Novel Graph-Based Algorithms for Soft-Output Detection over Dispersive Channels

Dario Fertonani, Alan Barbieri, and Giulio Colavolpe
Università di Parma, SPADiC Lab, Dipartimento di Ingegneria dell’Informazione, Parma - ITALY
Email: fertonani@tlc.unipr.it

Abstract—We address the design of low-complexity algorithms for soft-output detection over channels impaired by intersymbol interference. Unlike most works with similar aims, which assume the presence of the whitened matched filter at the receiver (Forney approach), algorithms that can directly work on the matched filter output (Ungerboeck approach) are considered. We introduce a novel (cyclic) factor graph describing the channel and, by applying the sum-product algorithm to it, we derive soft-output detection schemes that can provide impressive complexity reductions with respect to the benchmark algorithms, since their complexity is linear, instead of exponential, in the channel memory. Finally, we report simulation results proving that the performance of the proposed algorithms makes them appealing for turbo equalization in various practical scenarios.

I. INTRODUCTION

In several communication systems of practical interest, the received signal is affected by intersymbol interference (ISI), due either to frequency-selective transmission media or to the (intentional) use of modulation pulses that generate ISI [1]. To date, the receivers known as the most effective for energy-efficient communications over ISI channels are those employing turbo equalization [2], [3], which are based on the exchange of information between two (or more) soft-input soft-output (SISO) devices that iteratively refine the quality of their outputs. In this paper, we focus on the SISO detector [3], whose aim is to compute the a posteriori probability (APP) of each transmitted symbol. The target APPs are a by-product outcome of the algorithm for maximum-a-posteriori (MAP) symbol detection, which can be thus considered as the reference SISO device [2], [3].

Sufficient statistics for MAP detection are obtained by sampling at symbol rate the output of a filter matched to the received pulse [1]. Two approaches for MAP sequence detection have been known since the early Seventies, due to Forney [4] and Ungerboeck [5]. While the latter approach directly works on the samples at the output of the matched filter, the former is based on a filter that whitens them [4], leading to a Markovian channel model. Due to this property, the Forney approach can be extended to MAP symbol detection simply by resorting to the general forward-backward algorithm (FBA) derived in [6]. On the other hand, the extension of the Ungerboeck approach to MAP symbol detection is more involved and has been carried out only recently [7]. The two algorithms, which are rigorously equivalent and thus perform identically [7], are FBAs that work on the same trellis and differ for the branch metric only [7]. Unfortunately, the trellis size exponentially grows with the channel memory, so that the implementation of suboptimal algorithms that provide a convenient performance/complexity tradeoff is mandatory in several practical scenarios. Most low-complexity algorithms in the literature maintain the three-stage structure of the FBA, and obtain complexity reduction by performing a simplified trellis search (for instance, see [8]–[10]). These techniques often provide a satisfactory performance when the Forney approach is adopted, but, for reasons discussed in [10], [11], do not result effective when the Ungerboeck approach is adopted.

The design of low-complexity algorithms specifically addressed to the Ungerboeck approach is considered in this paper—beside the theoretical relevance, this is of interest because the implementation of the whitening filter is critical in various cases [12]. We introduce a novel factor graph (FG) [13] describing the channel, and derive a couple of SISO algorithms by applying the sum-product algorithm (SPA) [13] to it. The proposed algorithms are characterized by a complexity that grows linearly (instead of exponentially) with the channel memory, and can thus provide impressive reduction factors with respect to the FBA. We remark that the considered FG is exact, since no approximation is required for its derivation, but contains cycles, and thus cannot lead to an algorithm for exact MAP symbol detection [13]. On the other hand, since the shortest cycles in the FG have length six [13], the proposed algorithms provide very good convergence properties, such that they can well approximate exact MAP symbol detection in most scenarios of practical interest. In particular, we prove, by reporting extrinsic information transfer (EXIT) charts [14] and simulations of turbo equalizer, that the proposed solutions can be often considered, in terms of performance/complexity tradeoff, as the most convenient ones.

II. CHANNEL MODEL AND OPTIMAL SISO DETECTION

We consider linearly-modulated transmissions over an ISI channel that also introduces additive white Gaussian noise (AWGN). Assuming ideal synchronization, we can write the low-pass equivalent of the received signal as [1]

\[ r(t) = \sum_k c_k h(t - kT) + w(t) \]

(1)

where \( \{c_k\} \) is a sequence of complex-valued modulation symbols, all belonging to the same \( M \)-ary alphabet, \( T \) is the signaling interval, \( h(t) \) is the impulse response corresponding to the cascade of the transmission filter and the channel, and \( w(t) \) is a zero-mean circularly-symmetric Gaussian process with power spectral density \( 2N_0 \). The receiver is assumed to perfectly know the shaping pulse \( h(t) \) and the value of \( N_0 \).

