Self-Iterating Soft Equalizer

Seongwook Jeong, *Member, IEEE,* and Jaekyun Moon, *Fellow, IEEE*

Abstract—A self-iterating soft equalizer (SISE) consisting of a few relatively weak constituent equalizers is shown to provide robust performance even in severe intersymbol interference (ISI) channels that exhibit deep nulls and valleys within the signal band. Constituent equalizers are allowed to exchange soft information in the absence of interleavers based on the method that are designed to handle significant correlation among their soft outputs. The resulting SISE works well as a stand-alone equalizer or as the equalizer component of a turbo equalization system. The performance advantages over existing methods are validated with bit-error-rate (BER) simulations and extrinsic information transfer (EXIT) chart analysis. It is shown that in turbo equalizer setting the SISE achieves performance closer to the maximum *a posteriori* probability equalizer than any other known schemes in very severe ISI channels.

Index Terms—Equalizers, intersymbol interference, iterative methods, maximum *a posteriori* probability estimation, turbo equalization.

I. INTRODUCTION

TURBO equalization is a well-established technique that is highly effective in combating intersymbol interference (ISI) via iterative exchange of soft decisions between a soft-in soft-out (SISO) equalizer and a SISO error-correction decoder separated by an interleaver [1]. The Bahl-Cocke-Jelinek-Raviv (BCJR) algorithm [2] and the soft-output Viterbi algorithm (SOVA) [3] provide excellence performance as the equalizer component of turbo equalization systems, but both schemes require implementation complexity that grows exponentially with the length of the ISI channel.

Numerous suboptimal, low-complexity turbo equalization schemes have been proposed to mitigate the high computational complexity of the BCJR and SOVA methods. See, for example, [4]–[17]. Some of these schemes utilize reduced-trellis approaches while others rely on filter-based methods such as the linear equalizer (LE) or the decision feedback equalizer (DFE). The SISO version of the LE has been discussed in [4], [7]–[12]. The design is based on applying the classical minimum-mean-squared-error (MMSE) criterion while utilizing second order statistics of the input symbols

S. Jeong was with the Dept. of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA. He is now with LSI Corp., San Jose, CA 95131 USA (e-mail: jeong030@umn.edu).

J. Moon is with the Dept. of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon, 305-701, Republic of Korea (e-mail: jmoon@kaist.edu).

estimated from the extrinsic information made available by the outer decoder. The authors of [8] have shown that using Proakis' well-known example channel with ISI tap weights [1 2 3 2 1], the turbo equalizer based on the SISO LE performs as well as one based on the BCJR equalizer at error rates below 10^{-4}. While the original formulation of the equalizer in [7], [8] gives rise to a time-varying filter, a low-complexity quasi-time-invariant filter, whose tap setting changes only once in every iteration stage, has also been suggested [7]. This low-complexity method does not result in a significant performance loss [7]. The same authors also investigated the SISO DFE but concluded that it was inferior to the LE counterpart of [8]. The authors of [13] considered a more severe ISI channel of the form [1 2 3 4 3 2 1], a natural extension of the previously considered ISI patterns by Proakis, and showed that for this channel, there is a substantial performance gap between the LE-based turbo equalizer of [8] and the BCJR-based turbo equalizer. They showed that using reduced-trellis search based on soft-decision-feedback can close this performance gap with the BCJR-based turbo equalizer. This approach was also shown to be effective for the magnetic recording channel [14].

Another meaningful development on suboptimal equalization is to employ two DFEs (or DFE variants) running in opposite directions and combine their extrinsic information [17]–[21]. This "bi-directional" DFE (BiDFE) algorithm takes advantage of different decision error and noise distributions at the outputs of the forward and time-reversed DFEs. Moreover, unlike the LE or the DFE, the BiDFE algorithm can be designed to avoid performance degradation even when the filter taps are constrained to be time-invariant [17]. The BiDFE can be considered as a parallel concatenated scheme with two DFEs producing somewhat correlated yet significantly different extrinsic information. The time-reversal operation applied to the reverse DFE can be viewed as a type of interleaving that attempts to make two input streams going into the forward and reverse DFEs appear independent. In [17], an effective log-likelihood-ratio (LLR) combining strategy for the two bi-directional DFEs has been proposed. The work of [17] also showed a new way of formulating extrinsic information that can effectively suppress error propagation in the DFE leading to improved performance of the DFE-based turbo equalizer relative to the LE-based one. When each DFE employs the extrinsic information formulation method of [17] and a LLR combining strategy of [17] is used, the BiDFE is shown to perform considerably better than the LE.

In this paper, we focus on a new equalizer structure that employs the LE, the DFE or the BiDFE as constituent modules, devising a strategy that allows iterative exchange of soft information among the constituent equalizers. Unlike typical turbo

processing methods, no interleaver exists between the SISO equalizer modules, and a special strategy to combat the correlation between successive module outputs must be devised. Interleaving in the usual sense is not possible in our case. Placing a shuffling device between two component equalizers at the receiver side would imply that two sets of channel output sequences - one corresponding to the original input sequence and the other corresponding to the shuffled input sequence - are available. This would require transmission of a redundant set of data and is clearly not sensible in practice. The significant correlation between the equalizer module outputs is a direct consequence of the lack of interleaving. It is shown that the extrinsic information of one module becomes the *a priori* information for the next module in concatenation via a specific scaling law that depends on the correlation between the input and output information sets of the first module. This equalizer is viewed as a self-iterating soft equalizer (SISE) consisting, in general, of multiple suboptimal constituent equalizers which are concatenated with no interleavers placed between them. The rationale behind this particular equalizer structure is that the suboptimal equalizers such as the LE, the DFE and the BiDFE all have their own advantages and disadvantages, and one equalizer should be able to benefit from the presence of the other equalizers. For example, the LE does not have the error propagation problem which the DFE suffers from, whereas the DFE often shows significantly better performance than the LE when feedback decisions are correct; and the BiDFE provides solid performance even with time-invariant filters, although its complexity is roughly twice as high as that of the DFE. We show that through simulation and analysis this type of "self-iteration" among constituent equalizer components with different characteristics can improve upon traditional equalization schemes based on a single equalizer component.

The remainder of the paper is organized as follows. In Section II, a brief statement of the system model is given. In Section III, we show a proper way to generate *a priori* information from the extrinsic information out of other constituent equalizers when the information between the equalizers could be significantly correlated, and then propose self-iterating soft equalizer design for uncoded systems. We also provide turbo equalization algorithms based on the SISE in Section IV. In Section V, we briefly review the individual suboptimal equalizers with the several filter types utilizing *a priori* information which will be employed in the proposed SISE algorithms. In Section VI, numerical results and analysis are given. Finally, we draw conclusions in Section VII.

II. SYSTEM MODEL

Given the transmitted sequence of coded bits $\{x_k\}$, the ISI channel output at time n is

$$r_n = \sum_{k=0}^{L_h} h_k x_{n-k} + w_n \tag{1}$$

where w_n is additive white Gaussian noise (AWGN) with variance N_0 and $\{h_k\}$ is the energy-normalized channel impulse response with length $L_h + 1$. In this paper, it is assumed that the transmitted symbol is a binary input with the average power equal to 1, i.e., $x_n \in \{\pm 1\}$ and $\mathrm{E}(x_n^2) = 1$, and

the ISI channel coefficients and noise samples are real-valued. This restriction is not necessary for the algorithm development here but makes the presentation simpler. We also assume that the channel response is time-invariant and deterministic. The schemes investigated in this paper can be applied to random channels, where a channel response changes from one transmission to the next in some random fashion, provided that the channel remains static over a given transmission period and that channel estimation can be done at the receiver side.

In turbo equalization, the *a priori* log-likelihood ratio (LLR) of x_n to the equalizer is defined as

$$L_a(x_n) \triangleq \ln \frac{\Pr(x_n = +1)}{\Pr(x_n = -1)}$$

where the probabilities denoted by $\Pr(\cdot)$ are in reality just estimates. The probabilities are all set to 1/2 initially, and then, as the turbo iteration ensues, the extrinsic information generated by the outer decoder is used as the *a priori* information to the equalizer.

Based on these *a priori* LLR values for the symbols, the equalizer generates its own extrinsic information, which will in turn be passed to the decoder. Let y_n be the equalizer filter output sequence corresponding to the observation sequence r_n applied at the input. In an effort to produce the extrinsic information $L_e(x_n)$ that should not depend on the *a priori* LLR of the current symbol x_n, $L_a(x_n)$ *is set to zero during the computation of* y_n [8]. Then, the equalizer's extrinsic information is directly related to the equalizer output y_n as:

$$
\begin{aligned}
L_e(x_n) &\triangleq \ln \frac{\Pr(x_n = +1 \mid y_n)}{\Pr(x_n = -1 \mid y_n)} \bigg|_{L_a(x_n)=0} \\
&= \ln \frac{\mathrm{p}(y_n \mid x_n = +1)\Pr(x_n = +1)}{\mathrm{p}(y_n \mid x_n = -1)\Pr(x_n = -1)} \bigg|_{L_a(x_n)=0} \\
&= \ln \frac{\mathrm{p}(y_n \mid x_n = +1)}{\mathrm{p}(y_n \mid x_n = -1)}. \tag{2}
\end{aligned}
$$

where $\mathrm{p}(\cdot)$ is the probability density function.

III. SELF-ITERATING SOFT EQUALIZER ALGORITHM

In this section, we discuss the SISE algorithm. Basically, the SISE we focus on in this paper is a SISO equalizer which consists of one main suboptimal SISO equalizer and μ branch suboptimal SISO equalizers. The proposed SISE is illustrated in Fig. 1.

The key procedure in this algorithm is that the received data sequence is equalized by the main equalizer and its extrinsic information is passed to the branch equalizers as their *a priori* information. The extrinsic information generated in the branch equalizers is also passed back to the main equalizer to be used as its *a priori* information for the next stage. Note that since this equalization algorithm can perform iteratively without the decoder (hence the name "self-iterating" equalizer), it can be used in uncoded systems as well. The terms "main" and "branch" here do not necessarily imply the difference in the complexity levels or computational powers between the constituent equalizers. Rather, the distinction simply indicates the scheduling strategy, i.e., the main equalizer is the one that makes the initial decision in the serial concatenation of
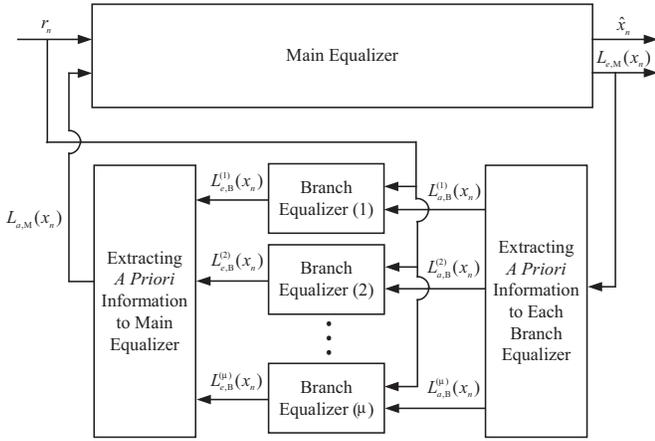
Fig. 1.   Self-iterating soft equalizer.

the constituent equalizers. In fact, different arrangements of the constituent equalizers are possible, including full parallel concatenation, full serial concatenation and combined parallel/serial concatenation, along with many different scheduling strategies. For the particular SISE structure shown in Fig. 1, for example, it can be seen that while there is self-iteration between the main equalizer and the block of branch equalizers, no self-iterations are assumed among the branch equalizers. While the concept and methods developed in this paper are general, for the performance analysis and simulation results to be presented, we shall focus on a serial concatenation of one main equalizer and one branch equalizer.

