Bit-Level Equalization and Soft Detection for Gray-Coded Multilevel Modulation

Darryl Dexu Lin and Teng Joon Lim *Senior Member, IEEE,*

Abstract

This paper investigates iterative soft-in soft-out (SISO) detection in coded multiple access channels, with Gray-coded M-ary quadrature amplitude modulation (QAM) for the channel symbols. The proposed solution may be summarized as a generic iterative detection scheme called *Bit-Level Equalization and Soft Detection* (BLESD), which is an extension of a unified variational inference framework for binary SISO detection proposed in our prior work. This new strategy fundamentally differs from the conventional symbol detector, in that data symbols are transparent to the new detector. Rather, soft estimates of the bits that make up the symbols are directly and naturally obtained by the detector in terms of posterior probabilities given the channel observation, facilitating efficient message-passing in joint detection and decoding. Case studies that illustrate the applications of the proposed scheme are presented for turbo multiuser detection (MUD) for multiple-access interference (MAI) channels and turbo equalization for inter-symbol interference (ISI) channels.

Index Terms

Free Energy, Gray Encoding, M-ary Modulation, Soft-In Soft-Out (SISO) Detection, Turbo Equalization, Turbo Multiuser Detection, Variational Inference.

I. Introduction

At the centre of physical layer wireless receiver design are the challenges of data detection and error control code (ECC) decoding. While traditionally these two tasks were treated separately, the discovery of the turbo principle enables the detector and decoder to exchange soft information in an iterative manner [1], resulting in dramatically improved bit-error-rate (BER) performance without any substantial complexity increase. Such a turbo receiver structure, depending on the areas of application, is called

D. D. Lin was with the Department of Electrical and Computer Engineering, University of Toronto, 10 King's College Road, Toronto, ON, Canada M5S 3G4. He is now with Qualcomm Incorporated, 5775 Morehouse Drive, San Diego, CA, 92121 USA (e-mail: darryl.lin@utoronto.ca). T. J. Lim is with the Department of Electrical and Computer Engineering, University of Toronto, 10 King's College Road, Toronto, ON, Canada M5S 3G4 (e-mail: tj.lim@utoronto.ca). This work was presented in part at the IEEE International Symposium on Information Theory 2006, Seattle, Washington, USA, July 2006, and the IEEE International Symposium on Information Theory 2007, Nice, France, June 2007.

turbo multiuser detector [2], [3], turbo equalizer [4], [5], or turbo MIMO (multiple-input multiple-output) equalizer [6], [7]. A key component in the turbo receiver is a practical soft-in soft-out (SISO) detector/equalizer, which has to be able to receive and generate soft estimates with small computational overhead.

To provide a comprehensive theory guiding the design of practical SISO detectors [2], [3], [5], [8], in [9], [10] we proposed a general framework, adopting the machine learning concept of *variational inference* [11], which illuminates the commonalities in many if not all practical turbo receivers. In [12], Nissilä and Pasupathy independently established a similar theory that interprets turbo MIMO equalizers as outcomes of variational optimization. These works describe SISO detectors as special cases of the variational inference framework simply by minimizing the "free energy" expressions corresponding to various postulates. With the help of new insights provided by probabilistic modelling, existing schemes may be improved systematically with clear and unified optimality objectives, while new schemes can also be found, and out-perform existing ones in certain applications.

In this paper, we investigate another application of the variational inference view for SISO detection, where we will extend the commonly-assumed binary phase shift keying (BPSK) model to the more challenging (and useful) realm of multilevel modulation. The proposed solution is called *Bit-Level Equalization and Soft Detection* (BLESD). As its name suggests, the BLESD scheme performs detection at the bit level, even when M-ary symbols are transmitted, contrary to the conventional procedure for handling multilevel modulation, where decisions are first made at the symbol level. In [13] and [14], the BPSK assumption of turbo multiuser detection (MUD) and equalization is extended to M-ary phase shift keying (M-PSK) with ease by exploiting the uniform-symbol-energy property of M-PSK symbols. However, the same technique does not apply to general quadrature amplitude modulation (QAM). To overcome this difficulty, Dejonghe and Vandendorpe proposed a more general solution in [15], enabling SISO detection at the symbol level for arbitrary multilevel modulation schemes, while allowing *extrinsic information* (EXT) to be obtained for each channel bit.

Our proposed BLESD approach is different, in that channel symbols are bypassed, thus directly generating bit EXT at the detector output. Detection algorithms for Gray-coded M-ary QAM will be systematically formulated and optimized. We focus our attention on Gray-coded QAM modulation for its simplicity and importance in wireless communications, which amounts to the bit-interleaved coded modulation (BICM) [16] scheme when combined with joint decoding.

The rest of the paper will be organized as follows. Section II presents the general signal model encapsulating both the multi-access interference (MAI) and inter-symbol interference (ISI) channels. Section III provides an overview for the variational inference framework for BPSK-based SISO detection. Section IV contains the contributions of this paper. A multi-linear transformation that describes the nonlinear bit-to-symbol Gray mapping in closed form is first introduced, followed by the derivation of a set of BLESD schemes based on variational free energy minimization. Section V addresses implementation challenges specific to turbo MUD and equalization, and presents simulation examples for both cases. Finally, Section VI concludes the paper. The mathematical notation in this paper follows that of [10].

II. Signal Model

We consider a real-valued linear vector channel, over which pulse-amplitude modulated (PAM) data symbols d_k, $k = 1, \cdots, K$, are transmitted:

$$\mathbf{r} = \mathbf{H}\mathbf{d} + \mathbf{n} \in \mathbb{R}^{N\times 1}, \tag{1}$$

where $\mathbf{H} \in \mathbb{R}^{N\times K}$ is the channel matrix. Each data symbol d_k in $\mathbf{d} = [d_1, d_2, \cdots, d_K]^T$ is a result of a Gray mapping of $\log_2 M = L$ information bits $\{b_{l,k}\}_{l=1}^{L}$. $\mathbf{n}$ is a white Gaussian noise vector with distribution $p(\mathbf{n}) = \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$. Since the complex-valued signal model for QAM signalling can be readily transformed to a real-valued model by concatenating the real and imaginary parts of the signal, (1) is sufficient for the study of QAM signalling as well.

The channel matrix $\mathbf{H}$ has different definitions in MAI and ISI channels:

1) *Multiuser CDMA*: In the context of synchronous CDMA in a flat-fading channel,

$$\mathbf{H} \triangleq \mathbf{S}\mathbf{A} = [A_1\,\mathbf{s}_1, \cdots, A_K\,\mathbf{s}_K] \in \mathbb{R}^{N\times K}, \tag{2}$$

where $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_K]$ is the normalized spreading code matrix, and $\mathbf{A} = \mathsf{diag}([A_1, \cdots, A_K]^T)$ contains channel gains of K active users. Typically, $K < N$, since the number of users is usually assumed to be smaller than the spreading gain.

2) *Multipath Channel*: In an ISI channel with channel impulse response (CIR) $\mathbf{h} = [h_0, \cdots, h_M]^T$ of length $M+1$, the received signal is the linear convolution of the CIR and the transmitted symbol sequence. In the interest of limiting the processing delay, we will take the sliding window approach at each time instance, and let $\mathbf{H}$ be the channel convolution matrix:

$$\mathbf{H} \triangleq \begin{bmatrix} h_M & \cdots & h_0 & 0 & \cdots & \cdots & 0 \\ 0 & h_M & \cdots & h_0 & 0 & \cdots & 0 \\ & & \ddots & & \ddots & & \\ 0 & \cdots & \cdots & 0 & h_M & \cdots & h_0 \end{bmatrix} \in \mathbb{R}^{N\times K}. \tag{3}$$

where $K = N + M > N$.

It is important to note, however, that despite sharing the vector channel model of (1), the MUD and equalization problems require slightly different mathematical treatments due to the different dimensionalities of $\mathbf{H}$ in the two cases. This should not detract from the common foundation underlying the multiple-access-channel problems that we are trying to establish.

III. Binary SISO Detection via Variational Inference

The general layout of a turbo receiver is well-known [3]. It consists of a *decoder component* and a *detector component* that exchange soft estimates about the channel bits. The decoder is used to resolve the ECC (outer code), and the detector is responsible for the channel distortion (inner code). In the case of a convolutional code being used as the outer code, the optimal design of the decoder component is an BCJR decoder[17].

The design of the detector component is more challenging, since the optimal *a posteriori* probability (APP) detector is known to be NP complete. Consequently, as summarized in Section I, much research has been devoted to searching for suboptimal SISO detectors that are able to accept prior information and generate soft EXT in terms of log-likelihood ratios (LLR) of channel bits.

Writing the BPSK channel model as $\mathbf{r} = \mathbf{H}\mathbf{b} + \mathbf{n}$, where $\mathbf{b} \in \{\pm 1\}^K$ are the channel bits, the role of a SISO detector is to approximate the LLR, for $\{b_k\}_{k=1}^K$, of the form

$$\Lambda_I(b_k) = \log \frac{p(\mathbf{r}|b_k = 1)}{p(\mathbf{r}|b_k = -1)}, \tag{4}$$

where $\Lambda_I(b_k)$ denotes the output EXT of the *Inner code*, and

$$p(\mathbf{r}|b_k) = \sum_{\{b_i\}_{i\neq k}} p(\mathbf{r}|\mathbf{b}) \prod_{i=1, i\neq k}^{K} p(b_i) \propto \frac{p(b_k|\mathbf{r})}{p(b_k)}. \tag{5}$$

In (5), $p(b_i)$ denotes the prior probability of b_i, which, in the turbo receiver context, comes from the output EXT of the *Outer code* decoder.

To avoid the exponential complexity of evaluating $p(b_k|\mathbf{r})$, the VFEM framework [9], [10] simplifies the problem by postulating a distribution $Q(\mathbf{b})$ that resembles $p(\mathbf{b}|\mathbf{r})$ but with a more convenient form. Our goal then becomes minimizing the Kullback-Leibler (KL) divergence between $Q(\mathbf{b})$ and $p(\mathbf{b}|\mathbf{r})$, also known as the *variational free energy*, up to an additive constant:

$$\mathcal{F}(\lambda) = \int_{\mathbf{b}} Q(\mathbf{b}; \lambda) \log \frac{Q(\mathbf{b}; \lambda)}{p(\mathbf{r}|\mathbf{b})p(\mathbf{b})} d\mathbf{b}. \tag{6}$$

In (6), $Q(\mathbf{b})$ is written as $Q(\mathbf{b}; \lambda)$ to denote the dependence of $Q(\mathbf{b})$ on λ explicitly, where λ contains a set of parameters that specify $Q(\mathbf{b})$. In the rest of the paper, we will however drop the dependence of the Q function on λ in accordance with the usual convention for writing probability distributions.

The distributions $p(\mathbf{b})$ and $Q(\mathbf{b})$ are called the postulated prior and posterior distribution, respectively, variations of which induce different detector types. For instance, setting $p(\mathbf{b})$ and $Q(\mathbf{b})$ to continuous

Gaussian distributions leads to MMSE-type SISO detectors [3], [5] (a.k.a. *Gaussian SISO* detectors), and setting $p(\mathbf{b})$ and $Q(\mathbf{b})$ to discrete binary distributions produces interference-cancellation-type SISO detectors [2], [8] (a.k.a. *discrete SISO* detectors) when $\mathcal{F}$ is minimized iteratively. In what follows, we will briefly review the derivation of these two detector types.

