Performance Assessment of MIMO-BICM Demodulators Based on Mutual Information

Peter Fertl, *Member, IEEE*, Joakim Jaldén, *Member, IEEE*, and Gerald Matz, *Senior Member, IEEE*

***Abstract*—We provide a comprehensive performance comparison of soft-output and hard-output demodulators in the context of non-iterative multiple-input multiple-output bit-interleaved coded modulation (MIMO-BICM). Coded bit error rate (BER), widely used in literature for demodulator comparison, has the drawback of depending strongly on the error correcting code being used. This motivates us to propose the mutual information of the equivalent modulation channel (comprising modulator, wireless channel, and demodulator) as a code-independent performance measure. We present extensive numerical results for spatially independent identically distributed (i.i.d.) ergodic and quasi-static fading channels under perfect and imperfect channel state information. These results reveal that the performance ranking of MIMO demodulators is rate-dependent and provide new insights regarding MIMO-BICM system design, i.e., the choice of antenna configuration, symbol constellation, and demodulator for a given target rate.**

***Index Terms*—Bit-interleaved coded modulation (BICM), log-likelihood ratio, multiple-input–multiple-output (MIMO), mutual information, performance limits, soft demodulation.**

I. Introduction

A. Background

BIT-INTERLEAVED CODED MODULATION (BICM) [2], [3] has been conceived as a pragmatic approach to coded modulation. It has received a lot of attention in wireless communications due to its bandwidth and power efficiency and its robustness against fading. For single-antenna systems, BICM with Gray labeling can approach channel capacity [2], [4]. These advantages have motivated extensions of BICM to multiple-input multiple-output (MIMO) systems [5]–[7].

In MIMO-BICM systems, the optimum demodulator is the soft-output maximum a posteriori (MAP) demodulator, which provides the channel decoder with log-likelihood ratios (LLRs) for the code bits. Due to its high computational complexity, numerous alternative demodulators have been proposed. Applying the max-log approximation [7] to the MAP demodulator reduces complexity without significant performance loss and leads to the minimization of a Euclidean distance. Sphere decoding implementations of the max-log MAP detector have been presented in [8]–[10]; sphere decoder variants in which the Euclidean norm is replaced with the ℓ^∞ norm have been proposed in [11], [12]. However, the complexity of sphere decoding grows exponentially with the number of transmit antennas [9]. An alternative demodulator is based on semidefinite relaxation (SDR) and has polynomial worst-case complexity [13], [14].

Several demodulators use a list of candidate data vectors to obtain approximate LLRs. The list size offers a performance/complexity tradeoff and can be generated using: i) tree search techniques as with the list sphere decoder (LSD) [15]; ii) lattice reduction (LR) [16]–[20]; or iii) bit flipping [21].

MIMO demodulators with still smaller complexity consist of a linear equalizer followed by per-layer scalar soft demodulators. This approach has been studied using zero-forcing (ZF) equalization [22], [23] and minimum mean-square error (MMSE) equalization [24], [25]. The soft interference canceler (SoftIC) proposed in [26] iteratively performs parallel MIMO interference cancelation by subtracting an interference estimate which is computed using soft symbols from the preceding iteration.

Hard-output MIMO demodulators provide tentative decisions for the code bits but no associated reliability information. Among the best-known schemes here are maximum likelihood (ML), ZF, and MMSE demodulation [27] and successive interference cancelation (SIC) [28]–[30].

B. Contributions

In the context of MIMO-BICM, the performance of the MIMO demodulators listed above has mostly been assessed in terms of coded bit error rate (BER) using a specific channel code. These BER results depend strongly on the channel code and hence render an impartial demodulator comparison difficult.

In this paper, we advocate an information-theoretic approach for assessing the performance of (soft and hard) MIMO demodulators in the context of non-iterative[1] (single-shot) BICM receivers (see also [1]). Inspired by [5], we propose the mutual

Manuscript received October 16, 2010; revised June 07, 2011 and October 10, 2011; accepted October 30, 2011. Date of publication December 01, 2011; date of current version February 10, 2012. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Francesco Verde. This work was supported by the STREP project MASCOT (IST-026905) within the Sixth Framework Programme of the European Commission and by FWF Grant N10606 "Information Networks". Part of the material in this paper was presented at the IEEE 9th Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2008), Recife, Brazil, July 2008.

P. Fertl is with BMW Group Research and Technology, 80992 Munich, Germany (e-mail: peter.fertl@bmw.de).

J. Jaldén is with the Signal Processing Lab, ACCESS Linnaeus Center, KTH Royal Institute of Technology, Osquldas väg 10, SE-100 44 Stockholm, Sweden (e-mail: jalden@kth.se).

G. Matz is with the Institute of Telecommunications, Vienna University of Technology, 1040 Vienna, Austria (e-mail: gmatz@nt.tuwien.ac.at).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2011.2177823

[1]A performance assessment of *soft-in* soft-out demodulators in iterative BICM receivers requires a completely different approach and is thus beyond the scope of this paper.

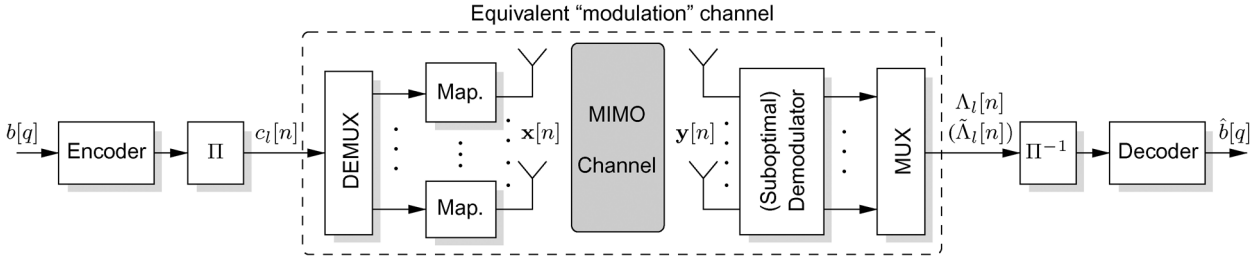


Fig. 1. Block diagram of a MIMO-BICM system.

information between the modulator input bits and the associated MIMO demodulator output as a *code-independent*[2] performance measure. Under the assumption of independent identically distributed (i.i.d.) uniform code bits, this quantity represents the capacity (i.e., maximum information rate allowing for error-free information recovery) of an equivalent "modulation" channel that comprises modulator and demodulator in addition to the physical channel. This approach establishes a systematic framework for the assessment of MIMO demodulators. We note that ZF-based and max-log demodulation have been compared in a similar spirit in [23].

Using Monte Carlo simulations, this paper provides extensive performance evaluations and comparisons for the aforementioned MIMO demodulators in terms of mutual information, considering different system configurations in fast and quasi-static spatially i.i.d. fading. We also investigate the performance loss of the various demodulators under imperfect channel state information (CSI). Due to lack of space, only a part of our numerical results is shown here. Further results for other antenna configurations, symbol constellations, and bit mappings can be found in a supporting document [31]. We also note that our main insights carry over to spatially correlated channels and to frequency-selective channels (see Section IX).

Our results allow for several conclusions. Most importantly, we found that no universal performance ranking of MIMO demodulators exists, i.e., the ranking depends on the information rate or, equivalently, on the signal-to-noise ratio (SNR). As an example, soft MMSE outperforms hard ML at low rates while at high rates it is the other way around. We also verify this surprising observation in terms of BER simulations using low-density parity-check (LDPC) codes. Finally, we use our numerical results to develop practical guidelines for the design of MIMO-BICM systems, i.e., which antenna configuration, symbol constellation, and demodulator to choose in order to achieve a certain rate with minimum SNR.

C. Paper Organization

The rest of this paper is organized as follows. Section II discusses the MIMO-BICM system model and Section III proposes mutual information as performance measure. In Sections IV and V, we assess the mutual information achievable with various MIMO-BICM demodulators for the case of fast fading. Section VI analyzes the impact of imperfect CSI on the demodulator performance, and Section VII investigates the rate-versus-outage tradeoff of selected demodulators in quasi-static environments. In Section VIII, we summarize key observations and infer practical system design guidelines. Finally, conclusions are provided in Section IX.

II. System Model

A. MIMO-BICM Transmission Model

A block diagram of our MIMO-BICM model is shown in Fig. 1. The information bits $b[q]$ are encoded using an error-correcting code and is then passed through a *bitwise* interleaver Π. The interleaved code bits are demultiplexed into M_{T} antenna streams ("layers"). In each layer $k = 1, \ldots, M_{\mathrm{T}}$, groups of Q code bits $c_k^{(i)}[n]$, $i = 1, \ldots, Q$, (n denotes symbol time) are mapped via a one-to-one function $\mu(\cdot)$ to (complex) data symbols $x_k[n]$ from a symbol alphabet $\mathcal{A}$ of size $|\mathcal{A}| = 2^Q$. Specifically, $x_k[n] = \mu(c_k^{(1)}[n], \ldots, c_k^{(Q)}[n])$, where $\{c_k^{(1)}[n], \ldots, c_k^{(Q)}[n]\} = \mu^{-1}(x_k[n])$ is referred to as the bit label of $x_k[n]$. The transmit vector is given by[3] $\mathbf{x}[n] \triangleq (x_1[n] \ldots x_{M_{\mathrm{T}}}[n])^{\mathrm{T}} \in \mathcal{A}^{M_{\mathrm{T}}}$ and satisfies the power constraint $\mathbb{E}\{\|\mathbf{x}[n]\|^2\} = E_s$. It carries $R_0 = QM_{\mathrm{T}}$ interleaved code bits $c_l[n]$, $l = 1, \ldots, R_0$, with $c_k^{(i)}[n] = c_{(k-1)Q+i}[n]$. We will for simplicity write $\mathbf{x}[n] = \mu(c_1[n], \ldots, c_{R_0}[n])$ and $\mathbf{c}[n] \triangleq (c_1[n] \ldots c_{R_0}[n])^{\mathrm{T}} = \mu^{-1}(\mathbf{x}[n])$ as shorthand for the mapping $\mathbf{x}[n] = (\mu(c_1^{(1)}[n], \ldots, c_1^{(Q)}[n]) \ldots \mu(c_{M_{\mathrm{T}}}^{(1)}[n], \ldots, c_{M_{\mathrm{T}}}^{(Q)}[n]))^{\mathrm{T}}$ and its inverse.

Assuming flat fading, the receive vector $\mathbf{y}[n] \triangleq (y_1[n] \ldots y_{M_{\mathrm{R}}}[n])^{\mathrm{T}}$ (M_{R} denotes the number of receive antennas) is given by

$$\mathbf{y}[n] = \mathbf{H}[n]\mathbf{x}[n] + \mathbf{v}[n], \qquad n = 1, \ldots, N \tag{1}$$

where $\mathbf{H}[n]$ is the $M_{\mathrm{R}} \times M_{\mathrm{T}}$ channel matrix, and $\mathbf{v}[n] \triangleq (v_1[n] \ldots v_{M_{\mathrm{R}}}[n])^{\mathrm{T}}$ is a noise vector with i.i.d. circularly symmetric complex Gaussian elements with zero mean and variance σ_v^2. In most of what follows, we will omit the time index n for convenience.

At the receiver, the optimum demodulator uses the received vector $\mathbf{y}$ and the channel matrix $\mathbf{H}$ to calculate LLRs Λ_l for all code bits c_l, $l = 1, \ldots, R_0$, carried by $\mathbf{x}$. In practice, the use of suboptimal demodulators or of a channel estimate $\hat{\mathbf{H}}$ will result in approximate LLRs $\tilde{\Lambda}_l$. The LLRs are passed through

[2]Coding is of course needed to achieve capacity. However, capacity is a code-independent measure in that it is not dependent on any particular code (it is the supremum of the achievable rates over all codes).

[3]The superscripts $^{\mathrm{T}}$ and $^{\mathrm{H}}$ denote transposition and Hermitian transposition, respectively. Furthermore, $\mathcal{A}^{M_{\mathrm{T}}} \triangleq \mathcal{A} \times \ldots \times \mathcal{A}$ is the M_{T}-fold Cartesian product of $\mathcal{A}$, $\mathbb{E}\{\cdot\}$ denotes expectation, and $\|\cdot\|$ is the ℓ^2 (Euclidean) norm.

the deinterleaver Π^{-1} and then on to the channel decoder that delivers the detected bits $\hat{b}[q]$.

B. Optimum Soft MAP Demodulation

Assuming i.i.d. uniform code bits (as guaranteed, e.g., by an ideal interleaver), the optimum soft MAP demodulator calculates the exact LLR for c_l based on $(\mathbf{y}, \mathbf{H})$ according to [7]

$$\Lambda_l \triangleq \log \frac{p(c_l = 1|\mathbf{y}, \mathbf{H})}{p(c_l = 0|\mathbf{y}, \mathbf{H})} = \log \frac{\sum\limits_{\mathbf{x} \in \mathcal{X}_l^1} \exp\left(-\frac{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2}{\sigma_v^2}\right)}{\sum\limits_{\mathbf{x} \in \mathcal{X}_l^0} \exp\left(-\frac{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2}{\sigma_v^2}\right)}. \tag{2}$$

Here, $p(c_l|\mathbf{y}, \mathbf{H})$ is the probability mass function (pmf) of the code bits conditioned on $\mathbf{y}$ and $\mathbf{H}$, $\mathcal{X}_l^1$ and $\mathcal{X}_l^0$ denote the complementary sets of transmit vectors for which $c_l = 1$ and $c_l = 0$, respectively (note that $\mathcal{A}^{M_\mathrm{T}} = \mathcal{X}_l^1 \cup \mathcal{X}_l^0$). Unfortunately, computation of (2) has complexity $\mathcal{O}(|\mathcal{A}|^{M_\mathrm{T}}) = \mathcal{O}(2^{R_0})$, i.e., exponential in the number of transmit antennas. For this reason, several suboptimal demodulators have been proposed which promise near-optimal performance while requiring a lower computational complexity. The aim of this work is to provide a fair performance comparison of these demodulators.

III. Capacity

In order for the information rates discussed below to have interpretations as ergodic capacities, we consider a fast fading scenario where the channel $\mathbf{H}[n]$ is a stationary, finite-memory process. We recall that the ergodic capacity with Gaussian inputs is given by [32]

$$C_\mathrm{G} = \mathbb{E}_\mathbf{H}\left\{\log_2 \det\left(\mathbf{I} + \frac{E_s}{M_\mathrm{T}\sigma_v^2}\mathbf{H}\mathbf{H}^\mathrm{H}\right)\right\} \tag{3}$$

(here, $\mathbf{I}$ denotes the identity matrix). The non-ergodic regime (slow fading) is discussed in Section III-D.