Unlike the extrinsic information between the decoder and the equalizer in usual turbo equalization setting, the extrinsic information between the main equalizer and the branch equalizers have correlation because no interleaving can be applied and their equalization processes are all based on the common received data sequence. Again, note that placing a shuffler between two component equalizers at the receiver side would imply that there is a redundant set of received samples available corresponding to the shuffled input sequence. This represents a highly wasteful system and would not be practical. It has been suggested that the high correlation between the *a priori* information and the extrinsic information of a module even in a turbo system can cause the performance degradation [22], [23]. In this section, we show a proper way to construct the *a priori* information for the main equalizer based on the extrinsic information generated by the branch equalizers when their outputs are correlated with the main equalizer output. The same method can be applied in processing the extrinsic information out of the main equalizer to obtain the *a priori* information for the branch equalizers when the main equalizer's soft output is correlated with those of the branch equalizers.

A. Generation of Uncorrelated A Priori Information

First let us assume that there is one main equalizer and one branch equalizer in an uncoded system. We will later extend the proposed algorithm to the case of multiple branch equalizers. We assume that the main equalizer generates the extrinsic LLR sequence, $L_{e,\mathrm{M}}(x_n)$, which is used to generate

the *a priori* LLR sequence, $L_{a,\mathrm{B}}(x_n)$, to be used by the branch equalizer. The branch equalizer in turn produces its own extrinsic LLR sequence, $L_{e,\mathrm{B}}(x_n)$, with the given $L_{a,\mathrm{B}}(x_n)$ sequence. In typical iterative processing, the $L_{e,\mathrm{M}}(x_n)$ sequence simply becomes the $L_{a,\mathrm{B}}(x_n)$ sequence (after interleaving) and the $L_{e,\mathrm{B}}(x_n)$ sequence obtained based on the $L_{a,\mathrm{B}}(x_n)$ sequence and channel observation, in turn, forms the $L_{a,\mathrm{M}}(x_n)$ sequence after proper deinterleaving. In the problem at hand, no interleaving/deinterleaving is allowed and there may be significant correlation between the two sequences, $L_{a,\mathrm{B}}(x_n)$ and $L_{e,\mathrm{B}}(x_n)$. The question is how we should generate $L_{a,\mathrm{M}}(x_n)$ to pass onto the main equalizer, given this correlation. The answer turns out to be a specific scaling law between $L_{a,\mathrm{B}}(x_n)$ and $L_{e,\mathrm{B}}(x_n)$, as already described in [24]. Below, we provide an improved derivation/justification for the same scaling law.

Modeling the *a priori* LLR and the extrinsic LLR as the output of an equivalent AWGN channel [26], we start by writing the unbiased versions of these LLRs associated with the branch equalizer as the transmitted symbol x corrupted by AWGN:

$$Y_{a,\mathrm{B}} = x + u_{a,\mathrm{B}},$$
$$Y_{e,\mathrm{B}} = x + u_{e,\mathrm{B}}$$

where time index n is dropped for notational simplicity; the process remains identical as n evolves. The noise terms $u_{a,\mathrm{B}}$ and $u_{e,\mathrm{B}}$ are assumed to be zero-mean Gaussian random variables which are independent of the transmitted data x but correlated with each other with correlation coefficient ρ_B. Note that $Y_{e,\mathrm{B}}$ for a specific time point is obtained with the given sequence of $Y_{a,\mathrm{B}}$ samples.

If $Y_{a,\mathrm{B}}$ and $Y_{e,\mathrm{B}}$ are uncorrelated, the *a priori* information to the main equalizer should be given as $L_{a,\mathrm{M}}(x) = L_{e,\mathrm{B}}(x) = \ln\{\mathrm{p}\,(Y_{e,\mathrm{B}}|x=+1)/\mathrm{p}\,(Y_{e,\mathrm{B}}|x=-1)\}$. However, since the two outputs are correlated and only the extrinsic information should be fed back to the main equalizer, the *a priori* information to the main equalizer can be defined as

$$L_{a,\mathrm{M}}(x) \triangleq \ln \frac{\mathrm{p}\,(Y_{e,\mathrm{B}}, Y_{a,\mathrm{B}}|x=+1)}{\mathrm{p}\,(Y_{e,\mathrm{B}}, Y_{a,\mathrm{B}}|x=-1)} - \ln \frac{\mathrm{p}\,(Y_{a,\mathrm{B}}|x=+1)}{\mathrm{p}\,(Y_{a,\mathrm{B}}|x=-1)}. \tag{3}$$

Notice that the equation (3) can reduce to $L_{a,\mathrm{M}}(x) = \ln\{\mathrm{p}\,(Y_{e,\mathrm{B}}|x=+1)/\mathrm{p}\,(Y_{e,\mathrm{B}}|x=-1)\}$ when $Y_{e,\mathrm{B}}$ and $Y_{a,\mathrm{B}}$ are independent. The first term in (3) signifies that both $Y_{a,\mathrm{B}}$ and $Y_{e,\mathrm{B}}$ may contain useful information that can be passed onto the main equalizer, whereas the subtraction of the second term is needed to suppress the information that has originated from the main equalizer itself. Replacing the likelihood function involving the correlated signals $Y_{a,\mathrm{B}}$ and $Y_{e,\mathrm{B}}$ with the likelihood function associated with their whitened versions,

$$
\begin{aligned}
J &= \mathrm{E}\left[\left|\left(\frac{1-\rho_B\lambda^{-1}}{1-\rho_B^2}-\alpha\right)L_{e,\mathrm{B}}(x)+\left(\frac{\rho_B^2-\rho_B\lambda}{1-\rho_B^2}\right)L_{a,\mathrm{B}}(x)\right|^2\right] \\
&= \mathrm{E}\left[\left|\left(\frac{1-\rho_B\lambda^{-1}}{1-\rho_B^2}-\alpha\right)\frac{2Y_{e,\mathrm{B}}}{N_{e,\mathrm{B}}}+\left(\frac{\rho_B^2-\rho_B\lambda}{1-\rho_B^2}\right)\frac{2Y_{a,\mathrm{B}}}{N_{a,\mathrm{B}}}\right|^2\right] \\
&= \frac{4}{N_{e,\mathrm{B}}^2 N_{a,\mathrm{B}}^2(1-\rho_B^2)^2\lambda^2}\,\cdot \\
&\quad \Big[\alpha^2\left\{\lambda^2\left(1-\rho_B^2\right)^2 N_{a,\mathrm{B}}^2(1+N_{e,\mathrm{B}})\right\} \\
&\quad -2\alpha\left\{\lambda\left(1-\rho_B^2\right)(\lambda-\rho_B)\left(N_{a,\mathrm{B}}^2(1+N_{e,\mathrm{B}})-\lambda\rho_B N_{a,\mathrm{B}} N_{e,\mathrm{B}}(1+\rho_B\sqrt{N_{a,\mathrm{B}} N_{e,\mathrm{B}}})\right)\right\} \\
&\quad +(\lambda-\rho_B)^2\left(\lambda^2\rho_B^2 N_{e,\mathrm{B}}^2(1+N_{a,\mathrm{B}})+N_{a,\mathrm{B}}^2(1+N_{e,\mathrm{B}})-2\lambda\rho_B N_{a,\mathrm{B}} N_{e,\mathrm{B}}(1+\rho_B\sqrt{N_{a,\mathrm{B}} N_{e,\mathrm{B}}})\right)\Big]
\end{aligned}
\tag{7}
$$

we can further write:

$$
\begin{aligned}
L_{a,\mathrm{M}}(x) &= \ln\frac{\mathrm{p}\big(Y_{e,\mathrm{B}}',Y_{a,\mathrm{B}}'|x=+1\big)}{\mathrm{p}\big(Y_{e,\mathrm{B}}',Y_{a,\mathrm{B}}'|x=-1\big)}-\ln\frac{\mathrm{p}\big(Y_{a,\mathrm{B}}|x=+1\big)}{\mathrm{p}\big(Y_{a,\mathrm{B}}|x=-1\big)} \\
&= \ln\frac{\mathrm{p}\big(Y_{e,\mathrm{B}}'|x=+1\big)}{\mathrm{p}\big(Y_{e,\mathrm{B}}'|x=-1\big)}+\ln\frac{\mathrm{p}\big(Y_{a,\mathrm{B}}'|x=+1\big)}{\mathrm{p}\big(Y_{a,\mathrm{B}}'|x=-1\big)} \\
&\quad -\ln\frac{\mathrm{p}\big(Y_{a,\mathrm{B}}|x=+1\big)}{\mathrm{p}\big(Y_{a,\mathrm{B}}|x=-1\big)} \\
&= \frac{\big(N_{a,\mathrm{B}}-\rho_B\sqrt{N_{a,\mathrm{B}} N_{e,\mathrm{B}}}\big)}{\big(1-\rho_B^2\big)N_{a,\mathrm{B}}}L_{e,\mathrm{B}}(x) \\
&\quad +\frac{\big(N_{e,\mathrm{B}}-\rho_B\sqrt{N_{e,\mathrm{B}} N_{a,\mathrm{B}}}\big)}{\big(1-\rho_B^2\big)N_{e,\mathrm{B}}}L_{a,\mathrm{B}}(x)-L_{a,\mathrm{B}}(x) \quad (4) \\
&= \frac{\big(N_{a,\mathrm{B}}-\rho_B\sqrt{N_{a,\mathrm{B}} N_{e,\mathrm{B}}}\big)}{\big(1-\rho_B^2\big)N_{a,\mathrm{B}}}L_{e,\mathrm{B}}(x) \\
&\quad +\frac{\big(\rho_B^2 N_{e,\mathrm{B}}-\rho_B\sqrt{N_{e,\mathrm{B}} N_{a,\mathrm{B}}}\big)}{\big(1-\rho_B^2\big)N_{e,\mathrm{B}}}L_{a,\mathrm{B}}(x) \tag{5}
\end{aligned}
$$

where $N_{a,\mathrm{B}}\triangleq\mathrm{Var}(u_{a,\mathrm{B}})$ and $N_{e,\mathrm{B}}\triangleq\mathrm{Var}(u_{e,\mathrm{B}})$, and $Y_{a,\mathrm{B}}'$ and $Y_{e,\mathrm{B}}'$ are the whitened versions of $Y_{a,\mathrm{B}}$ and $Y_{e,\mathrm{B}}$. Let $\mathbf{R}$ be the noise correlation matrix:

$$
\begin{aligned}
\mathbf{R} &\triangleq \begin{bmatrix} \mathrm{Var}(u_{a,\mathrm{B}}) & \mathrm{E}(u_{a,\mathrm{B}} u_{e,\mathrm{B}}) \\ \mathrm{E}(u_{e,\mathrm{B}} u_{a,\mathrm{B}}) & \mathrm{Var}(u_{e,\mathrm{B}}) \end{bmatrix} \\
&= \begin{bmatrix} N_{a,\mathrm{B}} & \rho_B\sqrt{N_{a,\mathrm{B}} N_{e,\mathrm{B}}} \\ \rho_B\sqrt{N_{a,\mathrm{B}} N_{e,\mathrm{B}}} & N_{e,\mathrm{B}} \end{bmatrix}.
\end{aligned}
$$

The combining weights for $L_{e,\mathrm{B}}(x)$ and $L_{a,\mathrm{B}}(x)$ in (4) are derived from the eigenvectors of $\mathbf{R}$. Write $\mathbf{R}$ as $\mathbf{R}=\mathbf{G}\mathbf{\Lambda}\mathbf{G}^{-1}$ where $\mathbf{G}$ is a unitary matrix consisting of the eigenvectors of $\mathbf{R}$ and $\mathbf{\Lambda}$ is a diagonal matrix whose elements are the eigenvalues of $\mathbf{R}$. It is well-known that the transformation $[Y_{e,\mathrm{B}}'\ Y_{a,\mathrm{B}}']^T=\mathbf{G}^T[Y_{e,\mathrm{B}}\ Y_{a,\mathrm{B}}]^T$ yields outputs whose correlation matrix is given by $\mathbf{\Lambda}$.

