Turbo Equalization for Gray-Coded M-ary QAM with Bit-Level Soft Decisions

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*— This paper investigates the soft-in soft-out (SISO) equalization of multilevel QAM symbols in coded inter-symbol interference (ISI) channels. Unlike the conventional approach of performing equalization at the symbol level, the proposed scheme targets the channel bits directly. This solution can be seen to belong to a family of SISO detection schemes which we call *Bit-Level Equalization and Soft Detection* (BLESD), stemming from the minimization of variational free energy given different postulates about the prior and posterior distributions of the channel bits. Simulation results demonstrate that the bit-level approach outperforms the symbol-level alternative in terms of error rate in the Porat-Friedlander channel.**

I. Introduction

Efficient data detection and error control code (ECC) decoding are essential in future wireless communication systems, which require high spectral efficiency, low power consumption, and variable data rates. Utilizing the turbo principle, the seminal work of [1], [2], among others, introduced a new philosophy for receiver design, in which the detector and decoder exchange soft information in an iterative manner, resulting in dramatically improved bit-error-rate (BER) performance without substantial complexity increase. This design methodology results in turbo multiuser detectors, turbo equalizers, or turbo MIMO (multiple-input multiple-output) receivers in various scenarios.

In this paper, we focus our attention on the turbo equalization of inter-symbol interference (ISI) channels. The developed algorithms may be readily extended to turbo multiuser detection and turbo MIMO equalization settings, since all three scenarios may be described by a common Gaussian linear channel model:

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

where $\mathbf{r} \in \mathbb{R}^{N\times 1}$ is the received signal, $\mathbf{H} \in \mathbb{R}^{N\times K}$ is the channel matrix, $\mathbf{b} = [b_1, \cdots, b_K]^T \in \mathbb{R}^{K\times 1}$ represents the transmitted channel bits, and $\mathbf{n} \in \mathbb{R}^{N\times 1}$ is a white Gaussian noise vector with distribution $p(\mathbf{n}) = \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$.

A key component in the turbo receiver is a soft-in soft-out (SISO) detector/equalizer, which computes a low-complexity approximation to the log-likelihood ratio (LLR) of $\{b_k\}_{k=1}^K$:

$$\Lambda_I(b_k) = \log \frac{p(\mathbf{r}|b_k = 1)}{p(\mathbf{r}|b_k = -1)}. \tag{2}$$

Exact computation of (2) requires a complexity exponential in K and thus the optimal turbo receiver does not scale well with the number of interfering symbols. In (2), $\Lambda_I(b_k)$ denotes the output extrinsic information (EXT) of the *Inner code* (the ISI channel), as opposed to the *Outer code* (the channel code).

The seminal works by Wang and Poor [2], and Tüchler, Singer and Koetter [3] successfully used the simple minimum mean-squared error (MMSE) principle in the design of suboptimal SISO detectors. On the other hand, [4] and [5] proposed powerful turbo multiuser detectors using, respectively, parallel and successive interference cancellation schemes as the detector component.

To provide a comprehensive theory guiding the design of practical SISO detectors mentioned above, in [6], [7] we proposed a general framework, adopting the machine learning concept of *variational inference* [8], which illuminates the commonalities in many if not all practical turbo receivers. This theory describes SISO detectors as special cases of the variational inference framework obtained from minimizing the "free energy" expressions corresponding to various postulates about the prior and posterior distributions of the channel bits. To understand this framework in simple terms, let us consider the approximation of $p(\mathbf{b}|\mathbf{r})$. Since the direct search for the maximum of $p(\mathbf{b}|\mathbf{r})$ entails exponential complexity, we seek to obtain its closest estimate in terms of minimal Kullback-Leibler (KL) divergence. Let $Q(\mathbf{b})$ represent our estimate of $p(\mathbf{b}|\mathbf{r})$. The *variational free energy* is the KL divergence between $Q(\mathbf{b})$ and $p(\mathbf{b}|\mathbf{r})$ up to an additive constant:

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

where $\lambda = \{\lambda_1, \cdots, \lambda_K\}$ contains parameters that specify $Q(\mathbf{b})$. In (3), $p(\mathbf{b})$ 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 [2], [3] (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 [4], [5] (a.k.a. discrete SISO detectors) when $\mathcal{F}$ is minimized iteratively.

In this paper, we investigate another application of the variational inference view for SISO detection. We will extend the commonly-assumed BPSK model to the realm of bit-interleaved coded modulation (BICM) [9]. The proposed solution, called *Bit-Level Equalization and Soft Detection* (BLESD), was first introduced in [10] in the multiuser detection context. As its name suggests, the BLESD scheme performs detection at the bit level, even when Gray-coded M-ary symbols are transmitted, contrary to the conventional procedure for handling multilevel modulation, where decisions are first made at the symbol level. However, only the discrete SISO detector is developed in [10], which reduces $\mathcal{F}$ through successive interference cancellation. This paper derives the Gaussian SISO version of the BLESD scheme, in order to produce an MMSE-type detector comparable to existing turbo equalizers. A journal paper offering a comprehensive discussion covering the content of both [10] and this paper is also available [11].

