Multiuser Detection of M-QAM Symbols via Bit-Level Equalization and Soft Detection

Darryl Dexu Lin and Teng Joon Lim
The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto
10 King's College Road, Toronto, Ontario, Canada M5S 3G4
{linde, limtj}@comm.utoronto.ca

***Abstract*— Building upon a unified framework for CDMA multiuser detection proposed in our prior work, we investigate the detection of M-QAM symbols in a multiuser CDMA channel. The solution proposed may be summarized as a generic iterative detection scheme for coded interference channels called *Bit-Level Equalization and Soft Detection* (BLESD). It is shown that this novel approach avoids the exponential complexity of *a posteriori probability* (APP) detection by optimizing a closely-related, but much more manageable, objective function called *variational free energy*. It also fundamentally differs from the conventional symbol detector, in that data symbols are transparent to the new detector. Instead, soft estimates of the bits that make up the symbols are directly and naturally obtained at the detector output, in terms of posterior probabilities given the channel observation, facilitating efficient message-passing in joint detection and decoding.**

I. Introduction

The evolution of multiuser detection (MUD) research has seen detectors being derived through many different approaches, such as minimizing the mean-squared error (MMSE), or multistage interference cancelation (IC). Within the past decade, there has been a growing interest in coded CDMA systems, where the need for joint detection and decoding leads to a different class of multiuser detectors, namely turbo multiuser detectors. Practical turbo multiuser detectors proposed in [1] and [2] are among the most celebrated ones, due to their simplicity and remarkable performance.

Recent attempts to provide a unified approach to study the wide range of multiuser detectors include [3], [4] and [5]. [3] generalizes iterative multiuser joint decoding as an approximate sum-product algorithm in a factor graph containing both the multiuser channel and code constraints. Such a generalization leads to elegant performance analysis through *density evolution*. [4] and [5] view the uncoded linear and optimal multiuser detectors as *posterior mean estimators* of the *Bayes retrochannel* such that the bit error rate (BER) may be evaluated through techniques from statistical mechanics. In a recent work of ours [6], a new framework for studying MUD algorithms was proposed, utilizing the concept of *variational free energy minimization* (VFEM) [7]. This framework allows us to rigorously address the design challenge of the MUD component within the iterative multiuser joint decoding problem highlighted in [3]. It also complements [4] and [5] by including non-linear (and iterative) detectors as special cases.

Most important among the contributions of [6], we showed that the detector section of the turbo multiuser detectors proposed in [1] and [2] can both be rigorously justified with the same VFEM routine (as summarized in Section III). Furthermore, because the minimization of variational free energy naturally incorporates the generalized EM algorithm [8] in its formulation, it was shown that a joint data detection and parameter estimation algorithm may be easily formulated as a *variational EM* algorithm.

With the theory developed in [6], this paper will focus on the application of the VFEM framework in MUD, in particular, on generalizing turbo multiuser detectors from BPSK signalling to M-QAM modulation. Though heuristic approaches to turbo MUD with M-QAM modulation can be formulated using traditional concepts such as interference cancelation and MMSE detection, by approximately obtaining symbol APPs (under some convenient assumptions) and then converting these to bit APPs, they are mostly not theoretically justifiable. Our approach yields a novel and low-complexity scheme that is based on the theory of VFEM, that is applicable to all interference channels with square QAM modulation.

II. Signal Model

We consider a flat-fading synchronous DS-CDMA wireless link with K users. After sampling the chip matched filter output at chip rate, the received signal can be written in the well-known vector form:

$$\mathbf{r} = \mathbf{H}\mathbf{d} + \mathbf{n}, \tag{1}$$

In (1), the channel matrix $\mathbf{H} = \mathbf{S}\mathbf{A} = [A_1\mathbf{s}_1, \cdots, A_K\mathbf{s}_K]$, where $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_K]$ is the normalized spreading code matrix, and $\mathbf{A} = \text{diag}([A_1, \cdots, A_k]^T)$ contains channel gains of K active users. In contrast to the conventional assumption of BPSK modulation, we assume the general M-QAM modulated symbols $\mathbf{d} = [d_1, d_2, \cdots, d_K]^T$. Each symbol d_k, $k = 1, \cdots, K$, is a result of Gray mapping of $2L$ information bits $\{b_{l,k}\}_{l=1}^{2L}$. $\mathbf{n}$ is a circularly symmetric complex white Gaussian noise vector with distribution $p(\mathbf{n}) = \mathcal{CN}(\mathbf{0}, 2\sigma^2\mathbf{I})$.

To simplify expressions, in the subsequent derivation we will consider only one-dimensional pulse-amplitude modulated (PAM) symbols $\{d_k\}_{k=1}^{K}$. The analysis for complex QAM modulation follows straightforwardly through a simple transformation that doubles the dimension of the signal model, and will be omitted. Therefore, only real signal and noise vectors will be considered in the rest of the paper, i.e. $p(\mathbf{n}) = \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$.

III. Multiuser Detection via Variational Free Energy Minimization

This section summarizes the key results of [6] to pave the way for the main discussion that will follow.

