An Efficient MMSE-Based Demodulator for MIMO Bit-Interleaved Coded Modulation

Dominik Seethaler, Gerald Matz, and Franz Hlawatsch*

Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology
Gusshausstrasse 25/389, A-1040 Vienna, Austria
phone: +43 1 58801 38958, fax: +43 1 58801 38999, email: dominik.seethaler@tuwien.ac.at

*Currently on leave with Laboratoire des Signaux et Systèmes, École Supérieure d'Électricité
3 Rue Joliot-Curie, F-91190 Gif-sur-Yvette, France

Abstract— In *bit-interleaved coded modulation* (BICM) systems employing maximum-likelihood decoding, a demodulator (demapper) calculates a log-likelihood ratio (LLR) for each coded bit, which is then used as a bit metric for Viterbi decoding. In the MIMO case, the computational complexity of LLR calculation tends to be excessively high, even if the log-sum approximation is used. Thus, there is a strong demand for efficient suboptimum MIMO-BICM demodulation algorithms with near-optimum performance.

Here, we propose an efficient MIMO-BICM demodulator that is derived by means of a Gaussian approximation for the post-detection interference. Our derivation results in an MMSE equalizer followed by per-layer LLR calculation (i.e., LLRs are calculated separately for each layer). The novel demodulator can be interpreted as an MMSE analogue of a recently proposed ZF-equalization based demodulator, as well as an extension of ZF-equalization based demodulation to correlated post-detection interference. Because it performs per-layer LLR calculation, it has the same (low) computational complexity as the ZF-equalization based demodulator. Simulation results demonstrate that the performance of our demodulator is close to that of LLR calculation using all layers jointly, and significantly better than that of the ZF-equalization based demodulator.

I. INTRODUCTION

Bit-interleaved coded modulation (BICM) [1] is a promising scheme for multiple-input multiple-output (MIMO) wireless communications. Specifically, MIMO-BICM has been shown to outperform space-time trellis coding in fast-fading enviroments [2]. This is important because independent fast-fading MIMO channels provide a model for the individual subcarriers of MIMO systems using orthogonal frequency division multiplexing (OFDM) with frequency interleaving (e.g., [2, 3]), and for block-fading channels with temporal interleaving.

In BICM systems employing maximum-likelihood decoding, a demodulator (demapper) has to provide the channel decoder with bit metrics that are given by a log-likelihood ratio (LLR) for each coded bit. In the MIMO case, LLR calculation is usually done by means of a complexity-reducing log-sum approximation [2]; this will be referred to as *optimum log-sum approximation (OLSA) demodulation*. Because the LLRs are calculated jointly for all layers, OLSA demodulation tends to be excessively complex [3]. Thus, there is a strong demand for efficient MIMO-BICM demodulation algorithms whose performance is close to that of OLSA demodulation.

Such a demodulation algorithm with low complexity but near-optimum performance is proposed in this paper. Our derivation of the novel demodulator is based on a Gaussian approximation for the post-detection interference, which is inspired and motivated by [4]. We obtain a structure that consists of a minimum mean-square error (MMSE) equalizer with subsequent *per-layer* LLR calculation. Thus, in contrast to OLSA demodulation, calculation of the LLRs is carried out for each layer separately. The novel demodulator can be seen as an MMSE analogue of the zero-forcing (ZF) based demodulator recently proposed in [3], which uses ZF equalization followed by per-layer LLR calculation.

We will also show that the ZF-based demodulator, too, can be derived with a Gaussian approximation for the post-detection interference; however, in contrast to our approach, the interference has to be assumed uncorrelated. Thus, our demodulator can also be seen as an extension of ZF-based demodulation to correlated post-detection interference, which explains its significant performance advantage over the ZF-based demodulator. Because of the per-layer LLR calculation it employs, our demodulator has the same (low) computational complexity as the ZF-based demodulator.

This paper is organized as follows. In the remainder of this section, we describe the system model and briefly review existing demodulation schemes. The novel demodulator is derived and discussed in Section II. Finally, Section III presents simulation results for fast-fading MIMO channels. It is demonstrated that our demodulator achieves near-OLSA performance for different alphabets and a wide range of SNRs, and that it significantly outperforms ZF-based demodulation.