A. Gaussian SISO Detector

The Gaussian SISO detectors are induced by the following set of postulated distributions[1] :

$$\begin{cases} p(\mathbf{b}) &= \mathcal{N}(\tilde{\mathbf{b}}, \mathbf{W}), \quad b_k \in \mathbb{R} \\ p(\mathbf{r}|\mathbf{b}) &= \mathcal{N}(\mathbf{H}\mathbf{b}, \sigma^2 \mathbf{I}) \\ Q(\mathbf{b}) &= \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad b_k \in \mathbb{R} \end{cases} \tag{7}$$

where $\tilde{\mathbf{b}} = [\tilde{b}_1, \cdots, \tilde{b}_K]^T$ are the average bit estimates from the BCJR decoder ($\tilde{b}_k = \mathsf{E}[b_k]$), and $\mathbf{W} = \mathsf{diag}([1-\tilde{b}_1^2, \cdots, 1-\tilde{b}_K^2]^T)$ (since $\mathsf{V}[b_k] = 1-\tilde{b}_k^2$). In particular, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the parameters of the variational free energy. Minizing the varitional free energy we have,

$$\boldsymbol{\mu} = \tilde{\mathbf{b}} + (\mathbf{H}^T\mathbf{H} + \sigma^2 \mathbf{W}^{-1})^{-1}\mathbf{H}^T(\mathbf{r} - \mathbf{H}\tilde{\mathbf{b}}) \tag{8}$$

$$\boldsymbol{\Sigma} = (\sigma^{-2}\mathbf{H}^T\mathbf{H} + \mathbf{W}^{-1})^{-1}. \tag{9}$$

Since $Q(b_k)$ by definition is an approximation to $p(b_k|\mathbf{r})$, the output EXT of the SISO detector may be computed via (4) and (5), yielding

$$\Lambda_I(b_k) = \frac{\mathbf{e}_k^T(\mathbf{W} + \sigma^2(\mathbf{H}^T\mathbf{H})^{-1})^{-1}((\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{r} - \tilde{\mathbf{b}}_k)}{1 - [\mathbf{W}(\mathbf{W} + \sigma^2(\mathbf{H}^T\mathbf{H})^{-1})^{-1}]_{k,k}}. \tag{10}$$

B. Discrete SISO Detector

The discrete SISO detectors, on the other hand, are induced by the following set of postulated distributions:

$$\begin{cases} p(\mathbf{b}) &= \prod_{k=1}^K \xi_k^{\frac{1+b_k}{2}}(1-\xi_k)^{\frac{1-b_k}{2}}, \quad b_k \in \{\pm 1\} \\ p(\mathbf{r}|\mathbf{b}) &= \mathcal{N}(\mathbf{H}\mathbf{b}, \sigma^2\mathbf{I}) \\ Q(\mathbf{b}) &= \prod_{k=1}^K \gamma_k^{\frac{1+b_k}{2}}(1-\gamma_k)^{\frac{1-b_k}{2}}, \quad b_k \in \{\pm 1\} \end{cases} \tag{11}$$

where 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

[1] In [10], we showed that in addition to different choices of postulated distributions, the scheduling scheme for message passing is also an important factor in SISO detector design. Without loss of generality, we will assume the so-called flooding schedule in the paper. It is worth noting that other scheduling schemes lead to equally valid, though slightly different, detector expressions.

on $\mathbf{r}$ (*mean-field approximation*) has to be made. The unknown parameter in (6), λ, corresponds to $\{\gamma_k\}_{k=1}^K$.

As demonstrated in [10], postulated probabilities of this type lead to a SISO detector resembling the successive interference cancellation (SIC) detector. Solving $Q(\mathbf{b})$ yields two recursive equations:

$$\Lambda_I(b_k) = \frac{2}{\sigma^2}\left[\mathbf{h}_k^T\mathbf{r} - \boldsymbol{\beta}_k^T\mathbf{m}\right] \tag{12}$$

$$m_k = \tanh\left[\frac{\Lambda_I(b_k)+\Lambda_O(b_k)}{2}\right], \tag{13}$$

where $\mathbf{m} = [m_1, \cdots, m_K]^T$ is the posterior mean of $\mathbf{b}$, i.e., $m_k = 2\gamma_k - 1$, and $\mathbf{h}_k$ and $\boldsymbol{\beta}_k$ are the k-th column vectors of $\mathbf{H}$ and $\mathbf{H}^T\mathbf{H} - \mathsf{diag}(\mathbf{H}^T\mathbf{H})$, respectively. $\Lambda_O(b_k)$ is the LLR obtained from the BCJR decoder, representing the prior probability of b_k. This recursive relation leads to SIC-like nonlinear iterations, where $\Lambda_I(b_k)$ is found upon convergence.

IV. Bit-Level Equalization and Soft Detection for M-QAM

Having revisited the variational inference interpretation of SISO detection for BPSK modulation, in this section, we will investigate new detection methods for Gray-coded M-ary modulation utilizing the same theoretical framework.

Various turbo receiver designs for multilevel modulation have been reported in the past. For instance, [13], [14] proposed SISO detector formulations for M-PSK, and [15] investigated the solution for arbitrary modulation schemes. These approaches share the commonality that they all amount to computing the symbol likelihoods using some form of MMSE filtering and converting symbol likelihoods to bit likelihoods.

This paper proposes an unconventional approach, in which bits modulated within channel symbols are detected directly, thus eliminating the need for soft symbol/bit mapping and demapping. The change of paradigm is made possible by treating the nonlinear bit-to-symbol mapping as part of the channel, and by using the variational inference framework (which subsumes the MMSE filtering approach).

This section progresses in the following manner: Section IV-A proposes a mapping from $\mathbf{b}$ to $\mathbf{d}$ for Gray-coded 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. Here $\mathbf{b}_l = [b_{l,1}, \cdots, b_{l,K}]^T$ is a vector containing the l-th bits of $\{d_k\}_{k=1}^K$, where the channel symbol d_k is made up of the Gray mapping of L channel bits $b_{1,k}, \cdots, b_{L,k}$. Section IV-B and Section IV-C, from the Gaussian SISO and discrete SISO perspectives, respectively, optimize $\mathcal{F}$ given the postulated prior, channel transition and posterior distributions, and reveal how practical SISO detection algorithms can be extracted as a result.

A. Gray Mapping and Multi-Linear Transformation

Consider a 4-PAM symbol $d_{[2]}$ that consists of two information bits, $b_1, b_2 \in \{1, -1\}$ (The subscript of $d_{[2]}$ indicates that each symbol contains two bits). If the value of $d_{[2]}$ is determined by the equation $d_{[2]} = b_2(2 + b_1) = b_2 b_1 + 2b_2$, then such a bit-to-symbol mapping is a Gray mapping, because the four values that $d_{[2]}$ takes on, -3, -1, 1 and 3, correspond to (b_2, b_1) pairs $(-1, 1)$, $(-1, -1)$, $(1, -1)$ and $(1, 1)$. In fact, this simple Gray mapping strategy may be generalized to arbitrary 2^L-PAM constellations:

Definition 1: A mapping of L bits to a 2^L-PAM constellation point $d_{[L]}$ following the equation

$$d_{[L]} = \sum_{l=1}^{L} 2^{l-1} b_L b_{L-1} \cdots b_l = \sum_{l=1}^{L} 2^{l-1} \prod_{p=l}^{L} b_p, \tag{14}$$

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

Note that the $\{b_l\}_{l=1}^{L} \to d_{[L]}$ mapping formula may also be written in a recursive form as

$$d_{[L]} = \begin{cases} b_L & L = 1 \\ b_L(2^{L-1} + d_{[L-1]}) & L > 1 \end{cases} \tag{15}$$

Using (15), the Gray mapping property of the above labeling scheme can be proven by induction [18]. However, the construction is not unique. It can be shown that, in (15), if we were to change the sign before the term 2^{L-1} or $d_{[L-1]}$, it would remain a Gray mapping. In particular, a sign-inverted mapping of the form $d_{[L]} = -\sum_{l=1}^{L} 2^{l-1} b_L b_{L-1} \cdots b_l$ corresponds to the conventional Gray mapping found in the literature. Without loss of generality, we shall use (14) for convenience. Note that $d_{[L]}$ 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 [19], which has useful properties for use in variational inference.

Now that a simple closed-form expression that describes the Gray bit-to-symbol mapping is available to us, this implies that our channel models may be written in terms of $b_{l,k}$, instead of d_k. It is then possible to design detectors for $b_{l,k}$, rather than d_k. Since $b_{l,k}$ is a binary random variable, the objective function associated with it should be much easier to handle. We will now derive SISO detection algorithms for M-ary symbols, making use of the variational inference framework outlined in Section III.

B. Gaussian SISO Detector for 2^L-PAM Modulation

1) Postulated Distributions: Prior Distribution: Because of interleaving, we may assume the L code bits that make up each symbol to be independent. Therefore,

$$\begin{aligned} p(\mathbf{d}) &= \textstyle\prod_{l=1}^{L} p(\mathbf{b}_l) \\ &= \textstyle\prod_{l=1}^{L} \mathcal{N}(\tilde{\mathbf{b}}_l, \mathbf{W}_l), \end{aligned} \tag{16}$$

where $\tilde{\mathbf{b}}_l = [\tilde{b}_{l,1}, \cdots, \tilde{b}_{l,K}]^T$ represents the mean estimates from the BCJR decoder of the l-th channel bits of all users.

Channel Transition Distribution: The channel transition distribution or likelihood function is $p(\mathbf{r}|\mathbf{d}) = \mathcal{N}(\mathbf{Hd}, \sigma^2\mathbf{I})$. 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. Recognizing $\mathbf{d} = \sum_{l=1}^{L} 2^{l-1} \prod_{p=l}^{L} \mathbf{b}_p$ from (14), then

$$\begin{aligned} p(\mathbf{r}|\mathbf{d}) &= p(\mathbf{r}|\mathbf{b}_1, \cdots, \mathbf{b}_L) \\ &= \mathcal{N}\left(\mathbf{H} \cdot \sum_{l=1}^{L} 2^{l-1} \prod_{p=l}^{L} \mathbf{b}_p, \sigma^2\mathbf{I}\right), \end{aligned} \tag{17}$$

where the notation $\prod$ represents a series of Schur (element-wise) products, i.e.,

$$\prod_{p=l}^{L} \mathbf{b}_p = \mathbf{b}_l \circ \mathbf{b}_{l+1} \circ \cdots \circ \mathbf{b}_L. \tag{18}$$

In the context of MUD, we place the l-th bit of all users in one vector $\mathbf{b}_l$, and eventually perform joint detection not only among K users, but also among the L bits in each user's symbol.

Posterior Distribution: Following the methodology for Gaussian SISO detector design, we restrict each vector $\mathbf{b}_l$ to have a Gaussian posterior distribution. Here we adopt the mean-field approximation and assume the independence of $\{\mathbf{b}_l\}_{l=1}^{L}$. This assumption is critical to reducing the computational complexity of the BLESD algorithm. We thus have

$$\begin{aligned} Q(\mathbf{d}) &= \prod_{l=1}^{L} Q(\mathbf{b}_l) \\ &= \prod_{l=1}^{L} \mathcal{N}(\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l). \end{aligned} \tag{19}$$

2) Free Energy Evaluation: The variational free energy expression for channel symbols $\{d_k\}_{k=1}^{K}$ may be written as:

$$\begin{aligned} \mathcal{F} &= \int_{\mathbf{d}} Q(\mathbf{d}) \log \frac{Q(\mathbf{d})}{p(\mathbf{r}|\mathbf{d})p(\mathbf{d})} d\mathbf{d} \\ &= \int_{\mathbf{d}} Q(\mathbf{d}) \log Q(\mathbf{d}) d\mathbf{d} - \int_{\mathbf{d}} Q(\mathbf{d}) \log p(\mathbf{r}|\mathbf{d}) d\mathbf{d} - \int_{\mathbf{d}} Q(\mathbf{d}) \log p(\mathbf{d}) d\mathbf{d}. \end{aligned} \tag{20}$$

The task of free energy evaluation then condenses to the computation of the integral expressions in (20) given $p(\mathbf{d})$, $p(\mathbf{r}|\mathbf{d})$ and $Q(\mathbf{d})$ defined in (16), (17) and (19), respectively. This is mathematically involved because the multi-linear transformation connecting the channel bits to Gray-mapped symbols necessitates the development of a new set of matrix algebra relations involving the Schur product. To preserve the clarity of the subsequent presentation, we move the complete derivation to Appendix A. While it is sufficient to work with the final $\mathcal{F}$ expression in (21) directly to arrive at desired Gaussian SISO detectors, readers are encouraged to refer to Appendix A for additional insights.