Prior to this work, MMSE-based symbol-level turbo equalization schemes for multilevel modulation already exist. In [12], the BPSK turbo equalizer is extended to M-PSK modulation with ease by exploiting the uniform-symbol-energy property of M-PSK symbols. However, the same technique does not apply to general QAM modulations. To overcome this difficulty, Dejonghe and Vandendorpe proposed a more general solution in [13], enabling SISO equalization at the symbol level for arbitrary multilevel modulation schemes, while allowing *extrinsic information* (EXT) to be obtained for each channel bit. We will use [13] as the benchmark to be compared against our proposed scheme in Section IV.

II. Signal Model

For simplicity of notation, in the rest of the paper we will only consider a real-valued channel model, containing pulse-amplitude modulated (PAM) symbols. This is viable because extensions to complex QAM symbols follow simply through a transformation that doubles the signal dimension.

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 $\{d_t\}_{t=1}^{T_{end}}$. In the interest of limiting the processing delay, we will take the sliding window approach at each time instance t, and define

$$\begin{array}{rcl}\mathbf{r}_t & = & [r_{t-N_1}, \cdots, r_t, \cdots, r_{t+N_2}]^T \in \mathbb{R}^{N\times 1},\\ \mathbf{d}_t & = & [d_{t-N_1-M}, \cdots, d_t, \cdots, d_{t+N_2}]^T \in \mathbb{R}^{K\times 1},\\ \mathbf{n}_t & = & [n_{t-N_1}, \cdots, n_t, \cdots, n_{t+N_2}]^T \in \mathbb{R}^{N\times 1}.\end{array} \tag{4}$$

We can easily show that the received signal can be written in a matrix form $\mathbf{r}_t = \mathbf{H}\mathbf{d}_t + \mathbf{n}_t$, where

$$\mathbf{H} \triangleq \begin{bmatrix} h_M & \cdots & h_0 & 0 & \cdots & \cdots & 0 \\ 0 & h_M & \cdots & h_0 & 0 & \cdots & 0 \\ & & \ddots & & \ddots & & \\ 0 & \cdots & \cdots & 0 & h_M & \cdots & h_0 \end{bmatrix} \in \mathbb{C}^{N\times K}. \tag{5}$$

In (5), $N = N_1+N_2+1$ is the equalizer length, and $K = N+M$. The sliding window channel model requires the window to be shifted once every symbol interval. Thus our goal is, for each window $\mathbf{r}_t$, to estimate the n-th element of $\mathbf{d}_t$, where $n = N_1 + M + 1$. To derive the general equations to achieve this, we may drop the subscript t and obtain:

$$\mathbf{r} = \mathbf{H}\mathbf{d} + \mathbf{n}. \tag{6}$$

In (6), d_k is a result of Gray mapping of L information bits $\{b_{l,k}\}_{l=1}^L$. It is proven in [10] that the nonlinear Gray mapping between d_k and $\{b_{l,k}\}_{l=1}^L$ can be written analytically as a multi-linear function:

$$d_k = \textstyle\sum_{l=1}^L 2^{l-1} b_{L,k} b_{L-1,k} \cdots b_{l,k} = \sum_{l=1}^L 2^{l-1} \prod_{p=l}^L b_{p,k}, \tag{7}$$

where $b_{l,k} \in \{-1, +1\}$. For instance, consider a 4-PAM symbol d_k that consists of two information bits, $b_{1,k}, b_{2,k} \in \{1, -1\}$. If the value of d_k is determined by the equation $d_k = b_{2,k}(2 + b_{1,k}) = b_{2,k}b_{1,k} + 2b_{2,k}$, then such a bit-to-symbol mapping is a Gray mapping, because the four values that d_k takes on, -3, -1, 1 and 3, correspond to $(b_{2,k}, b_{1,k})$ pairs $(-1, 1)$, $(-1, -1)$, $(1, -1)$ and $(1, 1)$. This implies that our channel models may now be written in terms of $b_{l,k}$, instead of d_k. It is then possible to design equalizers for $b_{l,k}$ directly, rather than d_k.

III. Gaussian SISO Equalizer for 2^L-PAM

As emphasized in [7], the most important steps in deriving variational-inference-based SISO detectors may be summarized as the following variational free energy minimization (VFEM) routine:

1) Postulate distributions for $p(\mathbf{d})$, $p(\mathbf{r}|\mathbf{d})$ and $Q(\mathbf{d})$;
2) Derive an analytical expression for $\mathcal{F}(\lambda_1, \cdots, \lambda_K)$;
3) Minimize $\mathcal{F}(\lambda_1, \cdots, \lambda_K)$ over $\{\lambda_k\}_{k=1}^K$.