A. Capacity of MIMO Coded Modulation

In a coded modulation (CM) system with equally likely transmit vectors $\mathbf{x} \in \mathcal{A}^{M_\mathrm{T}}$ and no CSI at the transmitter, the average mutual information in bits per channel use (bpcu) is given by (cf. [5])

$$C_\mathrm{CM} \triangleq I(\mathbf{x}; \mathbf{y}|\mathbf{H}) = R_0 - \mathbb{E}_{\mathbf{x},\mathbf{y},\mathbf{H}}\left\{\log_2 \frac{\sum\limits_{\mathbf{x}' \in \mathcal{A}^{M_\mathrm{T}}} f(\mathbf{y}|\mathbf{x}', \mathbf{H})}{f(\mathbf{y}|\mathbf{x}, \mathbf{H})}\right\}. \tag{4}$$

Here, we used the conditional probability density function (pdf) [cf. (1)]

$$f(\mathbf{y}|\mathbf{x}, \mathbf{H}) = \frac{1}{(\pi\sigma_v^2)^{M_\mathrm{R}}} \exp\left(-\frac{\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2}{\sigma_v^2}\right). \tag{5}$$

In the following, we will refer to C_CM as *CM capacity* [2] (sometimes, C_CM is alternatively termed constellation-constrained capacity). It is seen from (4) that $C_\mathrm{CM} \leq R_0$; in fact, the last term in (4) may be interpreted as a penalty term resulting from the noise and MIMO interference.

Using the fact that the mapping between the symbol vector $\mathbf{x}$ and the associated bit label $\{c_1, \ldots, c_{R_0}\}$ is one-to-one and applying the chain rule for mutual information [33, p. 24] to (4) leads to

$$\begin{aligned} C_\mathrm{CM} &= I(c_1, \ldots, c_{R_0}; \mathbf{y}|\mathbf{H}) = \sum_{l=1}^{R_0} I(c_l; \mathbf{y}|c_1, \ldots, c_{l-1}, \mathbf{H}) \\ &= \sum_{l=1}^{R_0} \mathbb{H}(c_l|\mathbf{H}) - \mathbb{H}(c_l|\mathbf{y}, c_1, \ldots, c_{l-1}, \mathbf{H}). \end{aligned} \tag{6}$$

Here, $\mathbb{H}(\cdot)$ denotes the entropy function. The single-antenna equivalent of (6) served as a motivation for multilevel coding and multistage decoding, which can indeed achieve CM capacity [4]. Multilevel coding for multiple antenna systems has been considered in [34].

B. Capacity of MIMO-BICM

We assume an *ideal, infinite-length* bit interleaver,[4] which allows us to treat the BICM system as a set of R_0 independent parallel memoryless binary-input channels as in [2, Section III.A]. Using the assumption of i.i.d. uniform code bits, the maximum rate achievable with BICM is given by (cf. [5])

$$\begin{aligned} C_\mathrm{BICM} &\triangleq \sum_{l=1}^{R_0} I(c_l; \mathbf{y}|\mathbf{H}) = \sum_{l=1}^{R_0} \mathbb{H}(c_l|\mathbf{H}) - \mathbb{H}(c_l|\mathbf{y}, \mathbf{H}) \\ &= R_0 - \sum_{l=1}^{R_0} \mathbb{E}_{\mathbf{x},\mathbf{y},\mathbf{H}}\left\{\log_2 \frac{\sum\limits_{\mathbf{x}' \in \mathcal{A}^{M_\mathrm{T}}} f(\mathbf{y}|\mathbf{x}', \mathbf{H})}{\sum\limits_{\mathbf{x}' \in \mathcal{X}_l^{c_l}} f(\mathbf{y}|\mathbf{x}', \mathbf{H})}\right\} \end{aligned} \tag{7}$$

where[5] $c_l = (\mathbf{c})_l = (\mu^{-1}(\mathbf{x}))_l$ denotes the lth bit in the label of $\mathbf{x}$. Since conditioning reduces entropy [33, p. 29], a comparison of (6) and (7) reveals that [34]

$$C_\mathrm{BICM} \leq C_\mathrm{CM}.$$

The gap $C_\mathrm{CM} - C_\mathrm{BICM}$ increases with $|\mathcal{A}|$ and M_T and depends strongly on the symbol labeling [7]. For single-antenna BICM systems with Gray labeling, this gap has been shown to be negligible [2], [4]; however, for MIMO systems (see Section IV) and at low SNRs in the wideband regime [3] it can be significant. The capacity loss can be attributed to the fact that the BICM receiver neglects the dependencies between the transmitted code bits. Under the unrealistic assumption of perfectly known channel SNR, multilevel coding with multistage decoding in principle can avoid such a capacity loss but otherwise suffers from error propagation [4], [34]. A hybrid version of CM and BICM whose complexity and performance is between the two was presented in [34]. Furthermore, augmenting BICM with

[4]In practice, this means that the interleaver needs to be much longer than the codewords transmitted over the channel.

[5]By $(\mathbf{x})_k$ and $(\mathbf{X})_{k,l}$ we respectively denote the kth element of the vector $\mathbf{x}$ and the element in row k and column l of the matrix $\mathbf{X}$.

space-time mappings can be beneficial (cf. [34] and [35]) but is not considered here due to space limitations.

It can be shown that the log-likelihood ratio Λ_l in (2) is a sufficient statistic [36] for c_l given $\mathbf{y}$ and $\mathbf{H}$. Therefore, (7) can be rewritten as

$$C_{\mathrm{BICM}} = \sum_{l=1}^{R_0} I(c_l; \Lambda_l). \tag{8}$$

Hence, C_{BICM} can be interpreted as the capacity of an equivalent channel with inputs c_l and outputs Λ_l (cf. Fig. 1). This channel is characterized by the conditional pdf $\prod_l f(\Lambda_l|c_l)$, which usually is hard to obtain analytically, however.

C. Capacity and Demodulator Performance

Motivated by the fact that C_{BICM} is the capacity of BICM using the optimum MAP demodulator, we propose to measure the performance of suboptimal MIMO-BICM demodulators via the capacity of the associated equivalent "modulation" channel with *binary* inputs c_l and the approximate LLRs $\tilde{\Lambda}_l$ as *continuous* outputs (cf. Fig. 1). This channel is described by the conditional pdf $\prod_l f(\tilde{\Lambda}_l|c_l)$. With i.i.id. uniform input bits, its capacity equals the mutual information between c_l and $\tilde{\Lambda}_l$, which can be shown to be given by

$$\begin{aligned} C &\triangleq \sum_{l=1}^{R_0} I(c_l; \tilde{\Lambda}_l) \\ &= R_0 - \sum_{l=1}^{R_0} \sum_{b=0}^{1} \int_{-\infty}^{\infty} \frac{1}{2} f(\tilde{\Lambda}_l|c_l=b) \log_2 \frac{f(\tilde{\Lambda}_l)}{\frac{1}{2} f(\tilde{\Lambda}_l|c_l=b)} d\tilde{\Lambda}_l \end{aligned} \tag{9}$$

where $f(\tilde{\Lambda}_l) = \frac{[f(\tilde{\Lambda}_l|c_l=0)+f(\tilde{\Lambda}_l|c_l=1)]}{2}$. We emphasize that the capacity C provides a performance measure for MIMO (soft) demodulators that is independent of the outer channel code. In fact, it has an operational meaning as the highest rate achievable (in the sense of asymptotically vanishing error probability) in a BICM system with independent parallel channels (assumption of an ideal infinite-length interleaver, cf. [2, Section III.A]), using the specific demodulator which produces $\tilde{\Lambda}_l$. Since $\tilde{\Lambda}_l$ is derived from $\mathbf{y}$ and $\mathbf{H}$, the data processing inequality [33, page 34] implies that $C \leq C_{\mathrm{BICM}}$ with equality if $\tilde{\Lambda}_l$ is a one-to-one function of Λ_l. The performance of a soft demodulator can thus be measured in terms of the gap $C_{\mathrm{BICM}} - C$. Of course, the information theoretic performance measure in (9) does not take into account complexity issues and it has to be expected that a reduction of the gap $C_{\mathrm{BICM}} - C$ in general can only be achieved at the expense of increasing computational complexity.

We caution the reader that the rates in (8) and (9) are sums of mutual informations for the individual code bits $c_1, \ldots, c_{R_0}$ carried by one symbol vector. Indeed, the pdfs $f(\Lambda_l|c_l)$ and $f(\tilde{\Lambda}_l|c_l)$ in general depend on the code bit position l, even though for certain systems (e.g. 4-QAM modulation) the code bit protection and LLR statistics are independent of the bit position l for reasons of symmetry.

Achieving (8) and (9) thus requires channel decoders that take the bit position into account. When the channel code fails to use this information, the rate loss is small provided that the mapping protects different code bits c_l roughly equally against noise and interference.

D. Nonergodic Channels

In the case of quasi-static or slow fading [37], the channel $\mathbf{H}$ is random but constant over time, i.e., each codeword can extend over only one channel realization. Here, the ergodic capacity of the modulation channel is no longer operationally meaningful [37], [38]. Instead we consider the *outage probability*

$$P_{\mathrm{out}}(R) \triangleq \mathbf{P}\{R_{\mathbf{H}} < R\}, \tag{10}$$

where $R_{\mathbf{H}}$ is a random variable defined as

$$R_{\mathbf{H}} \triangleq \sum_{l=1}^{R_0} I_{\mathbf{H}}(c_l; \tilde{\Lambda}_l).$$

Here, $I_{\mathbf{H}}(c_l; \tilde{\Lambda}_l)$ denotes the conditional mutual information, which is evaluated with $f(\tilde{\Lambda}_l|c_l, \mathbf{H})$ in place of $f(\tilde{\Lambda}_l|c_l)$ [cf. (9)]. Note that the ergodic system capacity C in (9) equals $C = \mathbb{E}_{\mathbf{H}}\{R_{\mathbf{H}}\}$. The outage probability $P_{\mathrm{out}}(R)$ can be interpreted as the smallest codeword error probability achievable at rate R [38]. A closely related concept is the ϵ-*capacity* of the equivalent modulation channel, defined as

$$C_\epsilon \triangleq \sup\{R \,|\, \mathbf{P}\{R_{\mathbf{H}} < R\} < \epsilon\}. \tag{11}$$

The ϵ-capacity may be interpreted as the maximum rate for which a codeword error probability less than ϵ can be achieved. Rates smaller than C_ϵ are referred to as ϵ-achievable rates [38]. If $P_{\mathrm{out}}(R)$ is a continuous and increasing function of R (which is usually the case in practice), it holds that $P_{\mathrm{out}}(C_\epsilon) = \epsilon$.

E. Generalized Mutual Information

The operational meaning of our performance measure as the largest achievable rate for a BICM system using a given demodulator requires the assumption of an ideal infinite-length interleaver. With a finite-length interleaver, the parallel channels (i.e., the different bits in a given symbol vector) are not independent in general and, hence, (9) is not meaningful a priori. In this case, achievable rates can be characterized in terms of the *generalized mutual information (GMI)* which is obtained by treating the BICM receiver as a mismatched decoder [3], [39], [40]. More precisely, for any rate (strictly) below the GMI and for any given probability of error, there exists a code with finite length that achieves this rate and probability of error using the product decoder metric associated with BICM and an interleaver no longer than the code length. For the case of optimum soft MAP demodulation [cf. (2)], the BICM capacity using the independent parallel-channel model coincides with the GMI [40]. We recently provided a non-straightforward extension of this result by showing that the GMI of a BICM system with suboptimal demodulators augmented with *scalar LLR correction* (see Section IV-D) coincides with the capacity in (9) obtained for the parallel-channel model [41].[6] Without LLR correction, the GMI (and hence the achievable rates) of a BICM system with

[6]We note that the LLR correction leaves the mutual information which underlies system capacity unchanged.

finite-length interleavers is less than C in (9). Scalar LLR correction has been used previously to provide the binary decoder with accurate reliability information [36], [42]–[45]. The GMI of a BICM system with finite interleaver and LLR-corrected suboptimal demodulators can thus efficiently be computed by evaluating (9) [41]; this provides additional justification for the use of (9) as a code-independent performance measure for approximate demodulators. We note that a GMI-based analysis of BICM with mismatched decoding metrics that generalizes our work in [41] has recently been presented in [46].

IV. Baseline MIMO-BICM Demodulators

In this section, we first review max-log and hard ML demodulation as well as linear MIMO demodulators and then we provide results illustrating their performance in terms of system capacities. These demodulators serve as baseline systems for later demodulator performance comparisons in Section V. We note that max-log and hard ML MIMO demodulators have the highest complexity among all soft and hard demodulation schemes, respectively, whereas linear MIMO demodulators are most efficient computationally. Due to space limitations, we only state the complexity order of each demodulator in the following and we give references that provide more detailed complexity analyses.

A. Max-Log and Hard ML Demodulator

Applying the max-log approximation to (2) simplifies the LLR computation to a minimum distance problem and results in the approximate LLRs [7]

$$\tilde{\Lambda}_l = \frac{1}{\sigma_v^2}\left[\min_{\mathbf{x}\in\mathcal{X}_l^0}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2 - \min_{\mathbf{x}\in\mathcal{X}_l^1}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2\right]. \quad (12)$$

This expression can be implemented easier than (2) since it avoids the logarithm and exponential functions. However, computation of $\tilde{\Lambda}_l$ in (12) still requires two searches over sets of size $\frac{|\mathcal{A}|^{M_\mathrm{T}}}{2} = 2^{R_0-1}$. Sphere decoder implementations of (12) are presented in [8], [10].

Hard vector ML demodulation amounts to the minimum distance problem

$$\hat{\mathbf{x}}_\mathrm{ML} = \arg\min_{\mathbf{x}\in\mathcal{A}^{M_\mathrm{T}}}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2. \quad (13)$$

This optimization problem can be solved by exhaustive search or using a sphere decoder; in both cases, the average and worst-case computational complexity scales exponentially with the number of transmit antennas. The detected code bits $\hat{c}_l$ corresponding to (13) are obtained via the one-to-one mapping between code bits and symbol vectors, i.e., $\hat{\mathbf{c}} = (\hat{c}_1 \dots \hat{c}_{R_0})^\mathrm{T} = \mu^{-1}(\hat{\mathbf{x}}_\mathrm{ML})$. It can be shown that the code bits $\hat{c}_l$ obtained by the hard ML detector correspond to the sign of the corresponding max-log LLRs in (12), i.e., $\hat{c}_l = \mathrm{u}(\tilde{\Lambda}_l)$ where $\mathrm{u}(\cdot)$ denotes the unit step function. When it comes to computing the system capacity with hard-output demodulators, the only difference to soft-output demodulation is the *discrete* nature of the outputs $\hat{c}_l$ of the equivalent "modulation" channel, which here becomes a binary channel. Consequently, the integral over $\tilde{\Lambda}_l$ in (9) is replaced with a summation over $\hat{c}_l \in \{0,1\}$.