The complete free energy expression for Gaussian SISO detection is assembled as follows:

TABLE I
PARAMETERS FOR BPSK AND 4-PAM GAUSSIAN SISO BLESD DETECTOR.

BPSK ($L=1$)		4-PAM: ($L=2$)			
$\mathbf{\Psi}_1 =$	$\mathbf{H}^T\mathbf{H}$	$\mathbf{\Psi}_1 =$	$(\mathbf{\Sigma}_2 + \boldsymbol{\mu}_2\boldsymbol{\mu}_2^T) \circ (\mathbf{H}^T\mathbf{H})$	$\mathbf{\Psi}_2 =$	$(\mathbf{\Sigma}_1 + \boldsymbol{\mu}_1\boldsymbol{\mu}_1^T) \circ (\mathbf{H}^T\mathbf{H}) + 4\mathbf{H}^T\mathbf{H}$
$\mathbf{\Xi}_1 =$	$\mathbf{0}$	$\mathbf{\Xi}_1 =$	$\mathbf{0}$	$\mathbf{\Xi}_2 =$	$2(\boldsymbol{\mu}_1\mathbf{1}^T + \mathbf{1}\boldsymbol{\mu}_1^T) \circ (\mathbf{H}^T\mathbf{H})$
$\mathbf{\Omega}_1 =$	$\mathbf{I}$	$\mathbf{\Omega}_1 =$	$\mathsf{diag}(\boldsymbol{\mu}_2)$	$\mathbf{\Omega}_2 =$	$\mathsf{diag}(\boldsymbol{\mu}_1 + 2\mathbf{1})$
$\boldsymbol{\phi}_1 =$	$\mathbf{0}$	$\boldsymbol{\phi}_1 =$	$2 \cdot [(\mathbf{\Sigma}_2 + \boldsymbol{\mu}_2\boldsymbol{\mu}_2^T) \circ (\mathbf{H}^T\mathbf{H})]\mathbf{1}$	$\boldsymbol{\phi}_2 =$	$\mathbf{0}$

$$\begin{aligned} & \mathcal{F}(\boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_L, \mathbf{\Sigma}_1, \cdots, \mathbf{\Sigma}_L) \\ = & \sum_{l=1}^L \left[-\tfrac{1}{2}\sum_{l=1}^L \log|\mathbf{\Sigma}_l| + \tfrac{1}{2}\mathsf{tr}(\mathbf{W}_l^{-1}\mathbf{\Sigma}_l) + \tfrac{1}{2}\boldsymbol{\mu}_l^T\mathbf{W}_l^{-1}\boldsymbol{\mu}_l - \tilde{\mathbf{b}}_l^T\mathbf{W}_l^{-1}\boldsymbol{\mu}_l \right] + \tfrac{1}{2\sigma^2}\left\{ \sum_{l=1}^L 2^{l-1}\,\mathsf{tr}\left[\textstyle\prod_{p=l}^L (\mathbf{\Sigma}_p + \boldsymbol{\mu}_p\boldsymbol{\mu}_p^T)(\mathbf{H}^T\mathbf{H}) \right] \right. \\ & \left. + \sum_{1\le i<j\le L} 2^{i+j-1}\,\mathsf{tr}\left[\textstyle\prod_{p=j}^L (\mathbf{\Sigma}_p + \boldsymbol{\mu}_p\boldsymbol{\mu}_p^T) \cdot \mathbf{H}^T\mathbf{H} \cdot \mathsf{diag}(\prod_{p=i}^{j-1}\boldsymbol{\mu}_p) \right] \right\} - \tfrac{1}{2\sigma^2}\left\{ 2\mathbf{r}^T\mathbf{H}(\textstyle\sum_{l=1}^L 2^{l-1}\prod_{p=l}^L \boldsymbol{\mu}_p) \right\} \end{aligned} \quad (21)$$

3) Free Energy Minimization: Taking the derivative of $\mathcal{F}(\{\boldsymbol{\mu}_l\}_{l=1}^L, \{\mathbf{\Sigma}_l\}_{l=1}^L)$ w.r.t. $\boldsymbol{\mu}_l$ and $\mathbf{\Sigma}_l$, $1 \le l \le L$, and equating to zero yields:

$$\boldsymbol{\mu}_l = \tilde{\mathbf{b}}_l + \left(\mathbf{\Psi}_l + \mathbf{\Xi}_l + \sigma^2\mathbf{W}_l^{-1}\right)^{-1}\left[\mathbf{\Omega}_l\mathbf{H}^T\mathbf{r} - \boldsymbol{\phi}_l - (\mathbf{\Psi}_l + \mathbf{\Xi}_l)\tilde{\mathbf{b}}_l\right], \quad (22)$$

$$\mathbf{\Sigma}_l = \left(\sigma^{-2}\mathbf{\Psi}_l + \sigma^{-2}\mathbf{\Xi}_l + \mathbf{W}_l^{-1}\right)^{-1} \quad (23)$$

where

$$\begin{cases} \mathbf{\Psi}_l = & \sum_{i=1}^l \left\{ 2^{2i-2}\left[(\mathbf{H}^T\mathbf{H}) \circ \prod_{p=i,p\neq l}^L (\mathbf{\Sigma}_p + \boldsymbol{\mu}_p\boldsymbol{\mu}_p^T) \right] \right\} + I(l=L)\cdot 2^{2l-2}\mathbf{H}^T\mathbf{H} \\ \mathbf{\Xi}_l = & \sum_{1\le i<j\le l, i\neq j} 2^{i+j-2}\left\{ I(j=L)\cdot\left[\mathbf{H}^T\mathbf{H}\cdot\mathsf{diag}(\prod_{p=i}^{j-1}\boldsymbol{\mu}_p) \right] + I(j=L)\cdot\left[\mathbf{H}^T\mathbf{H}\cdot\mathsf{diag}(\prod_{p=i}^{j-1}\boldsymbol{\mu}_p) \right]^T \right. \\ & \left. + \left[\mathbf{H}^T\mathbf{H}\cdot\mathsf{diag}(\prod_{p=i}^{j-1}\boldsymbol{\mu}_p) \circ \prod_{p=j,p\neq l}^L (\mathbf{\Sigma}_p + \boldsymbol{\mu}_p\boldsymbol{\mu}_p^T) \right] + \left[\mathbf{H}^T\mathbf{H}\cdot\mathsf{diag}(\prod_{p=i}^{j-1}\boldsymbol{\mu}_p) \circ \prod_{p=j,p\neq l}^L (\mathbf{\Sigma}_p + \boldsymbol{\mu}_p\boldsymbol{\mu}_p^T) \right]^T \right\} \\ \mathbf{\Omega}_l = & \sum_{i=1}^l \left\{ 2^{i-1}\,\mathsf{diag}(\prod_{p=i,p\neq l}^L \boldsymbol{\mu}_p) \right\} + I(l=L)\cdot 2^{l-1}\mathbf{I} \\ \boldsymbol{\phi}_l = & \sum_{1\le i\le l<j\le L} 2^{i+j-2}\left\{ \mathsf{diag}(\prod_{p=i,p\neq l}^{j-1}\boldsymbol{\mu}_p)\left[(\mathbf{H}^T\mathbf{H}) \circ \prod_{p=j}^L (\mathbf{\Sigma}_p + \boldsymbol{\mu}_p\boldsymbol{\mu}_p^T) \right]\mathbf{1} \right. \\ & \left. + I(i=l=j-1)\cdot\left[(\mathbf{H}^T\mathbf{H}) \circ \prod_{p=j}^L (\mathbf{\Sigma}_p + \boldsymbol{\mu}_p\boldsymbol{\mu}_p^T) \right]\mathbf{1} \right\} \end{cases} \quad (24)$$

In (24), $I(A)$ is an indicator function which equals 1 if A is true and 0 otherwise. Also, we let $\prod_{n\in\mathcal{S}} \mathbf{X}_n = \mathbf{0}$ for $\mathcal{S} = \emptyset$. In other words, the Schur product over an empty set of matrices equals zero.

4) Examples: We will now show how to obtain Gaussian SISO detectors for BPSK and 4-PAM modulation. Table I contains a list of parameters resulting from evaluating (24) for $L=1$ and $L=2$. Substituting the parameters corresponding to $L=1$ into (22) and (23), we have

$$\begin{aligned} \boldsymbol{\mu} &= \tilde{\mathbf{b}} + (\mathbf{H}^T\mathbf{H} + \sigma^2\mathbf{W}^{-1})^{-1}(\mathbf{H}^T\mathbf{r} - \mathbf{H}^T\mathbf{H}\tilde{\mathbf{b}}) \\ \mathbf{\Sigma} &= (\sigma^{-2}\mathbf{H}^T\mathbf{H} + \mathbf{W}^{-1})^{-1}. \end{aligned} \quad (25)$$

These are identical to $Q(\mathbf{b})$ evaluated in Section III-A. It verifies that, not surprisingly, the Gaussian SISO detector for BPSK modulation discussed in [10] is a special case of the general BLESD scheme when $L = 1$.

For $L = 2$, substituting the parameters in Table I into (22) and (23) yields expressions corresponding to $Q(\mathbf{b}_1)$ and $Q(\mathbf{b}_2)$:

$$\begin{aligned} \boldsymbol{\mu}_1 &= \tilde{\mathbf{b}}_1 + [\mathbf{R}_1 + \sigma^2 \mathbf{W}_1^{-1}]^{-1} [\mathbf{H}_1^T \mathbf{r} - \mathbf{R}_1(\tilde{\mathbf{b}}_1 + 2\mathbf{1})] \\ \boldsymbol{\Sigma}_1 &= (\sigma^{-2}\mathbf{R}_1 + \mathbf{W}_1^{-1})^{-1} \\ \boldsymbol{\mu}_2 &= \tilde{\mathbf{b}}_2 + [\mathbf{R}_2 + \sigma^2 \mathbf{W}_2^{-1}]^{-1} [\mathbf{H}_2^T \mathbf{r} - \mathbf{R}_2\tilde{\mathbf{b}}_2] \\ \boldsymbol{\Sigma}_2 &= (\sigma^{-2}\mathbf{R}_2 + \mathbf{W}_2^{-1})^{-1} \end{aligned} \tag{26}$$

where

$$\begin{aligned} \mathbf{H}_1 &= \mathsf{diag}(\boldsymbol{\mu}_2)\mathbf{H}; & \mathbf{R}_1 &= (\boldsymbol{\Sigma}_2 + \boldsymbol{\mu}_2\boldsymbol{\mu}_2^T) \circ (\mathbf{H}^T\mathbf{H}); \\ \mathbf{H}_2 &= \mathsf{diag}(\boldsymbol{\mu}_1 + 2\mathbf{1})\mathbf{H}; & \mathbf{R}_2 &= [\boldsymbol{\Sigma}_1 + (\boldsymbol{\mu}_1 + 2\mathbf{1})(\boldsymbol{\mu}_1 + 2\mathbf{1})^T] \circ (\mathbf{H}^T\mathbf{H}). \end{aligned} \tag{27}$$

Notice the clear similarity between the equations for 4-PAM and those of BPSK. An interesting intuitive interpretation may be drawn. For instance, in 4-PAM, bit 1 of all users, $\mathbf{b}_1$, sees an effective channel $\mathsf{diag}(\boldsymbol{\mu}_2)\mathbf{H}$ and effective channel correlation matrix $(\boldsymbol{\Sigma}_2 + \boldsymbol{\mu}_2\boldsymbol{\mu}_2^T) \circ (\mathbf{H}^T\mathbf{H})$. Similarly, $\mathbf{b}_2$ sees $\mathsf{diag}(\boldsymbol{\mu}_1 + 2\mathbf{1})\mathbf{H}$ and $[\boldsymbol{\Sigma}_1 + (\boldsymbol{\mu}_1 + 2\mathbf{1})(\boldsymbol{\mu}_1 + 2\mathbf{1})^T] \circ (\mathbf{H}^T\mathbf{H})$. Such a similarity implies that techniques for reducing the computational cost of SISO detection for BPSK, which are discussed in [3] and [14], also apply to the Gaussian SISO BLESD approach.