Since $\mathbf{d}$ is a known function of $\mathbf{b}$ through the multi-linear function (7), we can also view the distributions in step (1) as functions of $\mathbf{b}$, which means that in principle we can infer $\mathbf{b}$ without first inferring $\mathbf{d}$. This bit-inference idea is central to this paper, and will be explained presently.

A. Variational Free Energy Minimization

1) Postulated Distributions: Similar to the BPSK case [6], we make the following postulates for the 2^L-PAM signals:

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

$$\begin{array}{rcl} p(\mathbf{d}) & = & \prod_{l=1}^L p(\mathbf{b}_l) \\ & = & \prod_{l=1}^L \mathcal{N}(\tilde{\mathbf{b}}_l, \mathbf{W}_l), \end{array} \tag{8}$$

where $\tilde{\mathbf{b}}_l = [\tilde{b}_{l,1}, \cdots, \tilde{b}_{l,K}]^T$ represents the mean estimates from the APP decoder of the l-th channel bits of all users. $\mathbf{W}_l = \mathrm{diag}([1-\tilde{b}_{l,1}^2, \cdots, 1-\tilde{b}_{l,K}^2]^T)$ is the covariance matrix.

Channel Transition Distribution: The channel transition distribution or likelihood function is $p(\mathbf{r}|\mathbf{d}) = \mathcal{N}(\mathbf{H}\mathbf{d}, \sigma^2\mathbf{I})$. The multi-linear bit-to-symbol mapping in (7) ensures that the transition distribution may be written in terms of the channel bits. Recognizing $\mathbf{d} = \sum_{l=1}^L 2^{l-1} \prod_{p=l}^L \mathbf{b}_p$, then

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

where the notation $\prod$ represents a series of Schur (element-wise) products, i.e. $\prod_{p=l}^L \mathbf{b}_p = \mathbf{b}_l \circ \mathbf{b}_{l+1} \circ \cdots \circ \mathbf{b}_L$. In other words, we place the l-th bit of all symbols in one vector $\mathbf{b}_l$,

$$
\begin{aligned}
&\mathcal{F}(\boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_L, \boldsymbol{\Sigma}_1, \cdots, \boldsymbol{\Sigma}_L) \\
&= \sum_{l=1}^{L} \Big[-\tfrac{1}{2} \sum_{l=1}^{L} \log |\boldsymbol{\Sigma}_l| + \tfrac{1}{2} \mathrm{tr}(\mathbf{W}_l^{-1} \boldsymbol{\Sigma}_l) + \tfrac{1}{2} \boldsymbol{\mu}_l^T \mathbf{W}_l^{-1} \boldsymbol{\mu}_l - \tilde{\mathbf{b}}_l^T \mathbf{W}_l^{-1} \boldsymbol{\mu}_l \Big] + \tfrac{1}{2\sigma^2} \Big\{ \sum_{l=1}^{L} 2^{l-1} \mathrm{tr} \Big[\bigodot_{p=l}^{L} (\boldsymbol{\Sigma}_p + \boldsymbol{\mu}_p \boldsymbol{\mu}_p^T) \cdot (\mathbf{H}^T \mathbf{H}) \Big] \\
&\quad + \sum_{1 \le i < j \le L} 2^{i+j-1} \mathrm{tr} \Big[\bigodot_{p=j}^{L} (\boldsymbol{\Sigma}_p + \boldsymbol{\mu}_p \boldsymbol{\mu}_p^T) \cdot \mathbf{H}^T \mathbf{H} \cdot \mathrm{diag}(\bigodot_{p=i}^{j-1} \boldsymbol{\mu}_p) \Big] \Big\} - \tfrac{1}{2\sigma^2} \Big\{ 2\mathbf{r}^T \mathbf{H} (\sum_{l=1}^{L} 2^{l-1} \bigodot_{p=l}^{L} \boldsymbol{\mu}_p) \Big\}
\end{aligned}
\tag{12}
$$