A. Variational Inference and Free Energy Minimization

It is well-known that the optimal multiuser detector corresponds to Bayesian inference (exact inference) performed on the channel observation. Taking the channel model in (1) for example, the jointly optimal estimate of $\mathbf{d}$ maximizes

$$p(\mathbf{d}|\mathbf{r}) = \frac{p(\mathbf{r}|\mathbf{d})p(\mathbf{d})}{\sum_{\mathbf{d}} p(\mathbf{r}|\mathbf{d})p(\mathbf{d})}, \tag{2}$$

where $p(\mathbf{d})$ is the prior distribution of $\mathbf{d}$, and $p(\mathbf{r}|\mathbf{d})$ is determined by the channel model. In joint detection and decoding of coded CDMA using turbo message passing, $p(\mathbf{d})$ may be in the form of extrinsic information from the BCJR decoder, and $p(\mathbf{r}|\mathbf{d}) = \mathcal{N}(\mathbf{Hd}, \sigma^2\mathbf{I})$. Unfortunately, exact inference entails exponential complexity due to the summation over $\mathbf{d}$ in the denominator of (2).

Variational inference offers a low-complexity but sub-optimal alternative. It simplifies the solution by postulating a distribution $Q(\mathbf{d})$ that resembles $p(\mathbf{d}|\mathbf{r})$ but with a more convenient form. Our goal then becomes minimizing the Kullback-Leibler (KL) divergence between $Q(\mathbf{d})$ and $p(\mathbf{d}|\mathbf{r})$, also known as the *variational free energy*, up to an additive constant:

$$\mathcal{F}(\boldsymbol{\lambda}) = \int_{\mathbf{d}} Q(\mathbf{d}) \log \frac{Q(\mathbf{d})}{p(\mathbf{r}|\mathbf{d})p(\mathbf{d})} d\mathbf{d}, \tag{3}$$

where $\boldsymbol{\lambda}$ contains parameters that specify $Q(\mathbf{d})$. The following is the general procedure for VFEM-based MUD:

1) Postulate distributions for $p(\mathbf{d})$, $p(\mathbf{r}|\mathbf{d})$ and $Q(\mathbf{d})$;
2) Evaluate closed-form expression for $\mathcal{F}(\boldsymbol{\lambda})$;
3) Minimize $\mathcal{F}(\boldsymbol{\lambda})$ (exactly or iteratively) over $\boldsymbol{\lambda}$.

B. Relation to Existing Turbo Multiuser Detectors

The procedure above bears close resemblance to the routine of deriving thermodynamic state equations in statistical mechanics, with the second step being the most challenging. Well-selected (and reasonable) postulates for $p(\mathbf{d})$, $p(\mathbf{r}|\mathbf{d})$ and $Q(\mathbf{d})$ may lead to simplifications. [1] and [2] are examples of good selections, although they were originally not viewed from this perspective. Here we provide a brief summary of [1] and [2] in the context of variational inference. Details pertaining to the derivation may be found in [6]. Since BPSK modulation is assumed, the binary signal vector $\mathbf{b} = [b_1, \cdots, b_K]^T$ replaces $\mathbf{d}$ in the postulated distributions.

- *Discrete SISO Multiuser Detector* [1]:

$$\begin{cases} p(\mathbf{b}) &= \prod_{k=1}^K \xi_k^{\frac{1+b_k}{2}}(1-\xi_k)^{\frac{1-b_k}{2}}, \quad b_k \in \{\pm 1\} \\ p(\mathbf{r}|\mathbf{b}) &= \mathcal{N}(\mathbf{Hb}, \sigma^2\mathbf{I}) \\ Q(\mathbf{b}) &= \prod_{k=1}^K \gamma_k^{\frac{1+b_k}{2}}(1-\gamma_k)^{\frac{1-b_k}{2}}, \quad b_k \in \{\pm 1\} \end{cases} \tag{4}$$

  ξ_k and γ_k are the prior and posterior probabilities of b_k being 1. The advantage of this approach is that the binary nature of b_k is retained, but an approximation on the independence of $\{b_k\}_{k=1}^K$ conditioned on $\mathbf{r}$ (*mean-field approximation*) has to be made. The parameters in the free energy expression, $\boldsymbol{\lambda}$, correspond to $\{\gamma_k\}_{k=1}^K$.

- *Gaussian SISO Multiuser Detector* [2]:

$$\begin{cases} p(\mathbf{b}) &= \mathcal{N}(\tilde{\mathbf{b}}, \mathbf{W}), \quad b_k \in \mathbb{R} \\ p(\mathbf{r}|\mathbf{b}) &= \mathcal{N}(\mathbf{Hb}, \sigma^2\mathbf{I}) \\ Q(\mathbf{b}) &= \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad b_k \in \mathbb{R} \end{cases} \tag{5}$$

  $\tilde{\mathbf{b}} = [\tilde{b}_1, \cdots, \tilde{b}_K]^T$ are the soft bit estimates from the BCJR decoder, and $\mathbf{W} = \mathrm{diag}([1-\tilde{b}_1^2, \cdots, 1-\tilde{b}_K^2]^T)$, where $\mathrm{diag}(\mathbf{x})$ is a diagonal matrix with the vector $\mathbf{x}$ on its diagonal. The advantage of this approach is that it resembles the familiar MMSE detector, but allows soft prior information. The assumption $b_k \in \mathbb{R}$ requires an additional step of converting a continuous distribution $Q(\mathbf{b})$ into a discrete distribution over $\{-1, +1\}^K$, in contrast to the discrete SISO MUD in which $Q(\mathbf{b})$ already has the right sample space. The free energy parameters in this formulation are $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

These two multiuser detectors are used in the MUD section of a turbo MUD receiver, and both may be designed through the variational inference routine outlined in Section III-A. It is worth noting that uncoded MUD can also be derived similarly, but with uniform prior distributions, rather than biased ones.