C. Discrete SISO Detector for 2^L-PAM Modulation

Now we consider a different class of SISO detectors, using discrete prior and posterior distributions in the VFEM formulation.

1) Postulated Distributions: Prior Distribution: Assuming the independence of $b_{l,k}$, $l = 1, \cdots, L$, due to interleaving, we may write, for $b_{l,k} \in \{\pm 1\}$

$$\begin{aligned} p(\mathbf{d}) &= \textstyle\prod_{l=1}^L p(\mathbf{b}_l) \\ &= \textstyle\prod_{l=1}^L \prod_{k=1}^K \xi_{l,k}^{\frac{1+b_{l,k}}{2}} (1 - \xi_{l,k})^{\frac{1-b_{l,k}}{2}} \\ &= \textstyle\prod_{l=1}^L \prod_{k=1}^K (\frac{1+\tilde{b}_{l,k}}{2})^{\frac{1+b_{l,k}}{2}} (\frac{1-\tilde{b}_{l,k}}{2})^{\frac{1-b_{l,k}}{2}}, \end{aligned} \tag{28}$$

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 third equality, such that $\tilde{b}_{l,k}$ represents the prior mean of $b_{l,k}$, i.e., $\tilde{b}_{l,k} = 2\xi_{l,k} - 1$.

Channel Transition Distribution: Again we make use of the multi-linear bit-to-symbol mapping introduced in Section IV-A. The channel transition distribution will be the same as (17):

$$\begin{aligned} p(\mathbf{r}|\mathbf{d}) &= p(\mathbf{r}|\mathbf{b}_1, \cdots, \mathbf{b}_L) \\ &= \mathcal{N}\left(\mathbf{H} \cdot \sum_{l=1}^{L} 2^{l-1} \prod_{p=l}^{L} \mathbf{b}_p, \sigma^2 \mathbf{I}\right), \end{aligned} \tag{29}$$

Posterior Distribution: We make a mean-field approximation [10] and assume that $b_{l,k}$'s are independent over both l and k conditioned on observations. In particular,

$$\begin{aligned} Q(\mathbf{d}) &= \prod_{l=1}^{L} Q(\mathbf{b}_l) \\ &= \prod_{l=1}^{L} \prod_{k=1}^{K} \gamma_{l,k}^{\frac{1+b_{l,k}}{2}} (1-\gamma_{l,k})^{\frac{1-b_{l,k}}{2}} \\ &= \prod_{l=1}^{L} \prod_{k=1}^{K} (\tfrac{1+m_{l,k}}{2})^{\frac{1+b_{l,k}}{2}} (\tfrac{1-m_{l,k}}{2})^{\frac{1-b_{l,k}}{2}}, \end{aligned} \tag{30}$$

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 posterior mean of $b_{l,k}$.

2) Free Energy Evaluation: In Appendix B, we present the detailed derivation of $\mathcal{F}$ given $p(\mathbf{d})$, $p(\mathbf{r}|\mathbf{d})$ and $Q(\mathbf{d})$ defined in (28), (29) and (30), respectively. The complete free energy expression for discrete SISO detection is therefore assembled as follows:

$$\begin{aligned} & \mathcal{F}(\mathbf{m}_1, \cdots, \mathbf{m}_L) \\ = & \sum_{l=1}^{L} \sum_{k=1}^{K} \left\{ \tfrac{1+m_{l,k}}{2} \log \tfrac{1+\tilde{b}_{l,k}}{1+m_{l,k}} + \tfrac{1-m_{l,k}}{2} \log \tfrac{1-\tilde{b}_{l,k}}{1-m_{l,k}} \right\} - \tfrac{1}{2\sigma^2} \left\{ 2\mathbf{r}^T \mathbf{H} (\sum_{l=1}^{L} 2^{l-1} \prod_{p=l}^{L} \mathbf{m}_p) \right\} \\ & + \tfrac{1}{2\sigma^2} \left\{ (\sum_{l=1}^{L} 2^{l-1} \prod_{p=l}^{L} \mathbf{m}_p)^T [\mathbf{H}^T \mathbf{H} - \mathsf{diag}(\mathbf{H}^T \mathbf{H})] (\sum_{l=1}^{L} 2^{l-1} \prod_{p=l}^{L} \mathbf{m}_p) \right\} \\ & + \tfrac{1}{2\sigma^2} \left\{ \mathbf{1}^T \mathsf{diag}(\mathbf{H}^T \mathbf{H}) (\sum_{0<i\le j<L} 2^{i+j} \prod_{p=i}^{j} \mathbf{m}_p) \right\} \end{aligned} \tag{31}$$

3) Free Energy Minimization: Taking the derivative of $\mathcal{F}(\{\mathbf{m}_l\}_{l=1}^{L})$ w.r.t. $\mathbf{m}_l, 1 \le l \le L$, and equating to zero yields:

$$\begin{bmatrix} \log \frac{1+m_{l,1}}{1-m_{l,1}} \\ \vdots \\ \log \frac{1+m_{l,K}}{1-m_{l,K}} \end{bmatrix} = \begin{bmatrix} \log \frac{1+\tilde{b}_{l,1}}{1-\tilde{b}_{l,1}} \\ \vdots \\ \log \frac{1+\tilde{b}_{l,K}}{1-\tilde{b}_{l,K}} \end{bmatrix} + \tfrac{2}{\sigma^2} \left\{ \mathbf{\Delta}_l^T \mathbf{H}^T \mathbf{r} - \mathbf{E}_l \, \mathsf{diag}(\mathbf{H}^T \mathbf{H}) \mathbf{1} - \mathbf{\Delta}_l^T [\mathbf{H}^T \mathbf{H} - \mathsf{diag}(\mathbf{H}^T \mathbf{H})] (\mathbf{\Delta}_l \mathbf{m}_l + \boldsymbol{\zeta}_l) \right\} \tag{32}$$

where $\mathbf{\Delta}_l = \mathsf{diag}(\boldsymbol{\delta}_l)$, $\mathbf{E}_l = \mathsf{diag}(\boldsymbol{\epsilon}_l)$ and

$$\begin{cases} \boldsymbol{\delta}_l = \sum_{i=1}^{l} \left\{ 2^{i-1} \prod_{n=i, n\neq l}^{L} \mathbf{m}_n \right\} + I(l=L) \cdot 2^{l-1} \cdot \mathbf{1} \\ \boldsymbol{\epsilon}_l = \sum_{0<i\le l\le j<L, i\neq j} \left\{ 2^{i+j-1} \prod_{n=i, n\neq l}^{j} \mathbf{m}_n \right\} + I(0<l<L) \cdot 2^{2l-1} \cdot \mathbf{1} \\ \boldsymbol{\zeta}_l = \sum_{i=l+1}^{L} 2^{i-1} \prod_{n=i}^{L} \mathbf{m}_n. \end{cases} \tag{33}$$

$\mathcal{F}(\mathbf{m}_1, \cdots, \mathbf{m}_L)$ 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 (32) it is seen that setting $\partial\mathcal{F}/\partial m_{l,k} = 0$ leads to

$$\log\frac{1+m_{l,k}}{1-m_{l,k}} = \log\frac{1+\tilde{b}_{l,k}}{1-\tilde{b}_{l,k}} + \frac{2}{\sigma^2}\left\{\delta_{l,k}\cdot\mathbf{h}_k^T\mathbf{r} - \epsilon_{l,k}\cdot\boldsymbol{\rho}_k^T\mathbf{1} - \delta_{l,k}\cdot\boldsymbol{\beta}_k^T(\boldsymbol{\delta}_l\circ\mathbf{m}_l + \boldsymbol{\zeta}_l)\right\}, \tag{34}$$

where $\mathbf{h}_k$, $\boldsymbol{\rho}_k$, and $\boldsymbol{\beta}_k$ are the kth column of $\mathbf{H}$, $\mathsf{diag}(\mathbf{H}^T\mathbf{H})$, and $\mathbf{H}^T\mathbf{H} - \mathsf{diag}(\mathbf{H}^T\mathbf{H})$, respectively. Notice that the right hand side of (34) is independent of $m_{l,k}$, and

$$\Lambda_O(b_{l,k}) = \log\frac{1+\tilde{b}_{l,k}}{1-\tilde{b}_{l,k}} \tag{35}$$

represents the LLR of the prior probability. Therefore, the detector output EXT becomes

$$\begin{aligned}\Lambda_I(b_{l,k}) &= \log\tfrac{1+m_{l,k}}{1-m_{l,k}} - \log\tfrac{1+\tilde{b}_{l,k}}{1-\tilde{b}_{l,k}} \\ &= \tfrac{2}{\sigma^2}\left\{\delta_{l,k}\cdot\mathbf{h}_k^T\mathbf{r} - \epsilon_{l,k}\cdot\boldsymbol{\rho}_k^T\mathbf{1} - \delta_{l,k}\cdot\boldsymbol{\beta}_k^T(\boldsymbol{\delta}_l\circ\mathbf{m}_l + \boldsymbol{\zeta}_l)\right\}.\end{aligned} \tag{36}$$

4) Examples: We will now show how discrete SISO detection schemes can be formulated for BPSK and 4-PAM modulation. Evaluating (33) for $L = 1$ and $L = 2$ results in a set of parameters as listed in Table II.

TABLE II
PARAMETERS FOR BPSK AND 4-PAM DISCRETE SISO BLESD DETECTOR.

BPSK ($L=1$)		4-PAM: ($L=2$)			
$\boldsymbol{\delta}_1 =$	$\mathbf{1}$	$\boldsymbol{\delta}_1 =$	$\mathbf{m}_2$	$\boldsymbol{\delta}_2 =$	$\mathbf{m}_1 + 2\mathbf{1}$
$\boldsymbol{\epsilon}_1 =$	$\mathbf{0}$	$\boldsymbol{\epsilon}_1 =$	$2\mathbf{1}$	$\boldsymbol{\epsilon}_2 =$	$\mathbf{0}$
$\boldsymbol{\zeta}_1 =$	$\mathbf{0}$	$\boldsymbol{\zeta}_1 =$	$2\mathbf{m}_2$	$\boldsymbol{\zeta}_2 =$	$\mathbf{0}$

The update equations for a BPSK detector are determined after substituting the parameters corresponding to $L = 1$ into (34):

$$\log\frac{1+m_k}{1-m_k} = \log\frac{1+\tilde{b}_k}{1-\tilde{b}_k} + \frac{2}{\sigma^2}\left[\mathbf{h}_k^T\mathbf{r} - \boldsymbol{\beta}_k^T\mathbf{m}\right]. \tag{37}$$

Alternatively, (37) may be written as a recursive relation

$$\Lambda_I(b_k) = \frac{2}{\sigma^2}\left[\mathbf{h}_k^T\mathbf{r} - \boldsymbol{\beta}_k^T\mathbf{m}\right] \tag{38}$$

$$m_k = \tanh\left[\frac{\Lambda_I(b_k) + \Lambda_O(b_k)}{2}\right], \tag{39}$$

which is identical to the original discrete SISO detector for BPSK described in Section III-B. Substituting the parameters corresponding to $L = 2$ into (34), we have

$$\Lambda_I(b_{2,k}) = \frac{2}{\sigma^2}\left[(m_{1,k}+2)\cdot \mathbf{h}_k^T\mathbf{r} - (m_{1,k}+2)\cdot \boldsymbol{\beta}_k^T(\mathbf{m}_2\circ\mathbf{m}_1 + 2\mathbf{m}_2)\right] \tag{40}$$

$$\Lambda_I(b_{1,k}) = \frac{2}{\sigma^2}\left[m_{2,k}\cdot \mathbf{h}_k^T\mathbf{r} - 2\boldsymbol{\rho}_k^T\mathbf{1} - m_{2,k}\cdot \boldsymbol{\beta}_k^T(\mathbf{m}_2\circ\mathbf{m}_1 + 2\mathbf{m}_2)\right], \tag{41}$$

where, for $l = 1, 2$

$$m_{l,k} = \tanh\left[\frac{\Lambda_I(b_{l,k}) + \Lambda_O(b_{l,k})}{2}\right]. \tag{42}$$

These equations for 4-PAM modulation resemble the BPSK counterpart, but require the sequential update of $\Lambda_I(b_{2,k})$ as well as $\Lambda_I(b_{1,k})$.