A. MIMO-BICM System Model

We consider a flat-fading MIMO channel with M_T transmit antennas and $M_\mathrm{R} \geq M_\mathrm{T}$ receive antennas (briefly termed an $(M_\mathrm{T}, M_\mathrm{R})$ channel). We assume a spatial multiplexing system where for any given time instant n, the kth data stream $d_k[n]$ is directly transmitted on the kth transmit antenna (or *layer*). This leads to the well-known baseband model

$$\mathbf{r}[n] = \mathbf{H}[n]\mathbf{d}[n] + \mathbf{w}[n], \qquad n = 0, \ldots, N-1, \quad (1)$$

with the transmitted data symbol vector $\mathbf{d}[n] \triangleq (d_1[n] \cdots d_{M_\mathrm{T}}[n])^T$, the $M_\mathrm{R} \times M_\mathrm{T}$ channel matrix $\mathbf{H}[n]$, the received
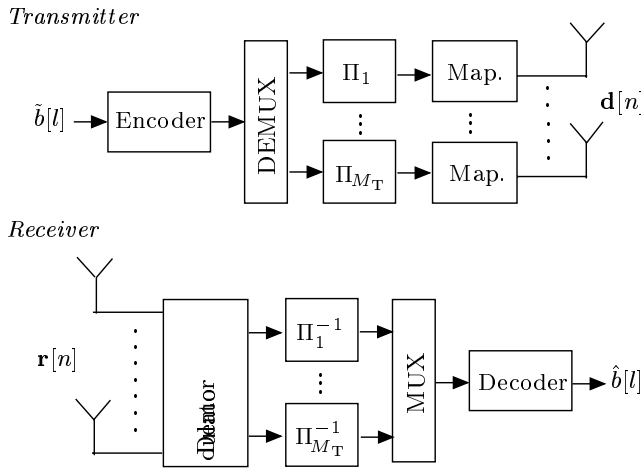
Fig. 1. *MIMO system using bit-interleaved coded modulation.*

vector $\mathbf{r}[n] \triangleq \left(r_1[n]\cdots r_{M_{\mathrm{R}}}[n]\right)^T$, and the noise vector $\mathbf{w}[n] \triangleq \left(w_1[n]\cdots w_{M_{\mathrm{R}}}[n]\right)^T$. The data symbols $d_k[n]$ are drawn from a complex symbol alphabet $\mathcal{A}$ and are assumed zero-mean with unit variance. The noise components $w_k[n]$ are assumed independent and circularly symmetric complex Gaussian with variance σ_w^2.

According to the MIMO-BICM system model of [2] (see Fig. 1), a sequence of information bits $\tilde{b}[l]$ is encoded using a convolutional code and cyclically demultiplexed into M_{T} layers. The coded bits of the kth layer are interleaved (using an interleaver Π_k) and subsequently mapped (using Gray labeling [1]) onto complex data symbols $d_k[n] \in \mathcal{A}$. At the receiver, the demodulator uses the received vector $\mathbf{r}[n]$ and knowledge of the channel $\mathbf{H}[n]$ to calculate an LLR $\Lambda_k^{(i)}$ for each coded bit $b_k^{(i)}$ of the symbol vector $\mathbf{d}[n]$. The resulting sequences of LLRs are deinterleaved (using deinterleavers Π_k^{-1}) and multiplexed into a single stream, which is then used for soft-in Viterbi decoding (e.g., [5]).

In the following, we concentrate on the demodulator that calculates the LLRs for a given symbol time index n. For simplicity of notation, we will omit the time index and, thus, write (1) as $\mathbf{r} = \mathbf{Hd} + \mathbf{w}$. Let $b_k^{(i)}$ with $i = 1, \ldots, \log_2|\mathcal{A}|$ denote the coded bits of the kth layer, to which the symbol $d_k \in \mathcal{A}$ is associated via Gray labeling. We assume that the code bits $b_k^{(i)}$ are equally likely and statistically independent.

B. Review of MIMO-BICM Demodulation Algorithms

OLSA Demodulator. The LLR of $b_k^{(i)}$ is given by

$$\Lambda_k^{(i)} \triangleq \log\left(\frac{f(\mathbf{r}|b_k^{(i)}=1)}{f(\mathbf{r}|b_k^{(i)}=0)}\right)$$

$$= \log\left(\frac{\sum_{\mathbf{d}:d_k\in\mathcal{A}_1^i} e^{-\frac{1}{\sigma_w^2}\|\mathbf{r}-\mathbf{Hd}\|^2}}{\sum_{\mathbf{d}:d_k\in\mathcal{A}_0^i} e^{-\frac{1}{\sigma_w^2}\|\mathbf{r}-\mathbf{Hd}\|^2}}\right), \qquad (2)$$

where $\mathcal{A}_b^i \subset \mathcal{A}$ denotes the set of all symbols $a \in \mathcal{A}$ whose label has $b \in \{0,1\}$ in bit position i. With OLSA demodulation,

the log-sum approximation is used to calculate the following approximate LLRs (e.g., [2]):

$$\Lambda_k^{(i)} \approx \frac{1}{\sigma_w^2}\left[\min_{\mathbf{d}:d_k\in\mathcal{A}_0^i}\|\mathbf{r}-\mathbf{Hd}\|^2 - \min_{\mathbf{d}:d_k\in\mathcal{A}_1^i}\|\mathbf{r}-\mathbf{Hd}\|^2\right], \quad (3)$$

which form the input to a soft-in Viterbi decoder. OLSA demodulation achieves near-optimum performance. However, its computational complexity is exponential in M_{T} because the LLRs are calculated jointly for all layers, requiring the computation of $|\mathcal{A}|^{M_{\mathrm{T}}}$ distances.