A possible sufficient statistic for this detection problem is given by the sequence of samples \( \{x_k\} \), where [1]

\[ x_k = \int_{-\infty}^{\infty} r(t) h^*(t - kT) dt. \]

(2)
The sequence \( \{x_k\} \) is the output of a filter, matched to the shaping pulse \( h(t) \), fed by the received signal \( r(t) \), and sampled at symbol rate. Defined the sequence \( \{g_i\} \) such that

\[
g_i = \int_{-\infty}^{\infty} h(t) h^*(t-iT) dt , \quad \forall i \in \mathbb{Z}
\]

and the parameter \( L \geq 0 \) as the smallest integer such that

\[
g_i = 0 , \quad \forall i : |i| > L
\]

the following discrete-time channel model results [1]

\[
x_k = \sum_{i=-L}^{L} g_i c_{k-i} + n_k
\]

where the noise samples \( \{n_k\} \) are zero-mean circularly-symmetric Gaussian random variables with autocorrelation

\[
E\{n_k n_{k-i}\} = 2N_0 g_i , \quad \forall i \in \mathbb{Z}
\]

\( E\{\} \) being the expectation operator. The noise samples are thus correlated (or colored) when \( L > 0 \), that is when ISI is present. We point out that, according to (3) and (4), the existence of a finite value of \( L \) is not assured when the shaping pulse has an infinite support (for instance, when it is band-limited). On the other hand, since for all shaping pulses of interest the following limit holds [1]

\[
\lim_{i \to \pm \infty} g_i = 0
\]

it is always possible to choose a finite value of \( L \) that makes the model (5) as accurate as desired.

We are interested in performing SISO detection of the modulation symbols \( c_{k}^K \) —the border effects in (5) are here managed by assuming that the symbols \( c_{k}^K \) are preceded (and followed) by at least \( L \) symbols.1 Also, we assume that the modulation symbols are independent, so that the following factorization holds

\[
\mathbb{P}(c_{1-L}^K|c_{L}^K) = \prod_{k=1-L}^{K+L} \mathbb{P}(c_k)
\]

\( \mathbb{P}(\cdot) \) denoting a probability mass function, and \( \mathbb{P}(c_k) \) being the a priori probability that, at time epoch \( k, c_k \) is transmitted. As recalled in Section I, two approaches for optimal SISO detection are known. The former approach, referred to as Forney approach, is based on a whitening filtering of the received sequence \( x_{1-L}^K \), while the latter approach, referred to as Ungerboeck approach, directly works on the samples \( x_{1-L}^K \). Referring to [7] for the details, we remark that MAP symbol detection can be exactly carried out, in both cases, by means of a FBA working on a trellis whose state can assume \( M^L \) values per time epoch. Hence, large values of \( L \) make the implementation of the FBA impractical and motivate the search for suboptimal algorithms with a convenient performance/complexity tradeoff.

1We use the notation \( v_k^n \) for denoting the elements of a sequence \( \{v_k\} \) from index \( k = a \) to index \( k = b \).

\[\text{Fig. 1. Three sections of the factor graph corresponding to (13), when } L = 2. \text{ The dotted lines are connected to nodes related to other time epochs. The bold-marked cycle has length six.}\]

III. A NOVEL FRAMEWORK FOR SISO DETECTION

We introduce a novel framework for SISO detection, addressed to the Ungerboeck approach and based on a suitable factorization of the conditional probability density function (PDF) \( p(x_{1-L}^K|c_{1-L}^K) \). Namely, we exploit the fact that the PDF \( p(x_{1-L}^K|c_{1-L}^K) \) is proportional to [5], [7]

\[
\prod_{k=1}^{K} \exp\left\{ \frac{1}{N_0} \text{Re} \left\{ x_k c_k - \frac{1}{2} g_0 |c_k|^2 - \sum_{i=1}^{L} g_i c_i c_{k-i} \right\} \right\}
\]