B. Linear Demodulators

In the following, $\Lambda_k^{(i)}$ is the LLR corresponding to $c_k^{(i)}$ (the ith bit in the bit label of the kth symbol x_k). Soft demodulators with extremely low complexity can be obtained by using a linear (ZF or MMSE) equalizer followed by *per-layer* max-log LLR calculation according to

$$\tilde{\Lambda}_k^{(i)} = \frac{1}{\sigma_k^2}\left[\min_{x\in\mathcal{A}_i^0}|\hat{x}_k - x|^2 - \min_{x\in\mathcal{A}_i^1}|\hat{x}_k - x|^2\right] \quad (14)$$

for $i = 1,\dots,Q$, $k = 1,\dots,M_\mathrm{T}$. Here, $\mathcal{A}_i^b \subset \mathcal{A}$ denotes the set of (scalar) symbols whose bit label at position i equals b, $\hat{x}_k$ is an estimate of the symbol in layer k provided by the equalizer, and σ_k^2 is an equalizer-specific weight (see below). We emphasize that calculating LLRs separately for each layer results in a significant complexity reduction. In fact, calculating the symbol estimates $\hat{x}_k$ using a ZF or MMSE equalizer requires $\mathcal{O}(M_\mathrm{R}M_\mathrm{T}^2)$ operations; furthermore, the complexity of evaluating (14) for all code bits scales as $\mathcal{O}(M_\mathrm{R}M_\mathrm{T}2^Q) = \mathcal{O}(M_\mathrm{R}M_\mathrm{T}|\mathcal{A}|)$, i.e., linearly in the number of antennas [24], [25].

Equalization-based hard bit decisions $\hat{c}_k^{(i)}$ can be obtained by quantization of the equalizer output $\hat{x}_k$ with respect to $\mathcal{A}$ (denoted by $\mathcal{Q}(\cdot)$), followed by the demapping, i.e., $(\hat{c}_k^{(1)} \dots \hat{c}_k^{(Q)})^\mathrm{T} = \mu^{-1}(\mathcal{Q}(\hat{x}_k))$. Again, the detected code bits correspond to the sign of the LLRs, i.e., $\hat{c}_k^{(i)} = \mathrm{u}(\hat{\Lambda}_k^{(i)})$.

1) ZF-Based Demodulator [22], [23]: Here, the first stage consists of ZF equalization, i.e.

$$\hat{\mathbf{x}}_\mathrm{ZF} = (\mathbf{H}^\mathrm{H}\mathbf{H})^{-1}\mathbf{H}^\mathrm{H}\mathbf{y} = \mathbf{x} + \tilde{\mathbf{v}} \quad (15)$$

where the post-equalization noise vector $\tilde{\mathbf{v}}$ has correlation matrix

$$\mathbf{R}_{\tilde{\mathbf{v}}} = \mathbb{E}\{\tilde{\mathbf{v}}\tilde{\mathbf{v}}^\mathrm{H}\} = \sigma_v^2(\mathbf{H}^\mathrm{H}\mathbf{H})^{-1}. \quad (16)$$

Subsequently, approximate bit LLRs are obtained according to (14) with symbol estimate $\hat{x}_k = (\hat{\mathbf{x}}_\mathrm{ZF})_k$ and weight factor $\sigma_k^2 = (\mathbf{R}_{\tilde{\mathbf{v}}})_{k,k}$.

2) MMSE-Based Demodulator [24], [25]: Here, the first stage is an MMSE equalizer that can be written as [cf. (15) and (16)]

$$\hat{\mathbf{x}}_\mathrm{MMSE} = \mathbf{W}\hat{\mathbf{x}}_\mathrm{ZF}, \quad \text{with} \quad \mathbf{W} = \left(\mathbf{I} + \frac{M_\mathrm{T}}{E_s}\mathbf{R}_{\tilde{\mathbf{v}}}\right)^{-1}. \quad (17)$$

Approximate LLRs are then calculated according to (14) with

$$\hat{x}_k = \frac{(\hat{\mathbf{x}}_\mathrm{MMSE})_k}{W_k} \quad \text{and} \quad \sigma_k^2 = \frac{E_s}{M_\mathrm{T}}\frac{1-W_k}{W_k}$$

where $W_k = (\mathbf{W})_{k,k}$. Here, $\hat{x}_k$ denotes the output of the *unbiased* MMSE equalizer, which is preferable to a biased MMSE equalizer for nonconstant modulus modulation schemes such as 16-QAM and 64-QAM [47]; in the remainder we will thus restrict to unbiased soft and hard MMSE demodulators.

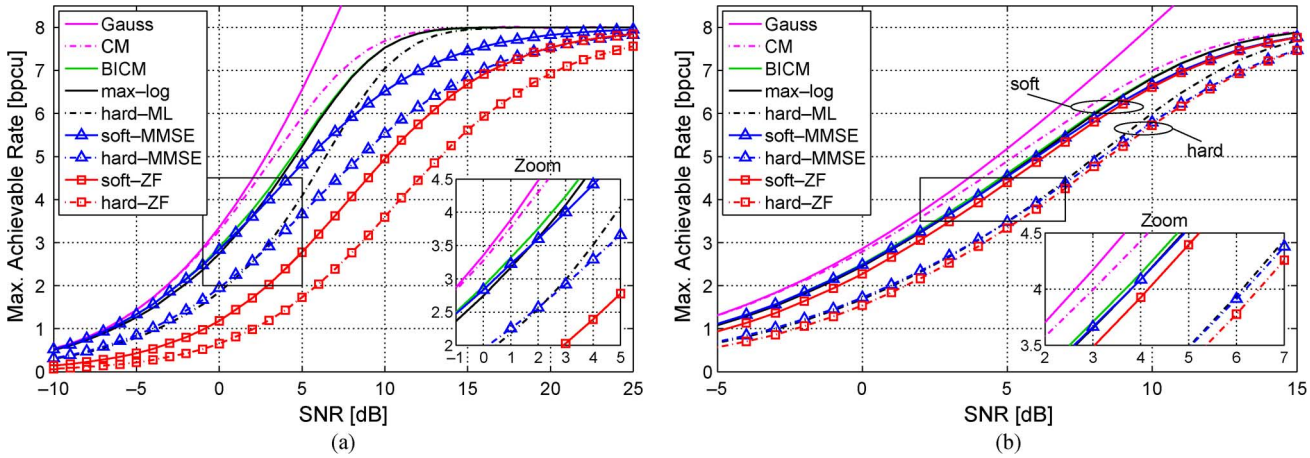


Fig. 2. Numerical capacity results for (a) a 4×4 MIMO system with 4-QAM, and (b) a 2×4 MIMO system with 16-QAM (in both cases with Gray labeling).

C. Capacity Results

We next compare the performance of the above baseline demodulators (i.e., max-log, hard ML, MMSE, and ZF) in terms of their maximum achievable rate C [ergodic capacity, see (9)]. In addition, the CM capacity C_{CM} in (4), the MIMO-BICM capacity C_{BICM} in (7) [corresponding to the capacity of the optimum soft MAP demodulator in (2)], and the Gaussian input channel capacity C_{G} in (3) are shown as benchmarks. Throughout the paper, all capacity results have been obtained for spatially i.i.d. Rayleigh fading, with all fading coefficients normalized to unit variance.

The pdfs required for evaluating (9) are generally hard to obtain in closed form. Thus, we measured these pdfs using Monte Carlo simulations and then evaluated all integrals numerically. Based on the results in [48], we numerically optimized the binning (used to measure the pdfs) in order to reduce the bias and variance of the mutual information estimates (see Appendix A for more details). The capacity results (obtained with 10^5 fading realizations) are shown in Fig. 2 in bits per channel use versus SNR $\rho \triangleq \frac{E_s}{\sigma_v^2}$. In the following, in some of the plots we show insets that provide zooms of the capacity curves around a rate of $\frac{R_0}{2}$ bpcu.

Fig. 2(a) pertains to the case of $M_{\mathrm{T}} \times M_{\mathrm{R}} = 4 \times 4$ MIMO with Gray-labeled 4-QAM (here, $R_0 = 8$). At a target rate of 4 bpcu, the SNR required for CM and Gaussian capacity is virtually the same, whereas that for BICM is larger by about 1.3 dB. The SNR penalty of using max-log demodulation instead of soft MAP is about 0.3 dB. Furthermore, hard ML demodulation requires a 2.1 dB higher SNR to achieve this rate than max-log demodulation; for soft and hard MMSE demodulation the SNR gaps to max-log are 0.2 dB and 3.1 dB, respectively, while for soft and hard ZF demodulation they respectively equal 5.1 dB and 8.1 dB. An interesting observation in this scenario is the fact that at low rates, soft and hard MMSE demodulation slightly outperform max-log and hard ML demodulation, respectively, whereas at high rates MMSE demodulation degrades to ZF performance. Hard MMSE demodulation can outperform hard ML demodulation since the latter minimizes the vector symbol error probability whereas our mutual information is defined on the bit level. Surprisingly, at low rates soft MMSE essentially coincides with BICM capacity. Moreover, soft MMSE demodulation outperforms hard ML demodulation at low-to-medium rates whereas at high rates it is the other way around (the crossover can be seen at about 5.8 bpcu). These observations reveal the somewhat unexpected fact that the demodulator performance ranking is not universal but depends on the target rate (or equivalently, the target SNR), even if the number of antennas, the symbol constellation, and the labeling are fixed. Similar observations apply to 16-QAM instead of 4-QAM and to set-partitioning labeling instead of Gray labeling (see [31]). Apart from a general shift of all curves to higher SNRs, the larger constellation and/or the different labeling strategy causes an increase of the gap between CM capacity and BICM capacity. The gaps between hard ML, hard MMSE, and soft ZF demodulation are significantly smaller, though, in this case (soft ZF outperforms hard MMSE for rates above 6.2 bpcu and approaches hard ML for rates around 6 bpcu). When decreasing the antenna configuration to a 2×2 system, we observed that soft ZF outperforms hard ML demodulation for low-to-medium rates, e.g., by about 1.7 dB at 4 bpcu with 16-QAM [31].

The excellent performance of MMSE demodulation at low rates may be attributed to the following two facts: (i) It is known that the MMSE filter is information lossless for Gaussian signaling. At low SNR the mutual information of the channel depends on the covariance of the input but is otherwise insensitive to the actual input distribution [49]. This suggest that at low SNR the MMSE filter is approximately information lossless also for non-Gaussian input alphabets, thereby at least partially explaining the excellent low-rate performance of MMSE demodulation. (ii) The soft MMSE demodulator was derived in [25] using a Gaussian approximation for the self-interference plus noise. This Gaussian approximation is more accurate at low SNR since here the Gaussian channel noise dominates the (non-Gaussian) interference.

The situation changes for the case of a 2×4 MIMO system with Gray-labeled 16-QAM (again $R_0 = 8$), shown in Fig. 2(b). The increased SNR gap between CM and BICM capacity implied by the larger constellation is compensated by having more receive than transmit antennas (this agrees with observations in [7]). In addition, the performance differences

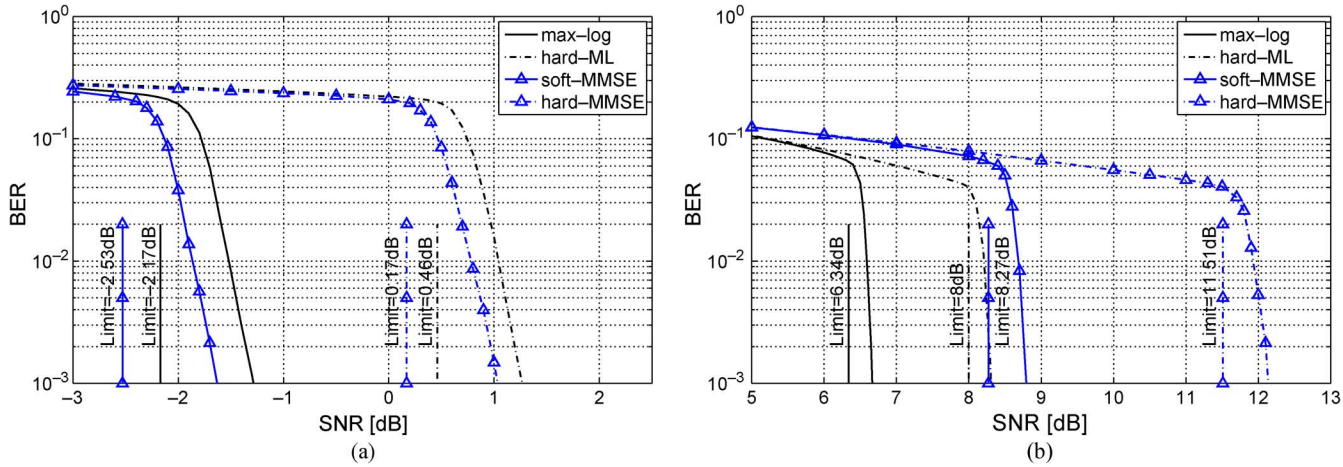


Fig. 3. BER versus SNR for a 4 × 4 MIMO system with Gray-labeled 4-QAM and LDPC codes of (a) rate 1/4 and (b) rate 3/4.

between the individual demodulators are significantly reduced, revealing an essential distinction being between soft and hard demodulators. Having $M_{\mathrm{R}} > M_{\mathrm{T}}$ helps the linear demodulators approach their non-linear counterparts even at larger rates, i.e., soft ZF/MMSE perform close to max-log and hard ZF/MMSE perform close to hard ML, with an SNR gap of about 2.3 dB between hard and soft demodulators. Note that in this scenario soft MMSE and soft ZF both outperform hard ML demodulation at all rates.