ZF-Equalization Based Demodulator. A recently proposed alternative demodulation scheme [3] uses ZF equalization with subsequent *per-layer* LLR calculation (i.e., LLRs are calculated separately for each layer). The ZF-equalized received vector is given by

$$\mathbf{y}_{\mathrm{ZF}} = \mathbf{H}^{\#}\mathbf{r}, \qquad (4)$$

where $\mathbf{H}^{\#} \triangleq (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H$ is the pseudo-inverse [6] of $\mathbf{H}$. One obtains

$$\mathbf{y}_{\mathrm{ZF}} = \mathbf{d} + \tilde{\mathbf{w}}, \qquad (5)$$

i.e., the transmitted data vector $\mathbf{d}$ plus a transformed noise vector $\tilde{\mathbf{w}} \triangleq \mathbf{H}^{\#}\mathbf{w}$ whose covariance matrix is

$$\mathbf{R}_{\tilde{\mathbf{w}}} = \sigma_w^2(\mathbf{H}^H\mathbf{H})^{-1}. \qquad (6)$$

Motivated by (5) and (6), the ZF-based demodulator was proposed in [3] in an ad-hoc manner as

$$\Lambda_{k,\mathrm{ZF}}^{(i)} \triangleq \frac{1}{\sigma_{\tilde{w},k}^2}\left[\min_{a\in\mathcal{A}_0^i}|y_{\mathrm{ZF},k}-a|^2 - \min_{a\in\mathcal{A}_1^i}|y_{\mathrm{ZF},k}-a|^2\right], \quad (7)$$

where $\sigma_{\tilde{w},k}^2 \triangleq (\mathbf{R}_{\tilde{\mathbf{w}}})_{k,k}$ denotes the noise variance after ZF equalization for the kth layer and $y_{\mathrm{ZF},k} \triangleq (\mathbf{y}_{\mathrm{ZF}})_k$. Because each layer is treated separately, only $M_{\mathrm{T}}|\mathcal{A}|$ scalar distances have to be calculated. Thus, the complexity of (7) is only $\mathcal{O}(M_{\mathrm{T}}^3)$ (due to the calculation of $\mathbf{y}_{\mathrm{ZF}}$), which is much smaller than the complexity of the OLSA demodulator in (3). However, the performance is significantly poorer (see Section III).

II. MMSE-BASED BICM DEMODULATION

We will now derive and discuss the proposed demodulation technique. In our derivation, we will use a Gaussian approximation for the post-detection interference to calculate approximate LLR values. This is motivated and inspired by [4], where an iteratively updated Gaussian approximation for the post-detection interference was used in the context of multiuser detection. Furthermore, we previously used a Gaussian interference approximation for developing a "dynamic" nulling-and-cancelling MIMO detector with improved performance [7].

A. The Gaussian Approximation

We start by reformulating[1] the LLR in terms of $\mathbf{y}_{\mathrm{ZF}}$ instead of $\mathbf{r}$. The exponent in (2) can be written as (e.g., [8])

---

[1]The following derivation could also be performed without going into the ZF domain; however, in that case the calculations would be more involved.

$$\frac{1}{\sigma_w^2}\|\mathbf{r}-\mathbf{Hd}\|^2 = (\mathbf{y}_{\mathrm{ZF}}-\mathbf{d})^H \mathbf{R}_{\tilde{\mathbf{w}}}^{-1}(\mathbf{y}_{\mathrm{ZF}}-\mathbf{d})$$
$$+ \|\mathbf{r}\|^2 - \|\mathbf{Hy}_{\mathrm{ZF}}\|^2,$$

and we can thus express the LLR (2) as

$$\Lambda_k^{(i)} = \log\left(\frac{\sum_{\mathbf{d}:d_k\in\mathcal{A}_1^i} e^{-(\mathbf{y}_{\mathrm{ZF}}-\mathbf{d})^H \mathbf{R}_{\tilde{\mathbf{w}}}^{-1}(\mathbf{y}_{\mathrm{ZF}}-\mathbf{d})}}{\sum_{\mathbf{d}:d_k\in\mathcal{A}_0^i} e^{-(\mathbf{y}_{\mathrm{ZF}}-\mathbf{d})^H \mathbf{R}_{\tilde{\mathbf{w}}}^{-1}(\mathbf{y}_{\mathrm{ZF}}-\mathbf{d})}}\right)$$
$$= \log\left(\frac{f(\mathbf{y}_{\mathrm{ZF}}|b_k^{(i)}=1)}{f(\mathbf{y}_{\mathrm{ZF}}|b_k^{(i)}=0)}\right).$$