Introducing a new variable $\lambda\triangleq\sqrt{N_{a,\mathrm{B}}/N_{e,\mathrm{B}}}$, (5) can be rewritten as

$$
L_{a,\mathrm{M}}(x)=\left(\frac{1-\rho_B\lambda^{-1}}{1-\rho_B^2}\right)L_{e,\mathrm{B}}(x)+\left(\frac{\rho_B^2-\rho_B\lambda}{1-\rho_B^2}\right)L_{a,\mathrm{B}}(x).
\tag{6}
$$

This equation allows us to construct the *a priori* LLR for the main equalizer based on two correlated LLRs $L_{e,\mathrm{B}}(x)$ and

$L_{a,\mathrm{B}}(x)$, along with the values of λ and ρ_B which can be estimated easily. In our simulation, however, we observe that occasionally $L_{e,\mathrm{B}}(x_n)$ and $L_{a,\mathrm{B}}(x_n)$ take opposite polarities, due to large local noise values. When this happens, the polarity of $L_{a,\mathrm{M}}(x_n)$ may also be inconsistent with that of $L_{e,\mathrm{B}}(x_n)$, depending on the particular values of λ and ρ_B as well as the magnitudes of $L_{e,\mathrm{B}}(x_n)$ and $L_{a,\mathrm{B}}(x_n)$. It turns out that these events degrade the overall performance significantly. Luckily, we find that a small modification handles this issue effectively and provides robust performance. The modification is based on imposing a constraint that the polarities of $L_{a,\mathrm{M}}(x_n)$ and $L_{e,\mathrm{B}}(x_n)$ remain the same for all n while following the rule set forth by (6) as much as possible. How do we achieve this? One possible trick is to force a linear relationship, $\hat{L}_{a,\mathrm{M}}(x)=\alpha L_{e,\mathrm{B}}(x)$ where α is a positive scaling factor that in general depends on ρ_B, and then find the value of α that would minimize the mean-squared error (MSE) between $L_{a,\mathrm{M}}(x)$ as given in (6) and its forced approximation $\hat{L}_{a,\mathrm{M}}(x)$.

Denoting the MSE by J, we can write it as (7) where the second equality is due to the fact that for Gaussian variables, the log-likelihood function can be expressed simply as a scaled version of the observation variable. Evaluating the derivative of (7) with respect to the variable α and setting it equal to zero, we obtain the desired α:

$$
\begin{aligned}
\alpha = &\frac{\lambda-\rho_B}{\lambda\left(1-\rho_B^2\right)N_{a,\mathrm{B}}(1+N_{e,\mathrm{B}})}\,\cdot \\
&\left\{N_{a,\mathrm{B}}(1+N_{e,\mathrm{B}})-\lambda\rho_B N_{e,\mathrm{B}}(1+\rho_B\sqrt{N_{a,\mathrm{B}} N_{e,\mathrm{B}}})\right\}.
\end{aligned}
\tag{8}
$$

In [17], when the outputs of the forward and reversed DFEs were combined, it was shown that the sensitivity of the combiner output to the error in estimating the LLR correlation coefficient was reduced greatly if the variances of the two DFE outputs were assumed equal. Here we adopt the same strategy and assume that $N_{a,\mathrm{B}}=N_{e,\mathrm{B}}=N$ or $\lambda=1$ to reduce the sensitivity of the combiner output to the error in estimating ρ_B. This assumption also reduces complexity, as λ and N no

$$\hat{\rho}_B = \frac{\sum_n \left\{ \left(L_{a,\mathrm{B}}(x_n) - \mathrm{sign}\left[L_{a,\mathrm{B}}(x_n)\right] m_{a,\mathrm{B}} \right) \left(L_{e,\mathrm{B}}(x_n) - \mathrm{sign}\left[L_{e,\mathrm{B}}(x_n)\right] m_{e,\mathrm{B}} \right) \right\}}{\sqrt{\sum_n \left(L_{a,\mathrm{B}}(x_n) - \mathrm{sign}\left[L_{a,\mathrm{B}}(x_n)\right] m_{a,\mathrm{B}} \right)^2} \sqrt{\sum_n \left(L_{e,\mathrm{B}}(x_n) - \mathrm{sign}\left[L_{e,\mathrm{B}}(x_n)\right] m_{e,\mathrm{B}} \right)^2}} \qquad (12)$$

longer need be estimated. Accordingly, (8) becomes

$$\begin{aligned}
\alpha &= \frac{1 + N - \rho_B(1 + \rho_B N)}{(1 + \rho_B)(1 + N)} \\
&= \frac{1 - \rho_B}{1 + \rho_B} + \frac{\rho_B(1 - \rho_B)N}{(1 + \rho_B)(1 + N)} \qquad (9) \\
&\simeq \frac{1 - \rho_B}{1 + \rho_B}. \qquad (10)
\end{aligned}$$

While the last approximation is valid when the noise correlation coefficient, ρ_B, is high or low and/or the noise variance is small, $N \ll 1$, empirical results indicate that (10) gives robust performance at all situations.

In summary, we set the *a priori* LLR for the main equalizer as a proper correlation-dependent scaling of the branch equalizer output LLR:

$$L_{a,\mathrm{M}}(x) = \left(\frac{1 - \rho_B}{1 + \rho_B} \right) L_{e,\mathrm{B}}(x). \qquad (11)$$

Notice that here $\rho_B = 0$ yields $L_{a,\mathrm{M}}(x) = L_{e,\mathrm{B}}(x)$ whereas $\rho_B = 1$ leads to $L_{a,\mathrm{M}}(x) = 0$. This is intuitively pleasing since when the extrinsic LLR of the branch equalizer is completely uncorrelated with the *a priori* LLR applied at the input of the branch equalizer, the former should simply be passed on as the *a priori* LLR of the main equalizer. On the other hand, a complete correlation would indicate that the extrinsic information out of the branch equalizer brings no new information to the main equalizer and the *a priori* LLRs should all be set to zero.

Equation (11) is also valid in the opposite direction, i.e., when the *a priori* LLR of the branch equalizer is constructed from the extrinsic LLR of the main equalizer. That is, (11) is valid with the subscripts "B" and "M" swapped.

B. Estimation of Noise Correlation Coefficient

Assuming that the noise is stationary, the correlation coefficient between $u_{a,\mathrm{B}}$ and $u_{e,\mathrm{B}}$ (or $L_{a,\mathrm{B}}(x)$ and $L_{e,\mathrm{B}}(x)$) can be estimated through time-averaging $L_{a,\mathrm{B}}(x)$ and $L_{e,\mathrm{B}}(x)$ over some reasonably large finite window, as shown in (12) where $m_{a,\mathrm{B}} = \mathrm{E}(L_{a,\mathrm{B}}(x_n)|x_n = +1)$ and $m_{e,\mathrm{B}} = \mathrm{E}(L_{e,\mathrm{B}}(x_n)|x_n = +1)$; due to symmetry it is also assumed that $\mathrm{E}(L_{a,\mathrm{B}}(x_n)|x_n = +1) = -\mathrm{E}(L_{a,\mathrm{B}}(x_n)|x_n = -1)$ and $\mathrm{E}(L_{e,\mathrm{B}}(x_n)|x_n = +1) = -\mathrm{E}(L_{e,\mathrm{B}}(x_n)|x_n = -1)$.

We can estimate the conditional means through time-averaging:

$$\begin{aligned}
\hat{m}_{a,\mathrm{B}} &= \frac{1}{2} \left\{ \overline{\overline{(L_{a,\mathrm{B}}(x_n)|L_{a,\mathrm{B}}(x_n) \geq 0)}} \right. \\
&\qquad \left. - \overline{\overline{(L_{a,\mathrm{B}}(x_n)|L_{a,\mathrm{B}}(x_n) < 0)}} \right\} \\
\hat{m}_{e,\mathrm{B}} &= \frac{1}{2} \left\{ \overline{\overline{(L_{e,\mathrm{B}}(x_n)|L_{e,\mathrm{B}}(x_n) \geq 0)}} \right. \\
&\qquad \left. - \overline{\overline{(L_{e,\mathrm{B}}(x_n)|L_{e,\mathrm{B}}(x_n) < 0)}} \right\}
\end{aligned}$$

where $\overline{\overline{u}}$ denotes the time-average of u. Note that the signs of the LLRs between the main equalizer output and the branch equalizer output might be different; in estimating the correlation coefficient, we only consider those LLR samples for which $\mathrm{sign}\left[L_{a,\mathrm{B}}(x_n)\right]$ and $\mathrm{sign}\left[L_{e,\mathrm{B}}(x_n)\right]$ are identical.

C. Extension to the Case of Multiple Branch Equalizers

Let us assume that there are one main equalizer and two branch equalizers. Then, the *a priori* LLR to the main equalizer from each branch equalizer can be formulated as

$$L_{a,\mathrm{M}}^{(1)}(x) = \left(\frac{1 - \rho_B^{(1)}}{1 + \rho_B^{(1)}} \right) L_{e,\mathrm{B}}^{(1)}(x)$$

$$L_{a,\mathrm{M}}^{(2)}(x) = \left(\frac{1 - \rho_B^{(2)}}{1 + \rho_B^{(2)}} \right) L_{e,\mathrm{B}}^{(2)}(x)$$

where superscript (i) points to a specific branch equalizer i and $\rho_B^{(i)}$ is the correlation coefficient between $L_{e,\mathrm{B}}^{(i)}(x)$ and $L_{a,\mathrm{B}}^{(i)}(x)$. Again, since the *a priori* information (or extrinsic information) can be modeled as an equivalent AWGN channel output, under the assumption that the noise variances are the same, the whitened/combined *a priori* information to the main equalizer can be shown to be

$$L_{a,\mathrm{M}}(x) = \frac{1}{1 + \xi} \left(L_{a,\mathrm{M}}^{(1)}(x) + L_{a,\mathrm{M}}^{(2)}(x) \right) \qquad (13)$$

where ξ is the noise correlation coefficient between $L_{a,\mathrm{M}}^{(1)}(x)$ and $L_{a,\mathrm{M}}^{(2)}(x)$, which can be also estimated through time-averaging based on an equation similar to (12). It is straightforward to extend the SISE algorithm to a system consisting of μ branch equalizers via noise-whitening transformation.

D. SISE Algorithm

Finally, the proposed SISE algorithm for an uncoded system can be summarized as follows:

- Initialize the *a priori* information of the main equalizer, i.e., $L_{a,\mathrm{M}}(x_n) = 0$ and $L_{a,\mathrm{M}}^{(i)}(x_n) = 0$ for all time index n and branch index i.
- For the specified number of self-iterations,
  1) Generate the extrinsic information of the main equalizer, $L_{e,\mathrm{M}}(x_n)$, with the *a priori* information $L_{a,\mathrm{M}}(x_k)$ for all $k \neq n$ (the process of generating the extrinsic information in a given equalizer is described in Section V).
  2) Compute the noise correlation coefficients, $\rho_{\mathrm{M}}^{(i)}$, between $L_{e,\mathrm{M}}(x_n)$ and $\alpha_i L_{a,\mathrm{M}}^{(i)}(x_n)$ where α_i is the combiner weight used on the ith branch equalizer output in constructing $L_{a,\mathrm{M}}(x_n)$ in the previous cycle, $L_{a,\mathrm{M}}(x_n) = \sum_i \alpha_i L_{a,\mathrm{M}}^{(i)}(x_n)$, and set $L_{a,\mathrm{B}}^{(i)}(x_n) = [(1 - \rho_{\mathrm{M}}^{(i)})/(1 + \rho_{\mathrm{M}}^{(i)})] L_{e,\mathrm{M}}(x_n)$ for all i.