The discrete SISO detectors derived above have a common feature. That is, the free energy expression in (31) is non-convex w.r.t. $\mathbf{m}$, implying that the coordinate descent method to decrease $\mathcal{F}(\mathbf{m}_1, \cdots, \mathbf{m}_L)$ does not guarantee global convergence. To alleviate this problem, the detector needs to be initialized to a point close to the global minimum of $\mathcal{F}$, thereby reducing the algorithm's chance of being trapped in local minima. [10] proposed a decorrelating-decision-feedback (DDF) SISO detector (the SISO version of the well-known DDF detector [20]) as the front end of BPSK-based discrete SISO detectors. In this paper, the DDF SISO detector will also be used in the first outer iteration of the discrete SISO BLESD turbo receiver to overcome local minima of $\mathcal{F}$. We refer the reader to [10] for details.

V. Numerical Results

In the previous section, the Gaussian SISO and discrete SISO detector formulations for Gray-coded PAM symbols were derived for the general linear channel model in (1). This section will test the proposed discrete SISO algorithm for turbo MUD and Gaussian SISO algorithm for turbo equalization. For conciseness, the simulation examples are only selected to exhibit the power of the paradigm shift brought by the BLESD approach, and are not exhaustive by any means.

A. Turbo Multiuser Detection

The discrete SISO detection formulation was presented in Section IV-C. Table III outlines the turbo MUD algorithm implementing discrete SISO detection for 4-PAM/16-QAM. It should be noted that in the actual simulation, the first outer iteration needs to employ a decorrelating decision-feedback (DDF) SISO detector, as described in [10].

In Fig. 1 we present simulations for the discrete SISO MUD, with 16-QAM channel symbols. We use a setting similar to that in [3], by assuming a flat-fading synchronous system with $K = 8$ users and equal spreading code cross-correlation $\rho = 0.5$. All users transmit packets containing 128 information bits each, which are encoded by the same rate-$1/2$ convolutional code with generator 101 and 111. All channels have

TABLE III

TURBO MUD OF 4-PAM/16-QAM IMPLEMENTING GAUSSIAN SISO DETECTION.

Discrete SISO MUD of 4-PAM/16-QAM

Initialization: $\Lambda_O(b_{l,k}) = 0$ and $\Lambda_I(b_{l,k}) = 0$ for $l = 1, 2,\ k = 1, \cdots, K$
FOR $j = 1 : J$ (*Outer Iteration*)
  $\tilde{m}_{l,k} = \tanh[(\Lambda_O(b_{l,k}) + \Lambda_I(b_{l,k}))/2]$ for $l = 1, 2,\ k = 1, \cdots, K$
  FOR $g = 1 : G$ (*Inner Iteration*)
    FOR $k = 1 : K$
      $\Lambda_I(b_{2,k}) = \frac{2}{\sigma^2}[(m_{1,k} + 2) \cdot \mathbf{h}_k^T \mathbf{r} - (m_{1,k} + 2) \cdot \boldsymbol{\beta}_k^T (\mathbf{m}_2 \circ \mathbf{m}_1 + 2\mathbf{m}_2)]$
      $m_{2,k} = \tanh[(\Lambda_O(b_{2,k}) + \Lambda_I(b_{2,k}))/2]$
    END
    FOR $k = 1 : K$
      $\Lambda_I(b_{1,k}) = \frac{2}{\sigma^2}[m_{2,k} \cdot \mathbf{h}_k^T \mathbf{r} - 2\boldsymbol{\rho}_k^T \mathbf{1} - m_{2,k} \cdot \boldsymbol{\beta}_k^T (\mathbf{m}_2 \circ \mathbf{m}_1 + 2\mathbf{m}_2)]$
      $m_{1,k} = \tanh[(\Lambda_O(b_{1,k}) + \Lambda_I(b_{1,k}))/2]$
    END
  END
  $\Lambda_O(b_{l,k}) \overset{\text{Decoding}}{\Longleftarrow} \Lambda_I(b_{l,k})$ for $l = 1, 2,\ k = 1, \cdots, K$ (*BCJR Decoding*)
END

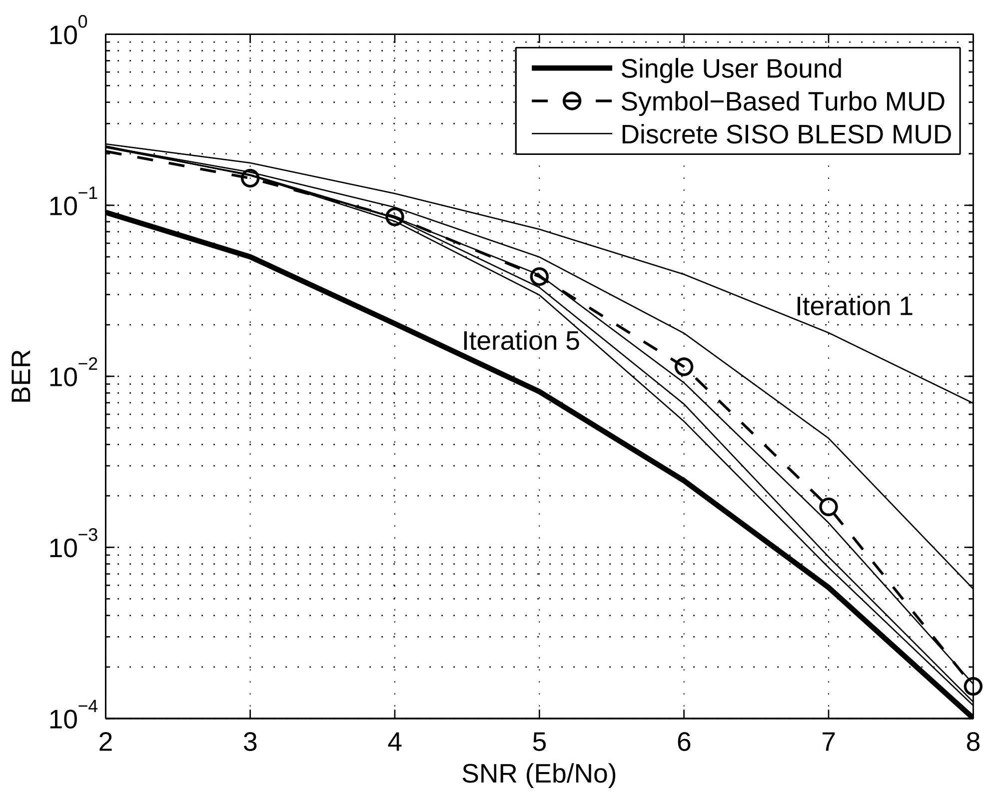

Fig. 1. BER performance of discrete-SISO BLESD and symbol-based turbo MUD for 16-QAM modulation ($K = 8$, $\rho = 0.5$).

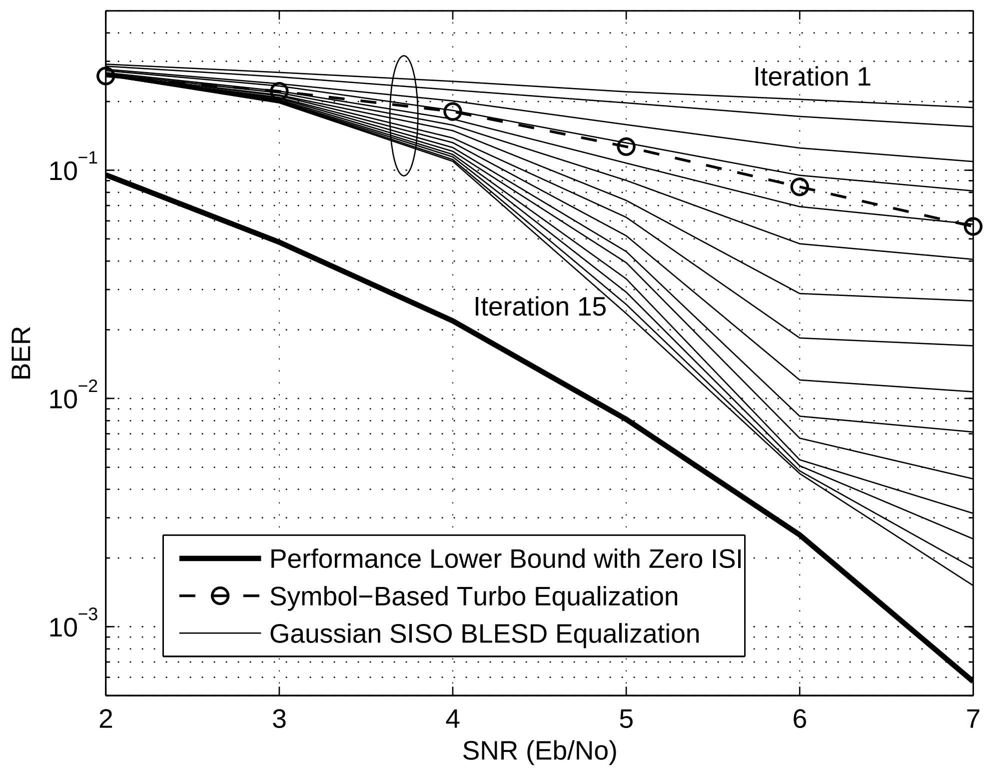

Fig. 2. BER performance of Gaussian-SISO BLESD and symbol-based turbo equalization for 16-QAM modulation.

perfect power control, i.e., $A_k = e^{j\phi_k}$, where ϕ_k is sampled from a uniform distribution, but is perfectly known at the receiver. It is seen that the discrete SISO BLESD receivers perform close to the single user bound (SUB). This is done amid both strong interference among users, as well as interaction between channel bits that make up each symbol. We use $G = 6$ inner iterations for each outer iteration. The symbol-based detection scheme in [15] was also implemented. Its performance after the fourth iteration (after which no further improvement is obtained) is plotted along with the proposed schemes. Under the assumed simulation settings, the discrete SISO BLESD detector demonstrates significantly reduced BER and faster convergence.

B. Turbo Equalization

The Gaussian SISO algorithm presented in Section IV-B needs to be modified to conform to the equalization setting. Appendix C shows how the conversion is done for 4-PAM/16-QAM modulated signals.

In Fig. 2, we compare the performance of the proposed BLESD equalizer with the symbol-level equalizer [15]. We assume a system employing 16-QAM modulation, with each packet containing 2048 information bits encoded by a rate $1/2$ convolutional code with generator 111 and 101. Similar to [14], the Porat-Friedlander channel [21] is chosen, which has $M+1=5$ complex taps, with CIR $\mathbf{h}=[2-0.4j, 1.5+1.8j, 1, 1.2-1.3j, 0.8+1.6j]^T$. We fix the detector window size to be $N_1=0$ and $N_2=9$ (We adopt the same notation as [15], using N_1 and N_2 to specify the sliding window size). In the BLESD equalizer, $G=4$ inner iterations are used for each outer iteration. Under this severe ISI channel, the symbol-based equalizer performs poorly (no improvement after three iterations). In contrast, the Gaussian SISO BLESD equalizer remains effective even in this channel condition.