$$
\begin{cases}
\boldsymbol{\phi}_l = \sum_{1 \le i \le l < j \le L} 2^{i+j-2} \Big\{ \mathrm{diag}(\bigodot_{p=i, p \ne l}^{j-1} \boldsymbol{\mu}_p) \Big[(\mathbf{H}^T \mathbf{H}) \circ \bigodot_{p=j}^{L} (\boldsymbol{\Sigma}_p + \boldsymbol{\mu}_p \boldsymbol{\mu}_p^T) \Big] \mathbf{1} + I(i = l = j-1) \Big[(\mathbf{H}^T \mathbf{H}) \circ \bigodot_{p=j}^{L} (\boldsymbol{\Sigma}_p + \boldsymbol{\mu}_p \boldsymbol{\mu}_p^T) \Big] \mathbf{1} \Big\} \\
\boldsymbol{\Psi}_l = \sum_{i=1}^{l} \Big\{ 2^{2i-2} \Big[(\mathbf{H}^T \mathbf{H}) \circ \bigodot_{p=i, p \ne l}^{L} (\boldsymbol{\Sigma}_p + \boldsymbol{\mu}_p \boldsymbol{\mu}_p^T) \Big] \Big\} + I(l = L) \cdot 2^{2l-2} \mathbf{H}^T \mathbf{H} \\
\boldsymbol{\Xi}_l = \sum_{1 \le i < j \le l, i \ne j} 2^{i+j-2} \Big\{ I(j = L) \cdot \Big[\mathbf{H}^T \mathbf{H} \cdot \mathrm{diag}(\bigodot_{p=i}^{j-1} \boldsymbol{\mu}_p) \Big] + I(j = L) \cdot \Big[\mathbf{H}^T \mathbf{H} \cdot \mathrm{diag}(\bigodot_{p=i}^{j-1} \boldsymbol{\mu}_p) \Big]^T \\
\qquad + \Big[\mathbf{H}^T \mathbf{H} \cdot \mathrm{diag}(\bigodot_{p=i}^{j-1} \boldsymbol{\mu}_p) \circ \bigodot_{p=j, p \ne l}^{L} (\boldsymbol{\Sigma}_p + \boldsymbol{\mu}_p \boldsymbol{\mu}_p^T) \Big] + \Big[\mathbf{H}^T \mathbf{H} \cdot \mathrm{diag}(\bigodot_{p=i}^{j-1} \boldsymbol{\mu}_p) \circ \bigodot_{p=j, p \ne l}^{L} (\boldsymbol{\Sigma}_p + \boldsymbol{\mu}_p \boldsymbol{\mu}_p^T) \Big]^T \Big\} \\
\boldsymbol{\Omega}_l = \sum_{i=1}^{l} \Big\{ 2^{i-1} \mathrm{diag}(\bigodot_{p=i, p \ne l}^{L} \boldsymbol{\mu}_p) \Big\} + I(l = L) \cdot 2^{l-1} \mathbf{I}
\end{cases}
\tag{14}
$$

and eventually perform joint equalization not only among K symbols, but also among the L bits in each symbol.

Posterior Distribution: 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 $\{\mathbf{b}_l\}_{l=1}^{L}$ given channel observations. We thus have

$$
\begin{aligned}
Q(\mathbf{d}) &= \prod_{l=1}^{L} Q(\mathbf{b}_l) \\
&= \prod_{l=1}^{L} \mathcal{N}(\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l).
\end{aligned}
\tag{10}
$$

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

$$
\begin{aligned}
\mathcal{F} &= \int_{\mathbf{d}} Q(\mathbf{d}) \log Q(\mathbf{d}) d\mathbf{d} - \int_{\mathbf{d}} Q(\mathbf{d}) \log p(\mathbf{r}|\mathbf{d}) d\mathbf{d} \\
&\quad - \int_{\mathbf{d}} Q(\mathbf{d}) \log p(\mathbf{d}) d\mathbf{d} \\
&= \mathsf{E}\left[\log Q(\mathbf{d})\right] - \mathsf{E}\left[\log p(\mathbf{r}|\mathbf{d})\right] - \mathsf{E}\left[\log p(\mathbf{d})\right].
\end{aligned}
\tag{11}
$$

The task of free energy evaluation then condenses to the computation of the integral expressions in (11) given $p(\mathbf{d})$, $p(\mathbf{r}|\mathbf{d})$ and $Q(\mathbf{d})$ defined in (8), (9) and (10), 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 I. The final free energy expression for Gaussian SISO BLESD is assembled in (12). While it is sufficient to work with (12) directly to arrive at desired Gaussian SISO equalizers, readers are encouraged to refer to Appendix I for additional insights.

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

$$
\begin{aligned}
\boldsymbol{\mu}_l &= \tilde{\mathbf{b}}_l + \left(\boldsymbol{\Psi}_l + \boldsymbol{\Xi}_l + \sigma^2 \mathbf{W}_l^{-1}\right)^{-1} \Big[\boldsymbol{\Omega}_l \mathbf{H}^T \mathbf{r} - \boldsymbol{\phi}_l - (\boldsymbol{\Psi}_l + \boldsymbol{\Xi}_l) \tilde{\mathbf{b}}_l \Big] \\
\boldsymbol{\Sigma}_l &= \left(\sigma^{-2} \boldsymbol{\Psi}_l + \sigma^{-2} \boldsymbol{\Xi}_l + \mathbf{W}_l^{-1}\right)^{-1}
\end{aligned}
\tag{13}
$$

In (13), $\boldsymbol{\phi}_l$, $\boldsymbol{\Psi}_l$, $\boldsymbol{\Xi}_l$, and $\boldsymbol{\Omega}_l$ are constants specified in (14), where $I(A)$ is an indicator function which equals 1 if A is true and 0 otherwise. Also, we define $\bigodot_{n \in \mathcal{S}} \mathbf{X}_n = \mathbf{0}$ for $\mathcal{S} = \emptyset$. In other words, the Schur product over an empty set of matrices equals zero. In the subsequent section, we will focus on the BPSK ($L = 1$) and 4-PAM ($L = 2$) cases.