3) Generate the extrinsic information of each branch equalizer, $L_{e,B}^{(i)}(x_n)$, with the given *a priori* information $L_{a,B}^{(i)}(x_k)$ for all $k \neq n$, for all i.

4) Compute the noise correlation coefficient, $\rho_B^{(i)}$, between $L_{e,B}^{(i)}(x_n)$ and $L_{a,B}^{(i)}(x_n)$ and set $L_{a,M}^{(i)}(x_n) = [(1 - \rho_B^{(i)})/(1 + \rho_B^{(i)})]L_{e,B}^{(i)}(x_n)$ for all i.

5) Generate the *a priori* information for the main equalizer, $L_{a,M}(x_n)$, from $L_{a,M}^{(i)}(x_n)$ via the extended equation of (13).

IV. APPLICATION TO TURBO EQUALIZATION

In this section, we propose turbo equalization based on the previously developed SISE algorithm. Various iterative equalization algorithms are possible based on this structure, but two main algorithms are considered here.

A. SISE 1 and SISE 2 Algorithms

The first algorithm passes the uncorrelated extrinsic information of the main equalizer to the branch equalizers and then to the decoder in turn. The information flow of the SISE 1 algorithm is shown in Fig. 2(a). While the time sequence is not clear in the figure, we note that the main equalizer passes the soft information to the decoder *after* one self-iteration is performed with its branch equalizer,

Due to the sequential nature of the self-iteration steps, the SISE 1 algorithm has a long latency issue; the second algorithm (SISE 2) is also proposed to get around this issue. In contrast to the first algorithm, SISE 2 *simultaneously* passes the extrinsic information and the correlation-compensated extrinsic information of the main equalizer to the decoder and to the branch equalizers, respectively. Thus, in this case, the self-iteration step is performed in parallel with the outer turbo iteration. The information flow of SISE 2 is shown in Fig. 2(b).

B. Comparison of Complexity and Latency

Let the computational complexity of the main equalizer, the branch equalizers, and the decoder be C_M, C_B, and C_D, respectively. For each outer iteration performed, the amount of computation for the conventional turbo equalization is $C_M + C_D$, whereas it is $2C_M + C_B + C_D$ and $C_M + C_B + C_D$ for SISE 1 and SISE 2, respectively. Moreover, assuming the processing time for the main equalizer, the parallel branch equalizers, and the decoder is all equal to T, the total processing time for each outer iteration is $2T$, $4T$, and $2T$ for the conventional system, SISE 1, and SISE 2, respectively. As will be shown later, in addition to having complexity/latency advantage, SISE 2 also has performance advantage over SISE 1 in many channel situations and thus seems to be the preferred choice. Also, while SISE 2 requires higher complexity (by C_B) than existing turbo equalizers when C_M is fixed in both cases, it will be shown that SISE 2 often enables substantial error rate reduction in channel conditions where the existing turbo equalizer cannot provide any performance improvement regardless of how large C_M is allowed to grow.
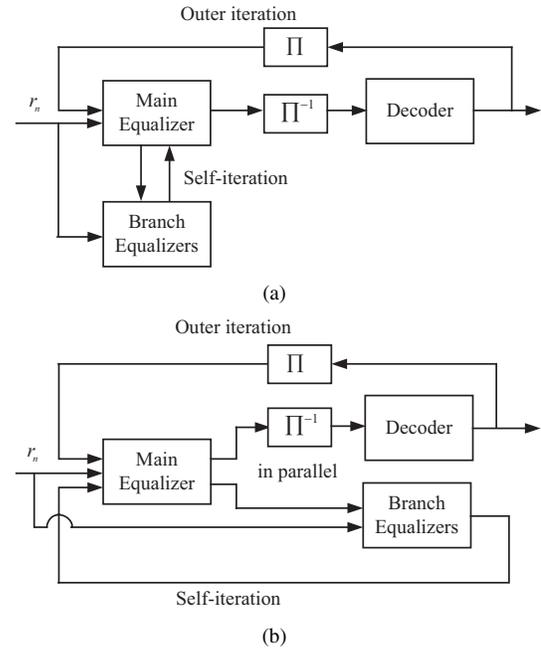

Fig. 2. Information flow: (a) Turbo SISE 1 (b) Turbo SISE 2.

V. INDIVIDUAL CONSTITUENT EQUALIZERS FOR SISE

Each constituent equalizer generates its own extrinsic LLR sequence given the estimated *a priori* LLR sequence as well as the channel observation sequence available at its input. While any equalizer can be used as a constituent equalizer of the SISE, we assume that the LE, the DFE or the BiDFE plays the role of each constituent equalizer in this paper. For practical reasons, we also assume that a constituent equalizer can be based on either a time-invariant (TI) filter or a quasi-time-invariant (QTI) filter with tap weight setting changing once per outer turbo iteration or self-iteration. As shown by the authors of [7], [8], the classical MMSE equalizer design can be modified by incorporating the mean and variance of the channel input symbols estimated via the extrinsic symbol information generated by the decoder. In Section V-A, we provide brief overviews on these equalizers, and in Section V-B, we discuss performance potentials of QTI versus TI filter structures in the limit of large numbers of iterations.

A. MMSE Filters Utilizing A Priori Information

1) Linear Equalizer of [8]: Here we basically summarize the LE approach of [8]. In the process, however, we clarify specific parameter settings for the simulation results presented in this paper. Let the filter coefficient vector be $\mathbf{c} = [c_{L_c} \cdots c_0 \cdots c_{-L_f}]^T$ and define the observation vector $\mathbf{r} \triangleq [r_{n-L_c} \cdots r_n \cdots r_{n+L_f}]^T$. The filter output at time n, $y_n = \mathbf{r}^T \mathbf{c}$, represents an estimate for the input symbol x_n. The corresponding estimation error $e_n = x_n - \mathbf{r}^T \mathbf{c}$, however, may have a non-zero mean if the average symbol value is non-zero, which is the case when the *a priori* probabilities exist for the input symbols. To force the mean of the equalizer error to zero, the LE output is modified to $y_n = (\mathbf{r} - \bar{\mathbf{r}})^T \mathbf{c} + \bar{x}_n$ so that the error is given by $e_n = x_n - \bar{x}_n - (\mathbf{r} - \bar{\mathbf{r}})^T \mathbf{c}$. The overbar denotes the statistical mean and indicates a vector of means

when placed over a vector. It is straightforward to find the filter weights that minimize the mean-squared-error. Assuming that the input symbols are independent, the MSE-minimizing tap weights at time n are

$$\mathbf{c}_{opt} = z_n \Big[\mathbf{H}\mathbf{A}\mathbf{H}^T + N_0\mathbf{I} \Big]^{-1} \mathbf{h} \qquad (14)$$

where z_n is the input symbol variance: $z_n \triangleq \overline{(x_n - \overline{x}_n)^2} = \overline{x_n^2} - [\overline{x}_n]^2 = 1 - [\overline{x}_n]^2$, $\mathbf{H}$ is the $L \times (L + L_h)$ channel response matrix with $L = 1 + L_c + L_f$:

$$\mathbf{H} = \begin{bmatrix} h_{L_h} & \cdots & h_1 & h_0 & 0 & \cdots & 0 \\ 0 & h_{L_h} & \cdots & h_1 & h_0 & \cdots & 0 \\ \vdots & & \ddots & \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & h_{L_h} & \cdots & h_1 & h_0 \end{bmatrix}, \qquad (15)$$

$\mathbf{A} = \text{Diag}\,[z_{n-L_c-L_h} \cdots z_n \cdots z_{n+L_f}]$, and $\mathbf{h} = [\,\underbrace{0 \cdots 0}_{L_c\ \text{zeros}}\ h_0 \cdots h_{L_f}]^T$ is the $(1 + L_c)^{\text{th}}$ column of $\mathbf{H}$. Clearly the weight vector of (14) is time-varying. In the classical MMSE-LE solution, x_n is assumed to be equally likely, meaning $\overline{x}_n = 0$ or $z_n = 1$ for all n. This would have reduced (14) to a time-invariant solution. Now, to complete the solution useful for turbo equalization, recall from Section II that y_n is to be generated while suppressing the current *a priori* LLR $L_a(x_n)$ sample to zero. Setting $L_a(x_n) = 0$ is equivalent to setting $\overline{x}_n = 0$, which in turns leads to $z_n = 1$. The final MMSE filter solution then can be constructed by replacing z_n in (14) by 1.

In an effort to reduce the complexity associated with time-varying filter implementation, however, the filter coefficients can be made to change only once per turbo iteration. A possible approximation that can be made in achieving this is to replace every z_i in (14) by the time average: $\overline{\overline{z}} = 1/N \sum_{k=0}^{N-1} z_k$ for the given iteration stage (N is the codeword size) [7]. Let the corresponding QTI filter coefficients be $\mathbf{c}_{qti}$:

$$\mathbf{c}_{qti} = \Big[\mathbf{H}\overline{\overline{\mathbf{A}}}\mathbf{H}^T + N_0\mathbf{I} \Big]^{-1} \mathbf{h} \qquad (16)$$

where $\overline{\overline{\mathbf{A}}}$ is the matrix replacing z_i for all $i \neq n$ in $\mathbf{A}$ with $\overline{\overline{z}}$ while setting $z_n = 1$. Clearly, $\mathbf{c}_{qti}$ is a time-invariant vector for a given value of $\overline{\overline{z}}$. The equalizer output is then

$$y_n = (\mathbf{r} - \overline{\mathbf{r}})^T \mathbf{c}_{qti} \qquad (17)$$

with the computation of $\overline{r}_i = \sum_l h_l \overline{x}_{i-l}$ in vector $\overline{\mathbf{r}}$ still based on the time-varying mean $\overline{x}_k = (1 - \delta_{k-n}) \tanh\{L_a(x_k)/2\}$, where $\delta_0 = 1$ while $\delta_l = 0$ for $l \neq 0$.

Assuming that y_n so obtained can be approximated as Gaussian when conditioned on a given value of the binary input symbol, the extrinsic LLR can be computed from (2). Write the equalizer output as $y_n = \beta x_n + v_n$, where β is a scaling factor that can be easily shown to be $\beta = \mathbf{h}^T \mathbf{c}_{qti}$ and v_n is the combination of the random noise and residual ISI, which is approximated as zero-mean Gaussian with variance

$$\sigma_{v_n}^2 = \text{var}\left[(\mathbf{r} - \overline{\mathbf{r}})^T \mathbf{c}_{qti}\right]\Big|_{z_n=0}$$
$$= \mathbf{c}_{qti}^T \left[\mathbf{H}\mathbf{A}'\mathbf{H}^T + N_0\mathbf{I}\right] \mathbf{c}_{qti} \qquad (18)$$

where $\mathbf{A}' = \text{Diag}\,[z_{n-L_c-L_h} \cdots z_{n-1}\ 0\ z_{n+1} \cdots z_{n+L_f}]$. The extrinsic LLR is set as $L_e(x_n) = 2\beta y_n / \sigma_{v_n}^2$. Note that while

β is time-invariant, $\sigma_{v_n}^2$ is time-varying and so is $\overline{\mathbf{r}}$ in the computation of y_n. While $\sigma_{v_n}^2$ can also be approximated by a quasi-time-invariant quantity (by replacing z_i for all $i \neq n$ in $\mathbf{A}'$ of (18) with $\overline{\overline{z}}$) [7], our simulation results in this paper reflect the time-varying $\sigma_{v_n}^2$ of (18).