D. BER Performance

Even though we advocate a demodulator comparison in terms of mutual information, the crossover of some of the capacity curves prompts a verification in terms of the BER of soft and hard MMSE demodulation as well as max-log and hard ML demodulation. We consider a 4 × 4 MIMO-BICM system with Gray-labeled 4-QAM in conjunction with irregular LDPC codes[7] [50] of block length 64000. For the case of soft demodulation, the LDPC codes were designed for an additive white Gaussian noise (AWGN) channel whereas for the case of hard demodulators the design was for a binary symmetric channel. At the receiver, message-passing LDPC decoding [50] was performed. In the case of hard demodulation, the message-passing decoder was provided with the LLRs

$$\tilde{\Lambda}_l = (2\hat{c}_l - 1)\log\frac{1-p_0}{p_0} \tag{18}$$

where $p_0 = \mathrm{P}\{\hat{c}_l \neq c_l | c_l\}$, the crossover probability of the equivalent binary channel, was determined via Monte Carlo simulations. With the soft demodulators, we performed an LLR correction via a lookup table as in [45]. Using LLR correction for soft demodulators and (18) for hard output demodulators is critical in order to provide the channel decoder with accurate reliability information [36], [41]–[45].

The BERs obtained for code rates of 1/4 (2 bpcu) and 3/4 (6 bpcu) are shown in Fig. 3(a) and (b), respectively. Vertical lines indicate the respective capacity limits, i.e., the minimum SNR required for the target rate according to Fig. 2(a). It is seen that the LDPC code designs are less than 1 dB away from the capacity limits. At low rates soft MMSE performs best and hard ML performs worst whereas at high rates max-log and hard MMSE give the best and worst results, respectively. More specifically, at rate 1/4 soft MMSE outperforms max-log and hard ML demodulation by 0.3 and 2.9 dB, respectively [cf. Fig. 3(a)]; at rate 3/4 soft MMSE performs 0.5 dB poorer than hard ML and 2.1 dB poorer than max-log [cf. Fig. 3(b)]. These BER results confirm the capacity-based observation that there is no universal (i.e., rate- and SNR-independent) demodulator performance ranking. We note that the block error rate results in [51] imply similar conclusions, even though not explicitly mentioned in that paper.

[7]The LDPC code design was performed using the web-tool at http://lthcwww.epfl.ch/research/ldpcopt.

V. Other Demodulators

In the following, we study the mutual information of several other MIMO-BICM demodulators that differ in their underlying principle and their computational complexity. Unless stated otherwise, capacity results shown in this section pertain to a 4 × 4 MIMO system with 4-QAM using Gray labeling ($R_0 = 8$). The results for asymmetric 2 × 4 MIMO systems with 16-QAM (shown in [31] but not here) essentially confirm the general distinction between hard and soft demodulators observed in connection with Fig. 2(b).

A. List-Based Demodulators

In order to save computational complexity, (12) can be approximated by decreasing the size of the search set, i.e., replacing $\mathcal{A}^{M_{\mathrm{T}}}$ with a smaller set. Usually, this is achieved by generating a (nonempty) candidate list $\mathcal{L} \subseteq \mathcal{A}^{M_{\mathrm{T}}}$ and restricting the search in (12) to this list, i.e.

$$\tilde{\Lambda}_l = \frac{1}{\sigma_v^2}\left[\min_{\mathbf{x}\in\mathcal{L}\cap\mathcal{X}_l^0}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2 - \min_{\mathbf{x}\in\mathcal{L}\cap\mathcal{X}_l^1}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2\right]. \tag{19}$$

As the number of operations required to compute the metric for each candidate of the list is $\mathcal{O}(M_{\mathrm{T}}M_{\mathrm{R}})$, the overall computational complexity of the metric evaluations and minima searches in (19) scales as $\mathcal{O}(M_{\mathrm{T}}M_{\mathrm{R}}|\mathcal{L}|)$. Thus, the list size $|\mathcal{L}|$ allows

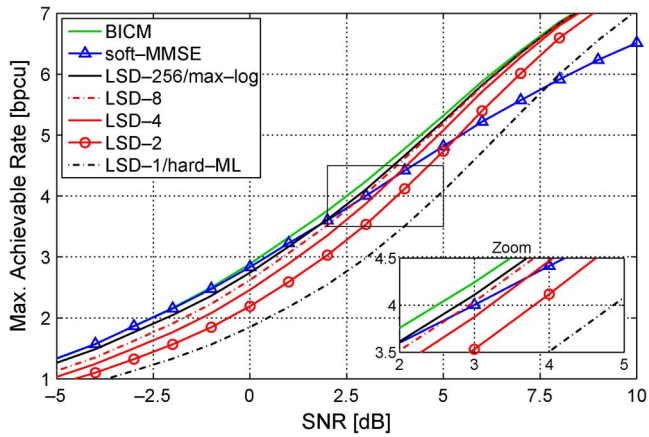


Fig. 4. Mutual information of LSD with list size $|\mathcal{L}| \in \{1, 2, 4, 8, 256\}$ (4×4 MIMO, 4-QAM, Gray labeling).

to trade off performance for complexity savings. A larger list size generally incurs higher complexity but yields more accurate approximations of the max-log LLRs. For a fixed list size, the performance further depends on how the list $\mathcal{L}$ is generated. In the following, we consider two types of list generation, one based on sphere decoding and the other on bit flipping.

1) List Sphere Decoder (LSD): The LSD proposed in [15] uses a simple modification of the hard-decision sphere decoder [52] to generate the candidate list $\mathcal{L}$ such that it contains the $|\mathcal{L}|$ symbol vectors $\mathbf{x}$ with the smallest ML metric $\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2$ [thus, by definition $\mathcal{L}$ contains the hard ML solution $\hat{\mathbf{x}}_{\mathrm{ML}}$ in (13)]. If the lth bit in the labels of *all* $\mathbf{x} \in \mathcal{L}$ equals 1, the set $\mathcal{L} \cap \mathcal{X}_l^0$ is empty and (19) cannot be evaluated. Since in this case there is strong evidence for $c_l = 1$ (at least if $|\mathcal{L}|$ is not too small), the LLR $\tilde{\Lambda}_l$ is set to a prescribed positive value $\check{\Lambda} \gg 0$. Analogously, $\tilde{\Lambda}_l = -\check{\Lambda}$ in case $\mathcal{L} \cap \mathcal{X}_l^1$ is empty.

While the LSD may offer significant complexity savings compared to max-log demodulation, statements about its computational complexity are difficult and depend strongly on the actual implementation of the sphere decoder as well as the choice of the list size (for details we refer to [15]). We note that the case $|\mathcal{L}| = 2^{R_0} = |\mathcal{A}^{M_{\mathrm{T}}}|$ implies $\mathcal{L} = \mathcal{A}^{M_{\mathrm{T}}}$; thus, $\mathcal{L} \cap \mathcal{X}_l^b = \mathcal{X}_l^b$ such that (19) equals the max-log demodulator in (12). The other extreme is a list size of one, i.e., $\mathcal{L} = \{\hat{\mathbf{x}}_{\mathrm{ML}}\}$ [cf. (13)], in which case either $\mathcal{L} \cap \mathcal{X}_l^0$ or $\mathcal{L} \cap \mathcal{X}_l^1$ is empty (depending on the bit label of $\hat{\mathbf{x}}_{\mathrm{ML}}$); here, $\tilde{\Lambda}_l = (2\hat{c}_l - 1)\check{\Lambda}$ where $\hat{\mathbf{c}} = (\hat{c}_1 \ldots \hat{c}_{R_0})^{\mathrm{T}} = \mu^{-1}(\hat{\mathbf{x}}_{\mathrm{ML}})$ and thus the LSD output is equivalent to hard ML demodulation (except for the choice of $\check{\Lambda}$, which is irrelevant, however, for capacity).

Capacity Results: Fig. 4 shows the maximum rates achievable with an LSD for various list sizes. BICM and soft MMSE capacity are shown for comparison. Note that with 4-QAM and $M_{\mathrm{T}} = 4$, $|\mathcal{L}| = 256$ and $|\mathcal{L}| = 1$ correspond to max-log and hard ML demodulation, respectively. It is seen that with increasing list size the gap between LSD and max-log decreases rapidly, specifically at high rates. In particular, the LSD with list sizes of $|\mathcal{L}| \geq 8$ is already quite close to max-log performance. However, at low rates LSD (even with large list sizes) is outperformed by soft MMSE: below 5.3, 3.7, and 2.8 dB the mutual information of soft MMSE is higher than that of LSD with list size 2, 4, and 8, respectively. Similar observations apply to other antenna configurations and symbol constellations (see [31]).

2) Bit Flipping Demodulators: Another way of generating the candidate list $\mathcal{L}$, proposed in [21], is to flip some of the bits in the label of the hard ML symbol vector estimate $\hat{\mathbf{x}}_{\mathrm{ML}}$ in (13). More generally, the ML solution $\hat{\mathbf{x}}_{\mathrm{ML}}$ can be replaced by a symbol vector $\hat{\mathbf{x}} \in \mathcal{A}^{M_{\mathrm{T}}}$ obtained with an arbitrary hard-output demodulator (e.g., hard ZF and MMSE demodulation). Let $\hat{\mathbf{c}} = \mu^{-1}(\hat{\mathbf{x}})$ denote the bit label of $\hat{\mathbf{x}}$. The candidate list then consists of all symbol vectors whose bit label has Hamming distance at most $D \leq R_0$ from $\hat{\mathbf{c}}$, i.e., $\mathcal{L} = \{\mathbf{x} : d_{\mathrm{H}}(\mu^{-1}(\mathbf{x}), \hat{\mathbf{c}}) \leq D\}$. Here, $d_{\mathrm{H}}(\mathbf{c}_1, \mathbf{c}_2)$ denotes the Hamming distance between two bit labels $\mathbf{c}_1$ and $\mathbf{c}_2$. This list can be generated by systematically flipping up to D bits in $\hat{\mathbf{c}}$ and mapping the results to symbol vectors. The resulting list size is given by $|\mathcal{L}| = \sum_{d=0}^{D} \binom{R_0}{d}$. Here, the structure of the list generated with bit flipping allows to reduce the complexity per candidate to $\mathcal{O}(M_{\mathrm{R}})$,[8] giving an overall complexity of $\mathcal{O}(M_{\mathrm{R}}|\mathcal{L}|)$ (plus the operations required for the initial estimate). For $D = R_0$, $\mathcal{L} = \mathcal{A}^{M_{\mathrm{T}}}$ and (19) reduces to max-log demodulation; furthermore, with $\hat{\mathbf{x}} = \hat{\mathbf{x}}_{\mathrm{ML}}$ and $D = 0$ there is $\mathcal{L} = \{\hat{\mathbf{x}}_{\mathrm{ML}}\}$ and (19) becomes equivalent to hard ML demodulation. In contrast to the LSD, bit flipping with $D > 0$ ensures that $\mathcal{L} \cap \mathcal{X}_l^0$ and $\mathcal{L} \cap \mathcal{X}_l^1$ are nonempty so that (19) can always be evaluated.

Capacity Results: Fig. 5 shows the maximum rates achievable with bit flipping demodulation where the initial symbol vector estimate is chosen either as the hard ML solution $\hat{\mathbf{x}}_{\mathrm{ML}}$ in (13) or the hard MMSE estimate $\mathcal{Q}(\hat{\mathbf{x}}_{\mathrm{MMSE}})$ (cf. (17)). For $D = 1$ ($|\mathcal{L}| = 9$), Fig. 5(a) reveals that flipping 1 bit (labeled 'flip-1') allows for significant performance improvements over the respective initial hard demodulator (about 2.1 dB at 2 bpcu). For rates below 5 bpcu, hard ML and hard MMSE initialization yield effectively identical results, with a maximum loss of 0.9 dB (at 3.5 bpcu) compared to soft MMSE. At higher rates, MMSE-based bit flipping outperforms soft MMSE demodulation slightly. For $D = 2$ ($|\mathcal{L}| = 37$), it can be seen from Fig. 5(b) that bit flipping demodulation performs close to max-log below 4 bpcu and that hard ML and hard MMSE initialization are very close to each other for almost all rates and SNRs; in fact, below 6.7 bpcu hard MMSE initialization performs slightly better than hard ML initialization while at higher rates ML initialization gives slightly better results. To maintain this behavior for larger constellations and more antennas, the maximum Hamming distance D may have to increase with increasing R_0 (see [31]).

B. LR-Aided Demodulation

LR is an important technique for improving the performance or complexity of MIMO demodulators for the case of QAM constellations [16], [17], [53]. The basic underlying idea is to view the columns of the channel matrix $\mathbf{H}$ as basis vectors of a point lattice. LR then yields an alternative basis which amounts to a transformation of the system model (1) prior to demodulation; the advantage of such an approach is that the transformed

[8]Changing the value of a particular bit changes only one symbol in the symbol vector. Thus, the residual $\mathbf{y} - \mathbf{H}\mathbf{x}$ in (19) can be easily updated by adding an appropriately scaled column of $\mathbf{H}$. This requires only $\mathcal{O}(M_{\mathrm{R}})$ instead of $\mathcal{O}(M_{\mathrm{T}}M_{\mathrm{R}})$ operations.

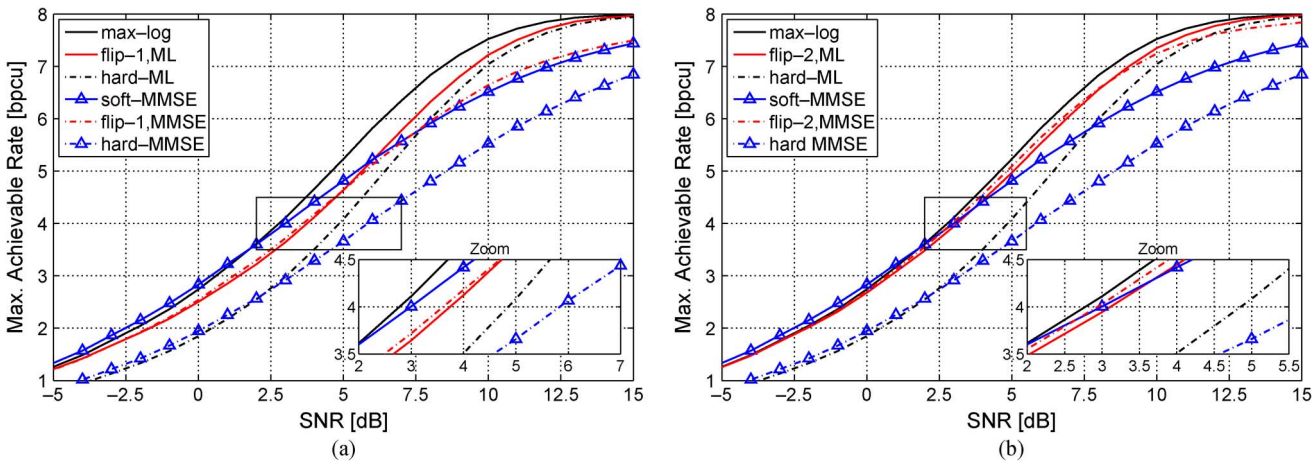


Fig. 5. Mutual information of bit flipping demodulator with (a) $D = 1$ and (b) $D = 2$ (4×4 MIMO, 4-QAM, Gray labeling).

channel matrix (i.e., the reduced basis) has improved properties (e.g., smaller condition number). An efficient algorithm to obtain a reduced basis was proposed by Lenstra, Lenstra, and Lovász (LLL) [54]. The overall computational complexity of LR-aided demodulation depends on the complexity of the LLL algorithm which is currently an active research topic. Bounds on the average computational complexity of the LLL algorithm have been provided in [55]. A comparison of different LR methods in the context of MIMO hard demodulation was provided in [56].