B. Examples: BPSK and 4-PAM

We will now show how to obtain Gaussian SISO equalizers for BPSK and 4-PAM modulations. Table I contains a list of parameters resulting from evaluating $\boldsymbol{\Psi}_l$, $\boldsymbol{\Xi}_l$, $\boldsymbol{\Omega}_l$ and $\boldsymbol{\phi}_l$ for $L = 1$ and $L = 2$. Substituting the parameters corresponding to $L = 1$ into (13), we have

$$
\begin{aligned}
\boldsymbol{\mu} &= \tilde{\mathbf{b}} + (\mathbf{H}^T \mathbf{H} + \sigma^2 \mathbf{W}^{-1})^{-1} (\mathbf{H}^T \mathbf{r} - \mathbf{H}^T \mathbf{H} \tilde{\mathbf{b}}) \\
\boldsymbol{\Sigma} &= (\sigma^{-2} \mathbf{H}^T \mathbf{H} + \mathbf{W}^{-1})^{-1}.
\end{aligned}
\tag{15}
$$

The expression for $\boldsymbol{\mu}$ found above corresponds to the MMSE filter output derived in [2] and [12].

For $L = 2$, substituting the parameters in Table I into (13) yields expressions corresponding to $Q(\mathbf{b}_1)$ and $Q(\mathbf{b}_2)$:

$$
\begin{aligned}
\boldsymbol{\mu}_1 &= \tilde{\mathbf{b}}_1 + [\mathbf{R}_1 + \sigma^2 \mathbf{W}_1^{-1}]^{-1} [\mathbf{H}_1^T \mathbf{r} - \mathbf{R}_1 (\tilde{\mathbf{b}}_1 + 2\mathbf{1})] \\
\boldsymbol{\Sigma}_1 &= (\sigma^{-2} \mathbf{R}_1 + \mathbf{W}_1^{-1})^{-1} \\
\boldsymbol{\mu}_2 &= \tilde{\mathbf{b}}_2 + [\mathbf{R}_2 + \sigma^2 \mathbf{W}_2^{-1}]^{-1} [\mathbf{H}_2^T \mathbf{r} - \mathbf{R}_2 \tilde{\mathbf{b}}_2] \\
\boldsymbol{\Sigma}_2 &= (\sigma^{-2} \mathbf{R}_2 + \mathbf{W}_2^{-1})^{-1}
\end{aligned}
\tag{16}
$$

where

$$
\begin{aligned}
\mathbf{H}_1 &= \mathrm{diag}(\boldsymbol{\mu}_2) \mathbf{H}; \\
\mathbf{R}_1 &= (\boldsymbol{\Sigma}_2 + \boldsymbol{\mu}_2 \boldsymbol{\mu}_2^T) \circ (\mathbf{H}^T \mathbf{H}); \\
\mathbf{H}_2 &= \mathrm{diag}(\boldsymbol{\mu}_1 + 2\mathbf{1}) \mathbf{H}; \\
\mathbf{R}_2 &= [\boldsymbol{\Sigma}_1 + (\boldsymbol{\mu}_1 + 2\mathbf{1})(\boldsymbol{\mu}_1 + 2\mathbf{1})^T] \circ (\mathbf{H}^T \mathbf{H}).
\end{aligned}
\tag{17}
$$

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 $\mathrm{diag}(\boldsymbol{\mu}_2)\mathbf{H}$ 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 $\mathrm{diag}(\boldsymbol{\mu}_1 + 2\mathbf{1})\mathbf{H}$ 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 equalization for BPSK, which are discussed in [2] and [12], also apply to the Gaussian SISO BLESD approach.

C. Practical Gaussian SISO Algorithm for 4-PAM

In this section we will resolve some practical challenges in converting the generic equations developed above to an equalization algorithm. Taking $Q(\mathbf{b}_2)$ for example:

$$
\boldsymbol{\mu}_2 = \tilde{\mathbf{b}}_2 + [\mathbf{R}_2 + \sigma^2 \mathbf{W}_2^{-1}]^{-1} [\mathbf{H}_2^T \mathbf{r} - \mathbf{R}_2 \tilde{\mathbf{b}}_2] \tag{18}
$$

$$
\boldsymbol{\Sigma}_2 = [\sigma^{-2} \mathbf{R}_2 + \mathbf{W}_2^{-1}]^{-1}. \tag{19}
$$

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

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

These equations cannot be used directly, since $\mathbf{W}_2 = \mathrm{diag}([1-\tilde{b}_{2,1}^2, \cdots, 1-\tilde{b}_{2,K}^2]^T)$ may become rank-deficient as the turbo iterations converge ($\tilde{b}_{2,k} \to \pm 1$). Thus (18) and (19) need to be converted to a different form.