2) Decision Feedback Equalizer of [17]: While MMSE-DFE design is well-established and not much could be added to the existing body of knowledge, it is worth clarifying the assumption made on the past decision statistics in the derivation of the forward and feedback filter taps in DFE. In [7], the DFE filter taps are obtained assuming that the variances associated with the past known symbols (due to perfect decisions) are all zero. Here we take a view that the statistics for the past decisions should be identical to the actual input symbol statistics, which is more in tune with the underlying principle of DFE design. It is in fact not necessary to assume zero-variances for the past symbols to derive the same results. We will also briefly summarize the technique of [17] for suppressing error propagation.

Let the forward and feedback filter coefficient vectors be $\mathbf{c} = [c_0\ c_{-1} \cdots c_{-L_f}]^T$ and $\mathbf{d} = [d_{L_d}\ d_{L_d-1} \cdots d_1]^T$, respectively. Given correct past decisions, past channel output (observation) samples are not correlated with the current and future observation samples while having no dependency on the current input. Thus, past observation samples do not provide any useful information for making decision on the current input symbol. Accordingly no causal forward taps are necessary. Also defining the composite vectors $\mathbf{u} \triangleq [x_{n-L_d} \cdots x_{n-1}\,|\,r_n \cdots r_{n+L_f}]^T \triangleq [\mathbf{x}_-^T\,|\,\mathbf{r}^T]^T$ and $\mathbf{f} \triangleq [-d_{L_d} \cdots - d_1\,|\,c_0 \cdots c_{-L_f}]^T \triangleq [-\mathbf{d}^T\,|\,\mathbf{c}^T]^T$, the ideal-feedback DFE output with a zero-mean output error can be expressed as $y_n = (\mathbf{u} - \overline{\mathbf{u}})^T \mathbf{f} + \overline{x}_n$. The MMSE solution can be obtained by solving $\overline{[(\mathbf{u} - \overline{\mathbf{u}})(\mathbf{u} - \overline{\mathbf{u}})^T]}\mathbf{f}_{opt} = \overline{(x_n - \overline{x}_n)(\mathbf{u} - \overline{\mathbf{u}})}$ or

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{D} \end{bmatrix} \begin{bmatrix} -\mathbf{d}_{opt} \\ \mathbf{c}_{opt} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ z_n \mathbf{h} \end{bmatrix}$$

where $\mathbf{A} = \overline{(\mathbf{x}_- - \overline{\mathbf{x}}_-)(\mathbf{x}_- - \overline{\mathbf{x}}_-)^T}$, $\mathbf{B} = \overline{(\mathbf{x}_- - \overline{\mathbf{x}}_-)(\mathbf{r} - \overline{\mathbf{r}})^T}$, $\mathbf{D} = \overline{(\mathbf{r} - \overline{\mathbf{r}})(\mathbf{r} - \overline{\mathbf{r}})^T}$ and $\mathbf{h} = [h_0 \cdots h_{L_f}]^T$. It is easy to show that the solutions are: $\mathbf{d}_{opt} = \mathbf{A}^{-1}\mathbf{B}\mathbf{c}_{opt}$ and $\mathbf{c}_{opt} = z_n(\mathbf{D} - \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{h}$.

Set $L_d = L_h$ (which lead to the optimal solution with the least number of coefficient taps) and define $\mathbf{A}_+ = \overline{(\mathbf{x}_+ - \overline{\mathbf{x}}_+)(\mathbf{x}_+ - \overline{\mathbf{x}}_+)^T}$ with $\mathbf{x}_+ = [x_n \cdots x_{n+L_f}]^T$. Making use of the relationship $\mathbf{r} = \mathbf{H}[(\mathbf{x}_- - \overline{\mathbf{x}}_-)^T\,|\,(\mathbf{x}_+ - \overline{\mathbf{x}}_+)^T]^T$, where $\mathbf{H}$ denotes the $L \times (L + L_d)$ channel response matrix of the form in (15) with $L = 1 + L_f$, we obtain

$$\mathbf{c}_{opt} = z_n \Big[\mathbf{H}_2 \mathbf{A}_+ \mathbf{H}_2^T + N_0\mathbf{I} \Big]^{-1} \mathbf{h},$$
$$\mathbf{d}_{opt} = \mathbf{H}_1^T \mathbf{c}_{opt} \qquad (19)$$

where the reduced channel matrices $\mathbf{H}_1$ and $\mathbf{H}_2$ take the first L_d columns and the remaining $1 + L_f$ columns of $\mathbf{H}$, respectively, and $\mathbf{A}_+ = \text{Diag}\,[z_n \cdots z_{n+L_f}]$. Note that this derivation does not resort to the assumption $\mathbf{A} = \mathbf{0}$ made in [7]. The resulting taps in (19) are time-varying; so, like in the case of LE, the filter coefficients are further constrained to be time-invariant at each iteration stage. This means that every

z_i in $\mathbf{A}_+$ in the computation of (19) is replaced by the time-average $\overline{\overline{z}} = (1/N) \sum_{k=0}^{N-1} z_k$ taken anew at each iteration stage. Let the corresponding QTI filter coefficient vectors be $\mathbf{c}_{qti}$ and $\mathbf{d}_{qti}$:

$$\mathbf{c}_{qti} = \left[\mathbf{H}_2 \overline{\overline{\mathbf{A}}}_+ \mathbf{H}_2^T + N_0 \mathbf{I} \right]^{-1} \mathbf{h},$$
$$\mathbf{d}_{qti} = \mathbf{H}_1^T \mathbf{c}_{qti}. \tag{20}$$

where $\overline{\overline{\mathbf{A}}}_+$ is the matrix replacing all z_i terms appearing in $\mathbf{A}_+$ with $\overline{\overline{z}}$ except setting $z_n = 1$.

The DFE output is

$$y_n = (\mathbf{r} - \bar{\mathbf{r}})^T \mathbf{c}_{qti} - (\hat{\mathbf{x}}_-)^T \mathbf{d}_{qti} \tag{21}$$

with the computation of $\overline{r}_i = \sum_l h_l \overline{x}_{i-l}$ in $\bar{\mathbf{r}}$ still based on the time-varying mean $\overline{x}_k = (1 - u_{k-n}) \tanh \{ L_a(x_k)/2 \}$, where $u_l = 1$ for $l \leq 0$ and $u_l = 0$ for $l > 0$. The vector $\hat{\mathbf{x}}_-$ consists of hard decisions on L_d past symbols.

Instead of making the usual assumption that the feedback decisions are correct and that y_n is conditionally Gaussian, the DFE design in [17] considers the possibility of incorrect decisions affecting y_n in formulating the extrinsic LLR. Write the DFE output as $y_n = \beta x_n + i_n + v_n$, where $\beta = \mathbf{h}^T \mathbf{c}_{qti}$ with $\mathbf{h} = [h_0 \cdots h_{L_f}]^T$, i_n is due to incorrect past decisions and v_n is the combination of the random noise and residual ISI, which is again approximated as Gaussian. In [17], the extrinsic LLR is constructed as a combination of two conditional extrinsic LLRs computed for two separate cases: one without decision errors and one with decisions errors. Specifically, we have

$$L_e(x_n) = \ln \left\{ \frac{\exp \left(L_e(x_n | i_n = 0) \right) \Pr(i_n = 0)}{1 + \exp \left(L_e(x_n | i_n = 0) \right)} \right.$$
$$\left. + \frac{\exp \left(L_e(x_n | i_n \neq 0) \right) \Pr(i_n \neq 0)}{1 + \exp \left(L_e(x_n | i_n \neq 0) \right)} \right\}$$
$$- \ln \left\{ \frac{\Pr(i_n = 0)}{1 + \exp \left(L_e(x_n | i_n = 0) \right)} \right.$$
$$\left. + \frac{\Pr(i_n \neq 0)}{1 + \exp \left(L_e(x_n | i_n \neq 0) \right)} \right\}. \tag{22}$$

As usual, $L_e(x_n | i_n = 0) = 2\beta y_n / \sigma_{v_n}^2$, but, $L_e(x_n | i_n \neq 0) = \ln \left\{ \sum_j \frac{\exp \left(L_e(x_n | i_n^{(j)}) \right) \Pr(i_n^{(j)})}{\left\{ 1 + \exp \left(L_e(x_n | i_n^{(j)}) \right) \right\} \Pr(i_n \neq 0)} \right\} - \ln \left\{ \sum_j \frac{\Pr(i_n^{(j)})}{\left\{ 1 + \exp \left(L_e(x_n | i_n^{(j)}) \right) \right\} \Pr(i_n \neq 0)} \right\} \simeq 2\varphi_n / (1 + |\varphi_n|)$ for the case where $i_n \neq 0$ where $i_n^{(j)}$ is the possible non-zero value of i_n corresponding to the jth error-pattern associated with the L_d past symbols and $\varphi_n = \beta (y_n - \overline{(i_n | i_n \neq 0)}) / \sigma_{v_n}^2$ [17]. The conditional mean $\overline{(i_n | i_n \neq 0)} = \overline{i}_n / \Pr(i_n \neq 0)$ can be estimated using the *a posteriori* LLRs associated with L_d past decisions. The variance of v_n is obtained as

$$\sigma_{v_n}^2 = \text{var} \left[(\mathbf{u} - \bar{\mathbf{u}})^T \mathbf{f}_{qti} \right] \big|_{z_n = 0}$$
$$= \mathbf{c}_{qti}^T \left[\mathbf{H}_2 \mathbf{A}'_+ \mathbf{H}_2^T + N_0 \mathbf{I} \right] \mathbf{c}_{qti} \tag{23}$$

where $\mathbf{f}_{qti} \triangleq [-\mathbf{d}_{qti}^T | \mathbf{c}_{qti}^T]^T$ and $\mathbf{A}'_+ = \text{Diag} [0 \; z_{n+1} \cdots z_{n+L_f}]$. Again, $\sigma_{v_n}^2$ is time-varying here.

3) Bi-directional DFE of [17]: The BiDFE of [17] is based on using two DFEs running in opposite directions. The extrinsic LLR of each DFE is obtained using the method described above and combined to yield the BiDFE's extrinsic LLR:

$$L_e(x_n) = \frac{1}{(1 + \rho_{bi})} \Big(L_{e,f}(x_n) + L_{e,b}(x_n) \Big) \tag{24}$$

where the subscripts f and b are for the forward and backward DFEs, respectively, and ρ_{bi} is the noise correlation coefficient between two DFEs and can be estimated using time-average, [17].

B. Performance Potentials of Suboptimal Equalizers Under Perfect A Priori Information

Let us consider the ideal condition for the equalizer where the perfect *a priori* information is available, i.e., $\overline{x}_n = x_n$ [or $L_a(x_n) = \pm \infty$] for all n. This condition may be satisfied when many iterations are performed at sufficiently high channel SNRs in turbo equalization. Under this condition, it has already been discussed in [7] that the LE based on the time-varying optimal filter provides the ideal matched filter bound. Since the DFE and the BiDFE cannot be worse than the LE under this ideal condition, it follows that all equalizers with time-varying filters provide the matched filter performance. It turns out that the same is true for the QTI-based equalizers as well, while the TI filters cannot achieve the matched filter performance even under the ideal condition.