IV. Bit-Level Equalization and Soft Detection

Heuristics for uncoded MUD schemes with higher order modulation may be derived with ease. For instance, MMSE multiuser detection for QAM symbols only requires simple modifications to the slicer after MMSE filtering. However, in a coded system, i.e. when the bit/symbol priors are non-uniform, a rigorous extension from BPSK to other modulation schemes is complex, partially due to the unpleasant non-linear bit-to-symbol mapping. In [9] and [10], the Gaussian SISO multiuser detector [2] is extended to M-PSK modulation with some loss of optimality, but good simulation results are reported. The same technique does not apply to general QAM modulations, however, due to the unequal symbol energy.

In this paper, we will tackle the problems with M-QAM. The approach to be taken applies to both discrete SISO MUD as well as Gaussian SISO MUD, although we only present the discrete SISO MUD due to space limitations. In contrast to [9] and [10], this extension follows strictly within the unifying framework of VFEM, and may in general be referred to as *Bit-Level Equalization and Soft Detection* (BLESD).

This section is organized as follows: Section IV-A proposes a mapping from $\mathbf{b}$ to $\mathbf{d}$ for Gray-mapped PAM symbols, facilitating expressing $p(\mathbf{r}|\mathbf{d})$ as $p(\mathbf{r}|\mathbf{b}_1, \cdots, \mathbf{b}_L)$ in a manageable closed form; Section IV-B evaluates $\mathcal{F}$ given the postulated prior, channel conditional and posterior distributions; Section IV-C demonstrates how a practical MUD algorithm is drawn from the free energy expression; Section IV-D outlines a further extension, also within the free energy minimization framework, on iteratively estimating noise variance jointly with data detection using the variational EM algorithm.

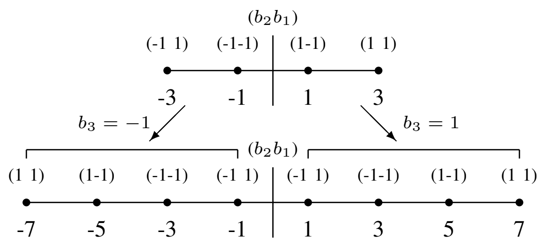

Fig. 1. The transformation from 4-PAM to 8-PAM Gray mapping.

A. Gray Mapping and Multi-Linear Transformation

Lemma 1: A mapping of L bits to a 2^L-PAM constellation point d_L following the equation

$$d_L = \sum_{l=1}^{L} 2^{l-1} b_L b_{L-1} \cdots b_l, \tag{6}$$

where $b_l \in \{-1, +1\}$, results in a Gray mapping strategy.

Proof: We now provide a proof by induction. Note that the $\{b_l\}_{l=1}^{L} \rightarrow d_L$ mapping formula may be written in a recursive form as

$$d_L = \begin{cases} b_L & L = 1 \\ b_L(2^{L-1} + d_{L-1}) & L > 1 \end{cases} \tag{7}$$

We need to show that given $\{b_l\}_{l=1}^{L-1} \rightarrow d_{L-1}$ is Gray mapping, $\{b_l\}_{l=1}^{L} \rightarrow d_L$ also preserves the Gray property. It can be seen from Fig. 1 (for $L = 3$) that the operation $b_L(2^{L-1} + d_{L-1})$ creates two mirror images of d_{L-1} on either side of the origin, depending on the value of b_L. But since the additional bit b_L takes on the same value (1 or -1) in either quadrant, the Gray property is preserved for adjacent constellation points in the same quadrant. Now consider the two adjacent constellation points that belong to different quadrants. Because they represent the same bits $\{b_l\}_{l=1}^{L-1}$ except for b_L, the Gray property also holds for them. Therefore, it is shown that the transformation from d_{L-1} to d_L by adding b_L preserves the Gray property. Hence the mapping construction governed by (6) is a Gray mapping scheme. ∎

The proposed Gray mapping construction is not unique. It can be shown that, in (7), if we were to change the sign before the term 2^{L-1} or d_{L-1}, it would remain a Gray mapping. Without loss of generality, we shall maintain the positive signs in the equation. (6) is a nonlinear function of $b_1, \cdots, b_L$, but is linear w.r.t. each variable individually. Thus it is called a *multi-linear* function [11], which is well-suited to variational inference.