Since LR algorithms are often formulated for equivalent real-valued models, we assume for now that all quantities are real-valued. Any lattice basis transformation is described by a unimodular transformation matrix $\mathbf{T}$, i.e., a matrix with integer entries and $\det(\mathbf{T}) = \pm 1$. Denoting the "reduced channel" by $\tilde{\mathbf{H}} = \mathbf{HT}$ and defining $\mathbf{z} = \mathbf{T}^{-1}\mathbf{x}$, the system model (1) can be rewritten as

$$\mathbf{y} = \mathbf{Hx} + \mathbf{v} = \tilde{\mathbf{H}}\mathbf{z} + \mathbf{v}. \tag{20}$$

Under the assumption $\mathbf{x} \in \mathbb{Z}^{M_{\mathrm{T}}}$ (which for QAM can be ensured by an appropriate offset and scaling), the unimodularity of $\mathbf{T}$ guarantees $\mathbf{z} \in \mathbb{Z}^{M_{\mathrm{T}}}$ and hence any demodulator can be applied to the better-behaved transformed system model on the right-hand side of (20). LR-aided soft demodulators (cf. [18]) are essentially list-based [19], [20], and often apply bit flipping (cf. Section V-A-2) to LR-aided hard-output demodulators. Here, we restrict to LR-aided hard and soft output MMSE demodulation [17].

Capacity Results: Fig. 6 shows the capacity results for hard and soft LR-aided MMSE demodulation. Soft outputs are obtained by applying bit flipping with $D = 1$ and $D = 2$ to the LR-aided hard MMSE demodulator output (cf. Section V-A-2). It is seen for 4×4 MIMO with 4-QAM ($R_0 = 8$) that LR with hard MMSE demodulation shows a significant performance advantage over hard MMSE for SNRs above 7.2 dB (rates higher than 4.5 bpcu). At rates higher than about 7.1 bpcu, LR-aided hard demodulation even outperforms soft MMSE. Bit flipping is helpful particularly at low-to-medium rates. For SNRs below

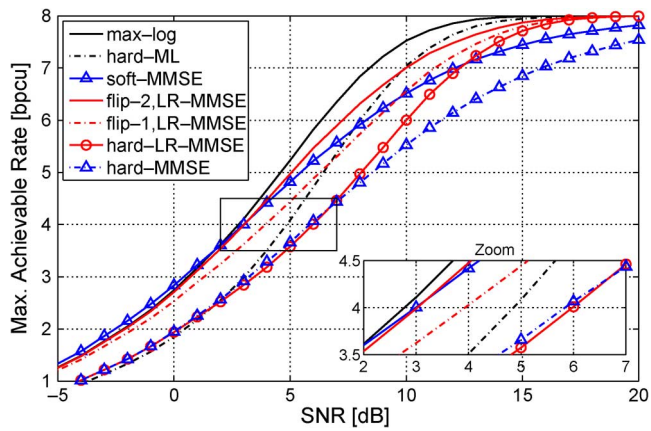


Fig. 6. Mutual information of LR-aided hard and soft MMSE demodulation (4×4 MIMO, 4-QAM, Gray labeling).

6.8 dB (rates lower than 5.2 bpcu) LR-aided soft demodulation with $D = 1$ performs better than hard ML. When flipping up to $D = 2$ bits, LR-aided soft demodulation approaches max-log performance and reveals a significant performance advantage over soft MMSE without LR in the high-rate regime.

C. Semidefinite Relaxation Demodulation

Based on convex optimization techniques, semidefinite relaxation (SDR) is an approach to approximately solve the hard ML problem (13) with polynomial worst-case complexity [13], [57]. We specifically consider hard-output and soft-output versions of an SDR demodulator that approximates max-log demodulation and has an overall worst-case complexity of $\mathcal{O}(R_0^{4.5})$ (see [14]). We note that this approach applies only to BPSK or 4-QAM alphabets and employs a randomization procedure described in detail in [13].

Capacity Results: In Fig. 7 we show the mutual information for a 4×4 MIMO system with 4-QAM ($R_0 = 8$) using hard and soft SDR demodulation (as described in [14]) and randomization with 25 trials. Surprisingly, hard and soft SDR demodulation here exactly match the performance of hard ML and max-log demodulation, respectively.

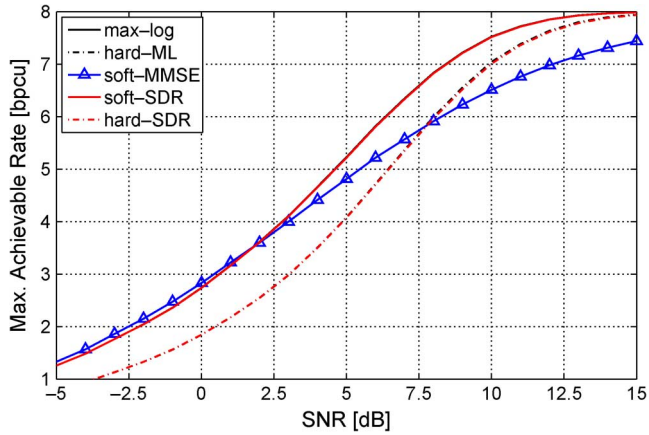


Fig. 7. Mutual information of hard and soft SDR demodulation (4 × 4 MIMO, 4-QAM, Gray labeling). The curves for hard and soft SDR are identical to those for hard ML and max-log, respectively.

D. Infinity-Norm Demodulator

The VLSI implementation complexity of the sphere decoder for hard ML demodulation is significantly reduced by replacing the ℓ^2 norm in (13) with the ℓ^∞ norm, i.e.,

$$\hat{\mathbf{x}}_\infty = \arg\min_{\mathbf{x}\in\mathcal{A}^{M_\mathrm{T}}} \|\mathbf{Q}^\mathrm{H}\mathbf{y} - \mathbf{R}\mathbf{x}\|_\infty. \tag{21}$$

Here, the ℓ^∞ norm is defined as $\|\mathbf{a}\|_\infty \triangleq \max\{|\mathrm{Re}\{a_1\}|, \ldots, |\mathrm{Re}\{a_M\}|, |\mathrm{Im}\{a_1\}|, \ldots, |\mathrm{Im}\{a_M\}|\}$ and $\mathbf{Q}$ and $\mathbf{R}$ are the $M_\mathrm{R} \times M_\mathrm{T}$ unitary and $M_\mathrm{T} \times M_\mathrm{T}$ upper triangular factors in the QR decomposition $\mathbf{H} = \mathbf{Q}\mathbf{R}$ of the channel matrix. The advantage of (21) is that expensive squaring operations are avoided and fewer nodes are visited during the tree search underlying the sphere decoder [11], [12]. If (21) has no unique solution, one solution is selected at random.

Soft outputs can be generated by using the ℓ^∞-norm sphere decoder to determine

$$\tilde{\mathbf{x}}_l^b = \arg\min_{\mathbf{x}\in\mathcal{X}_l^b} \|\mathbf{Q}^\mathrm{H}\mathbf{y} - \mathbf{R}\mathbf{x}\|_\infty$$

for $b \in \{0, 1\}$ and then evaluating the approximate LLRs using the ℓ^2 norm

$$\tilde{\Lambda}_l = \frac{1}{\sigma_v^2}\left[\left\|\mathbf{y} - \mathbf{H}\tilde{\mathbf{x}}_l^0\right\|^2 - \left\|\mathbf{y} - \mathbf{H}\tilde{\mathbf{x}}_l^1\right\|^2\right].$$

Capacity Results: Fig. 8 shows the system capacity for hard and soft ℓ^∞-norm demodulation. For the 4 × 4 case with 4-QAM in Fig. 8(a), hard and soft ℓ^∞-norm demodulation perform within 1 dB of hard ML and max-log, respectively. At rates below about 4 bpcu, ℓ^∞-norm demodulation is outperformed by MMSE demodulation, though. For the 2 × 4 case with 16-QAM depicted in Fig. 8(b), all soft-output baseline demodulators perform almost identical and the same is true for all hard-output baseline demodulators, i.e., there is only a distinction between soft and hard demodulation (cf. Fig. 2(b)). However, soft and hard ℓ^∞-norm demodulation perform significantly worse in this asymmetric setup, specifically at low-to-medium rates. At 2 bpcu, soft ℓ^∞-norm demodulation requires 1.75 dB higher SNR than max-log and soft MMSE and hard ℓ^∞-norm demodulation requires 2.3 dB higher SNR than hard ML/MMSE.

E. Successive and Soft Interference Cancelation

Successive interference cancelation (SIC) is a hard-output demodulation approach that became popular with the V-BLAST (*Vertical Bell Labs Layered Space-Time*) system [28]. Within one SIC iteration, only the layer with the largest post-equalization SNR is detected and its contribution to the receive signal is subtracted (canceled). A SIC implementation that replaces the ZF detector from [28] with an MMSE detector and orders the layers efficiently according to signal-to-interference-plus-noise-ratio (SINR) was presented in [29]. Note that this approach shows a complexity order of $\mathcal{O}(M_\mathrm{R}M_\mathrm{T}^2)$ which is the same as for linear MIMO demodulation. Suboptimal but more efficient SIC schemes are discussed in [30].

A parallel soft interference cancelation (SoftIC) scheme with reduced error propagation was proposed in [26]. SoftIC is an iterative method that iteratively performs (i) parallel MIMO interference cancelation based on soft symbols and (ii) computation of improved soft symbols using the output of the interference cancelation stage. The complexity of one SoftIC iteration depends linearly on the number of antennas. Here, we use a modification that builds upon bit-LLRs. Let $\tilde{\Lambda}_k^{(i)}[j]$ denote the LLR for the ith bit in layer k obtained in the jth iteration. Symbol probabilities can then be obtained as

$$P_k^{(j)}(x) = \prod_{i=1}^{Q} \frac{\exp\left(b_i(x)\tilde{\Lambda}_k^{(i)}[j]\right)}{1 + \exp\left(\tilde{\Lambda}_k^{(i)}[j]\right)}$$

with $b_i(x)$ denoting the ith bit in the label of $x \in \mathcal{A}$, leading to the soft symbol estimate

$$\tilde{\mathbf{x}}_k^{(j)} = \sum_{x\in\mathcal{A}} x P_k^{(j)}(x).$$

Soft interference cancelation for each layer then yields

$$\mathbf{y}_k^{(j)} = \mathbf{y} - \sum_{k'\neq k} \mathbf{h}_{k'}\tilde{x}_{k'}^{(j)} = \mathbf{h}_k x_k + \sum_{k'\neq k} \mathbf{h}_{k'}\left(x_{k'} - \tilde{x}_{k'}^{(j)}\right) + \mathbf{v} \tag{22}$$

where $\mathbf{h}_k$ denotes the kth column of $\mathbf{H}$. Finally, updated LLRs $\tilde{\Lambda}_k^{(i)}[j+1]$ are calculated from (22) based on a Gaussian assumption for the residual interference plus noise (for details we refer to [26]). In contrast to [26], we suggest to initialize the scheme with the LLRs obtained by a low-complexity soft demodulator, e.g., the soft ZF demodulator in Section IV-B. By carefully counting operations it can be shown that the complexity per iteration of the above SoftIC algorithm scales as $\mathcal{O}(2^Q M_\mathrm{T}(Q + M_\mathrm{R}))$.

Capacity Results: In Fig. 9(a), we display capacity results for (hard) MMSE-SIC with detection ordering as in [30] (therein referred to as 'MMSE-BLAST') and for SoftIC demodulation with 3 iterations (initialized using a soft ZF demodulator whose performance is shown for reference). Hard MMSE-SIC demodulation is seen to perform similarly to hard ML demodulation at low rates and even outperforms it slightly at very low rates. At high rates, MMSE-SIC shows a noticeable gap to hard ML but outperforms soft MMSE and SoftIC. SoftIC is superior to

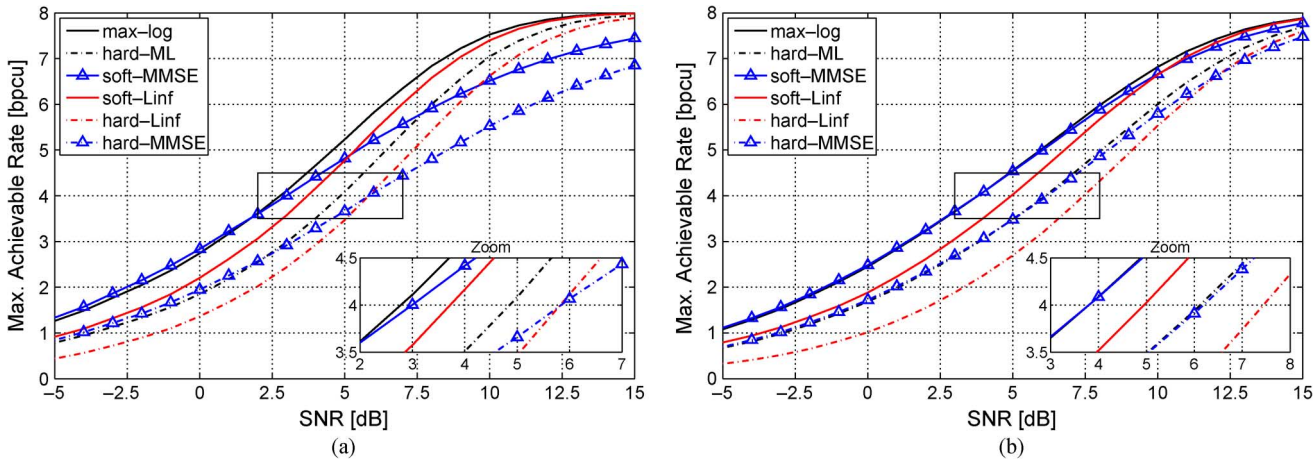


Fig. 8. Mutual information of hard and soft ℓ^∞-norm demodulation for (a) 4×4 MIMO with 4-QAM. (b) 2×4 MIMO with 16-QAM (Gray labeling in both cases).