Define the output SNR for the QTI-LE associated with an infinite number of turbo iterations as $\text{SNR}_{\infty,\text{QTI–LE}} = \lim_{\overline{\overline{z}} \to 0} \beta^2 / \sigma_{v_n}^2$, assuming repeated iterations lead to perfect *a priori* information (and thus perfect time-averaged *a priori* information as well). Apparently, as $z_k \to 0$ for all k, $\overline{\overline{z}} \to 0$. The filter weight vector of (16) then becomes $\mathbf{c}_{qti} \to \mathbf{h} / (1 + N_0)$ and the variance of v_n in (18) takes the form $\sigma_{v_n}^2 \to N_0 \mathbf{c}_{qti}^T \mathbf{c}_{qti}$. Since $\beta = \mathbf{h}^T \mathbf{c}_{qti}$, it is easy to see that $\text{SNR}_{\infty,\text{QTI–LE}} = \mathbf{h}^T \mathbf{h} / N_0 = 1/N_0$, which corresponds to the matched filter performance. Again, the DFE gives the same result under the QTI constraint. For the BiDFE, the noise correlation coefficient between the normal DFE and the time-reversed DFE is 1, which implies that they produce the same equalized output and there is no SNR advantage of combining [17]. Accordingly, the LE, the DFE and the BiDFE all produce the same equalized outputs and the output SNRs with the QTI filters are given by

$$\text{SNR}_{\infty,\text{QTI–LE}} = \text{SNR}_{\infty,\text{QTI–DFE}}$$
$$= \text{SNR}_{\infty,\text{QTI–BiDFE}}$$
$$= 1/N_0.$$

While all equalizer schemes can achieve the matched filter performance under the ideal condition, we stress that their realized performances are significantly different in practice [17].

For the TI filters, on the other hand, the maximum achievable output SNRs can be found as follows. First consider the LE. The filter weight vector is $\mathbf{c}_{tiLE} = [\mathbf{H}\mathbf{H}^T + N_0 \mathbf{I}]^{-1} \mathbf{h}$. Keep in mind that the equalizers based on TI filters do not

make use the *a priori* information. As $z_k \to 0$ for all k, $\sigma_{v_n}^2 \to N_0 \mathbf{c}_{tiLE}^T \mathbf{c}_{tiLE}$. The output SNR thus becomes

$$
\begin{aligned}
\mathrm{SNR}_{\infty,\mathrm{TI-LE}} &= \lim_{\overline{z} \to 0} \frac{\beta^2}{\sigma_{v_n}^2} = \frac{1}{N_0} \frac{[\mathbf{h}^T \mathbf{c}_{tiLE}]^2}{\mathbf{c}_{tiLE}^T \mathbf{c}_{tiLE}} \\
&= \frac{1}{N_0} \frac{[\mathbf{h}^T \mathbf{P} \mathbf{h}]^2}{\mathbf{h}^T \mathbf{P}^T \mathbf{P} \mathbf{h}}
\end{aligned}
\tag{25}
$$

where $\mathbf{P} = [\mathbf{H}\mathbf{H}^T + N_0\mathbf{I}]^{-1}$. For the DFE, we get a similar expression:

$$
\begin{aligned}
\mathrm{SNR}_{\infty,\mathrm{TI-DFE}} &= \frac{[\mathbf{h}^T \mathbf{c}_{tiDFE}]^2}{N_0 \mathbf{c}_{tiDFE}^T \mathbf{c}_{tiDFE}} \\
&= \frac{1}{N_0} \frac{[\mathbf{h}^T \mathbf{P}_2 \mathbf{h}]^2}{\mathbf{h}^T \mathbf{P}_2^T \mathbf{P}_2 \mathbf{h}}
\end{aligned}
\tag{26}
$$

where $\mathbf{P}_2 = [\mathbf{H}_2 \mathbf{H}_2^T + N_0\mathbf{I}]^{-1}$. For the BiDFE, it can be shown that $\mathrm{SNR}_{\infty,\mathrm{TI-BiDFE}} = \{2/(1+\rho)\}\mathrm{SNR}_{\infty,\mathrm{TI-DFE}}$ where ρ is the noise correlation coefficient between the normal DFE and the time-reversed DFE outputs under perfect *a priori* information and is given in [17] as a fucntion of the two DFE filter coefficients. It is easy to see that

$$
\begin{aligned}
\mathrm{SNR}_{\infty,\mathrm{TI-LE}} &\leq \mathrm{SNR}_{\infty,\mathrm{TI-DFE}} \\
&\leq \mathrm{SNR}_{\infty,\mathrm{TI-BiDFE}} \\
&\leq 1/N_0
\end{aligned}
$$

where the equalities hold in the memoryless AWGN channel.

In order to incorporate the above output SNR analysis with the EXIT chart analysis, one can compute the mutual information (MI) with these output SNRs. Because only the Gaussian noise term remains in the equalized output under the perfect *a priori* information, the MI is simply given as

$$
C_b(\mathrm{SNR}) \triangleq 1 - \int_{-\infty}^{\infty} \frac{e^{-\tau^2/2}}{\sqrt{2\pi}} \log_2 \left\{ 1 + e^{-2\tau\sqrt{\mathrm{SNR}} - 2\mathrm{SNR}} \right\} d\tau
\tag{27}
$$

where $C_b(\mathrm{SNR})$ is the symmetric information rate of the binary-input Gaussian channel for a given value of SNR. Note that the MI computed by substituting $\mathrm{SNR} = \mathrm{SNR}_\infty$ in (27) is the maximum attainable MI by an equalizer. The corresponding numerical results will be shown to be consistent with the the simulated MI trajectories as will be presented for some selected cases toward the end of Section VI.

VI. SIMULATION RESULTS AND DISCUSSION

In this section, simulation results of the proposed SISE equalization schemes for both uncoded and coded systems are presented. The transmitted symbols are modulated with binary phase-shift keying (BPSK) so $x_n \in \{\pm 1\}$. The message bit length is 2^{11}. We also assume that the noise is AWGN, and the noise variance and the channel impulse response are perfectly known to the receiver, except for one channel example in which we deliberately introduce mismatch between the actual channel response and the one assumed by the receiver. The channel we consider here are severe ISI channels that show deep nulls and valleys within the signal band. When the ISI is not severe, the LE scheme of [8] already provides near-optimal performance, as mentioned in the Introduction section.
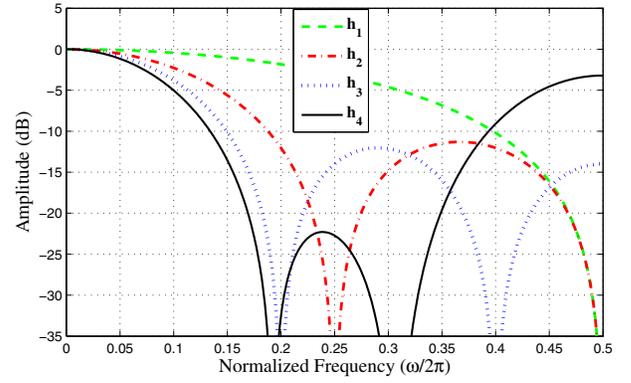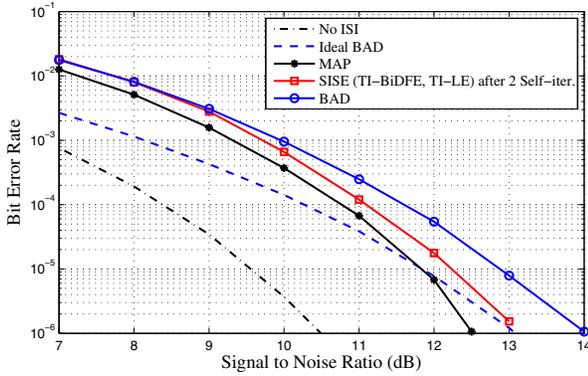

Fig. 3. Frequency Magnitude Response of the ISI Channels: $\mathbf{h_1} = (1/\sqrt{6})[1 \quad 2 \quad 1]^T$, $\mathbf{h_2} = (1/\sqrt{44})[1 \quad 2 \quad 3 \quad 4 \quad 3 \quad 2 \quad 1]^T$, $\mathbf{h_3} = (1/\sqrt{85})[1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 4 \quad 3 \quad 2 \quad 1]^T$, and $\mathbf{h_4} \approx 0.0795[1 \ 0.707 \ 4 \ 2.121 \ 7.25 \ 3 \ 7.25 \ 2.121 \ 4 \ 0.707 \ 1]^T$.

A. Uncoded System

The impulse response of the ISI channel $\mathbf{h_1} = (1/\sqrt{6})[1 \quad 2 \quad 1]^T$ discussed in [27] is used for the uncoded system. This channel with its system response (D-transform of the impulse response) given by $H_1(D) = (1/\sqrt{6})(1+D)^2$ has a second-order spectral null at the right edge of the Nyquist band, as shown in Fig. 3. Three different equalizer types are simulated for this channel. The SISE method is the self-iterating soft equalizer algorithm described in Section III-D. Specifically, the SISO BiDFE algorithm of [17] is adopted as the main equalizer and the SISO LE of [8] is used for the branch equalizer. Moreover, when the final hard decisions are released by the BiDFE, the arbitration criterion of [21] with window size 15 is employed; the symbol sequence is decided among the estimated sequences of two DFEs based on which candidate shows the smaller MSE in a window centered around the symbol of interest. Finally, two self-iterations are applied. The "BAD" method is the algorithm of [21] employing two classical DFEs running in opposite directions with an arbitration strategy as described above. The MAP equalizer is the optimal equalizer implemented via the well-known BCJR algorithm [2]. Each DFE in the BiDFE or BAD consists of 13 feedforward taps and 2 feedback taps ($L_c = 12$ and $L_d = 2$) while the LE uses 15 taps ($L_c = 7$ and $L_f = 7$) for $\mathbf{h_1}$. Fig. 4 shows the performance comparison. As the figure shows, the proposed SISE algorithm shows the superior performance to the BAD method of [21] and approaches the performance of "Ideal BAD" with perfect past decisions and *a priori* information in either direction. Increasing the number of filter taps for BAD did not give any noticeable performance gain.
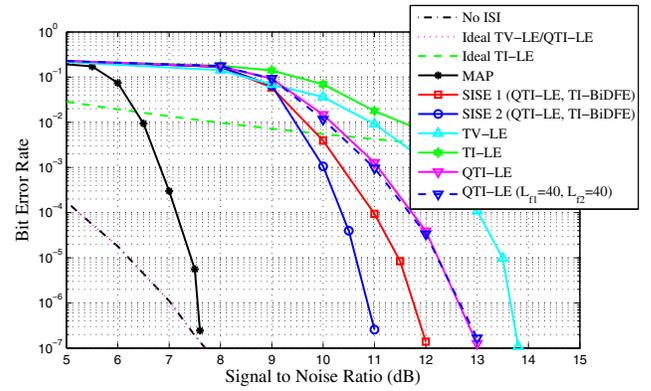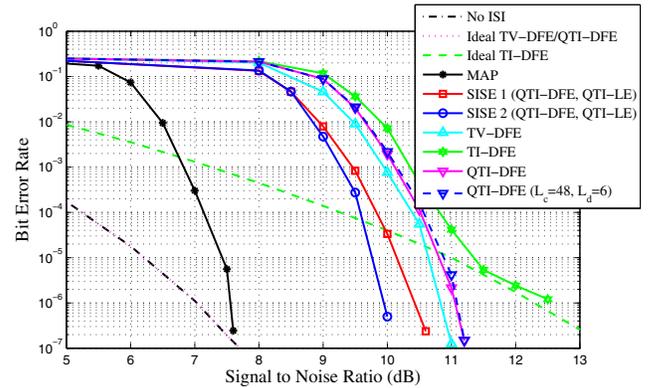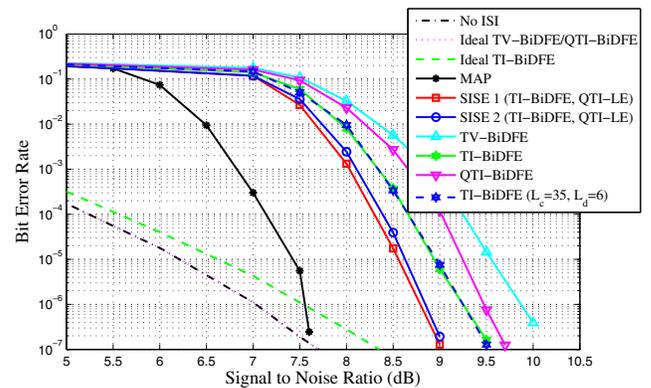
B. Coded System

In this subsection, simulation results of the SISE schemes in turbo equalizer setting are presented. The transmitted symbols are encoded with a recursive rate-$1/2$ convolutional code encoder with the parity generator polynomial $(1 + D^2)/(1 + D + D^2)$. The size of the coded data packet (and thus the size of the interleaver used for turbo iteration between the equalizer and the decoder) is 2^{12} bits. A

Fig. 4.   BER curves on Channel $\mathbf{h_1}$.