B. Free Energy Evaluation for 2^L-PAM Modulation

Similar to (4), we make the following postulates for the 2^L-PAM signals:

1) Prior Distribution: The prior distribution $p(\mathbf{d}) = \prod_{l=1}^{L} p(\mathbf{b}_l)$ represents the extrinsic information that comes from the BCJR decoder about the distribution of the channel bits. In the traditional multiuser detection viewpoint, this information may be used for interference cancelation in the detection stage. We will not explicitly do so, but as the subsequent derivation shows, the interference cancelation operation is naturally realized within VFEM. We may thus write

$$\begin{aligned} p(\mathbf{d}) &= \prod_{l=1}^{L} p(\mathbf{b}_l) \\ &= \prod_{l=1}^{L} \prod_{k=1}^{K} \xi_{l,k}^{\frac{1+b_{l,k}}{2}} (1-\xi_{l,k})^{\frac{1-b_{l,k}}{2}} \\ &= \prod_{l=1}^{L} \prod_{k=1}^{K} \left(\tfrac{1+\tilde{b}_{l,k}}{2}\right)^{\frac{1+b_{l,k}}{2}} \left(\tfrac{1-\tilde{b}_{l,k}}{2}\right)^{\frac{1-b_{l,k}}{2}}, \end{aligned} \tag{8}$$

where $\xi_{l,k}$ is the prior probability of the l-th bit of user k's symbol being 1. A change of variable is made in the second equality, such that $\tilde{b}_{l,k}$ represents the mean estimate of $b_{l,k}$, i.e. $\tilde{b}_{l,k} = 1 \cdot \xi_{l,k} + (-1) \cdot (1 - \xi_{l,k}) = 2\xi_{l,k} - 1$.

2) Channel Conditional Distribution: The channel conditional distribution $p(\mathbf{r}|\mathbf{d}) = \mathcal{N}(\mathbf{Hd}, \sigma^2\mathbf{I})$ is assumed to be Gaussian with noise variance σ^2 as Section II indicates. The multi-linear bit-to-symbol mapping developed in Section IV-A ensures that the conditional distribution may be written in terms of the channel bits. Realizing $\mathbf{d} = \sum_{l=1}^{L} 2^{l-1} \coprod_{p=l}^{L} \mathbf{b}_p$, we may write the channel conditional distribution as

$$\begin{aligned} p(\mathbf{r}|\mathbf{d}) &= p(\mathbf{r}|\mathbf{b}_1, \cdots, \mathbf{b}_L) \\ &= \mathcal{N}\left(\mathbf{H} \cdot \sum_{l=1}^{L} 2^{l-1} \coprod_{p=l}^{L} \mathbf{b}_p, \sigma^2\mathbf{I}\right), \end{aligned} \tag{9}$$

where the notation $\coprod$ represents the Schur product (element-wise product) between vectors: $\coprod_{p=l}^{L} \mathbf{b}_p = \mathbf{b}_l \circ \mathbf{b}_{l+1} \circ \cdots \mathbf{b}_L$. In other words, we place the l-th bit of all users in one vector $\mathbf{b}_l$, and perform multiuser detection not only among K users, but also among L bits in each user.

3) Posterior Distribution: The exact evaluation of the posterior bit probability $p(\mathbf{b}_1, \cdots, \mathbf{b}_L|\mathbf{r})$ corresponds to the jointly-optimal (JO) multiuser detector, but in general leads to exponential complexity (in K), even for BPSK signals. Here we make a *mean-field approximation* similar to its BPSK counterpart in (4), where the postulated posterior probability $Q(b_{l,k})$ is assumed to be independent over both l and k. This assumption is essential in reducing the computational complexity of the BLESD algorithm. In particular,

$$\begin{aligned} Q(\mathbf{d}) &= \prod_{l=1}^{L} Q(\mathbf{b}_l) \\ &= \prod_{l=1}^{L} \prod_{k=1}^{K} \gamma_{l,k}^{\frac{1+b_{l,k}}{2}} (1-\gamma_{l,k})^{\frac{1-b_{l,k}}{2}} \\ &= \prod_{l=1}^{L} \prod_{k=1}^{K} \left(\tfrac{1+m_{l,k}}{2}\right)^{\frac{1+b_{l,k}}{2}} \left(\tfrac{1-m_{l,k}}{2}\right)^{\frac{1-b_{l,k}}{2}}, \end{aligned} \tag{10}$$

where $\gamma_{l,k}$ is the posterior probability of $b_{l,k}$ being 1. A change of variable is also made here, such that $m_{l,k}$ represents the mean estimate of $b_{l,k}$. The formulations in (8), (9) and (10) enable us to perform inference on the bit level through $Q(b_{l,k})$, which has only one parameter $m_{l,k}$. On the other hand, inference on the symbol level requires updates of $Q(d_k)$, which has $L-1$ parameters, making the optimization of the resulting free energy difficult.

The variational free energy expression may be separated into three terms:

$$\begin{aligned} \mathcal{F} &= \int_{\mathbf{d}} Q(\mathbf{d}) \log \tfrac{Q(\mathbf{d})}{p(\mathbf{r}|\mathbf{d})p(\mathbf{d})} d\mathbf{d} \\ &= \int_{\mathbf{d}} Q(\mathbf{d}) \log Q(\mathbf{d}) d\mathbf{d} - \int_{\mathbf{d}} Q(\mathbf{d}) \log p(\mathbf{r}|\mathbf{d}) d\mathbf{d} \\ &\quad - \int_{\mathbf{d}} Q(\mathbf{d}) \log p(\mathbf{d}) d\mathbf{d}. \end{aligned} \tag{11}$$