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

$$\begin{aligned} & \mathbf{R}_2 + \sigma^2\mathbf{W}_2^{-1} \\ = \; & [\mathbf{\Sigma}_1 + (\boldsymbol{\mu}_1 + 2\mathbf{1})(\boldsymbol{\mu}_1 + 2\mathbf{1})^T] \circ (\mathbf{H}^T\mathbf{H}) + \sigma^2\mathbf{W}_2^{-1} \\ = \; & \mathbf{H}_2^T\mathbf{H}_2 + \mathbf{\Sigma}_1 \circ (\mathbf{H}^T\mathbf{H}) + \sigma^2\mathbf{W}_2^{-1} \\ = \; & \mathbf{H}_2^T\mathbf{H}_2 + \sigma^2\tilde{\mathbf{W}}_2^{-1}, \end{aligned} \tag{20}$$

where $\tilde{\mathbf{W}}_2 = (\mathbf{I} + \sigma^{-2}\mathbf{W}_2[\mathbf{\Sigma}_1 \circ (\mathbf{H}^T\mathbf{H})])^{-1}\mathbf{W}_2$. The derivation makes use of the fact that $\mathbf{\Sigma}_1$ and $\tilde{\mathbf{W}}_2$ are diagonal matrices, 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, (18) and (19) can be rewritten as

$$\begin{aligned} \boldsymbol{\mu}_2 &= \tilde{\mathbf{b}}_2 + [\mathbf{H}_2^T\mathbf{H}_2 + \sigma^2\tilde{\mathbf{W}}_2^{-1}]^{-1}[\mathbf{H}_2^T\mathbf{r} - \mathbf{R}_2\tilde{\mathbf{b}}_2] \\ \mathbf{\Sigma}_2 &= [\sigma^{-2}\mathbf{H}_2^T\mathbf{H}_2 + \tilde{\mathbf{W}}_2^{-1}]^{-1}. \end{aligned} \tag{21}$$

Having evaluated the posterior distribution $Q(\mathbf{b}_2)$, we recognize that $\mathcal{L}(b_{2,n}) = Q(b_{2,n})/p(b_{2,n})$ approximates the likelihood function $p(\mathbf{r}|b_{2,n})$ (The n-th symbol is the symbol of interest for every window position). It can be subsequently shown that the LLR of $b_{2,n}$ is

$$\Lambda_I(b_{2,n}) = \log\frac{\mathcal{L}(b_{2,n}=1)}{\mathcal{L}(b_{2,n}=-1)} = \frac{2\check{\mu}_{2,n}}{1-\check{\alpha}_{2,n}}, \tag{22}$$

where

$$\begin{aligned} \check{\mu}_{2,n} &= \sigma^{-2}\mathbf{e}_n^T[\mathbf{I} - \mathbf{H}_2^T(\mathbf{H}_2\tilde{\mathbf{W}}_2\mathbf{H}_2^T + \sigma^2\mathbf{I})^{-1}\mathbf{H}_2\tilde{\mathbf{W}}_2] \\ & \quad \cdot[\mathbf{H}_2^T\mathbf{r} - \mathbf{R}_2\tilde{\mathbf{b}}_{2,n}] \\ \check{\alpha}_{2,n} &= [\mathbf{H}_2^T(\mathbf{H}_2\tilde{\mathbf{W}}_2\mathbf{H}_2^T + \sigma^2\mathbf{I})^{-1}\mathbf{H}_2\tilde{\mathbf{W}}_2]_{n,n}. \end{aligned} \tag{23}$$

In (23), $\tilde{\mathbf{b}}_{2,n} = [\tilde{b}_{2,1}, \cdots, \tilde{b}_{2,n-1}, 0, \tilde{b}_{2,n+1}, \cdots, \tilde{b}_{2,K}]^T$. The LLR of $b_{1,n}$, $\Lambda_I(b_{1,n}) = \frac{2\check{\mu}_{1,n}}{1-\check{\alpha}_{1,n}}$ may be computed similarly.

IV. NUMERICAL RESULTS

Multiple iterations are required to evaluate (21) and the correction expressions for $\boldsymbol{\mu}_1$ and $\mathbf{\Sigma}_1$, since $(\boldsymbol{\mu}_1, \mathbf{\Sigma}_1)$ and $(\boldsymbol{\mu}_2, \mathbf{\Sigma}_2)$ are coupled due to Gray encoding. We call this the inner iteration, as opposed to the outer (turbo) iterations. Usually, only a small number of inner iterations are needed in each outer iteration.

In Fig. 1, we compare the performance of the proposed BLESD equalizer with the symbol-level equalizer [13]. 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 [12], the Porat-Friedlander channel [14] 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 $N_2 = 9$. In the BLESD equalizer, $G=4$ inner iterations are used for each outer iteration. Under this severe ISI channel, the symbol-based equalizer fails (no improvement after three iterations). In contrast, the Gaussian SISO BLESD equalizer remains effective even in this channel condition. The performance of the symbol-based equalizer after three iterations is also plotted for comparison.