Fig. 5.   LE-based BER curves on Channel $\mathbf{h_2}$ after 20 outer iterations.

Fig. 6.   DFE-based BER curves on Channel $\mathbf{h_2}$ after 20 outer iterations.

Fig. 7.   BiDFE-based BER curves on Channel $\mathbf{h_2}$ after 20 outer iterations.

random interleaver is employed. The impulse response of the severe ISI channel $\mathbf{h_2} = (1/\sqrt{44})[1 \quad 2 \quad 3 \quad 4 \quad 3 \quad 2 \quad 1]^T$ investigated in [17] as well as a more severe ISI channel of $\mathbf{h_3} = (1/\sqrt{85})[1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 4 \quad 3 \quad 2 \quad 1]^T$ are used for evaluating and comparing the performances of turbo equalizers. While both of these channels as well as $\mathbf{h_1}$ represent samples of a triangular time function, their spectral characteristics are quite different, as can be seen in Fig. 3. Their corresponding system polynomials are $H_2(D) = (1/\sqrt{44})(1 + D)^2(1 + D^2)^2$ and $H_3(D) = (1/\sqrt{85})(1 + D + D^2 + D^3 + D^4)^2$, which reveal that these channels possess second-order nulls either within the Nyquist band or at its edge. We also select an additional channel whose time-domain shape is quite different from these channels. Specifically, we place one second-order null and one third-order null well inside the Nyquist band according the system polynomial $H_4(D) = (\sqrt{32/5061}) (1 + D/\sqrt{2} + D^2)^3 (1 - D/\sqrt{2} + D^2)^2$. Its magnitude response in frequency is also included in Fig. 3. The corresponding impulse response is $\mathbf{h_4} = (\sqrt{32/5061})[1 \quad 1/\sqrt{2} \quad 4 \quad 3/\sqrt{2} \quad 29/4 \quad 17/4\sqrt{2} \quad 29/4 \quad 3/\sqrt{2} \quad 4 \quad 1/\sqrt{2} \quad 1]^T$.

In the error rate figures, curves labeled by "SISE 1" and "SISE 2" correspond to the turbo SISE algorithms described in Section IV. The label "(M, B)" in the legend denotes a specific 'M' algorithm as the main equalizer and a 'B' algorithm as the branch equalizer. For instance, "SISE 2 (QTI-LE, QTI-DFE)" denotes the turbo SISE 2 algorithm with the LE with a quasi-time-invariant filter as the main equalizer and the DFE with quasi-time-invariant filters as the branch equalizer. The label "TV-LE" denotes the MMSE linear equalizer with the time-varying filter (or the "exact MMSE" of [8]). The label "Ideal" indicates the performance of an equalizer with perfect *a priori* information.

The straightforward LLR mapping method [y_n-to-$L_e(x_n)$ conversion] of [8] is adopted for the LE while the LLR mapping method of [17] is used for the DFE and BiDFE. Moreover, the DFE (and each DFE in the BiDFE) consists of 21 feedforward taps and 6 feedback taps ($L_c = 20$ and $L_d = 6$) on $\mathbf{h_2}$; $L_c = 20$ and $L_d = 8$ on $\mathbf{h_3}$; and $L_c = 20$ and $L_d = 10$ on $\mathbf{h_4}$. The LE uses 27 taps ($L_c = 13$ and $L_f = 13$) for $\mathbf{h_2}$, 29 taps ($L_c = 14$ and $L_f = 14$) for $\mathbf{h_3}$, and 31 taps ($L_c = 15$ and $L_f = 15$) for $\mathbf{h_4}$. Again, "MAP" refers

to the optimal equalizer implemented via the BCJR algorithm. Finally, the decoder is implemented using the BCJR algorithm and 20 turbo outer iterations are applied to achieve the full potential of each turbo equalization system.

Figs. 5, 6, and 7 show the performance of several turbo equalizers on $\mathbf{h_2}$. The performance of the proposed iterative SISE algorithms is compared with the performance achieved when the equalizer consists only of a single main equalizer.

Fig. 5 shows performance comparison with LE-based single equalizers. Among the single LE schemes, the QTI filter design method gives the best performance, outperforming even the TV filter method (the reason for which is explained below). When the number of filter taps increases to 81 (40 causal and

40 anticausal taps) versus a total of 27, the QTI-LE method does not provide any performance gain, indicating that using 27 taps for this channel already realizes the QTI-LE's full potential. Of the two SISE methods showing superior performance to the single LE, SISE 2 is better and is an obvious winner given its lower complexity and latency compared to SISE 1. Both SISE schemes employ TI-BiDFE as the sole branch equalizer. Given that each DFE in BiDFE has a total of 27 taps, each SISE scheme requires 81 taps overall.

Fig. 6 tells a similar story but the comparison is now with the DFE schemes. Accordingly, the main equalizer of the SISE methods is also set up with the DFE, under the QTI filter design scheme. The branch equalizer is a QTI-LE. Among the single equalizers, this time, the TV filter method performs slightly better than the QTI filter scheme (again, increasing the number of QTI filter taps does not give any performance boost). Still, the SISE methods give the best performance, other than the MAP-based turbo equalizer, with SISE 2 once again coming out ahead. Fig. 7 presents results that compare schemes based on the BiDFE. The QTI-filter-based single equalizer is better than the TV filter method and, again, increasing the filter length does not improve performance. As for the branch equalizer, the SISE schemes use the QTI-LE. This time, SISE 1 performs better than SISE 2, albeit by a very small margin.

A comparison of the simulation results in Figs. 5 and 7 reveals that the performance of "SISE (TI-BiDFE, QTI-LE)" is considerably better than that of "SISE (QTI-LE, TI-BiDFE)". On the other hand, among the single equalizers, TI-BiDFE outperforms QTI-LE by a large margin. It turns out that the extrinsic LLR quality of the main equalizer is an important factor determining the overall performance of the SISE algorithms, as the extrinsic LLRs of the main equalizer are passed to the decoder as well as the branch equalizers. Therefore, in the design of the SISE algorithm, the equalizer showing the best BER performance (or LLR quality) should be chosen as the main equalizer.

When different filter types are compared, the QTI filters sometimes provide even better BER performance than the TV filters (as shown in Figs. 5 and 7), which contradict the simulation results of [7]. This is mainly due to the fact that when inaccurate *a priori* information arrives, the optimal TV filters more easily fail to produce reliable extrinsic information than the QTI filters. In other words, the TV (or exact MMSE) filters of [7] are designed on the premise that the incoming *a priori* information is accurate; when the incoming *a priori* information is not reliable as in severe ISI channels under consideration in this paper, they tend to generate low-quality LLRs. Furthermore, a few large incoming LLRs in wrong direction can degrade the overall turbo equalization performance quickly as we have observed during our simulation. In the case of the QTI filter, a small number of large-magnitude LLRs tend to get averaged-down. Also, recall that an equalizer with QTI filters can also achieve the matched filter performance if enough turbo iterations are performed. Between the LE and the DFE, this effect is more pronounced with the LE, although even in the case of the DFE (as seen in Figs. 6) the performance advantage of TV design is small over the QTI method. In the case of the BiDFE, this sensitivity plays an
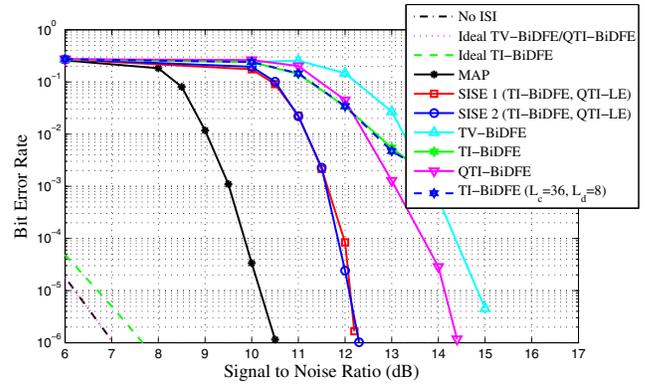

Fig. 8.   BiDFE-based BER curves on Channel $\mathbf{h_3}$ after 20 outer iterations.
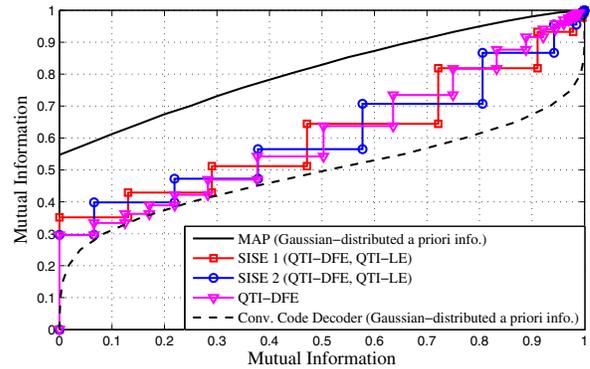

Fig. 9.   DFE-based EXIT chart on Channel $\mathbf{h_2}$ at SNR=10 dB.

even bigger role in determining the overall performance. As indicated in Fig. 7, TI-BiDFE shows better performance than BiDFE with any other filter types since it would be better not to update the filter taps at all when the *a priori* information is unreliable at low SNRs.

Fig. 8 shows comparison in the extremely severe ISI channel of $\mathbf{h_3}$. Among the single equalizers, TI-BiDFE schemes exhibit clear failure because the erroneously generated *a priori* LLRs from the decoder during the iterations cause more errors in the subsequent turbo iterations, while QTI-BiDFE is better than TV-BiDFE. Both SISE schemes show robust performance, again lagging in performance only by the optimal MAP-based scheme.

The performance of the turbo equalizers are analyzed by using the EXIT chart [26], a diagram demonstrating the MI transfer characteristics of the two constituent modules that exchange soft information. In the EXIT charts, the behavior of the channel equalizer is described with its input and output on the horizontal and vertical axis, respectively, while the behaviour of the decoder is described in the opposite way. The EXIT chart curves typically define a path for the MI trajectory to move up during iterative processing of soft information. Moreover, the number of stairs that a given MI trajectory (averaged over 1000 sample codeword blocks here) takes to reach the highest value indicates the necessary number of iterations toward convergence.