It is worth mentioning that the symbol-based schemes are also capable of equalizing the Porat-Friedlander channel, as demonstrated in [22] and [23]. However, they require a much stronger code, a longer interleaver, and a larger window size. The proposed BLESD equalizer achieves a significant advantage in these aspects over the symbol-based alternatives at the cost of increased detector complexity, as multiple inner iterations are required to minimize the free energy.

VI. Conclusions

The BLESD method for SISO detection of M-QAM symbols opens up a new way for the receiver to handle the mapping between bits and symbols. The resulting algorithms perform data detection on the bit level, as opposed to prior work which translate symbol EXT to bit EXT through approximate low-complexity techniques that may introduce greater errors because of their two-level structure.

The view of Gray-coded data symbols as a multi-linear combination of transmitted bits has potentially far-reaching implications. For instance, a layered coding strategy may be implemented to offer non-uniform coding protection, to take account of the non-uniform weight each bit takes up in the QAM symbol. With the encoding optimized over the "layers" in which the bits belong, the achievable information rate may be improved considerably [24]. This may be realized with multilevel modulation coupled with the BLESD methodology, where encoding and detection are both performed in a layer-by-layer fashion.

Appendix A
Free Energy Evaluation of Gaussian SISO Detector

We first develop a few matrix identities involving the Schur product:

Lemma 1:

$$\mathsf{tr}[\mathsf{diag}(\mathbf{x}) \cdot \mathbf{A} \cdot \mathsf{diag}(\mathbf{y}) \cdot \mathbf{B}^T] = \mathbf{x}^T(\mathbf{A} \circ \mathbf{B})\mathbf{y} \tag{43}$$

for square matrices $\mathbf{A}$, $\mathbf{B} \in \mathbb{R}^{N\times N}$, and vectors $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^{N\times 1}$.

Proof: Writing $\mathbf{A} = [A_{ij}]$ and $\mathbf{B} = [B_{ij}]$, it is easily verified that both sides of the equation are equal to $\sum_{i,j} x_i A_{ij} B_{ij} y_j$. ■

Lemma 2:

$$\mathsf{tr}[\mathbf{A} \cdot (\mathbf{B} \circ \mathbf{C})^T] = \mathsf{tr}[(\mathbf{A} \circ \mathbf{B}) \cdot \mathbf{C}^T] \tag{44}$$

for square matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C} \in \mathbb{R}^{N \times N}$.

Proof: Writing $\mathbf{A} = [A_{ij}]$, $\mathbf{B} = [B_{ij}]$ and $\mathbf{C} = [C_{ij}]$, it is easily verified that both sides of the equation are equal to $\sum_{i,j} A_{ij} B_{ij} C_{ij}$. ■

Now we are ready to derive some quadratic expectation properties for Schur products of Gaussian random vectors.

Lemma 3: Consider independent Gaussian random vectors $\mathbf{b}_1, \cdots, \mathbf{b}_U \in \mathbb{R}^{K \times 1}$, each with a distribution $\mathbf{b}_l \sim \mathcal{N}(\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)$, $l = 1, \cdots, U$. Then, for a real symmetric matrix $\mathbf{C}$,

$$\mathsf{E}\left[(\textstyle\prod_{l=1}^{U} \mathbf{b}_l)^T \mathbf{C} (\prod_{l=1}^{U} \mathbf{b}_l)\right] = \mathsf{tr}\left[\textstyle\prod_{l=1}^{U} (\boldsymbol{\Sigma}_l + \boldsymbol{\mu}_l \boldsymbol{\mu}_l^T) \cdot \mathbf{C}\right]. \tag{45}$$

Proof: This identity may be proven by induction. It is easily verified that for $U = 1$,

$$\begin{aligned} \mathsf{E}\left[\mathbf{b}_1^T \mathbf{C} \mathbf{b}_1\right] &= \boldsymbol{\mu}_1^T \mathbf{C} \boldsymbol{\mu}_1 + \mathsf{tr}[\mathbf{C}\boldsymbol{\Sigma}_1] \\ &= \mathsf{tr}\left[(\boldsymbol{\Sigma}_1 + \boldsymbol{\mu}_1 \boldsymbol{\mu}_1^T) \cdot \mathbf{C}\right]. \end{aligned} \tag{46}$$

Assuming (45) is true for $U = u$, i.e.,

$$\mathsf{E}\left\{(\textstyle\prod_{l=1}^{u} \mathbf{b}_l)^T \mathbf{C} (\prod_{l=1}^{u} \mathbf{b}_l)\right\} = \mathsf{tr}\left[\textstyle\prod_{l=1}^{u} (\boldsymbol{\Sigma}_l + \boldsymbol{\mu}_l \boldsymbol{\mu}_l^T) \cdot \mathbf{C}\right], \tag{47}$$

then for $U = u + 1$, we have

$$\begin{aligned} & \mathsf{E}\left\{(\textstyle\prod_{l=1}^{u+1} \mathbf{b}_l)^T \mathbf{C} (\prod_{l=1}^{u+1} \mathbf{b}_l)\right\} \\ = \; & \mathsf{E}\left\{(\textstyle\prod_{l=1}^{u} \mathbf{b}_l)^T \left[\mathbf{B}_{u+1} \mathbf{C} \mathbf{B}_{u+1}\right] (\prod_{l=1}^{u} \mathbf{b}_l)\right\} \\ = \; & \mathsf{E}\left\{\mathsf{tr}\left[\textstyle\prod_{l=1}^{u} (\boldsymbol{\Sigma}_l + \boldsymbol{\mu}_l \boldsymbol{\mu}_l^T) \cdot \mathbf{B}_u \mathbf{C} \mathbf{B}_u\right]\right\} \\ \overset{L.1}{=} \; & \mathsf{E}\left\{\mathbf{b}_u^T \left[\textstyle\prod_{l=1}^{u} (\boldsymbol{\Sigma}_l + \boldsymbol{\mu}_l \boldsymbol{\mu}_l^T) \circ \mathbf{C}\right] \mathbf{b}_u\right\} \\ = \; & \mathsf{tr}\left\{\left[\textstyle\prod_{l=1}^{u} (\boldsymbol{\Sigma}_l + \boldsymbol{\mu}_l \boldsymbol{\mu}_l^T) \circ \mathbf{C}\right] \cdot \boldsymbol{\Sigma}_{u+1}\right\} + \boldsymbol{\mu}_{u+1}^T \left[\textstyle\prod_{l=1}^{u} (\boldsymbol{\Sigma}_l + \boldsymbol{\mu}_l \boldsymbol{\mu}_l^T) \circ \mathbf{C}\right] \boldsymbol{\mu}_{u+1} \\ = \; & \mathsf{tr}\left\{(\boldsymbol{\Sigma}_{u+1} + \boldsymbol{\mu}_{u+1} \boldsymbol{\mu}_{u+1}^T) \cdot \left[\textstyle\prod_{l=1}^{u} (\boldsymbol{\Sigma}_l + \boldsymbol{\mu}_l \boldsymbol{\mu}_l^T) \circ \mathbf{C}\right]\right\} \\ \overset{L.2}{=} \; & \mathsf{tr}\left\{\textstyle\prod_{l=1}^{u+1} (\boldsymbol{\Sigma}_l + \boldsymbol{\mu}_l \boldsymbol{\mu}_l^T) \cdot \mathbf{C}\right\} \end{aligned} \tag{48}$$

where $\mathbf{B}_l = \mathsf{diag}(\mathbf{b}_l)$. ■

Lemma 4: Consider independent Gaussian random vectors $\mathbf{b}_1, \cdots, \mathbf{b}_V \in \mathbb{R}^{K \times 1}$, each with a distribution $\mathbf{b}_l \sim \mathcal{N}(\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)$, $l = 1, \cdots, V$. Then, for a real symmetric matrix $\mathbf{C}$ and $U < V$,

$$\begin{aligned} \mathsf{E}\left[(\prod_{l=1}^{U} \mathbf{b}_l)^T \mathbf{C} (\prod_{l=1}^{V} \mathbf{b}_l)\right] &= \mathsf{tr}\left[\prod_{l=1}^{U} (\boldsymbol{\Sigma}_l + \boldsymbol{\mu}_l \boldsymbol{\mu}_l^T) \cdot [\mathbf{C}\, \mathsf{diag}(\prod_{l=u+1}^{V} \mathbf{u}_l)]\right] \\ &= \mathbf{1}^T \left[\prod_{l=1}^{U} (\boldsymbol{\Sigma}_l + \boldsymbol{\mu}_l \boldsymbol{\mu}_l^T) \circ \mathbf{C}\right] \prod_{l=u+1}^{V} \boldsymbol{\mu}_l. \end{aligned} \tag{49}$$

Proof: Utilizing Lemma 3 yields

$$\begin{aligned} & \mathsf{E}\left\{(\prod_{l=1}^{U} \mathbf{b}_l)^T \mathbf{C} \prod_{l=1}^{V} \mathbf{b}_l\right\} \\ \stackrel{L.3}{=} \; & \mathsf{E}\left\{\mathsf{tr}\left[\prod_{l=1}^{U} (\boldsymbol{\Sigma}_l + \boldsymbol{\mu}_l \boldsymbol{\mu}_l^T) \cdot \mathbf{C} \cdot \mathsf{diag}(\prod_{l=u+1}^{V} \mathbf{b}_l)\right]\right\} \\ = \; & \mathsf{tr}\left[\prod_{l=1}^{U} (\boldsymbol{\Sigma}_l + \boldsymbol{\mu}_l \boldsymbol{\mu}_l^T) \cdot \mathbf{I} \cdot \mathbf{C} \cdot \mathsf{diag}(\prod_{l=u+1}^{V} \boldsymbol{\mu}_l)\right] \\ \stackrel{L.1}{=} \; & \mathbf{1}^T \left[\prod_{l=1}^{U} (\boldsymbol{\Sigma}_l + \boldsymbol{\mu}_l \boldsymbol{\mu}_l^T) \circ \mathbf{C}\right] \prod_{l=u+1}^{V} \boldsymbol{\mu}_l. \end{aligned} \tag{50}$$

■

The three integrals in (20) can now be evaluated in closed form:

$$\begin{aligned} \int_{\mathbf{d}} Q(\mathbf{d}) \log Q(\mathbf{d}) d\mathbf{d} = \; & \sum_{l=1}^{L} \int_{\mathbf{b}_l} Q(\mathbf{b}_l) \log Q(\mathbf{b}_l) d\mathbf{b}_l \\ = \; & -\tfrac{1}{2} \sum_{l=1}^{L} \log |\boldsymbol{\Sigma}_l|. \end{aligned} \tag{51}$$

$$\begin{aligned} \int_{\mathbf{d}} Q(\mathbf{d}) \log p(\mathbf{d}) d\mathbf{d} = \; & \sum_{l=1}^{L} \int_{\mathbf{b}_l} Q(\mathbf{b}_l) \log p(\mathbf{b}_l) d\mathbf{b}_l \\ = \; & \sum_{l=1}^{L} \mathsf{E}\left[-\tfrac{1}{2}(\mathbf{b}_l - \tilde{\mathbf{b}}_l)^T \mathbf{W}_l^{-1} (\mathbf{b}_l - \tilde{\mathbf{b}}_l)\right] \\ = \; & -\sum_{l=1}^{L} \left[\tfrac{1}{2} \mathsf{tr}(\mathbf{W}_l^{-1} \boldsymbol{\Sigma}_l) + \tfrac{1}{2} \boldsymbol{\mu}_l^T \mathbf{W}_l^{-1} \boldsymbol{\mu}_l - \tilde{\mathbf{b}}_l^T \mathbf{W}_l^{-1} \boldsymbol{\mu}_l\right]. \end{aligned} \tag{52}$$

$$\begin{aligned} & \int_{\mathbf{d}} Q(\mathbf{d}) \log p(\mathbf{r}|\mathbf{d}) d\mathbf{d} \\ = \; & -\tfrac{1}{2\sigma^2} \left\{\mathsf{E}\left[(\sum_{l=1}^{L} 2^{l-1} \prod_{p=l}^{L} \mathbf{b}_p)^T \mathbf{H}^T \mathbf{H} (\sum_{l=1}^{L} 2^{l-1} \prod_{p=l}^{L} \mathbf{b}_p)\right] - \mathsf{E}\left[2\mathbf{r}^T \mathbf{H} (\sum_{l=1}^{L} 2^{l-1} \prod_{p=l}^{L} \mathbf{b}_p)\right]\right\} \\ = \; & -\tfrac{1}{2\sigma^2} \left\{\sum_{l=1}^{L} 2^{l-1} \mathsf{tr}\left[\prod_{p=l}^{L} (\boldsymbol{\Sigma}_p + \boldsymbol{\mu}_p \boldsymbol{\mu}_p^T) \dot{(} \mathbf{H}^T \mathbf{H})\right] + \sum_{1 \le i < j \le L} 2^{i+j-1} \mathsf{tr}\left[\prod_{p=j}^{L} (\boldsymbol{\Sigma}_p + \boldsymbol{\mu}_p \boldsymbol{\mu}_p^T) \cdot \mathbf{H}^T \mathbf{H} \cdot \mathsf{diag}(\prod_{p=i}^{j-1} \boldsymbol{\mu}_p)\right]\right\} \\ & + \tfrac{1}{2\sigma^2} \left\{2\mathbf{r}^T \mathbf{H} (\sum_{l=1}^{L} 2^{l-1} \prod_{p=l}^{L} \boldsymbol{\mu}_p)\right\} \end{aligned} \tag{53}$$

Combining (51), (52) and (53) yields the complete free energy expression in (21).