\( \text{Re}\{\cdot\} \) denoting the real component of a complex number.2 Hence, if we define the functions

\[
F_k(c_k) = \exp\left\{ \frac{1}{N_0} \text{Re} \left\{ x_k c_k - \frac{1}{2} g_0 |c_k|^2 \right\} \right\}
\]

\[
H_i(c_k, c_{k-i}) = \exp\left\{ -\frac{1}{N_0} \text{Re} \left\{ g_i c_i c_{k-i} \right\} \right\}
\]

for each value of \( k \) in \( \{1, 2, \ldots, K\} \) and each value of \( i \) in \( \{1, 2, \ldots, L\} \), we can write

\[
p(x_{1-L}^K|c_{1-L}^K) \propto \prod_{k=1}^{K} \prod_{i=1}^{L} H_i(c_k, c_{k-i})
\]

Finally, the APP of the transmitted sequence is factorized as

\[
\mathbb{P}(c_{1-L}^K|x_{1-L}^K) \propto \prod_{k=1}^{K} \prod_{i=1}^{L} H_i(c_k, c_{k-i})
\]

by exploiting (8) and (12). Three sections of the FG corresponding to (13) are depicted in Fig. 1, for the case \( L = 2 \).

We point out that the marginalization of (13) required for computing the target APPs of the modulation symbols cannot be exactly carried out by applying the SPA to the FG in Fig. 1, since it contains cycles [13], as that bold-marked in the figure. On the other hand, when the length of the cycles is at least six, the SPA is generally expected to provide a good approximation of the exact marginalizations (see [13] for the general treatment, and [15], [16] for appealing applications). Hence, since the considered FG, irrespectively of the value of \( L \), does not contain any cycle of length lower than six, the SPA can be expected to effectively work. The interest for this

2Two quantities are here considered proportional when they differ for a positive multiplicative factor irrelevant for the detection process.
suboptimal framework will appear when its complexity will be discussed and compared with that of the optimal FBA.

For each value of \( k \) in \( \{1, 2, \ldots, K\} \) and each value of \( i \) in \( \{1, 2, \ldots, L\} \), we adopt the following notation for the various messages over the FG:

- \( O_k(c_k) \) is the message from the variable node \( c_k \) to the function node \( P_k \);
- \( \mu_{F,k,i}(c_k) \) is the message from the variable node \( c_k \) that goes forward to the function node \( H_i \) connected also with the variable node \( c_{k+i} \);
- \( \mu_{B,k,i}(c_k) \) is the message on the same edge as \( \mu_{F,k,i}(c_k) \) that goes backward to the variable node \( c_k \);
- \( \mu_{D,k,i}(c_k) \) is the message from the variable node \( c_k \) that goes downward to the function node \( H_i \) connected also with the variable node \( c_{k-i} \);
- \( \mu_{U,k,i}(c_k) \) is the message on the same edge as \( \mu_{D,k,i}(c_k) \) that goes upward to the variable node \( c_k \).

Fig. 2 clarifies the conventions the concepts of forward, backward, downward, and upward are related to. Once defined \( V_k(c_k) \) as the product of all messages incoming to the variable node \( c_k \), namely

\[
V_k(c_k) = P_k(c_k)F_k(c_k) \prod_{i=1}^{L} \mu_{U,k,i}(c_k)\mu_{B,k,i}(c_k) \tag{14}
\]

the application of the SPA leads to the following rules for message updating:

\[
O_k(c_k) = \frac{V_k(c_k)}{P_k(c_k)} \tag{15}
\]

\[
\mu_{F,k,i}(c_k) = \frac{V_k(c_k)}{\mu_{B,k,i}(c_k)} \tag{16}
\]

\[
\mu_{D,k,i}(c_k) = \frac{V_k(c_k)}{\mu_{U,k,i}(c_k)} \tag{17}
\]

\[
\mu_{U,k,i}(c_k) = \sum_{c_{k+i}} H_i(c_{k+i}, c_k)\mu_{F,k,i}(c_k) \tag{18}
\]

\[
\mu_{B,k,i}(c_k) = \sum_{c_{k-i}} H_i(c_{k}, c_{k-i})\mu_{D,k,i}(c_k) \tag{19}
\]

All messages \( \{\mu_{X,k,i}\} \), with \( X \in \{F, B, D, U\} \), should be initialized to the same positive value—the choice of this value is irrelevant [13]. Also, we point out the probabilistic meaning of the term \( V_k(c_k) \), which is proportional to the (approximated) APP \( P(c_k|x^K_1) \), and that of the term \( O_k(c_k) \), which is proportional to the (approximated) PDF \( p(x^K_1|c_k) \) and thus is the so-called extrinsic information produced by the algorithm [13].