In order to avoid excessive cluttering, only the trajectories of SISE 1, SISE 2 and the single equalizer are plotted in Figs.
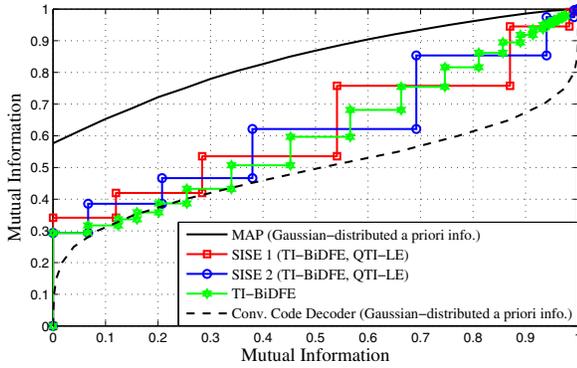
Fig. 10. BiDFE-based EXIT chart on Channel $\mathbf{h_3}$ at SNR=13 dB.
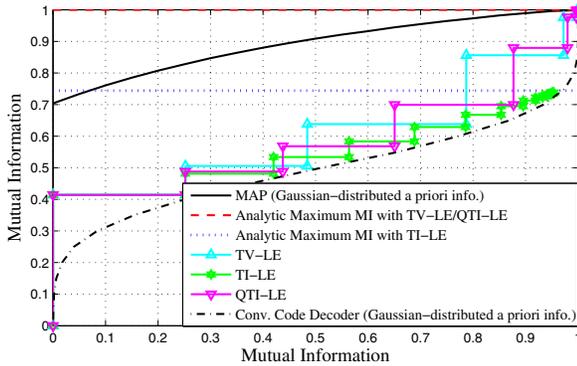

Fig. 11. EXIT Chart of LE with different filter types on Channel $\mathbf{h_2}$ at SNR=12 dB.

9 and 10. They describe the EXIT charts corresponding to $\mathbf{h_2}$ at a 10 dB SNR when QTI-DFE is used for the main equalizer and the EXIT chart on $\mathbf{h_3}$ at a 13 dB SNR when TI-BiDFE is used for the main equalizer, respectively. As the figures indicate, both SISE algorithms appear to widen the EXIT chart tunnels with the aid of the branch equalizers while the trajectory of the single main equalizer by itself tends to get stuck in the bottleneck regions. Accordingly, both SISE algorithms reach the maximum MI value at a considerably smaller number of steps than the single equalizer scheme, indicating a faster convergence for the SISE schemes.

The EXIT chart of the LE with various filter types on $\mathbf{h_2}$ at a 12 dB SNR is also shown in Fig. 11. As the figure shows, the trajectory of QTI-LE shows a clear path from 0 bits of MI to 1 bit of MI, in the same way TV-LE does while TI-LE fails to approach 1 bit of MI even though it keeps moving up towards its own maximum MI limit as the number of iterations increases. A word of caution in interpreting Fig. 11 is in order. The BER performance is not always consistent with the EXIT chart performance; the MI in the EXIT chart depends on the overall quality of the extrinsic LLRs and a few erroneously generated extrinsic LLRs have little effect on the MI, while this is not the case for BER performance.

Finally, the performance of SISE 2 with a QTI-LE main equalizer and a QTI-DFE branch equalizer is compared with that of the stand-alone QTI-LE for the $\mathbf{h_4}$ channel. The results are summarized in Fig. 12. The no-ISI reference curve is also
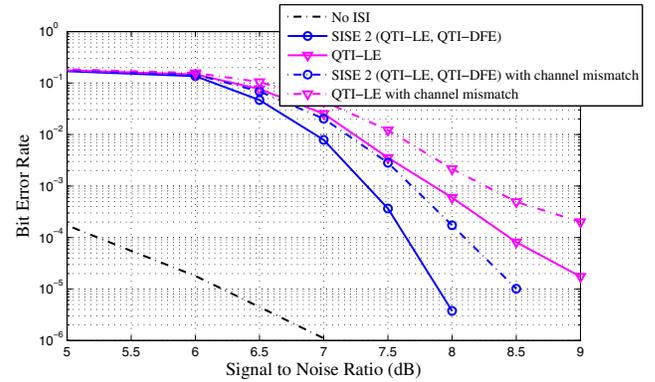

Fig. 12. LE-based BER curves on Channel $\mathbf{h_4}$ after 20 outer iterations: with and without channel mismatch.

shown. Also included in the plot are the error rate curves of the same equalizers in the presence of channel mismatch, which show each scheme's sensitivity to potential channel estimation errors in practice. Increasing the filter length for QTI-LE did not provide any performance again, so the QTI-LE curves shown in the figure represents the maximum performance potential of the QTI-LE. The performance gain of the SISE is clearly observed both with and without channel estimation errors. In plotting the channel-mismatched curves of Fig. 12, the equalizers operate with an erroneous assumption that the channel response is given by $\hat{\mathbf{h}}_4 = [\hat{h}_0 \ \hat{h}_1 \ \cdots \ \hat{h}_{10}]$ with $\hat{h}_i = \{1 + (0.1/\sqrt{SNR})\varepsilon_i\}h_i$ where ε_i is a zero-mean unit-variance Gaussian random variable. In this modeling, the channel estimation error is allowed to drop with increasing SNR. The channel estimation error changes from one channel tap to next, as well as from one packet transmission to next. The channel errors, however, remain fixed during the turbo/self-equalization process.

VII. CONCLUSIONS

In this paper, we proposed self-iterating soft equalizers which can be employed in turbo equalization systems to improve performance in very severe ISI channels. The proposed algorithms are designed to utilize the extrinsic information of other serially concatenated suboptimal equalizers by properly handling correlation on the information generated by other equalizers. The proposed algorithms show robust performance, even when the constituent suboptimal equalizers are individually weak. The proposed SISE schemes also provide good BER performance in uncoded systems.

REFERENCES

[1] C. Douillard, M. Jezequel, C. Berrou, A. Picart, P. Didier, and A. Glavieux, "Iterative correction of intersymbol interference: turbo equalization," *European Trans. Telecommun.*, vol. 6, no. 5, pp. 507–511, Sep.-Oct. 1995.

[2] 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.

[3] J. Hagenauer and P. Hoeher, "A Viterbi algorithm with soft-decision outputs and its applications," in *Proc. 1989 IEEE Global Telecommunications Conference*, pp. 1680–1686.

[4] 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, Jul. 1999.

[5] A. Chan and G. Wornell, "A class of block-iterative equalizers for intersymbol interference channels: fixed channel results," *IEEE Trans. Commun.*, vol. 49, no. 11, pp. 1966–1976, Nov. 2001.

[6] Z. Wu and J. Cioffi, "Low complexity iterative decoding with decision-aided equalization for magnetic recording channels," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 4, pp. 699–708, Apr. 2001.

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

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

[9] C. Laot, A. Glavieux, and J. Labat, "Turbo equalization: adaptive equalization and channel decoding jointly optimized," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 9, pp. 1744–1752, Sep. 2001.

[10] M. Honig, G. Woodward, and Y. Sun, "Adaptive iterative multiuser decision feedback detection," *IEEE Trans. Wireless Commun.*, vol. 3, no. 2, pp. 477–485, Mar. 2004.

[11] C. Laot, R. Le Bidan, and D. Leroux, "Low-complexity MMSE turbo equalization: a possible solution for EDGE," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 965–974, May 2005.

[12] Y. Sun, V. Tripathi, and M. Honig, "Adaptive turbo reduced-rank equalization for MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2789–2800, Nov. 2005.

[13] J. Moon and F. R. Rad, "Turbo equalization via constrained-delay APP estimation with decision feedback," *IEEE Trans. Commun.*, vol. 53, no. 12, pp. 2102–2113, Dec. 2005.

[14] F. R. Rad and J. Moon, "Turbo equalization utilizing soft decision feedback," *IEEE Trans. Magnetics*, vol. 41, no. 10, pp. 2998–3000, Oct. 2005.

[15] R. Lopes and J. Barry, "The soft-feedback equalizer for turbo equalization of highly dispersive channels," *IEEE Trans. Commun.*, vol. 54, no. 5, pp. 783–788, May 2006.

[16] S. Jeong and J. Moon, "Turbo equalization based on bi-directional DFE," in *Proc. 2010 IEEE International Conference on Communications*.

[17] S. Jeong and J. Moon, "Soft-in soft-out DFE and bi-directional DFE," *IEEE Trans. Commun.*, vol. 59, no. 10, pp. 2729–2741, Oct. 2011.

[18] P. Supnithi, R. Lopes, and S. McLaughlin, "Reduced-complexity turbo equalization for high-density magnetic recording systems," *IEEE Trans. Magnetics*, vol. 39, no. 5, pp. 2585–2587, Sep. 2003.

[19] J. Jiang, C. He, E. Kurtas, and K. Narayanan, "Performance of soft feedback equalization over magnetic recording channels," in *Proc. 2006 Intermag*, pp. 795.

[20] J. Balakrishnan and C. Johnson, Jr., "Bidirectional decision feedback equalizer: infinite length results," in *Proc. 2001 Asilomar Conf. on Signals, Systems, and Computers*, pp. 1450–1454.

[21] J. Nelson, A. Singer, U. Madhow, and C. McGahey, "BAD: bidirectional arbitrated decision-feedback equalization," *IEEE Trans. Commun.*, vol. 53, no. 2, pp. 214–218, Feb. 2005.

[22] L. Papke, P. Robertson, and E. Villebrun, "Improved decoding with the SOVA in a parallel concatenated (turbo-code) scheme," in *Proc. 1996 IEEE International Conference on Communications*, pp. 102–106.

[23] C. Huang and A. Ghrayeb, "A simple remedy for the exaggerated extrinsic information produced by the SOVA algorithm," *IEEE Trans. Wireless Commun.*, vol. 5, no. 5, pp. 996–1002, May 2006.

[24] S. Jeong and J. Moon, "Self-iterating soft equalizer," in *Proc. 2011 IEEE Global Telecommunications Conference*.

[25] S. Ariyavisitakul, "A decision feedback equalizer with time-reversal structure," *IEEE J. Sel. Areas Commun.*, vol. 10, no. 3, pp. 599–613, Apr. 1992.

[26] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, Oct. 2001.

[27] J. Proakis, *Digital Communications*, 4th ed. McGraw-Hill Higher Education, 2000, pp. 631.
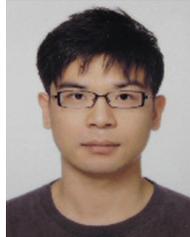
Seongwook Jeong received the B.S. degree with honor in electrical engineering from Seoul National University, Seoul, Korea, in 2004, the M.S. degree from the University of Southern California, Los Angeles in 2005, and the Ph.D. degree from the University of Minnesota, Minneapolis in 2011. From 2006 to 2011, he was a research assistant with the Department of Electrical and Computer Engineering, University of Minnesota. He also worked at Link-A-Media Devices and LSI Corp. during the summer of 2008 and 2011, respectively. Since December 2011, he has been a Senior Engineer with LSI Corp., San Jose, CA. His research interests include information theory, coding theory, and equalization with application to wireless communication and data storage systems.

Jaekyun Moon received the Ph.D. degree in electrical and computer engineering at Carnegie Mellon University, Pittsburgh, PA, USA. He is currently a Professor of electrical engineering at KAIST. From 1990 through early 2009, he was with the faculty of the Department of Electrical and Computer Engineering at the University of Minnesota, Twin Cities. He consulted as Chief Scientist for DSPG, Inc. from 2004 to 2007. He also worked as Chief Technology Officer at Link-A-Media Devices Corporation. His research interests are in the area of channel characterization, signal processing and coding for data storage and digital communication. Prof. Moon received the McKnight Land-Grant Professorship from the University of Minnesota. He received the IBM Faculty Development Awards as well as the IBM Partnership Awards. He was awarded the National Storage Industry Consortium (NSIC) Technical Achievement Award for the invention of the maximum transition run (MTR) code, a widely-used error-control/modulation code in commercial storage systems. He served as Program Chair for the 1997 IEEE Magnetic Recording Conference. He is also Past Chair of the Signal Processing for Storage Technical Committee of the IEEE Communications Society. In 2001, he cofounded Bermai, Inc., a fabless semiconductor start-up, and served as founding President and CTO. He served as a guest editor for the 2001 IEEE JSAC issue on Signal Processing for High Density Recording. He also served as an Editor for IEEE TRANSACTIONS ON MAGNETICS in the area of signal processing and coding for 2001-2006. He is an IEEE Fellow.