$$\begin{aligned}
&\int_{\mathbf{d}} Q(\mathbf{d}) \log p(\mathbf{r}|\mathbf{d}) d\mathbf{d} \\
&= -\tfrac{1}{2\sigma^2}\left\{E\left[(\textstyle\sum_{l=1}^{L} 2^{l-1} \prod_{p=l}^{L} \mathbf{b}_p)^T \mathbf{H}^T\mathbf{H}(\sum_{l=1}^{L} 2^{l-1} \prod_{p=l}^{L} \mathbf{b}_p)\right] - E\left[2\mathbf{r}^T\mathbf{H}(\sum_{l=1}^{L} 2^{l-1} \prod_{p=l}^{L} \mathbf{b}_p)\right]\right\} \\
&= -\tfrac{1}{2\sigma^2}\left\{(\textstyle\sum_{l=1}^{L} 2^{l-1} \prod_{p=l}^{L} \mathbf{m}_p)^T[\mathbf{H}^T\mathbf{H} - \mathrm{diag}(\mathbf{H}^T\mathbf{H})](\sum_{l=1}^{L} 2^{l-1} \prod_{p=l}^{L} \mathbf{m}_p) + \mathbf{1}^T\mathrm{diag}(\mathbf{H}^T\mathbf{H})(\sum_{0<i\le j<L} 2^{i+j} \prod_{p=i}^{j} \mathbf{m}_p)\right\} \\
&\quad + \tfrac{1}{2\sigma^2}\left\{2\mathbf{r}^T\mathbf{H}(\textstyle\sum_{l=1}^{L} 2^{l-1} \prod_{p=l}^{L} \mathbf{m}_p)\right\}
\end{aligned} \tag{16}$$

$$\begin{aligned}
&\mathcal{F}(\mathbf{m}_1, \cdots, \mathbf{m}_L) \\
&= \textstyle\sum_{l=1}^{L}\sum_{k=1}^{K} \frac{1+m_{l,k}}{2}\log\frac{1+\tilde{b}_{l,k}}{1+m_{l,k}} + \frac{1-m_{l,k}}{2}\log\frac{1-\tilde{b}_{l,k}}{1-m_{l,k}} - \frac{1}{2\sigma^2}\left\{2\mathbf{r}^T\mathbf{H}(\sum_{l=1}^{L} 2^{l-1} \prod_{p=l}^{L} \mathbf{m}_p)\right\} \\
&\quad + \tfrac{1}{2\sigma^2}\left\{(\textstyle\sum_{l=1}^{L} 2^{l-1} \prod_{p=l}^{L} \mathbf{m}_p)^T[\mathbf{H}^T\mathbf{H} - \mathrm{diag}(\mathbf{H}^T\mathbf{H})](\sum_{l=1}^{L} 2^{l-1} \prod_{p=l}^{L} \mathbf{m}_p) + \mathbf{1}^T\mathrm{diag}(\mathbf{H}^T\mathbf{H})(\sum_{0<i\le j<L} 2^{i+j} \prod_{p=i}^{j} \mathbf{m}_p)\right\}
\end{aligned} \tag{17}$$

$$\begin{bmatrix} \log\frac{1+m_{l,1}}{1-m_{l,1}} \\ \vdots \\ \log\frac{1+m_{l,K}}{1-m_{l,K}} \end{bmatrix} = \begin{bmatrix} \log\frac{1+\tilde{b}_{l,1}}{1-\tilde{b}_{l,1}} \\ \vdots \\ \log\frac{1+\tilde{b}_{l,K}}{1-\tilde{b}_{l,K}} \end{bmatrix} + \frac{1}{\sigma^2}\left\{2\boldsymbol{\Delta}_l^T\mathbf{H}^T\mathbf{r} - \mathbf{E}_l\mathrm{diag}(\mathbf{H}^T\mathbf{H})\mathbf{1} - 2\boldsymbol{\Delta}_l^T[\mathbf{H}^T\mathbf{H} - \mathrm{diag}(\mathbf{H}^T\mathbf{H})](\boldsymbol{\Delta}_l\mathbf{m}_l + \boldsymbol{\zeta}_l)\right\} \tag{18}$$

To evaluate $\mathcal{F}$, we require two matrix identities related to the Schur product, as summarized in Lemmas 2 and 3.

Lemma 2: Consider independent binary random vectors $\mathbf{b}_1, \cdots, \mathbf{b}_U \in \{\pm 1\}^{K\times 1}$ with independently distributed elements. Let the means of these vectors be $\{\mathbf{m}_l\}_{l=1}^{U}$, $-\mathbf{1} \preceq \mathbf{m}_l \preceq \mathbf{1}$. Then, for a symmetric matrix $\mathbf{C}$,

$$\begin{aligned}
&E\left[(\textstyle\prod_{l=1}^{U}\mathbf{b}_l)^T\mathbf{C}(\prod_{l=1}^{U}\mathbf{b}_l)\right] = \\
&(\textstyle\prod_{l=1}^{U}\mathbf{m}_l)^T\left[\mathbf{C} - \mathrm{diag}(\mathbf{C})\right](\prod_{l=1}^{U}\mathbf{m}_l) + \mathbf{1}^T\mathrm{diag}(\mathbf{C})\mathbf{1}.
\end{aligned} \tag{12}$$