Inserting

$$f(\mathbf{y}_{\mathrm{ZF}}|b_k^{(i)}=b) = \frac{1}{|\mathcal{A}_b^i|}\sum_{a\in\mathcal{A}_b^i} f(\mathbf{y}_{\mathrm{ZF}}|d_k=a)$$

and using the log-sum approximation, we obtain further

$$\Lambda_k^{(i)} \approx \max_{a\in\mathcal{A}_1^i}\log f(\mathbf{y}_{\mathrm{ZF}}|d_k=a) - \max_{a\in\mathcal{A}_0^i}\log f(\mathbf{y}_{\mathrm{ZF}}|d_k=a). \quad (8)$$

Here, $f(\mathbf{y}_{\mathrm{ZF}}|d_k=a)$ can be interpreted as the conditional probability density function (pdf) of the post-detection interference for the data symbol of interest, d_k. Because of the Gaussianity of the noise, $f(\mathbf{y}_{\mathrm{ZF}}|d_k=a)$ is a multivariate multimodal Gaussian mixture pdf. To obtain a computationally efficient approximation to (8), we now approximate $f(\mathbf{y}_{\mathrm{ZF}}|d_k=a)$ by a Gaussian pdf $\tilde{f}(\mathbf{y}_{\mathrm{ZF}}|d_k=a)$ with the same mean $\boldsymbol{\mu}_k \triangleq \mathrm{E}\{\mathbf{y}_{\mathrm{ZF}}|d_k=a\}$ and the same covariance $\mathbf{C}_k \triangleq \mathrm{cov}\{\mathbf{y}_{\mathrm{ZF}}|d_k=a\}$:

$$\tilde{f}(\mathbf{y}_{\mathrm{ZF}}|d_k=a) = \frac{1}{\pi^{M_{\mathrm{T}}}\det(\mathbf{C}_k)} e^{-(\mathbf{y}_{\mathrm{ZF}}-\boldsymbol{\mu}_k)^H \mathbf{C}_k^{-1}(\mathbf{y}_{\mathrm{ZF}}-\boldsymbol{\mu}_k)}. \quad (9)$$

To find expressions of $\boldsymbol{\mu}_k$ and $\mathbf{C}_k$, we reformulate $\mathbf{y}_{\mathrm{ZF}} = \mathbf{d} + \tilde{\mathbf{w}}$ in (5) as

$$\mathbf{y}_{\mathrm{ZF}} = d_k\mathbf{e}_k + \sum_{\substack{j=1\\j\neq k}}^{M_{\mathrm{T}}} d_j\mathbf{e}_j + \tilde{\mathbf{w}},$$

where $\mathbf{e}_k$ is the kth M_{T}-dimensional unit vector. We then obtain (recall that the d_k are independent with zero mean and $\mathrm{var}\{d_k\}=1$)

$$\boldsymbol{\mu}_k = a\mathbf{e}_k, \qquad \mathbf{C}_k = \mathbf{I}-\mathbf{e}_k\mathbf{e}_k^T + \mathbf{R}_{\tilde{\mathbf{w}}}. \quad (10)$$

The Gaussian pdf $\tilde{f}(\mathbf{y}_{\mathrm{ZF}}|d_k=a)$ is now completely determined, and the LLR in (8) is approximated according to

$$\Lambda_k^{(i)} \approx \Lambda_{k,\mathrm{MMSE}}^{(i)} \triangleq \max_{a\in\mathcal{A}_1^i}\log\tilde{f}(\mathbf{y}_{\mathrm{ZF}}|d_k=a)$$
$$- \max_{a\in\mathcal{A}_0^i}\log\tilde{f}(\mathbf{y}_{\mathrm{ZF}}|d_k=a). \quad (11)$$

The subscript "MMSE" in $\Lambda_{k,\mathrm{MMSE}}^{(i)}$ will be justified presently.