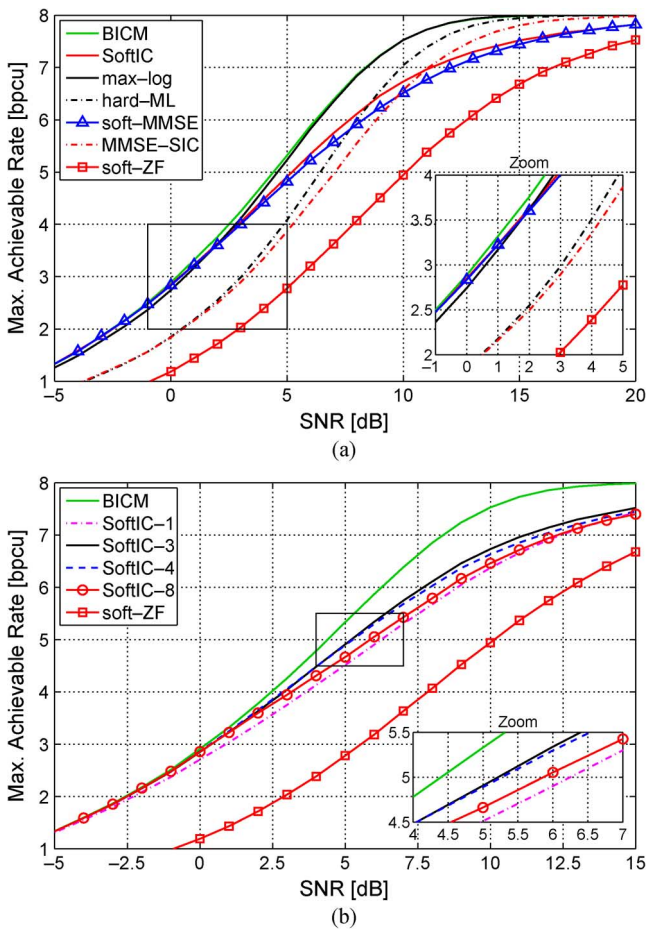


Fig. 9. Mutual information of MMSE-SIC and SoftIC (4×4 MIMO, 4-QAM, Gray labeling).

MMSE-SIC for rates of up to 7 bpcu (SNRs below 11 dB). At low rates, SoftIC even performs slightly better than max-log demodulation and essentially coincides with BICM capacity and soft MMSE. For the chosen system parameters, SoftIC closely matches soft MMSE at low rates and even outperforms it at high rates. This statement is not generally valid, however. For example, with 16-QAM SoftIC performs poorer than soft MMSE even at high rates (see [31]).

At high SNRs, we observed that SoftIC performance degrades if iterated too long (see Fig. 9(b), showing SoftIC with 1, 3, 4, and 8 iterations). This can be explained by the fact that at high SNRs the residual interference-plus-noise becomes very small and hence the LLR magnitudes grow unreasonably large. Our simulations showed that SoftIC performs best when terminated after 2 or 3 iterations.

VI. Imperfect Channel State Information (CSI)

We next investigate the ergodic capacity (9) for the case of imperfect CSI. In particular, we consider training-based estimation of the channel matrix $\mathbf{H}$ and the noise variance σ_v^2 and assess how the amount of training influences the performance of the various demodulators.

Training-Based Channel Estimation: To estimate the channel, the transmitter sends $N_{\mathrm{p}} > M_{\mathrm{T}}$ training vectors[9] which are arranged into an $M_{\mathrm{T}} \times N_{\mathrm{p}}$ training matrix $\mathbf{X}_{\mathrm{p}}$. We assume that $\mathbf{X}_{\mathrm{p}}$ has full rank and has Frobenius norm [58] $\|\mathbf{X}_{\mathrm{p}}\|_{\mathrm{F}}^2 = N_{\mathrm{p}} E_s$ such that the power per channel use for training is the same as for the data. Assuming that the channel stays constant for the duration of one block (which contains training and actual data), the $M_{\mathrm{R}} \times N_{\mathrm{p}}$ receive matrix $\mathbf{Y}_{\mathrm{p}}$ induced by the training is given by

$$\mathbf{Y}_{\mathrm{p}} = \mathbf{H}\mathbf{X}_{\mathrm{p}} + \mathbf{V}. \tag{23}$$

Here, $\mathbf{V}$ is an $M_{\mathrm{R}} \times N_{\mathrm{p}}$ i.i.d. Gaussian noise matrix.

Using (23), the least-squares (ML) channel estimate is computed as [59]

$$\hat{\mathbf{H}} = \mathbf{Y}_{\mathrm{p}}\mathbf{X}_{\mathrm{p}}^{\mathrm{H}} \left(\mathbf{X}_{\mathrm{p}}\mathbf{X}_{\mathrm{p}}^{\mathrm{H}}\right)^{-1}. \tag{24}$$

This estimate is unbiased and its mean square error equals

$$\mathbb{E}\left\{\|\hat{\mathbf{H}} - \mathbf{H}\|_{\mathrm{F}}^2\right\} = M_{\mathrm{R}}\sigma_v^2 \mathrm{tr}\left\{\left(\mathbf{X}_{\mathrm{p}}\mathbf{X}_{\mathrm{p}}^{\mathrm{H}}\right)^{-1}\right\} \geq \frac{M_{\mathrm{R}} M_{\mathrm{T}}^2}{N_{\mathrm{p}}} \frac{1}{\rho}$$

[9]While $N_{\mathrm{p}} = M_{\mathrm{T}}$ is sufficient to estimate $\mathbf{H}$, extra training is required for estimation of σ_v^2.

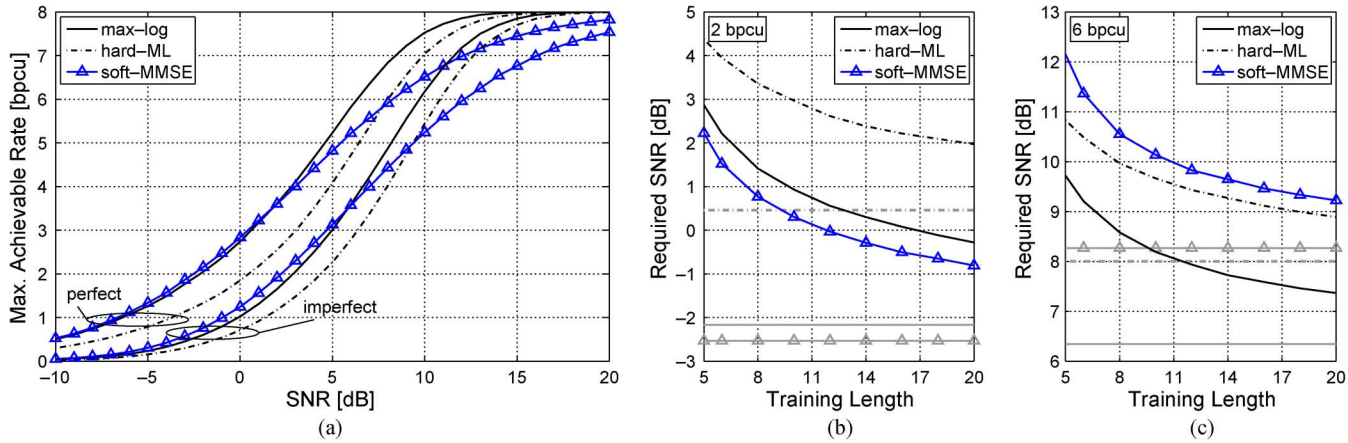


Fig. 10. Impact of imperfect CSI on baseline demodulators: (a) capacity versus SNR for $N_{\mathrm{p}} = 5$; (b), (c) required SNR versus training length N_{p} for a target rate of (b) $C = 2$ bpcu and (c) $C = 6$ bpcu (4×4 MIMO, 4-QAM, Gray labeling).

where the lower bound is attained with orthogonal training sequences, i.e., $\mathbf{X}_{\mathrm{p}}\mathbf{X}_{\mathrm{p}}^{\mathrm{H}} = \frac{N_{\mathrm{p}}E_s}{M_{\mathrm{T}}}\mathbf{I}$ (we recall that $\rho = \frac{E_s}{\sigma_v^2}$ denotes the SNR). The noise variance is then estimated as the mean power of $\mathbf{Y}_{\mathrm{p}}$ in the $(N_{\mathrm{p}} - M_{\mathrm{T}})$-dimensional orthogonal complement of the range space of $\mathbf{X}_{\mathrm{p}}^{\mathrm{H}}$, i.e.

$$\hat{\sigma}_v^2 = \frac{1}{M_{\mathrm{R}}(N_{\mathrm{p}} - M_{\mathrm{T}})}\|\mathbf{Y}_{\mathrm{p}} - \hat{\mathbf{H}}\mathbf{X}_{\mathrm{p}}\|_{\mathrm{F}}^2. \tag{25}$$

The noise variance estimate is unbiased and its MSE is independent of the transmit power

$$\mathbb{E}\left\{\left|\sigma_v^2 - \sigma_v^2\right|^2\right\} = \frac{\sigma_v^4}{M_{\mathrm{R}}(N_{\mathrm{p}} - M_{\mathrm{T}})}.$$

Capacity Results: We show results for the ergodic capacity of mismatched[10] max-log, hard ML, and soft MMSE demodulation where the true channel matrix and noise variance are replaced by $\hat{\mathbf{H}}$ in (24) and $\hat{\sigma}_v^2$ in (25), respectively. Throughout, a 4×4 MIMO system with 4-QAM and Gray labeling is considered $(R_0 = 8)$. Results for other demodulators with imperfect CSI are provided in [31].

Fig. 10(a) shows the maximum achievable rates versus SNR for a fixed orthogonal training sequence of length $N_{\mathrm{p}} = 5$ (the worst case with minimum amount of training). It is seen that imperfect CSI causes a significant performance degradation of all three demodulators, e.g., at 4 bpcu the SNR losses are 3.9 dB (max-log), 3.2 dB (hard ML), and 4 dB (soft MMSE). In this worst case setup (minimum training length), the performance advantage of soft MMSE over hard ML at low rates is slightly less pronounced; the intersection of hard ML and soft MMSE performance shifts from 5.8 bpcu (at an SNR of about 7.7 dB) for perfect CSI to 5 bpcu (at 9.4 dB) for imperfect CSI. However, the gap between soft MMSE and max-log is slightly larger at low rates, e.g., 0.7 dB at 2 bpcu. The performance losses for all demodulators tend to be smaller at high rates, which may be partly attributed to the fact that the CSI becomes more accurate with increasing SNR. In general it can be observed that the performance loss of hard ML is the smallest while soft MMSE and max-log performance deteriorates stronger; unlike max-log and soft MMSE, hard ML does not use the noise variance and hence is more robust to estimation errors in σ_v^2.

To investigate the impact of the amount of training, Fig. 10(b) and (c) depict the minimum SNRs required by the individual demodulators to achieve target rates of 2 bpcu and 6 bpcu, respectively, versus the training length N_{p}. It is seen that for all demodulators, the required SNR decreases rapidly with increasing amount of training. Yet, even for $N_{\mathrm{p}} = 20$ there is a significant gap of 1 to 2 dB to perfect CSI performance (indicated by horizontal gray lines with corresponding line style). Here, soft MMSE consistently performs better than max-log and hard ML at 2 bpcu. In contrast, hard ML outperforms soft MMSE at 6 bpcu, especially for very small training durations.

The results shown in Fig. 10 correspond to a worst case scenario where both, channel and noise variance, are imperfectly known. Further capacity results, specifically for the case of imperfect channel knowledge but perfect noise variance and for other demodulators discussed in Section V are provided in [31]. These results generally show that imperfect receiver CSI degrades the performance throughout for all investigated demodulation schemes. An interesting observations is that—in the MIMO setup considered—the LSD with list size $|\mathcal{L}| \geq 8$ consistently outperforms max-log for training duration $N_{\mathrm{p}} = 5$ [31]. For larger training durations, LSD with $|\mathcal{L}| = 8$ performs slightly poorer than but still very close to max-log.

VII. Quasi-Static Fading

In this section we provide a demodulator performance comparison for quasi-static fading MIMO channels based on the outage probability $P_{\mathrm{out}}(R)$ in (10) and the ϵ-capacity C_ϵ in (11). The setup considered (4×4 MIMO with Gray-labeled 4-QAM) is the same as before apart from the spatially i.i.d. Rayleigh fading channel which now is assumed to be quasi-static. The outage probability $P_{\mathrm{out}}(R)$ was measured using 10^5 blocks (affected by independent fading realizations), each consisting of 10^5 symbol vectors. To keep the presentation concise, we restrict to the baseline demodulators from Section IV.

Fig. 11(a) shows the outage probability versus SNR for rates of $R = 2$ bpcu and $R = 6$ bpcu. For $R = 2$ bpcu, soft

[10]One could also modify these demodulators in order to take into account the fact that the CSI is imperfect as e.g. in [60]; however, this is beyond the scope of this paper.

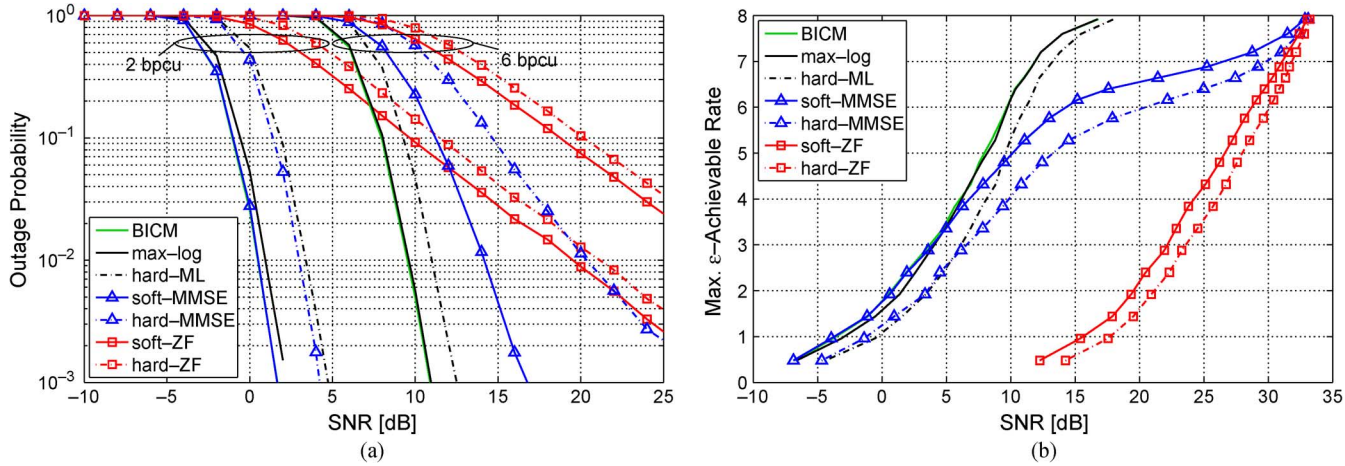


Fig. 11. Demodulator performance in quasi-static fading. (a) Outage probability versus SNR for $R = 2$ bpcu and $R = 6$ bpcu. (b) ϵ-capacity versus SNR for $\epsilon = 10^{-2}$ (4×4 MIMO, 4-QAM, Gray labeling).