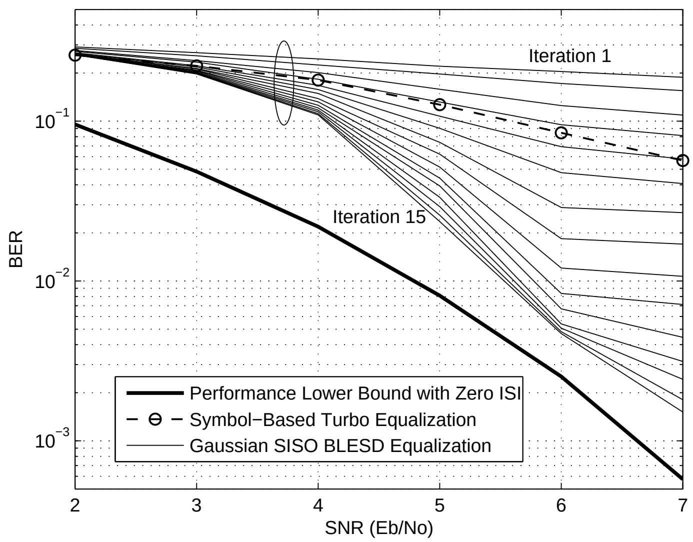

Fig. 1. BER performance of Gaussian-SISO BLESD and symbol-based turbo equalization for 16-QAM modulation.

It is worth mentioning that the symbol-based schemes are also capable of equalizing the Porat-Friedlander channel, as demonstrated in [15] and [16]. 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.

V. CONCLUSIONS

The Gaussian SISO BLESD equalization of multilevel QAM symbols points to a new way for an iterative receiver to handle the mapping between bits and symbols. The resulting algorithms perform equalization on the bit level, even when the transmitted bits are mapped onto Gray-coded symbols (as in BICM). The performance of the turbo receiver is thus improved as the conversion between symbol EXT and bit EXT in symbol-based equalization schemes is avoided.

APPENDIX I
FREE ENERGY EVALUATION

We first develop a few identities involving the Schur product:

Lemma 1:

$$\mathrm{tr}[\mathrm{diag}(\mathbf{x}) \cdot \mathbf{A} \cdot \mathrm{diag}(\mathbf{y}) \cdot \mathbf{B}^T] = \mathbf{x}^T(\mathbf{A} \circ \mathbf{B})\mathbf{y} \tag{24}$$

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:

$$\mathrm{tr}[\mathbf{A} \cdot (\mathbf{B} \circ \mathbf{C})^T] = \mathrm{tr}[(\mathbf{A} \circ \mathbf{B}) \cdot \mathbf{C}^T] \tag{25}$$

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 $\mathbf{C}$,

$$\mathsf{E}\left[(\textstyle\bigodot_{l=1}^U \mathbf{b}_l)^T \mathbf{C} (\bigodot_{l=1}^U \mathbf{b}_l)\right] = \mathrm{tr}\left[\textstyle\bigodot_{l=1}^U (\boldsymbol{\Sigma}_l + \boldsymbol{\mu}_l\boldsymbol{\mu}_l^T) \cdot \mathbf{C}\right]. \tag{26}$$

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

$$\begin{aligned} \mathsf{E}\left[\mathbf{b}_1^T \mathbf{C} \mathbf{b}_1\right] &= \boldsymbol{\mu}_1^T \mathbf{C} \boldsymbol{\mu}_1 + \mathrm{tr}[\mathbf{C}\boldsymbol{\Sigma}_1] \\ &= \mathrm{tr}\left[(\boldsymbol{\Sigma}_1 + \boldsymbol{\mu}_1\boldsymbol{\mu}_1^T) \cdot \mathbf{C}\right]. \end{aligned} \tag{27}$$

Assuming (26) is true for $U = u$, i.e.

$$\mathsf{E}\left\{(\textstyle\bigodot_{l=1}^u \mathbf{b}_l)^T \mathbf{C} (\bigodot_{l=1}^u \mathbf{b}_l)\right\} = \mathrm{tr}\left[\textstyle\bigodot_{l=1}^u (\boldsymbol{\Sigma}_l + \boldsymbol{\mu}_l\boldsymbol{\mu}_l^T) \cdot \mathbf{C}\right], \tag{28}$$