B. Calculation of $\Lambda_{k,\mathrm{MMSE}}^{(i)}$

We will next derive a simple expression for $\Lambda_{k,\mathrm{MMSE}}^{(i)}$ that can be calculated very efficiently. Inserting (9) and (10) into (11), we obtain

$$\Lambda_{k,\mathrm{MMSE}}^{(i)} = \min_{a\in\mathcal{A}_0^i} Q_k(a) - \min_{a\in\mathcal{A}_1^i} Q_k(a), \quad (12)$$

with the quadratic form

$$Q_k(a) \triangleq (\mathbf{y}_{\mathrm{ZF}}-a\mathbf{e}_k)^H \mathbf{C}_k^{-1}(\mathbf{y}_{\mathrm{ZF}}-a\mathbf{e}_k)$$
$$= \mathbf{y}_{\mathrm{ZF}}^H\mathbf{C}_k^{-1}\mathbf{y}_{\mathrm{ZF}} - 2\,\mathrm{Re}\left\{\mathbf{y}_{\mathrm{ZF}}^H\mathbf{C}_k^{-1}\mathbf{e}_k a\right\}$$
$$+ |a|^2\mathbf{e}_k^T\mathbf{C}_k^{-1}\mathbf{e}_k. \quad (13)$$

The first term of the expression (13) does not depend on a and thus can be disregarded in (12). It remains to develop the second and third terms.

Applying the matrix inversion lemma [6] to $\mathbf{C}_k^{-1} = \left(\mathbf{I} - \mathbf{e}_k\mathbf{e}_k^T + \mathbf{R}_{\tilde{\mathbf{w}}}\right)^{-1}$ and using (6), we obtain

$$\mathbf{C}_k^{-1} = \mathbf{W}\left[\mathbf{I} + \frac{\mathbf{e}_k\mathbf{e}_k^T\mathbf{W}}{1-W_{k,k}}\right], \quad (14)$$

with

$$\mathbf{W} \triangleq (\mathbf{I}+\mathbf{R}_{\tilde{\mathbf{w}}})^{-1} = \left[\mathbf{I}+\sigma_w^2(\mathbf{H}^H\mathbf{H})^{-1}\right]^{-1}$$

(this is termed *Wiener estimator* in [9]) and with $W_{k,k}$ denoting the kth diagonal element of $\mathbf{W}$. The Wiener estimator converts ZF equalization (4) into MMSE equalization [9, 10]:

$$\mathbf{y}_{\mathrm{MMSE}} \triangleq (\mathbf{H}^H\mathbf{H}+\sigma_w^2\mathbf{I})^{-1}\mathbf{H}^H\mathbf{r} = \mathbf{W}\mathbf{y}_{\mathrm{ZF}}.$$

Using this result and (14), we obtain for the second and third terms in (13)

$$\mathbf{y}_{\mathrm{ZF}}^H\mathbf{C}_k^{-1}\mathbf{e}_k = \mathbf{y}_{\mathrm{ZF}}^H\mathbf{W}\mathbf{e}_k\left[1+\frac{\mathbf{e}_k^T\mathbf{W}\mathbf{e}_k}{1-W_{k,k}}\right] = \frac{y_{\mathrm{MMSE},k}^*}{1-W_{k,k}},$$

$$\mathbf{e}_k^T\mathbf{C}_k^{-1}\mathbf{e}_k = \mathbf{e}_k^T\mathbf{W}\mathbf{e}_k\left[1+\frac{\mathbf{e}_k^T\mathbf{W}\mathbf{e}_k}{1-W_{k,k}}\right] = \frac{W_{k,k}}{1-W_{k,k}}.$$

Furthermore, since the eigenvalues of $\mathbf{W}$ satisfy $0 \leq \lambda_{W,k} < 1$, the diagonal elements of $\mathbf{W}$ must satisfy

$$0 \leq W_{k,k} < 1, \qquad 1 \leq \frac{1}{1-W_{k,k}} < \infty.$$

It is then easily verified that (12) simplifies to

$$\Lambda_{k,\mathrm{MMSE}}^{(i)} = \frac{W_{k,k}}{1-W_{k,k}}\left[\min_{a\in\mathcal{A}_0^i}\psi_k^2(a) - \min_{a\in\mathcal{A}_1^i}\psi_k^2(a)\right], \quad (15)$$

with the "unbiased distance"

$$\psi_k(a) \triangleq \left|\frac{y_{\mathrm{MMSE},k}}{W_{k,k}} - a\right|. \quad (16)$$

Here, compared to the conventional distance $|y_{\mathrm{MMSE},k}-a|$, the bias after MMSE equalization (defined as $\mathrm{E}\{y_{\mathrm{MMSE},k}-d_k|d_k\}$ [11]) is compensated through division of $y_{\mathrm{MMSE},k}$ by $W_{k,k}$, i.e., $\mathrm{E}\{\frac{y_{\mathrm{MMSE},k}}{W_{k,k}}|d_k\}=d_k$. It is important to note that, similarly to the ZF-based demodulator in (7) but in contrast to the OLSA demodulator in (3), the calculation of $\Lambda_{k,\mathrm{MMSE}}^{(i)}$ from $\mathbf{y}_{\mathrm{ZF}}$ is performed entirely within the kth layer.