Proof: See Appendix I. ■

Lemma 3: Consider independent binary random vectors $\mathbf{b}_1, \cdots, \mathbf{b}_V \in \{\pm 1\}^{K\times 1}$ with independently distributed elements. Let the means of these vectors be $\{\mathbf{m}_l\}_{l=1}^{V}$, $-\mathbf{1} \preceq \mathbf{m}_l \preceq \mathbf{1}$. Then, for a symmetric matrix $\mathbf{C}$ and $1 \le U < V$,

$$\begin{aligned}
&E\left[(\textstyle\prod_{l=1}^{U}\mathbf{b}_l)^T\mathbf{C}\prod_{l=1}^{V}\mathbf{b}_l\right] = \\
&(\textstyle\prod_{l=1}^{U}\mathbf{m}_l)^T\left[\mathbf{C} - \mathrm{diag}(\mathbf{C})\right]\prod_{l=1}^{V}\mathbf{m}_l + \mathbf{1}^T\mathrm{diag}(\mathbf{C})\prod_{l=U+1}^{V}\mathbf{m}_l.
\end{aligned} \tag{13}$$

Proof: The proof follows from applying Lemma 2, but its details will be omitted due to lack of space. ■

The three integrals in (11) can now be evaluated in closed form. The derivations are shown in equations (14), (15) and (16):

$$\begin{aligned}
&\int_{\mathbf{d}} Q(\mathbf{d})\log Q(\mathbf{d})d\mathbf{d} \\
&= \textstyle\sum_{l=1}^{L}\int_{\mathbf{b}_l} Q(\mathbf{b}_l)\log Q(\mathbf{b}_l)d\mathbf{b}_l \\
&= \textstyle\sum_{l=1}^{L}\sum_{k=1}^{K}\frac{1+m_{l,k}}{2}\log\frac{1+m_{l,k}}{2} + \frac{1-m_{l,k}}{2}\log\frac{1-m_{l,k}}{2}
\end{aligned} \tag{14}$$

$$\begin{aligned}
&\int_{\mathbf{d}} Q(\mathbf{d})\log p(\mathbf{d})d\mathbf{d} \\
&= \textstyle\sum_{l=1}^{L}\int_{\mathbf{b}_l} Q(\mathbf{b}_l)\log p(\mathbf{b}_l)d\mathbf{b}_l \\
&= \textstyle\sum_{l=1}^{L}\sum_{k=1}^{K}\frac{1+m_{l,k}}{2}\log\frac{1+\tilde{b}_{l,k}}{2} + \frac{1-m_{l,k}}{2}\log\frac{1-\tilde{b}_{l,k}}{2}
\end{aligned} \tag{15}$$

The complete free energy expression is assembled in (17).

C. BLESD via Free Energy Minimization

Taking the derivative of $\mathcal{F}(\mathbf{m}_1, \cdots, \mathbf{m}_L)$ w.r.t. $\mathbf{m}_l, 1 \le l \le L$, and equating to zero yields an equation of the form of (18). In (18), $\boldsymbol{\Delta}_l = \mathrm{diag}(\boldsymbol{\delta}_l)$, $\mathbf{E}_l = \mathrm{diag}(\boldsymbol{\epsilon}_l)$ and

$$\begin{cases}
\boldsymbol{\delta}_l = \sum_{i=1}^{l} 2^{i-1}\prod_{n=i,n\ne l}^{L}\mathbf{m}_n + I(l=L)\cdot 2^{l-1}\mathbf{1} \\
\boldsymbol{\epsilon}_l = \sum_{0<i\le l\le j<L, i\ne j}\prod_{n=i,n\ne l}^{j}\mathbf{m}_n + I(0<l<L)\cdot 2^{2l}\mathbf{1} \\
\boldsymbol{\zeta}_l = \sum_{i=l+1}^{L} 2^{i-1}\prod_{n=i}^{L}\mathbf{m}_n
\end{cases} \tag{19}$$

where $I(A)$ is an indicator function which equals 1 if A is true and 0 otherwise. Similar to the BPSK case, $\mathcal{F}(\mathbf{m}_1, \cdots, \mathbf{m}_L)$ cannot be minimized over $\{\mathbf{m}_l\}_{l=1}^{L}$ in one step, but iterative schemes, such as the coordinate descent method, are available to decrease the free energy iteratively.

From (18) it is seen that setting $\partial\mathcal{F}/\partial m_{l,k} = 0$ leads to

$$\begin{aligned}
\log\tfrac{1+m_{l,k}}{1-m_{l,k}} \;=\; &\log\tfrac{1+\tilde{b}_{l,k}}{1-\tilde{b}_{l,k}} + \tfrac{1}{\sigma^2}\left\{2\delta_{l,k}\cdot\mathbf{h}_k^T\mathbf{r} - \epsilon_{l,k}\cdot\boldsymbol{\rho}_k^T\mathbf{1}\right. \\
&\left.-2\delta_{l,k}\cdot\boldsymbol{\beta}_k^T(\boldsymbol{\delta}_l\circ\mathbf{m}_l + \boldsymbol{\zeta}_l)\right\},
\end{aligned} \tag{20}$$

where $\mathbf{h}_k$, $\boldsymbol{\rho}_k$, and $\boldsymbol{\beta}_k$ are the kth column of $\mathbf{H}$, $\mathrm{diag}(\mathbf{H}^T\mathbf{H})$, and $\mathbf{H}^T\mathbf{H} - \mathrm{diag}(\mathbf{H}^T\mathbf{H})$, respectively. Noticing that the right hand side of the equation is independent of $m_{l,k}$, a closed-form update for $m_{l,k}$ may be found by letting $\tilde{b}_{l,k} = 0$:

$$\begin{aligned}
m_{l,k} \;\leftarrow\; &\tanh\left\{\tfrac{1}{2\sigma^2}[2\delta_{l,k}\cdot\mathbf{h}_k^T\mathbf{r} - \epsilon_{l,k}\cdot\boldsymbol{\rho}_k^T\mathbf{1}\right. \\
&\left.-2\delta_{l,k}\cdot\boldsymbol{\beta}_k^T(\boldsymbol{\delta}_l\circ\mathbf{m}_l + \boldsymbol{\zeta}_l)]\right\}.
\end{aligned} \tag{21}$$