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

$$\begin{aligned} & \mathsf{E}\left\{(\textstyle\bigodot_{l=1}^{u+1} \mathbf{b}_l)^T \mathbf{C} (\bigodot_{l=1}^{u+1} \mathbf{b}_l)\right\} \\ = \; & \mathsf{E}\left\{(\textstyle\bigodot_{l=1}^{u} \mathbf{b}_l)^T \left[\mathbf{B}_{u+1}\mathbf{C}\mathbf{B}_{u+1}\right] (\bigodot_{l=1}^{u} \mathbf{b}_l)\right\} \\ = \; & \mathsf{E}\left\{\mathrm{tr}\left[\textstyle\bigodot_{l=1}^u (\boldsymbol{\Sigma}_l + \boldsymbol{\mu}_l\boldsymbol{\mu}_l^T) \cdot \mathbf{B}_u \mathbf{C} \mathbf{B}_u\right]\right\} \\ \overset{L.1}{=} \; & \mathsf{E}\left\{\mathbf{b}_u^T \left[\textstyle\bigodot_{l=1}^u (\boldsymbol{\Sigma}_l + \boldsymbol{\mu}_l\boldsymbol{\mu}_l^T) \circ \mathbf{C}\right] \mathbf{b}_u\right\} \\ = \; & \mathrm{tr}\left\{\left[\textstyle\bigodot_{l=1}^u (\boldsymbol{\Sigma}_l + \boldsymbol{\mu}_l\boldsymbol{\mu}_l^T) \circ \mathbf{C}\right] \cdot \boldsymbol{\Sigma}_{u+1}\right\} \\ & + \boldsymbol{\mu}_{u+1}^T \left[\textstyle\bigodot_{l=1}^u (\boldsymbol{\Sigma}_l + \boldsymbol{\mu}_l\boldsymbol{\mu}_l^T) \circ \mathbf{C}\right] \boldsymbol{\mu}_{u+1} \\ = \; & \mathrm{tr}\left\{(\boldsymbol{\Sigma}_{u+1} + \boldsymbol{\mu}_{u+1}\boldsymbol{\mu}_{u+1}^T) \cdot \left[\textstyle\bigodot_{l=1}^u (\boldsymbol{\Sigma}_l + \boldsymbol{\mu}_l\boldsymbol{\mu}_l^T) \circ \mathbf{C}\right]\right\} \\ \overset{L.2}{=} \; & \mathrm{tr}\left\{\textstyle\bigodot_{l=1}^{u+1} (\boldsymbol{\Sigma}_l + \boldsymbol{\mu}_l\boldsymbol{\mu}_l^T) \cdot \mathbf{C}\right\} \end{aligned} \tag{29}$$

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

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 $U < V$,

$$\begin{aligned} & \mathsf{E}\left[(\textstyle\bigodot_{l=1}^U \mathbf{b}_l)^T \mathbf{C} (\bigodot_{l=1}^V \mathbf{b}_l)\right] \\ = \; & \mathrm{tr}\left[\textstyle\bigodot_{l=1}^U (\boldsymbol{\Sigma}_l + \boldsymbol{\mu}_l\boldsymbol{\mu}_l^T) \cdot \left[\mathbf{C}\mathrm{diag}(\bigodot_{l=u+1}^V \mathbf{u}_l)\right]\right] \\ = \; & \mathbf{1}^T \left[\textstyle\bigodot_{l=1}^U (\boldsymbol{\Sigma}_l + \boldsymbol{\mu}_l\boldsymbol{\mu}_l^T) \circ \mathbf{C}\right] \bigodot_{l=u+1}^V \boldsymbol{\mu}_l. \end{aligned} \tag{30}$$

Proof: Utilizing Lemma 3 yields

$$\begin{aligned} & \mathsf{E}\left\{(\textstyle\bigodot_{l=1}^U \mathbf{b}_l)^T \mathbf{C} \bigodot_{l=1}^V \mathbf{b}_l\right\} \\ \overset{L.3}{=} \; & \mathsf{E}\left\{\mathrm{tr}\left[\textstyle\bigodot_{l=1}^U (\boldsymbol{\Sigma}_l + \boldsymbol{\mu}_l\boldsymbol{\mu}_l^T) \cdot \mathbf{C} \cdot \mathrm{diag}(\bigodot_{l=u+1}^V \mathbf{b}_l)\right]\right\} \\ = \; & \mathrm{tr}\left[\textstyle\bigodot_{l=1}^U (\boldsymbol{\Sigma}_l + \boldsymbol{\mu}_l\boldsymbol{\mu}_l^T) \cdot \mathbf{I} \cdot \mathbf{C} \cdot \mathrm{diag}(\bigodot_{l=u+1}^V \boldsymbol{\mu}_l)\right] \\ \overset{L.1}{=} \; & \mathbf{1}^T \left[\textstyle\bigodot_{l=1}^U (\boldsymbol{\Sigma}_l + \boldsymbol{\mu}_l\boldsymbol{\mu}_l^T) \circ \mathbf{C}\right] \bigodot_{l=u+1}^V \boldsymbol{\mu}_l. \end{aligned} \tag{31}$$

■

The three integrals in (11) can now be evaluated, yielding the complete free energy expression in (12).

REFERENCES

[1] 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.

[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] 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.

[4] 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.

[5] 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.

[6] 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.

[7] ——, "A variational inference framework for soft-in-soft-out detection in interference channels," *submitted to IEEE Trans. Inf. Theory.*

[8] 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.

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

[10] D. D. Lin and T. J. Lim, "Multiuser detection of M-QAM symbols via bit-level equalization and soft detection," in *Proc. IEEE Int. Symp. Information Theory (ISIT'06)*, July 2006, pp. 1909–1913.

[11] ——, "Bit-level equalization and soft detection for Gray-coded multilevel modulation," *submitted to IEEE Trans. Inf. Theory.*

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.