from the all-zero codeword. When compared to the all-zero codeword, the codeword $[111010010101110000000\ldots]$ has the worst performance for BICMB-OFDM. It corresponds to (31). On this codeword the code and the interleaver combination result in $\alpha_1 = 3$, and $\alpha_2 = 7$. Consequently, when $S = 2$, BICMB-OFDM achieves a maximum diversity order of $3NM + 7(N - 1)(M - 1)$ (19 for a $2 \times 2$ system). Note that, on Figure 5 up to an rms delay spread of 15 ns, BICMB-OFDM achieves the maximum diversity with full spatial multiplexing of 2. The $2 \times 2$ system over a 20 ns channel provides a maximum achievable diversity order of 20. Therefore, BICMB-OFDM achieves a diversity order of 19 for rms delay spreads of 20 ns, 25 ns, and 50 ns.

Figures 6 and 7 illustrate the simulation results for BICMB-OFDM and BICM-OFDM with spatial multiplexing (BICM-SM-OFDM) using ML, MMSE, and ZF receivers. In both figures, the spatial multiplexing order is set as 2. The simulations are carried over the IEEE channel models B and D [37],
Note that BICMB-OFDM employs CSI at both the transmitter and the receiver, while ML, MMSE, and ZF employ CSI at the receiver. As can be seen, BICMB-OFDM outperforms significantly high complexity, but best spatial multiplexing receiver, ML, by more than 3.5 dB. Note that the decoding complexity of BICMB-OFDM is substantially lower in complexity than ML receiver. BICMB-OFDM outperforms MMSE and ZF receivers at $10^{-5}$ BER by 6 dB and 7.5 dB, respectively. It is possible that the base station (or the access point) has more antennas than the receiver. BICMB-OFDM with 4 transmit and 2 receive antennas with spatial multiplexing of 2 outperforms ML receiver by 9 dB. Similar to BICMB results, the performance of the $4 \times 2$ and the $2 \times 4$ cases are identical for BICMB-OFDM. Therefore, the same high performance is available for both the downlink and the uplink.

Figure 8 presents the results for the $4 \times 4$ case transmitting 4 streams for BICMB-OFDM, MMSE, and ZF over IEEE channel models B and D. For the $S = 4$ case, when compared to the all-zero codeword, the codeword $[001110010100010101110000 \ldots]$ leads to the worst diversity order. The coefficients are given as $\alpha_1 = 1$, $\alpha_2 = 3$, $\alpha_3 = 2$, $\alpha_4 = 4$, which leads to a maximum diversity order of 55 for the $4 \times 4$ case. BICMB-OFDM outperforms MMSE by 11.5 dB, and ZF by 15 dB at $10^{-5}$ BER for IEEE channel model B.

VII. Conclusion

In this paper, we analyzed bit interleaved coded multiple beamforming (BICMB). BICMB utilizes the channel state information at the transmitter and the receiver. By doing so, BICMB achieves full spatial multiplexing of $\min(N, M)$, while maintaining full spatial diversity of $NM$ over $N$ transmit and $M$ receive antennas. We presented interleaver design guidelines to guarantee full diversity at full spatial multiplexing.

If the channel is frequency selective, then we combined BICMB with OFDM in order to combat ISI. BICMB-OFDM achieves full spatial multiplexing of $\min(N, M)$, while maintaining full spatial and frequency diversity of $NML$ for a $N \times M$ system over $L$-tap frequency selective channels when an appropriate convolutional code is used.

Simulation results also showed that, with perfect CSIT, BICMB and BICMB-OFDM outperform the optimal high complexity ML and easy-to-implement MMSE and ZF receivers substantially which do not employ CSIT. The substantial performance gains may point to practical systems with channel estimation errors and limited feedback whose performance or complexity gains are more modest but still significantly more than conventional systems.
VIII. ACKNOWLEDGEMENTS

The authors would like to thank the anonymous reviewers whose comments improved the quality of the paper.

REFERENCES


[38] IEEE 802.11-03/940r2:TGN Channel Models. IEEE. [Online]. Available: ftp://ieee:wireless@ftp.802wirelessworld.com/11/03/11-03-0940-02-000n-tgn-channel-models.doc

Fig. 1. BICMB with 4 transmit and 4 receive antennas and with different number of streams.

Fig. 2. BICMB transmitting min(N,M) streams with the 2 x 2, 3 x 3, and 4 x 4 cases.
Fig. 3. BICMB transmitting $\min(N, M)$ streams with the $2 \times 2$, and $4 \times 4$ cases using an interleaver meeting and not meeting the design criteria.

Fig. 4. BICMB vs MLD, ZF and MMSE for the $2 \times 2$ case.
Fig. 5. BICMB-OFDM transmitting 2 streams using 2 transmit and 2 receive antennas.

Fig. 6. BICMB-OFDM vs MLD, MMSE, and ZF transmitting 2 streams over IEEE Channel Model B.
Fig. 7. BICMB-OFDM vs MLD, MMSE, and ZF transmitting 2 streams over IEEE Channel Model D.

Fig. 8. BICMB-OFDM vs MMSE, and ZF transmitting 4 streams over IEEE Channel Models B, and D.