It can furthermore be shown that

$$\frac{W_{k,k}}{1-W_{k,k}} = \mathrm{SNR}_{\mathrm{MMSE},k}.$$

Here, $\text{SNR}_{\text{MMSE},k}$ is the MMSE post-detection SNR of the kth layer given by $\text{SNR}_{\text{MMSE},k} = \frac{1}{\text{MSE}_{\text{MMSE},k}} - 1$ (e.g., [12]), where $\text{MSE}_{\text{MMSE},k} \triangleq \text{E}\{|y_{\text{MMSE},k} - d_k|^2\}$ is the minimum MSE of the kth layer (e.g., [13]). Thus, (15) can be rewritten as

$$\Lambda_{k,\text{MMSE}}^{(i)} = \text{SNR}_{\text{MMSE},k}\left[\min_{a\in\mathcal{A}_0^i}\psi_k^2(a) - \min_{a\in\mathcal{A}_1^i}\psi_k^2(a)\right]. \quad (17)$$

C. Constant-Modulus Alphabets

For constant-modulus alphabets, i.e., $|a| = 1$ for all $a \in \mathcal{A}$, (17) simplifies to

$$\Lambda_{k,\text{MMSE}}^{(i)} = \frac{1}{\text{MSE}_{\text{MMSE},k}}\left[\min_{a\in\mathcal{A}_0^i}|y_{\text{MMSE},k} - a|^2 \right.$$
$$\left. - \min_{a\in\mathcal{A}_1^i}|y_{\text{MMSE},k} - a|^2\right], \quad (18)$$

which is now formulated in terms of the "biased distance" $|y_{\text{MMSE},k} - a|^2$. Even more simple expressions for $\Lambda_{k,\text{MMSE}}^{(i)}$ are obtained for BPSK and 4-QAM (or QPSK) modulation:

- BPSK with Gray labeling:

$$\Lambda_{k,\text{MMSE}}^{(1)} = \frac{4}{\text{MSE}_{\text{MMSE},k}}\text{Re}\{y_{\text{MMSE},k}\}; \quad (19)$$

- 4-QAM with Gray labeling:

$$\Lambda_{k,\text{MMSE}}^{(i)} = \begin{cases} \frac{2\sqrt{2}}{\text{MSE}_{\text{MMSE},k}}\,\text{Re}\{y_{\text{MMSE},k}\}, & i = 1, \\ \frac{2\sqrt{2}}{\text{MSE}_{\text{MMSE},k}}\,\text{Im}\{y_{\text{MMSE},k}\}, & i = 2. \end{cases}$$

A result formally similar to (19) was obtained in a multiuser context in [14], however using a different approach.

D. Discussion

The proposed MIMO-BICM demodulator was derived by using a Gaussian approximation for the post-detection interference in the calculation of the LLRs. Our final expression (17) (or (18) for constant-modulus alphabets), together with the expression of $\psi_k(a)$ in terms of $y_{\text{MMSE},k}$ (see (16)), shows that the proposed demodulator consists of *MMSE equalization* and subsequent *per-layer LLR calculation*. Because after MMSE equalization each layer is processed separately, the computational complexity of our demodulator is not higher than that of the ZF-based demodulator in (7).

The structure of our demodulator is similar to both the OLSA demodulator in (3) and the ZF-based demodulator in (7). All three demodulators compute a difference of two distances, where one distance corresponds to the respective bit being 0 and the other corresponds to that bit being 1. However, these distances are differently defined for the three demodulators. Furthermore, the pre-factors in the approximate LLR expressions are different, too:

- With OLSA demodulation (3), the pre-factor is the reciprocal noise variance, $1/\sigma_w^2$.
- With ZF-based demodulation (7), the pre-factor is the reciprocal post-equalization noise variance, $1/\sigma_{\tilde{w},k}^2$. We

note that $\sigma_{\tilde{w},k}^2$ can be shown to equal the ZF post-detection MSE, i.e. $\sigma_{\tilde{w},k}^2 = \text{MSE}_{\text{ZF},k} = \text{E}\{|y_{\text{ZF},k} - d_k|^2\}$.

- Finally, with our demodulator (17), the pre-factor is the MMSE post-detection SNR, $\text{SNR}_{\text{MMSE},k}$; in the special case of constant-modulus alphabets, the pre-factor in the alternative expression (18) is the reciprocal minimum post-detection MSE, $1/\text{MSE}_{\text{MMSE},k}$.

All pre-factors decrease with increasing noise power σ_w^2.