APPENDIX B
FREE ENERGY EVALUATION OF DISCRETE SISO DETECTOR

We require two identities related to the quadratic expectations of binary random vectors, as summarized in Lemmas 5 and 6 below.

Lemma 5: 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 real matrix $\mathbf{C}$,

$$\mathsf{E}\left[(\prod_{l=1}^{U} \mathbf{b}_l)^T \mathbf{C} (\prod_{l=1}^{U} \mathbf{b}_l)\right] = (\prod_{l=1}^{U} \mathbf{m}_l)^T [\mathbf{C} - \mathsf{diag}(\mathbf{C})] (\prod_{l=1}^{U} \mathbf{m}_l) + \mathbf{1}^T \mathsf{diag}(\mathbf{C}) \mathbf{1}. \tag{54}$$

Proof: This identity may be proven by induction. It is easily verified that for $U = 1$,

$$\mathsf{E}\left[\mathbf{b}_1^T \mathbf{C} \mathbf{b}_1\right] = \mathbf{m}_1^T[\mathbf{C} - \mathsf{diag}(\mathbf{C})]\mathbf{m}_1 + \mathbf{1}^T \mathsf{diag}(\mathbf{C})\mathbf{1}. \tag{55}$$

Assuming (54) is true for $U = u$, i.e.,

$$\begin{aligned} & \mathsf{E}\left\{(\not\prod_{l=1}^{u} \mathbf{b}_l)^T \mathbf{C} (\not\prod_{l=1}^{u} \mathbf{b}_l)\right\} \\ = \; & (\not\prod_{l=1}^{u} \mathbf{m}_l)^T \left[\mathbf{C} - \mathsf{diag}(\mathbf{C})\right] (\not\prod_{l=1}^{u} \mathbf{m}_l) + \mathbf{1}^T \mathsf{diag}(\mathbf{C})\mathbf{1}, \end{aligned} \tag{56}$$

then for $U = u + 1$, we have

$$\begin{aligned} & \mathsf{E}\left\{(\not\prod_{l=1}^{u+1} \mathbf{b}_l)^T \mathbf{C} (\not\prod_{l=1}^{u+1} \mathbf{b}_l)\right\} \\ = \; & \mathsf{E}\left\{(\not\prod_{l=1}^{u} \mathbf{m}_l)^T \left[\mathbf{B}_{u+1}\mathbf{C}\mathbf{B}_{u+1} - \mathsf{diag}(\mathbf{B}_{u+1}\mathbf{C}\mathbf{B}_{u+1})\right] (\not\prod_{l=1}^{u} \mathbf{m}_l) + \mathbf{1}^T \mathsf{diag}(\mathbf{B}_{u+1}\mathbf{C}\mathbf{B}_{u+1})\mathbf{1}\right\} \\ = \; & \mathsf{E}\left\{\mathbf{b}_{u+1}^T \left[(\prod_{l=1}^{u} \mathbf{M}_l)^T \mathbf{C} (\prod_{l=1}^{u} \mathbf{M}_l) - \mathsf{diag}((\prod_{l=1}^{u} \mathbf{M}_l)\mathbf{C}(\prod_{l=1}^{u} \mathbf{M}_l))\right] \mathbf{b}_{u+1} + \mathbf{b}_{u+1}^T \mathsf{diag}(\mathbf{C})\mathbf{b}_{u+1}\right\} \\ = \; & (\not\prod_{l=1}^{u+1} \mathbf{m}_l)^T \left[\mathbf{C} - \mathsf{diag}(\mathbf{C})\right] (\not\prod_{l=1}^{u+1} \mathbf{m}_l) + \mathbf{1}^T \mathsf{diag}(\mathbf{C})\mathbf{1}, \end{aligned} \tag{57}$$

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

Lemma 6: 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 real matrix $\mathbf{C}$ and $U < V$,

$$\mathsf{E}\left[(\not\prod_{l=1}^{U} \mathbf{b}_l)^T \mathbf{C} \not\prod_{l=1}^{V} \mathbf{b}_l\right] = (\not\prod_{l=1}^{U} \mathbf{m}_l)^T \left[\mathbf{C} - \mathsf{diag}(\mathbf{C})\right] \not\prod_{l=1}^{V} \mathbf{m}_l + \mathbf{1}^T \mathsf{diag}(\mathbf{C}) \not\prod_{l=U+1}^{V} \mathbf{m}_l. \tag{58}$$

Proof: Utilizing Lemma 5 yields

$$\begin{aligned} & \mathsf{E}\left\{(\not\prod_{l=1}^{U} \mathbf{b}_l)^T \mathbf{C} \not\prod_{l=1}^{V} \mathbf{b}_l\right\} \\ = \; & \mathsf{E}\left\{(\not\prod_{l=1}^{U} \mathbf{m}_l)^T \left[\mathbf{C} \prod_{l=U+1}^{V} \mathbf{B}_l - \mathsf{diag}(\mathbf{C} \prod_{l=U+1}^{V} \mathbf{B}_l)\right] \not\prod_{l=1}^{U} \mathbf{m}_l + \mathbf{1}^T \mathsf{diag}(\mathbf{C} \prod_{l=U+1}^{V} \mathbf{B}_l)\mathbf{1}\right\} \\ = \; & \mathsf{E}\left\{(\not\prod_{l=U+1}^{V} \mathbf{b}_l)^T \left[\mathbf{C} \prod_{l=1}^{U} \mathbf{M}_l - \mathsf{diag}(\mathbf{C} \prod_{l=1}^{U} \mathbf{M}_l)\right] (\not\prod_{l=U+1}^{V} \mathbf{b}_l) + \mathbf{1}^T \mathsf{diag}(\mathbf{C}) \not\prod_{l=U+1}^{V} \mathbf{b}_l\right\} \\ = \; & (\not\prod_{l=1}^{U} \mathbf{m}_l)^T \left[\mathbf{C} - \mathsf{diag}(\mathbf{C})\right] \not\prod_{l=1}^{V} \mathbf{m}_l + \mathbf{1}^T \mathsf{diag}(\mathbf{C}) \not\prod_{l=U+1}^{V} \mathbf{m}_l. \end{aligned} \tag{59}$$

■

The three integrals in (20) can now be evaluated in closed form:

$$\begin{aligned} \int_{\mathbf{d}} Q(\mathbf{d}) \log Q(\mathbf{d}) d\mathbf{d} = \; & \sum_{l=1}^{L} \int_{\mathbf{b}_l} Q(\mathbf{b}_l) \log Q(\mathbf{b}_l) d\mathbf{b}_l \\ = \; & \sum_{l=1}^{L} \sum_{k=1}^{K} \left\{ \frac{1+m_{l,k}}{2} \log \frac{1+m_{l,k}}{2} + \frac{1-m_{l,k}}{2} \log \frac{1-m_{l,k}}{2} \right\}. \end{aligned} \tag{60}$$

$$\begin{aligned} \int_{\mathbf{d}} Q(\mathbf{d}) \log p(\mathbf{d}) d\mathbf{d} = \; & \sum_{l=1}^{L} \int_{\mathbf{b}_l} Q(\mathbf{b}_l) \log p(\mathbf{b}_l) d\mathbf{b}_l \\ = \; & \sum_{l=1}^{L} \mathsf{E}\left[\sum_{k=1}^{K} \left\{ \frac{1+b_{l,k}}{2} \log \frac{1+\tilde{b}_{l,k}}{2} + \frac{1-b_{l,k}}{2} \log \frac{1-\tilde{b}_{l,k}}{2} \right\}\right] \\ = \; & \sum_{l=1}^{L} \sum_{k=1}^{K} \left\{ \frac{1+m_{l,k}}{2} \log \frac{1+\tilde{b}_{l,k}}{2} + \frac{1-m_{l,k}}{2} \log \frac{1-\tilde{b}_{l,k}}{2} \right\}. \end{aligned} \tag{61}$$

$$
\begin{aligned}
&\int_{\mathbf{d}} Q(\mathbf{d}) \log p(\mathbf{r}|\mathbf{d}) d\mathbf{d} \\
= \;& -\frac{1}{2\sigma^2}\left\{\mathsf{E}\left[(\textstyle\sum_{l=1}^{L} 2^{l-1} \prod_{p=l}^{L} \mathbf{b}_p)^T \mathbf{H}^T \mathbf{H} (\sum_{l=1}^{L} 2^{l-1} \prod_{p=l}^{L} \mathbf{b}_p)\right] - \mathsf{E}\left[2\mathbf{r}^T \mathbf{H} (\textstyle\sum_{l=1}^{L} 2^{l-1} \prod_{p=l}^{L} \mathbf{b}_p)\right]\right\} \\
= \;& -\frac{1}{2\sigma^2}\left\{(\textstyle\sum_{l=1}^{L} 2^{l-1} \prod_{p=l}^{L} \mathbf{m}_p)^T [\mathbf{H}^T\mathbf{H} - \mathsf{diag}(\mathbf{H}^T\mathbf{H})] (\sum_{l=1}^{L} 2^{l-1} \prod_{p=l}^{L} \mathbf{m}_p)\right\} \\
& -\frac{1}{2\sigma^2}\left\{\mathbf{1}^T \mathsf{diag}(\mathbf{H}^T\mathbf{H}) (\textstyle\sum_{0<i\le j<L} 2^{i+j} \prod_{p=i}^{j} \mathbf{m}_p) - 2\mathbf{r}^T\mathbf{H}(\sum_{l=1}^{L} 2^{l-1} \prod_{p=l}^{L} \mathbf{m}_p)\right\}
\end{aligned} \tag{62}
$$

Combining (60), (61) and (62) yields the complete free energy expression in (31).