MAP (labeled 'BICM' for consistency) and soft MMSE exactly coincide and outperform max-log by about 0.5 dB. In this low-rate regime, max-log performs about 2.5 dB better than hard ML. While max-log, hard ML, and soft MMSE demodulation all achieve full diversity, soft and hard ZF only have diversity order one, resulting in a huge performance loss (respectively 19 dB and 20.5 dB at $P_{\text{out}}(R) = 10^{-2}$). At $R = 6$ bpcu the situation is quite different: here, max-log coincides with soft MAP and hard ML looses only 1.4 dB (again, those three demodulators achieve full diversity). Hard and soft MMSE deteriorate at this rate and loose all diversity. At $P_{\text{out}}(R) = 10^{-2}$, the SNR loss of soft MMSE and soft ZF relative to max-log equals about 4.4 and 19 dB, respectively.

The degradation of soft MMSE with increasing rate is also visible in Fig. 11(b), which shows ϵ-capacity versus SNR for an outage probability of $\epsilon = 10^{-2}$. The ϵ-capacity qualitatively behaves similar as the ergodic capacity (cf. Fig. 2(a)): at low rates soft MMSE outperforms hard ML (by up to 2.8 dB for rates less than 4.7 bpcu) while at high rates it is the opposite way. Furthermore, for low rates soft MMSE essentially coincides with soft MAP whereas at high rates it approaches soft ZF performance.

A similar rate-dependent performance of MMSE demodulation has been observed in [61], [62]. There it was shown that with coding across the antennas, the diversity order of MMSE equalization equals $M_{\text{T}}M_{\text{R}}$ at low rates and $M_{\text{R}} - M_{\text{T}} + 1$ at high rates; in contrast, ZF equalization always achieves a diversity of $M_{\text{R}} - M_{\text{T}} + 1$. These results, interpreted in detail in [62, Section IV], match well our observations that the MMSE demodulator looses diversity for rates larger than 5 bpcu (see Fig. 11 and [31]).

A comparison of Fig. 11(b) and Fig. 2(a) suggests that there is a connection between the diversity of a demodulator in the quasi-static scenario and its SNR loss relative to optimum demodulation in the ergodic scenario. For all rates (SNRs), the max-log and hard ML demodulator both achieve constant (full) diversity in the quasi-static regime and maintain a roughly constant gap to soft MAP in the ergodic scenario. In contrast, with MMSE demodulation the diversity order in the quasi-static case and the SNR gap to soft MAP in the ergodic scenario both deteriorate with increasing rate/SNR.

VIII. Key Observations and Design Guidelines

Based on the previous results, we summarize key observations and provide system design guidelines.

Soft MMSE demodulation approaches BICM capacity and outperforms max-log at low rates, both in the ergodic and the quasi-static regime and for various system configurations (see also [31]). Moreover, soft MMSE is very attractive since it has the lowest complexity among all soft demodulators that we discussed. Therefore, soft MMSE demodulation is arguably the demodulator of choice when designing MIMO-BICM systems with outer codes of low to medium rate. Since soft ZF performs consistently poorer than soft MMSE at the same computational cost, there appears to be no reason to prefer soft ZF in practical implementations. The case for soft MMSE is particularly strong for asymmetric MIMO systems (i.e., $M_{\text{T}} < M_{\text{R}}$), where it performs close to BICM capacity for *all* rates. Fig. 12(a) compares a 4×4 MIMO system using 4-QAM (system I) and a 2×4 system using 16-QAM (system II), both using Gray labeling and achieving $R_0 = 8$. Whereas at low rates soft MMSE demodulation performs better with system I than with system II, it is the other way around for high rates. For example, at 7 bpcu system II requires 1.1 dB less SNR than system I, in spite of using fewer active transmit antennas. This observation is of interest when designing MIMO-BICM systems with adaptive modulation and coding. Specifically, with soft MMSE demodulation system I is preferable below 6 bpcu, whereas above 6 bpcu it is advantageous to deactivate two transmit antennas and switch to the 16-QAM constellation (system II). We note that with max-log and hard ML demodulation, system II performs consistently worse than system I. The only regime where soft MMSE suffers from a noticeable performance loss is symmetric systems at high rates (both, in the ergodic and quasi-static scenario). In the high-rate regime, hard and soft SDR are the only low-complexity demodulation schemes that are able to achieve hard ML and max-log performance, respectively. These

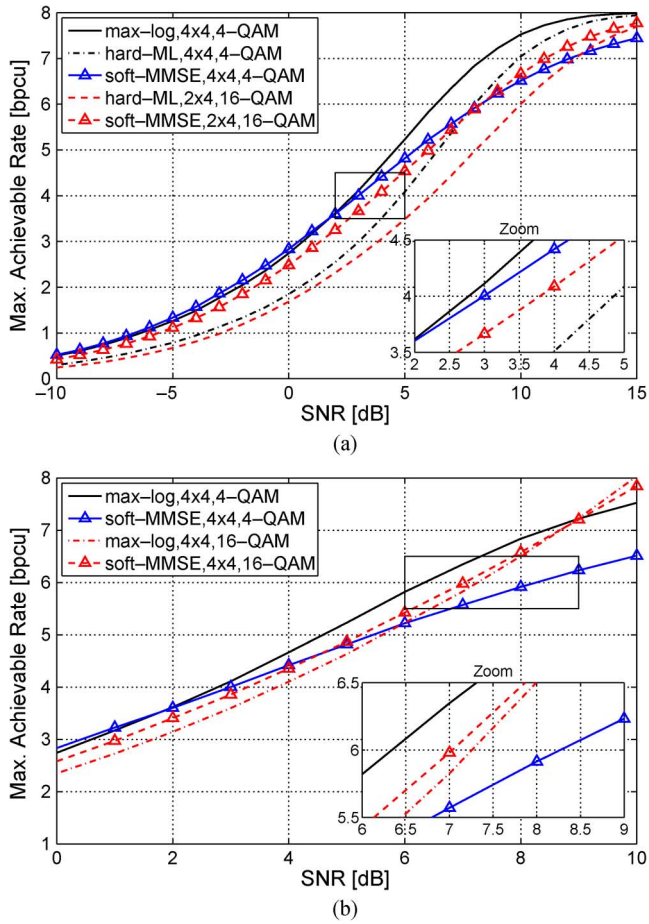


Fig. 12. Mutual information of baseline demodulators for (a) a 4×4 MIMO system using 4-QAM and a 2×4 MIMO system using 16-QAM. (b) A 4×4 MIMO system using 4-QAM and 16-QAM (Gray labeling in all cases).

observations apply also in the case of imperfect CSI. This suggests that hard and soft SDR demodulation are preferable over hard ML and max-log demodulation. With perfect CSI, also LSD, bit-flipping demodulation, and soft ℓ^∞-norm demodulation come reasonably close to max-log. The LSD has the additional advantage of being able to trade off performance for complexity reduction. Furthermore, note that for system I hard SDR demodulation (which approaches hard ML performance) outperforms most suboptimal soft demodulators for rates larger than 6 bpcu.

The above discussion suggests that in order to achieve a given target information rate, it may be preferable to adapt the number of antennas and the symbol constellation in a system with a low-rate code and a low-complexity MMSE demodulator instead of using a high-rate code and a computationally expensive nonlinear demodulator. Such a design approach has recently been advocated also in [62]. While RF complexity may be a limiting factor with respect to antenna number, increasing the constellation size is inexpensive. Fig. 12(b) compares soft MMSE and max-log for a 4×4 MIMO system with Gray-labeled 4-QAM ($R_0 = 8$) and 16-QAM ($R_0 = 16$). Below 3.5 bpcu, soft MMSE demodulation with 4-QAM is optimal; at higher rates, switching to 16-QAM allows the MMSE demodulator to perform within about 0.7 dB of max-log with 4-QAM while increasing the soft-MMSE complexity only little.

In case of imperfect receiver CSI, the performance of all demodulators deteriorates significantly (see also [31]), i.e., all capacity curves are shifted to higher SNRs (depending on the amount of training). Demodulators that take the noise variance into account require somewhat more training. An exception is the LSD which can outperform max-log in case of poor channel and noise variance estimates [31].

We conclude that at low rates linear soft demodulation is generally preferable due to its very low computational complexity. At high rates non-linear demodulators perform better, even when they deliver hard rather than soft outputs. If complexity is not an issue, soft SDR demodulation is advantageous since it approaches max-log performance and is largely superior to all other demodulators over a wide range of system parameters and SNRs. A notable exception is the low-complexity SoftIC demodulator which for some system configurations has the potential to outperform soft SDR (and max-log) at low rates.

IX. Conclusion

We provided a comprehensive performance assessment and comparison of soft and hard demodulators for non-iterative MIMO-BICM systems. Our comparison is based on the information-theoretic notion of mutual information, which quantifies the maximum achievable rate of the equivalent "modulation" channel that comprises modulator, physical channel, and demodulator. As a performance measure, mutual information has the main advantage of being independent of any outer code. Extensive simulation results show that a universal demodulator performance ranking is *not* possible and that the demodulator performance can depend strongly on the rate (or equivalently the SNR) at which a system operates. In addition to ergodic capacity results, we investigated the non-ergodic fading scenario in terms of outage probability and ϵ-capacity and analyzed the robustness of certain demodulators under imperfect channel state information. Our observations provide new insights into the design of MIMO-BICM systems (i.e., choice of demodulator, number of antennas, and symbol constellation). Moreover, our approach sheds light on issues that have not been apparent in the previously prevailing BER performance comparisons for specific outer codes. For example, a key observation is that with low-rate outer codes soft MMSE is preferable over other demodulators since it has low complexity but close-to-optimal performance.

In the paper, we have focused on spatially i.i.d. flat fading channels. Results for spatially correlated MIMO channels (not provided here due to lack of space) show that our general conclusions from the uncorrelated case still apply in the correlated case, i.e., the demodulator performance ranking is rate-dependent and at low rates soft MMSE performs better than hard ML and max-log. Furthermore, our insights also apply to frequency-selective channels with arbitrary power delay profile when the demodulators are used on a per-subcarrier basis in the context of MIMO-OFDM systems. Here, the corresponding ergodic BICM capacity is simply the sum of the ergodic BICM capacities for the individual subcarriers (irrespective of any subcarrier correlation induced by a particular channel delay spread).

APPENDIX A
MEASURING MUTUAL INFORMATION

Evaluating the mutual information in (9) involves the conditional LLR distributions $f(\tilde{\Lambda}_l|c_l)$. We approximate these pdfs by histograms obtained via Monte Carlo simulations. To achieve a small bias and variance of the mutual information estimate, the number and the size of the histogram bins as well as the sample size need to be carefully balanced [48]. Instead of LLRs, we use the bit probabilities

$$\phi_l = \frac{1}{1+e^{-\tilde{\Lambda}_l}} \in [0,1]. \tag{26}$$

Due to the one-to-one correspondence of $\tilde{\Lambda}_l$ and ϕ_l, the mutual information of the equivalent modulation channel equals that of the channel characterized by the conditional pdf $f(\phi_l|c_l)$; the latter is easier to approximate by a histogram with uniform bins. By performing Monte Carlo simulations in which N code bits, the noise, and the channel are randomly generated, we obtain a histogram with K uniform bins $\left[\frac{k-1}{K}, \frac{k}{K}\right]$, $k = 1, \ldots, K$, and the associated conditional relative frequencies $\Xi_{l,k}^b$ (i.e., the normalized number of probabilities ϕ_l lying in the kth bin conditioned on $c_l = b$). The mutual information in (9) is then approximated as

$$C \approx \hat{C} = R_0 - \sum_{l=1}^{R_0}\sum_{b=0}^{1}\sum_{k=1}^{K} \frac{1}{2}\Xi_{l,k}^b \log_2 \frac{\sum_{b'=0}^{1}\Xi_{l,k}^{b'}}{\Xi_{l,k}^b}. \tag{27}$$

The accuracy of this approximation depends on the number K of histogram bins (this determines the discretization error) and on the number N of samples (code bits) used to estimate the histogram (this determines the bias and variance of $\Xi_{l,k}^b$ and $\hat{C}$). The bias and the variance of $\hat{C}$ can be bounded as [48]

$$\begin{aligned} \mathbb{E}\{\hat{C}\} - C_{\mathrm{Q}} &\le \log_2\left(1 + \frac{K-1}{N}\right), \\ \mathbb{E}\left\{\left(\hat{C} - \mathbb{E}\{\hat{C}\}\right)^2\right\} &\le \frac{(\log_2 N)^2}{N}. \end{aligned} \tag{28}$$

Here, C_{Q} is the mutual information of the discretized channel, i.e., equal to (27) but with $\Xi_{l,k}^b$ replaced by $\mathbf{P}\left(\phi_l \in \left[\frac{k-1}{K}, \frac{k}{K}\right] | c_l = b\right)$. Hence, the bias in (28) quantifies the systematic error resulting from the empirical estimation of the histograms. We conclude that a large number K of bins is advantageous in order to keep the discretization error small; in view of (28), this necessitates a significantly larger number N of samples ($N \gg K$) in order to achieve a small estimation bias. Large N simultaneously ensures a small estimation variance. The price paid for accurate capacity estimates is computational complexity.

To design K and N, we first evaluated the BICM capacity in (7) by direct numerical integration using the pdfs in (5); then we estimated the same capacity via Monte Carlo simulations using the optimum soft MAP demodulator and increasingly larger K and N. With $K = 256$ and $N = 10^5$ the difference between direct evaluation and Monte Carlos simulations was on the order of 10^{-4} over a large SNR range. These numbers were then used to estimate the mutual information for all other demodulators.

ACKNOWLEDGMENT

The authors would like to thank G. Lechner for kindly providing his LDPC decoder implementation.