For constant-modulus alphabets, our demodulator (18) can be viewed as an MMSE analogue of the ZF-based demodulator (7), in which $y_{\text{ZF},k}$ is replaced by $y_{\text{MMSE},k}$ and $\sigma_{\tilde{w},k}^2 = \text{MSE}_{\text{ZF},k}$ is replaced by $\text{MSE}_{\text{MMSE},k}$ but *not* by the noise variance after MMSE equalization (note that in contrast to ZF equalization where $\sigma_{\tilde{w},k}^2 = \text{MSE}_{\text{ZF},k}$, the noise variance after MMSE equalization is different from $\text{MSE}_{\text{MMSE},k}$). In fact, using the noise variance after MMSE equalization instead of $\text{MSE}_{\text{MMSE},k}$ in the pre-factor of (18)—as motivated by the analogy to (7)—would result in a significant performance degradation.

For general alphabets, our demodulator calculates the per-layer distances $\psi_k(a)$ in (17) using the *unbiased* MMSE equalized components $y_{\text{MMSE},k}/W_{k,k}$ instead of $y_{\text{MMSE},k}$ (cf. (16)). For alphabets that are not constant-modulus, use of (18) instead of (17) would cause a slight performance degradation.

If we develop our demodulator by using the Gaussian approximation for the post-detection interference in (8), however *with the correlations in $\mathbf{y}_{\text{ZF}}$ neglected*, the ZF-based demodulator (7) is obtained. Indeed, it can be verified that with all nondiagonal elements of $\mathbf{C}_k = \mathbf{I} - \mathbf{e}_k\mathbf{e}_k^T + \mathbf{R}_{\tilde{\mathbf{w}}}$ set equal to zero, (12) becomes equivalent to (7). In this sense, then, our demodulator is an extension of the ZF-based demodulator to correlated post-detection interference. This provides an explanation of the significant performance advantage of our demodulator over the ZF-based demodulator (as demonstrated in Section III). Fortunately, this performance advantage is obtained with no increase in complexity.

III. SIMULATION RESULTS

We now assess the performance of our demodulator by means of simulation results. We considered a MIMO-BICM system using a rate-$1/2$ 64-state convolutional code with octal generators $(133, 171)$ and random interleaving. The MIMO channel had iid Gaussian matrix entries with unit variance. To simulate fast fading, the channel was independently generated for each time instant. We considered the proposed demodulator as well as the ZF-based and OLSA demodulators for comparison. The Viterbi decoder used a traceback depth of 35 and employed 6 bits for trellis termination.

Fig. 2 shows the bit-error rate (BER) of this MIMO-BICM system using the various demodulators versus the SNR[2] for a $(2, 2)$ channel and a $(3, 3)$ channel. Each figure part shows three sets of curves corresponding to BPSK, 4-QAM (QPSK), and 16-QAM modulation using Gray labeling.

---

[2]The SNR is defined as $\text{E}\{\|\mathbf{H}\mathbf{d}\|^2\}/\text{E}\{\|\mathbf{w}\|^2\} = M_{\text{T}}/\sigma_w^2$.
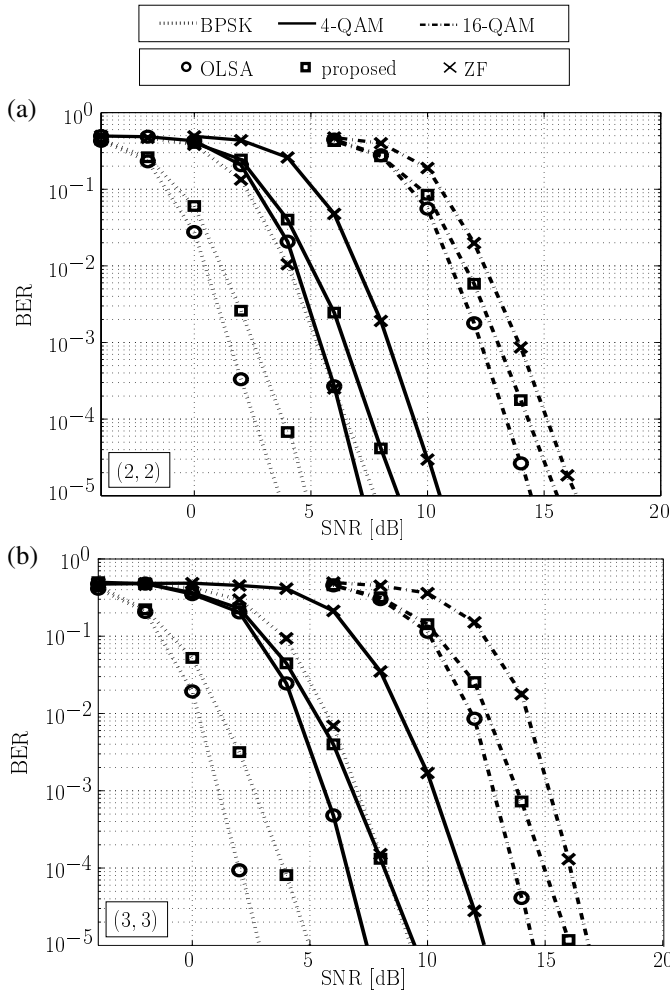
Fig. 2. *BER performance of the proposed MMSE-based demodulator, the ZF-based demodulator, and the OLSA demodulator for BPSK, 4-QAM (QPSK), and 16-QAM modulation: (a) (2,2) channel, (b) (3,3) channel.*

The following conclusions can be drawn from these results.