The variational free energy is then iteratively decreased by updating $m_{l,k}$ for each l and k. We call this the inner iterations (iterations within the detector section), as opposed to the outer (turbo) iterations that alternates between the detector section and decoder section. In each inner iteration (indexed by i), serial or parallel updates of $m_{l,k}$ is possible. It is easily verified that a parallel update scheme with $L = 1$ converges to the original discrete SISO MUD for BPSK [1].

D. Variational EM and Joint Parameter Estimation

As described in [6], the variational EM algorithm for parameter estimation may be incorporated in VFEM with the unknown system parameter θ included as an additional unknown in $\mathcal{F}$, i.e. we write the free energy as $\mathcal{F}(\theta, \boldsymbol{\lambda})$. The E step then solves for $\hat{\boldsymbol{\lambda}} = \arg\min_{\boldsymbol{\lambda}}\mathcal{F}(\hat{\theta}, \boldsymbol{\lambda})$, while the M step solves for $\hat{\theta} = \arg\min_{\theta}\mathcal{F}(\theta, \hat{\boldsymbol{\lambda}})$. Suppose the unknown parameter θ is the noise variance σ^2. We then obtain an iterative noise variance estimation algorithm embedded within the turbo MUD algorithm. This is analogous to the BPSK case presented in [6], and its details will be omitted here.

V. Simulations

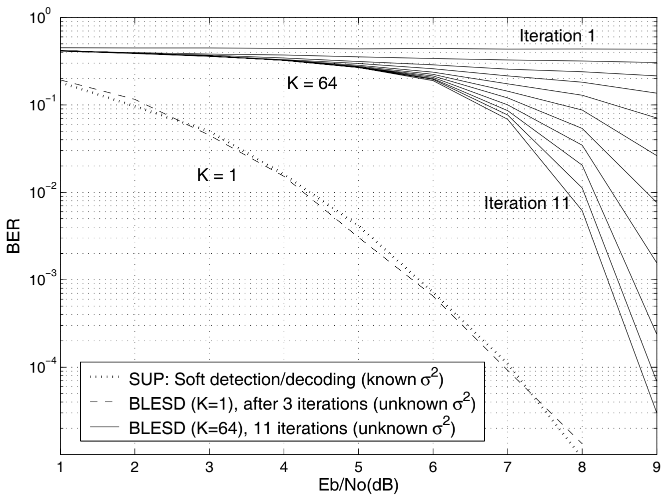

Fig. 2. BER vs. SNR performance of discrete SISO BLESD algorithm for 4-PAM modulation in turbo MUD.

In this section we briefly study the performance of the proposed MUD algorithm with iterative noise variance estimation. We assume a 4-PAM ($L = 2$) random spreading system of spreading gain $N = 64$ with $K = 64$ users. All users have equal power and employ the same rate $1/2$ convolutional code with generators 10011 and 11101. The variational EM algorithm is deployed to iteratively update the estimate for σ^2 in each outer iteration. In Fig. 2, the BER performance is plotted for different numbers of outer iteration, to show the convergence of the algorithm. We use a single inner iteration for each outer iteration, in which the parallel update of $\{m_{2,k}\}_{k=1}^K$ is followed by the parallel update of $\{m_{1,k}\}_{k=1}^K$. More specifically, the update equations are:

$$\begin{aligned} m_{2,k} \quad \leftarrow \quad & \tanh\{\tfrac{1}{\sigma^2}[(m_{1,k}+2)\cdot \mathbf{h}_k^T\mathbf{r} \\ & -(m_{1,k}+2)\cdot \boldsymbol{\beta}_k^T(\mathbf{m}_2 \circ \mathbf{m}_1 + 2\mathbf{m}_2)]\}; \end{aligned} \tag{22}$$

$$\begin{aligned} m_{1,k} \quad \leftarrow \quad & \tanh\{\tfrac{1}{\sigma^2}[m_{2,k}\cdot \mathbf{h}_k^T\mathbf{r} - 2\boldsymbol{\rho}_k^T\mathbf{1} \\ & -m_{2,k}\cdot \boldsymbol{\beta}_k^T(\mathbf{m}_2 \circ \mathbf{m}_1 + 2\mathbf{m}_2)]\}. \end{aligned} \tag{23}$$

The single user performance (SUP) refers to estimating the symbol probabilities based on the channel observation, converting them to bit probabilities, and applying BCJR decoding with soft bit priors. This obvious approach cannot be extended to the multiuser scenario, since the complexity of estimating symbol probabilities is exponential in K. We demonstrate that, with the help of BLESD, a 4-PAM system with loading-factor 1 is able to perform close to SUP at high SNR, even with unknown noise variance. Note that implementing BLESD for $K = 1$ results in a BER curve almost identical to the SUP curve after 3 iterations, revealing that the BLESD approach also has applications in coded AWGN channels, as it performs close to the conventional soft detection/decoding scheme, while not requiring the noise variance information.