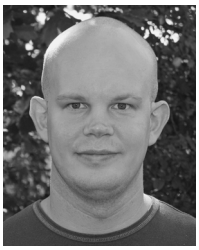
REFERENCES

[1] P. Fertl, J. Jaldén, and G. Matz, "Capacity-based performance comparison of MIMO-BICM demodulators," in *Proc. IEEE SPAWC-2008*, Recife, Brazil, Jul. 2008, pp. 166–170.
[2] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 5, pp. 927–945, May 1998.
[3] A. Guillén i Fàbregas, A. Martinez, and G. Caire, "Bit-interleaved coded modulation," *Found. Trends Commun. Inf. Theory*, vol. 5, no. 1–2, pp. 1–153, 2008.
[4] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1361–1391, July 1999.
[5] E. Biglieri, G. Taricco, and E. Viterbo, "Bit-interleaved time-space codes for fading channels," in *Proc. Conf. Inf. Sci. Syst.*, Princeton, NJ, Mar. 2000, pp. WA 4.1–WA 4.6.
[6] A. Stefanov and T. M. Duman, "Turbo-coded modulation for systems with transmit and receive antenna diversity over block fading channels: System model, decoding approaches, and practical considerations," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 5, pp. 958–968, May 2001.
[7] S. H. Müller-Weinfurtner, "Coding approaches for multiple antenna transmission in fast fading and OFDM," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2442–2450, Oct. 2002.
[8] J. Jaldén and B. Ottersten, "Parallel implementation of a soft output sphere decoder," in *Proc. 39th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Oct./Nov. 2005, pp. 581–585.
[9] J. Jaldén and B. Ottersten, "On the complexity of sphere decoding in digital communications," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1474–1484, Apr. 2005.
[10] C. Studer, A. Burg, and H. Bölcskei, "Soft-output sphere decoding: Algorithms and VLSI implementation," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 2, pp. 290–300, Feb. 2008.
[11] A. Burg, M. Borgmann, M. Wenk, M. Zellweger, W. Fichtner, and H. Bölcskei, "VLSI implementation of MIMO detection using sphere decoding algorithm," *IEEE J. Solid-State Circuits*, vol. 40, no. 7, pp. 1566–1577, Jul. 2005.
[12] D. Seethaler and H. Bölcskei, "Infinity-norm sphere-decoding," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Toronto, Canada, Jul. 2008, pp. 2002–2006.
[13] W. K. Ma, T. N. Davidson, K. M. Wong, Z. Q. Luo, and P. C. Ching, "Quasi-maximum-likelihood multiuser detection using semidefinite relaxation with application to synchronous CDMA," *IEEE Trans. Signal Process.*, vol. 50, no. 4, pp. 912–922, Apr. 2002.
[14] B. Steingrimsson, Z.-Q. Luo, and K. M. Wong, "Soft quasi-maximum-likelihood detection for multiple-antenna wireless channels," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2710–2719, Nov. 2003.
[15] 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, Mar. 2003.
[16] H. Yao and G. W. Wornell, "Lattice-reduction-aided detectors for MIMO communication systems," in *Proc. IEEE GLOBECOM-2002*, Taipei, Taiwan, Nov. 2002, vol. 1, pp. 424–428.
[17] D. Wübben, R. Böhnke, V. Kühn, and K. Kammeyer, "MMSE-based lattice-reduction for near-ML detection of MIMO systems," in *Proc. ITG Workshop on Smart Antennas*, Munich, Germany, Mar. 2004, pp. 106–113.
[18] C. Windpassinger, L. H.-J. Lampe, and R. F. H. Fischer, "From lattice-reduction-aided detection towards maximum-likelihood detection in MIMO systems," in *Proc. Wireless, Optical Commun. Conf.*, Banff, AB, Canada, Jul. 2003.
[19] P. Silvola, K. Hooli, and M. Juntii, "Suboptimal soft-output MAP detector with lattice reduction," *IEEE Signal Process. Lett.*, vol. 13, no. 6, pp. 321–324, Jun. 2006.
[20] V. Ponnampalam, D. McNamara, A. Lillie, and M. Sandell, "On generating soft outputs for lattice-reduction-aided MIMO detection," in *Proc. IEEE ICC-07*, Glasgow, U.K., Jun. 2007, pp. 4144–4149.
[21] R. Wang and G. B. Giannakis, "Approaching MIMO channel capacity with soft detection based on hard sphere decoding," *IEEE Trans. Commun.*, vol. 54, no. 4, pp. 587–590, Apr. 2006.

[22] M. Butler and I. Collings, "A zero-forcing approximate log-likelihood receiver for MIMO bit-interleaved coded modulation," *IEEE Commun. Lett.*, vol. 8, no. 2, pp. 105–107, Feb. 2004.
[23] M. R. McKay and I. B. Collings, "Capacity and performance of MIMO-BICM with zero-forcing receivers," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 74–83, Jan. 2005.
[24] I. B. Collings, M. R. G. Butler, and M. R. McKay, "Low complexity receiver design for MIMO bit-interleaved coded modulation," in *Proc. IEEE ISSTA 2004*, Sydney, Australia, Aug.–Sep. 2004, pp. 12–16.
[25] D. Seethaler, G. Matz, and F. Hlawatsch, "An efficient MMSE-based demodulator for MIMO bit-interleaved coded modulation," in *Proc. IEEE GLOBECOM-2004*, Dallas, TX, Dec. 2004, vol. IV, pp. 2455–2459.
[26] W.-J. Choi, K.-W. Cheong, and J. Cioffi, "Iterative soft interference cancellation for multiple antenna systems," in *Proc. IEEE WCNC-2000*, Chicago, IL, Sep. 2000, vol. 1, pp. 304–309.
[27] A. Paulraj, R. U. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
[28] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. URSI Int. Symp. on Signals, Syst. Electron.*, Pisa, Italy, Sep. 1998, pp. 295–300.
[29] B. Hassibi, "A fast square-root implementation for BLAST," in *Proc. 34th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov./Dec. 2000, pp. 1255–1259.
[30] R. Böhnke, D. Wübben, V. Kühn, and K. D. Kammeyer, "Reduced complexity MMSE detection for BLAST architectures," in *Proc. IEEE GLOBECOM-2003*, San Francisco, CA, Dec. 2003, vol. 4, pp. 2258–2262.
[31] P. Fertl, J. Jaldén, and G. Matz, Performance Assessment of MIMO-BICM Demodulators Based on System Capacity: Further Results Vienna Univ. Technol., Austria, Tech. Rep. #09-1, Mar. 2009 [Online]. Available: http://publik.tuwien.ac.at/files/PubDat_174303.pdf
[32] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–596, Nov. 1999.
[33] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[34] L. H.-J. Lampe, R. Schober, and R. F. H. Fischer, "Multilevel coding for mutliple-antenna transmission," *IEEE Trans. Wireless Commun.*, vol. 3, no. 1, pp. 203–208, Jan. 2004.
[35] Z. Hong and B. L. Hughes, "Robust space-time trellis codes based on bit-interleaved coded modulation," in *Proc. CISS-01*, Mar. 2001, vol. 2, pp. 665–670.
[36] M. van Dijk, A. J. E. M. Janssen, and A. G. C. Koppelaar, "Correcting systematic mismatches in computed log-likelihood ratios," *Eur. Trans. Telecomm.*, vol. 14, no. 3, pp. 227–244, Jul. 2003.
[37] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Boston, MA: Cambridge Univ. Press, 2005.
[38] A. Lapidoth and P. Naryan, "Reliable communication under channel uncertainty," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2148–2177, Oct. 1998.
[39] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai, "On information rates for mismatched decoders," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1953–1967, Nov. 1994.
[40] A. Martinez, A. Guillén i Fàbregas, G. Caire, and F. Willems, "Bit-interleaved coded modulation revisited: A mismatched decoding perspective," in *Proc. ISIT-2008*, Toronta, Canada, Jul. 2008, pp. 2337–2341.
[41] J. Jaldén, P. Fertl, and G. Matz, "On the generalized mutual information of BICM systems with approximate demodulation," in *Proc. IEEE Inf. Theory Workshop*, Cairo, Egypt, Jan. 2010.
[42] A. Burg, M. Wenk, and W. Fichtner, "VLSI implementation of pipelined sphere decoding with early termination," in *Proc. EUSIPCO-2006*, Florence, Italy, Sep. 2006.
[43] S. Schwandter, P. Fertl, C. Novak, and G. Matz, "Log-likelihood ratio clipping in MIMO-BICM systems: Information geometric analysis and impact on system capacity," in *Proc. IEEE ICASSP-2009*, Taipei, Taiwan, Apr. 2009, pp. 2433–2436.
[44] C. Novak, P. Fertl, and G. Matz, "Quantization for soft-output demodulators in bit-interleaved coded modulation systems," in *Proc. IEEE ISIT-2009*, Seoul, Korea, Jun./Jul. 2009, pp. 1070–1074.
[45] C. Studer and H. Bölcskei, "Soft-input soft-output single tree-search sphere decoding," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 4827–4842, Oct. 2010.
[46] T. T. Nguyen and L. Lampe, "Bit-interleaved coded modulation with mismatched decoding metrics," *IEEE Trans. Commun.*, to be published.
[47] 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.
[48] L. Paninski, "Estimation of entropy and mutual information," *Neural Comput.*, vol. 15, no. 6, pp. 1191–1253, Jun. 2003.
[49] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, Jun. 2002.
[50] T. J. Richardson and R. L. Urbanke, "The capacity of low-density parity check codes under message-passing decoding," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 599–618, Feb. 2001.
[51] C. Michalke, E. Zimmermann, and G. Fettweis, "Linear MIMO receivers vs. tree search detection: A performance comparison overview," in *Proc. IEEE PIMRC-06*, Helsinki, Finland, Sep. 2006, pp. 1–7.
[52] M. Damen, H. El Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2389–2402, Oct. 2003.
[53] D. Wübben, D. Seethaler, J. Jaldén, and G. Matz, "Lattice reduction: A survey with applications in wireless communications," *IEEE Signal Process. Mag.*, vol. 28, no. 3, pp. 70–91, May 2011.
[54] A. K. Lenstra, H. W. Lenstra, Jr., and L. Lovsz, "Factoring polynomials with rational coefficients," *Math. Ann.*, vol. 261, no. 4, pp. 515–534, Dec. 1982.
[55] J. Jaldén, D. Seethaler, and G. Matz, "Worst- and average-case complexity of LLL lattice reduction in MIMO wireless systems," in *Proc. IEEE ICASSP-2008*, Las Vegas, NV, Apr. 2008, pp. 2685–2688.
[56] D. Wübben and D. Seethaler, "On the performance of lattice reduction schemes for MIMO data detection," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 2007, pp. 1534–1538.
[57] P. H. Tan and L. K. Rasmussen, "The application of semidefinite programming for detection in CDMA," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 8, pp. 1442–1449, Aug. 2001.
[58] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1999.
[59] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: A study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884–893, Mar. 2006.
[60] G. Taricco and E. Biglieri, "Space-time decoding with imperfect channel estimation," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1874–1888, Jul. 2005.
[61] A. Hedayat and A. Nosratinia, "Outage and diversity of linear receivers in flat-fading MIMO channels," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5868–5873, Dec. 2007.
[62] K. R. Kumar, G. Caire, and A. L. Moustakas, "Asymptotic performance of linear receivers in MIMO fading channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4398–4418, Oct. 2009.

Peter Fertl (M'09) received the Dipl.-Ing. and Dr.techn. degrees in electrical/communication engineering from the Vienna University of Technology, Vienna, Austria, in 2005 and 2009, respectively.

From 2006 to 2009, he was a Research and Teaching Assistant at the Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology, investigating wireless communications with emphasis on MIMO multiuser signal processing and wireless relay networks. He also worked at Nokia Research Center, Helsinki, as a Visiting Researcher from July to September 2007. Since April 2010 he has been a researcher at BMW Group Research and Technology, Munich, working on wireless communications systems for automotive connectivity.



Joakim Jaldén (S'03–M'08) received the M.Sc. and Ph.D. degrees in electrical engineering from the Royal Institute of Technology (KTH), Stockholm, Sweden, in 2002 and 2007, respectively.

From July 2007 to June 2009, he held a Postdoctoral research position at the Vienna University of Technology, Vienna, Austria. He also studied at Stanford University, CA, from September 2000 to May 2002, and worked at ETH, Zürich, Switzerland, as a visiting researcher, from August to September, 2008. In July 2009, he joined the Signal Processing Lab within the School of Electrical Engineering, KTH, as an Assistant Professor.

Dr. Jaldén has been awarded the IEEE Signal Processing Society's 2006 Young Author Best Paper Award for his work on MIMO communications, and the first prize in the Student Paper Contest at the 2007 International Conference on Acoustics, Speech and Signal Processing (ICASSP). He is also a recipient of the Ingvar Carlsson Award issued in 2009 by the Swedish Foundation for Strategic Research.

Gerald Matz (S'95–M'01–SM'07) received the Dipl.-Ing. and Dr. techn. degrees in electrical engineering in 1994 and 2000, respectively, and the Habilitation degree for communication systems in 2004, all from the Vienna University of Technology, Vienna, Austria.

Since 1995, he has been with the Institute of Telecommunications, Vienna University of Technology, where he currently holds a tenured position as Associate Professor. From March 2004 to February 2005, he was on leave as an Erwin Schrödinger Fellow with the Laboratoire des Signaux et Systèmes, Ecole Supérieure d'Electricité, France. He was a Guest Researcher with the Communication Theory Lab, ETH Zurich, Switzerland, in 2007 and with the Ecole Nationale Supérieure d'Electrotechnique, d'Electronique, d'Informatique et d'Hydraulique de Toulouse (ENSEEIHT), France, in 2011. He has directed or actively participated in numerous research projects funded by the Austrian Science Fund (FWF), the Vienna Science and Technology Fund (WWTF), and the European Union. He has published about 150 papers in international journals, conference proceedings, and edited books. He is a coeditor of the book *Wireless Communications over Rapidly Time-Varying Channels* (New York: Academic, 2011). His research interests include wireless communication and sensor networks, statistical signal processing, and information theory.

Prof. Matz serves on the IEEE Signal Processing Society (SPS) Technical Committee on Signal Processing for Communications and Networking and on the IEEE SPS Technical Committee on Signal Processing Theory and Methods. He was an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2007–2010), for the IEEE SIGNAL PROCESSING LETTERS (2004–2008) and for the EURASIP journal SIGNAL PROCESSING (2007–2010). He was a Technical Program Co-Chair of EUSIPCO 2004 and has been on the Technical Program Committee of numerous international conferences. In 2006, he received the Kardinal Innitzer Most Promising Young Investigator Award.