- The performance of the proposed demodulator is very close to that of OLSA demodulation. For all considered symbol alphabets, at a target BER of 10^{-4}, our demodulator is within roughly 1 dB and 2 dB of OLSA performance for the $(2, 2)$ channel and the $(3, 3)$ channel, respectively.

- Our demodulator significantly outperforms ZF-based demodulation even though its computational complexity is the same. The performance gain over ZF-based demodulation is up to 4 dB at a target BER of 10^{-4}. In particular, for the $(3, 3)$ channel, our demodulator for 4-QAM modulation achieves roughly the same performance as the ZF-based demodulator for BPSK modulation even though the data rate (in bits per channel use) is doubled.

- The performance advantage of our demodulator over the ZF-based demodulator is strongest for small symbol alphabets. Furthermore, a comparison of the results obtained for the $(2, 2)$ channel and the $(3, 3)$ channel suggests that for an increasing number of antennas, both the

performance advantage over the ZF-based demodulator and the performance loss relative to the OLSA demodulator grow larger. On the other hand, the complexity savings achieved with our demodulator relative to the OLSA demodulator are larger for an increasing number of antennas and an increasing alphabet size.

IV. CONCLUSIONS

We have presented a novel, efficient demodulator for MIMO bit-interleaved coded modulation (BICM) systems. Our development was based on a Gaussian approximation for the post-detection interference and resulted in a demodulation technique consisting of MMSE equalization and subsequent per-layer log-likelihood ratio (LLR) calculation. The computational complexity of the novel MIMO-BICM demodulator is very low due to the per-layer processing it employs.

We showed that the proposed demodulator can be interpreted as an extension of a recently proposed ZF-equalization based demodulator to correlated post-detection interference. This extension yields a substantial performance improvement without an increase in computational complexity. We verified through simulations that the performance of our demodulator is close to that of LLR calculation using all layers jointly, and significantly better than that of the ZF-based demodulator.

REFERENCES

[1] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, pp. 927–945, May 1998.
[2] S. H. Müller-Weinfurtner, "Coding approaches for multiple antenna transmission in fast fading and OFDM," *IEEE Trans. Signal Processing*, vol. 50, pp. 2442–2450, Oct. 2002.
[3] M. Butler and I. Collings, "A zero-forcing approximate log-likelihood receiver for MIMO bit-interleaved coded modulation," *IEEE Commun. Letters*, vol. 8, no. 2, pp. 105–107, Feb. 2004.
[4] J. Luo, K. Pattipati, P. Willett, and F. Hasegawa, "A PDA approach to CDMA multiuser detection," in *Proc. IEEE GLOBECOM-2001*, vol. 2, San Antonio, TX, Nov. 2001, pp. 763–766.
[5] S. Wilson, *Digital Modulation and Coding*. Englewood Cliffs (NJ): Prentice Hall, 1996.
[6] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore: Johns Hopkins University Press, 1996.
[7] D. Seethaler, H. Artés, and F. Hlawatsch, "Dynamic nulling and cancelling with near-ML performance," in *Proc. IEEE ICASSP 2004*, vol. IV, Montreal, Canada, May 2004, pp. 777–780.
[8] B. M. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 389–399, March 2003.
[9] A. Klein, G. Kaleh, and P. W. Baier, "Zero forcing and minimum mean square error equalization for multiuser detection in code-division multiple-access channels," *IEEE Trans. Veh. Technol.*, vol. 45, no. 2, pp. 276–287, May 1996.
[10] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs (NJ): Prentice Hall, 1993.
[11] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu, and G. D. Forney, "MMSE decision-feedback equalizers and coding – Part I: Equalization results," *IEEE Trans. Commun.*, vol. 43, no. 10, pp. 2582–2594, Oct. 1995.
[12] R. W. Heath, S. Sandhu, and A. J. Paulraj, "Antenna selection for spatial multiplexing systems with linear receivers," *IEEE Commun. Letters*, vol. 5, no. 4, pp. 142–144, April 2001.
[13] B. Hassibi, "A fast square-root implementation for BLAST," in *Proc. 34th Asilomar Conf. Signals, Systems, Computers*, Pacific Grove, CA, Nov./Dec. 2000, pp. 1255–1259.
[14] X. Wang and H. V. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Comm.*, vol. 47, pp. 1046–1061, July 1999.