VI. Conclusions

In this paper, we proposed a multi-linear transformation to formulate the non-linear bit-to-symbol mapping of M-QAM, enabling the application of variational inference to M-QAM turbo MUD. The generalization of turbo multiuser detectors to M-QAM modulation through the BLESD algorithm has implications beyond the scope of multiuser CDMA, since this analysis can readily be carried over to a wide range of scenarios, such as MIMO or multipath fading channels.

Appendix I
Proof of Lemma 2 by Induction

It is easily verified that for $U = 1$,

$$E\left[\mathbf{b}_1^T\mathbf{C}\mathbf{b}_1\right] = \mathbf{m}_1^T[\mathbf{C} - \mathrm{diag}(\mathbf{C})]\mathbf{m}_1 + \mathbf{1}^T\mathrm{diag}(\mathbf{C})\mathbf{1}. \tag{24}$$

Assuming (12) is true for $U = u$, i.e.

$$\begin{aligned} & E\left\{(\textstyle\prod_{l=1}^u \mathbf{b}_l)^T\mathbf{C}(\prod_{l=1}^u \mathbf{b}_l)\right\} \\ = \quad & (\textstyle\prod_{l=1}^u \mathbf{m}_l)^T\left[\mathbf{C} - \mathrm{diag}(\mathbf{C})\right](\prod_{l=1}^u \mathbf{m}_l) + \mathbf{1}^T\mathrm{diag}(\mathbf{C})\mathbf{1}, \end{aligned} \tag{25}$$

we need to verify it for $U = u + 1$. We have

$$\begin{aligned} & E\left\{(\textstyle\prod_{l=1}^{u+1} \mathbf{b}_l)^T\mathbf{C}(\prod_{l=1}^{u+1} \mathbf{b}_l)\right\} \\ = \quad & E\left\{(\textstyle\prod_{l=1}^u \mathbf{m}_l)^T\left[\mathbf{B}_{u+1}\mathbf{C}\mathbf{B}_{u+1} - \mathrm{diag}(\mathbf{B}_{u+1}\mathbf{C}\mathbf{B}_{u+1})\right]\right. \\ & \left.\cdot(\textstyle\prod_{l=1}^u \mathbf{m}_l) + \mathbf{1}^T\mathrm{diag}(\mathbf{B}_{u+1}\mathbf{C}\mathbf{B}_{u+1})\mathbf{1}\right\} \\ = \quad & (\textstyle\prod_{l=1}^{u+1} \mathbf{m}_l)^T\left[\mathbf{C} - \mathrm{diag}(\mathbf{C})\right](\prod_{l=1}^{u+1} \mathbf{m}_l) + \mathbf{1}^T\mathrm{diag}(\mathbf{C})\mathbf{1}, \end{aligned} \tag{26}$$

where $\mathbf{B}_l = \mathrm{diag}(\mathbf{b}_l)$ and $\mathbf{M}_l = \mathrm{diag}(\mathbf{m}_l)$.

References

[1] P. D. Alexander, A. J. Grant, and M. C. Reed, "Iterative detection in code-division multiple-access with error control coding," *Euro. Trans. Telecommun.*, vol. 9, no. 5, pp. 419–426, Oct. 1998.

[2] X. Wang and H. V. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1046–1061, July 1999.

[3] J. Boutros and G. Caire, "Iterative multiuser joint decoding: Unified framework and asymptotic analysis," *IEEE Trans. Info. Theory*, vol. 48, no. 7, pp. 1772–1793, July 2002.

[4] T. Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. Info. Theory*, vol. 48, no. 11, pp. 2888–2910, Nov. 2002.

[5] D. Guo and S. Verdú, "Randomly spread CDMA: asymptotics via statistical physics," *IEEE Trans. Info. Theory*, vol. 51, no. 6, pp. 1983–2010, June 2005.

[6] D. D. Lin and T. J. Lim, "A variational free energy minimization interpretation of multiuser detection in CDMA," *Proc. IEEE Globecom'05*, vol. 3, pp. 1570–1575, Dec. 2005.

[7] D. MacKay, *Information Theory, Inference and Learning Algorithms*. Cambridge U.K.: Cambridge University Press, 2003.

[8] R. M. Neal and G. E. Hinton, "A view of the EM algorithm that justifies incremental, sparse, and other variants," in *Learning in Graphical Models*, M. I. Jordan, Ed. Kluwer Academic Publishers, 1998.

[9] B. Lu and X. Wang, "Iterative receivers for multiuser space-time coding systems," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 11, pp. 2322–2335, Nov. 2000.

[10] M. Tüchler, A. C. Singer, and R. Koetter, "Minimum mean squared error equalization using *a priori* information," *IEEE Trans. Signal Processing*, vol. 50, no. 3, pp. 673–683, Mar. 2002.

[11] J. Winn and C. M. Bishop, "Variational message passing," *Journal of Machine Learning Research 6*, pp. 661–694, Apr. 2005.