Appendix C
Gaussian SISO Equalizer of 4-PAM/16-QAM Modulation

Consider the expressions for $Q(\mathbf{b}_l)$, $l = 1, 2$, obtained in Section IV-B4. Taking $Q(\mathbf{b}_2)$ for example:

$$\boldsymbol{\mu}_2 = \tilde{\mathbf{b}}_2 + [\mathbf{R}_2 + \sigma^2 \mathbf{W}_2^{-1}]^{-1} [\mathbf{H}_2^T \mathbf{r} - \mathbf{R}_2 \tilde{\mathbf{b}}_2] \tag{63}$$

$$\boldsymbol{\Sigma}_2 = [\sigma^{-2} \mathbf{R}_2 + \mathbf{W}_2^{-1}]^{-1}, \tag{64}$$

where $\mathbf{H}_2 = \mathsf{diag}(\boldsymbol{\mu}_1 + 2\mathbf{1}) \cdot \mathbf{H}$ and $\mathbf{R}_2 = [\boldsymbol{\Sigma}_1 + (\boldsymbol{\mu}_1 + 2\mathbf{1})(\boldsymbol{\mu}_1 + 2\mathbf{1})^T] \circ (\mathbf{H}^T\mathbf{H})$. However, since $\mathbf{W}_l = \mathsf{diag}([1 - \tilde{b}_{l,1}^2, \cdots, 1 - \tilde{b}_{l,K}^2]^T)$ may become rank-deficient as the turbo iterations converge ($\tilde{b}_{l,k} \to \pm 1$), (63) and (64) need to be converted to a more suitable form that does not require the inversion of $\mathbf{W}_l$.

To proceed, we first notice that $\boldsymbol{\Sigma}_l$ can be approximated as a diagonal matrix. This is true under a mean-field approximation assuming the independence of $Q(b_{l,k})$, $l = 1, 2$, $k = 1, \cdots, K$. Writing $\boldsymbol{\Sigma}_1 = \mathsf{diag}([\sigma_{b_{1,1}}^2, \cdots, \sigma_{b_{1,K}}^2]^T)$, we have

$$
\begin{aligned}
\mathbf{R}_2 + \sigma^2 \mathbf{W}_2^{-1} &= [\boldsymbol{\Sigma}_1 + (\boldsymbol{\mu}_1 + 2\mathbf{1})(\boldsymbol{\mu}_1 + 2\mathbf{1})^T] \circ (\mathbf{H}^T\mathbf{H}) + \sigma^2 \mathbf{W}_2^{-1} \\
&= \mathbf{H}_2^T \mathbf{H}_2 + \boldsymbol{\Sigma}_1 \circ (\mathbf{H}^T\mathbf{H}) + \sigma^2 \mathbf{W}_2^{-1} \\
&= \mathbf{H}_2^T \mathbf{H}_2 + \sigma^2 \tilde{\mathbf{W}}_2^{-1},
\end{aligned} \tag{65}
$$

where the diagonal matrix $\tilde{\mathbf{W}}_2 = \left(\mathbf{I} + \sigma^{-2} \mathbf{W}_2 [\boldsymbol{\Sigma}_1 \circ (\mathbf{H}^T\mathbf{H})]\right)^{-1} \mathbf{W}_2$. The derivation makes use of the fact that $\boldsymbol{\Sigma}_1$ is a diagonal matrix, as well as the identity $[(\boldsymbol{\mu}_1 + 2\mathbf{1})(\boldsymbol{\mu}_1 + 2\mathbf{1})^T] \circ (\mathbf{H}^T\mathbf{H}) = \mathbf{H}_2^T \mathbf{H}_2$.

Now, recognizing that (63) and (64) can be rewritten as

$$\boldsymbol{\mu}_2 = \tilde{\mathbf{b}}_2 + [\mathbf{H}_2^T \mathbf{H}_2 + \sigma^2 \tilde{\mathbf{W}}_2^{-1}]^{-1} [\mathbf{H}_2^T \mathbf{r} - \mathbf{R}_2 \tilde{\mathbf{b}}_2] \tag{66}$$

$$\boldsymbol{\Sigma}_2 = [\sigma^{-2} \mathbf{H}_2^T \mathbf{H}_2 + \tilde{\mathbf{W}}_2^{-1}]^{-1}, \tag{67}$$

we may use the matrix inversion lemma to evaluate $[\mathbf{H}_2^T \mathbf{H}_2 + \sigma^2 \tilde{\mathbf{W}}_2^{-1}]^{-1}$, and obtain

$$\boldsymbol{\mu}_2 = \tilde{\mathbf{b}}_2 + \sigma^{-2} \tilde{\mathbf{W}}_2 \left[\mathbf{I} - \mathbf{H}_2^T (\mathbf{H}_2 \tilde{\mathbf{W}}_2 \mathbf{H}_2^T + \sigma^2 \mathbf{I})^{-1} \mathbf{H}_2 \tilde{\mathbf{W}}_2\right] (\mathbf{H}_2^T \mathbf{r} - \mathbf{R}_2 \tilde{\mathbf{b}}_2) \tag{68}$$

$$\boldsymbol{\Sigma}_2 = \tilde{\mathbf{W}}_2 - \tilde{\mathbf{W}}_2 \mathbf{H}_2^T [\mathbf{H}_2 \tilde{\mathbf{W}}_2 \mathbf{H}_2^T + \sigma^2 \mathbf{I}]^{-1} \mathbf{H}_2 \tilde{\mathbf{W}}_2. \tag{69}$$

It can be subsequently shown that the detector output LLR of $b_{2,n}$ is $\Lambda_I(b_{2,n}) = \frac{2\check{\mu}_{2,n}}{1-\check{\alpha}_{2,n}}$, where

$$\check{\mu}_{2,n} = \sigma^{-2}\mathbf{e}_n^T[\mathbf{I} - \mathbf{H}_2^T(\mathbf{H}_2\tilde{\mathbf{W}}_2\mathbf{H}_2^T + \sigma^2\mathbf{I})^{-1}\mathbf{H}_2\tilde{\mathbf{W}}_2] \cdot [\mathbf{H}_2^T\mathbf{r} - \mathbf{R}_2\tilde{\mathbf{b}}_{2,n}] \tag{70}$$

$$\check{\alpha}_{2,n} = [\mathbf{H}_2^T(\mathbf{H}_2\tilde{\mathbf{W}}_2\mathbf{H}_2^T + \sigma^2\mathbf{I})^{-1}\mathbf{H}_2\tilde{\mathbf{W}}_2]_{n,n}. \tag{71}$$

In (70), $\tilde{\mathbf{b}}_{2,n} = [\tilde{b}_{2,1}, \cdots, \tilde{b}_{2,n-1}, 0, \tilde{b}_{2,n+1}, \cdots, \tilde{b}_{2,K}]^T$.

Likewise, the LLR of $b_{1,n}$, $\Lambda_I(b_{1,n}) = \frac{2\check{\mu}_{1,n}}{1-\check{\alpha}_{1,n}}$ is specified by

$$\check{\mu}_{1,n} = \sigma^{-2}\mathbf{e}_n^T[\mathbf{I} - \mathbf{H}_1^T(\mathbf{H}_1\tilde{\mathbf{W}}_1\mathbf{H}_1^T + \sigma^2\mathbf{I})^{-1}\mathbf{H}_1\tilde{\mathbf{W}}_1] \cdot [\mathbf{H}_1^T\mathbf{r} - \mathbf{R}_1(\tilde{\mathbf{b}}_{1,n} + 2\mathbf{1})] \tag{72}$$

$$\check{\alpha}_{1,n} = [\mathbf{H}_1^T(\mathbf{H}_1\tilde{\mathbf{W}}_1\mathbf{H}_1^T + \sigma^2\mathbf{I})^{-1}\mathbf{H}_1\tilde{\mathbf{W}}_1]_{n,n}. \tag{73}$$

REFERENCES

[1] C. Berrou, A. Glavieux, and P. Titmajshima, "Near Shannon limit error-correction coding and decoding: Turbo codes," in *Proc. IEEE Int. Conf. Commun. (ICC'93)*, May 1993, pp. 1064–1070.

[2] 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.

[3] 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.

[4] C. Douillard, A. Picart, P. Didier, M. Jézéquel, C. Berrou, and A. Glavieux, "Iterative correction of intersymbol interference: Turbo equalization," *Euro. Trans. Telecommun.*, vol. 6, pp. 507–511, Sept. 1995.

[5] M. Tüchler, A. C. Singer, and R. Koetter, "Turbo equalization: Principles and new results," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 754–767, May 2002.

[6] S. L. Ariyavisitakul, "Turbo space-time processing to improve wireless channel capacity," *IEEE Trans. Commun.*, vol. 48, no. 8, pp. 1347–1359, Aug. 2000.

[7] M. Sellathurai and S. Haykin, "Turbo-BLAST for wireless communications: Theory and experiments," *IEEE Trans. Signal Processing*, vol. 50, no. 10, pp. 2538–2546, Oct. 2002.

[8] M. Kobayashi, J. Boutros, and G. Caire, "Successive interference cancellation with SISO decoding and EM channel estimation," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 8, pp. 1450–1460, Aug. 2001.

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

[10] D. D. Lin and T. J. Lim, "A variational inference framework for soft-in-soft-out detection in multiple access channels," *submitted to IEEE Trans. Inf. Theory*.

[11] M. I. Jordan, Z. Ghahramani, T. Jaakkola, and L. K. Saul, "An introduction to variational methods for graphical models," *Machine Learning*, vol. 37, no. 2, pp. 183–233, 1999.

[12] M. Nissilä and S. Pasupathy, "Soft-input soft-output equalizers for turbo receivers : A statistical physics perspective," *IEEE Trans. Commun.*, vol. 55, no. 7, pp. 1300–1307, July 2007.

[13] 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.

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

[15] A. Dejonghe and L. Vandendorpe, "Turbo-equalization for multilevel modulation: An efficient low-complexity scheme," in *Proc. IEEE Int. Conf. Communications (ICC'02)*, vol. 3, Apr. 2002, pp. 1863– 1867.

[16] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.

[17] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory*, vol. IT-20, pp. 284–287, Mar. 1974.

[18] D. D. Lin, "Variational inference methods for signal processing in wireless communications," Ph.D. dissertation, Univ. of Toronto, Toronto, Canada, Oct. 2007. [Online]. Available: http://individual.utoronto.ca/darryl/file/phd_thesis.pdf

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

[20] A. Duel-Hallen, "Decorrelating decision-feedback multiuser detector for synchronous code-division multiple-access channel," *IEEE Trans. Commun.*, vol. 41, no. 2, pp. 285–290, Feb. 1993.

[21] B. Porat and B. Friedlander, "Blind equalization of digital communication channels using high-order moments," *IEEE Trans. Signal Processing*, vol. 39, no. 2, pp. 522–526, Feb. 1991.

[22] A. Glavieux, C. Laot, and J. Labat, "Turbo equalization over a frequency selective channel," in *Proc. Int. Symp. on Turbo Codes*, Brest, France, Sept. 1997.

[23] F. Volgelbrush, R. Zukunft, and S. Haar, "16-QAM turbo equalization based on minimum mean squared error linear equalization," *Thirty-Sixth Asilomar Conference on Signals, Systems and Computers*, vol. 2, pp. 1943–1947, Nov. 2002.

[24] S. Shamai (Shitz) and A. Steiner, "A broadcast approach for a single user slowly fading MIMO channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2617–2635, Oct. 2003.

Darryl Dexu Lin received the B.A.Sc. degree (with Honours) in Engineering Science from the University of Toronto, Toronto, Canada, in 2001. At the same university, he received the M.A.Sc. and Ph.D. degree in Electrical Engineering in 2003 and 2007, respectively.

He is currently a senior systems engineer at Qualcomm Incorporated, San Diego, California. His research interests include probabilistic inference and its applications in wireless communications.

Teng Joon Lim (S'92-M'95-SM'02) joined the Department of Electrical & Computer Engineering of the University of Toronto in December 2000, where he is currently an Associate Professor. In the five years prior to that, he was a Member of Technical Staff at the Centre for Wireless Communications in Singapore, serving as the leader of the digital communications and signal processing group. He obtained the B.Eng. degree from the National University of Singapore in 1992, and the Ph.D. degree from the University of Cambridge in 1996.

His research interests are in wireless transceiver design, in particular multi-user detection, OFDM and OFDMA receiver structures, MIMO techniques, precoding in downlink channels, and cross-layer aspects of cooperative network design, and he has published widely in these areas. He is a senior member of the IEEE, and contributes regularly in organizing conferences, serving on technical program committees, and organizing seminars for the IEEE Toronto Communications Chapter.