Finally, we remark the main feature of the considered framework: unlike the FBA, which requires the propagation of \( M^K \)-ary messages, no state variable appears, so that all messages over the FG are \( M \)-ary. Moreover, we notice that the exponential dependence on the value of \( L \), which characterizes the complexity of the FBA, does not appear neither in (14), nor in the multiply-accumulate operations in (18) and (19). Hence, we can state that the complexity of the proposed framework increases linearly with the value of \( L \), since this value only impacts on the number of edges in the FG, while the function nodes \( \{H_i\} \) are connected with two variable nodes only, irrespectively of the value of \( L \). A FG with a similar property is obtained in the case of transmissions over correlated flat-fading channels [17]. For ISI channels, we mention the framework presented in [16], where a FG that does not contain state variables is also considered. On the other hand, the algorithms in [16], although effective for sparse channels, still present an exponential dependence on the value of \( L \) in the multiply-accumulate operations executed for message updating, since there exist function nodes connected with \( L \) variable nodes. To our knowledge, the proposed one is thus the only framework leading to linear-complexity algorithms for SISO detection over ISI channels. We will show in Section V that this property can lead to impressive complexity reductions.

IV. PROPOSED ALGORITHMS

A. Description of the Algorithms

Due to the presence of cycles in the considered FG, the SPA cannot lead to a unique schedule nor to a unique stopping criterion for the message passing [13]. Two of the various possible algorithms deriving from different schedules are described in the following—we investigated other algorithms that were just found to provide improvements over particular channels, and are thus not described here.

The former algorithm will be referred to as parallel-schedule SPA (PS-SPA). It basically exploits the fact that the lower part of the FG in Fig. 1 is formally identical to the FGS describing the low-density parity-check (LDPC) codes, and adopts the same flooding schedule used for standard LDPC decoding [15]. The PS-SPA can be formalized by the following sequence of steps:

1) update all messages \( \{\mu_{D,k,i}\} \) and \( \{\mu_{F,k,i}\} \);
2) update all messages \( \{\mu_{U,k,i}\} \) and \( \{\mu_{B,k,i}\} \);
3) update all terms \( \{V_k\} \);
4) if the stopping criterion is not satisfied go to step 1.

We will only consider stopping criteria based on the number of self-iterations, that is on the number of times step 4 is executed. We point out that all operations at the same step can be executed in parallel. Hence, no serial operation is required by the PS-SPA, whose latency thus does not depend on the value of \( K \). This feature makes the PS-SPA very attractive.
The latter algorithm will be referred to as serial-schedule SPA (SS-SPA). It is inspired to shuffled LDPC decoding [18] and to one of the schedules described in [16]. Let us define the forward recursion as the following sequence of steps, to be serially executed for each value of $k$ from 1 to $K$:

a. update the messages $\{\mu_{U,k,i}\}$, $i = 1, 2, \ldots, L$;

b. update the term $V_k$;

c. update the messages $\{\mu_{F,k,i}\}$, $i = 1, 2, \ldots, L$.

Let us also define the backward recursion as the following sequence of steps, to be serially executed for each value of $k$ from $K$ down to 1:

a. update the messages $\{\mu_{B,k,i}\}$, $i = 1, 2, \ldots, L$;

b. update the term $V_k$;

c. update the messages $\{\mu_{D,k,i}\}$, $i = 1, 2, \ldots, L$.

Finally, the SS-SPA can be formalized by the following sequence of steps:

1) run the forward recursion;
2) run the backward recursion;
3) if the stopping criterion is not satisfied go to step 1.

Again, we will only consider stopping criteria based on the number of self-iterations, that is on the number of times step 3 is executed. Due to the presence of serial recursions, the SS-SPA is characterized by a latency that linearly increases with the value of $K$, like the optimal FBA but unlike the PS-SPA.

Finally, we describe a technique for reducing the complexity of the PS-SPA and the SS-SPA. The idea is to neglect the channel coefficients $\{g_i\}$ with the lowest magnitudes. Formally, we introduce the integer parameter $L_E \leq L$, and define a new FG that does not include all $L$ function nodes $\{H_i\}$, but only the $L_E$ function nodes $\{H_i\}$ related to the coefficients $g_i$ with the largest magnitudes, so that a reduction factor equal to about $L/L_E$ results with respect to the original FG. We will show in Section V that this reduction factor can often be large without noticeably affecting the detection performance.

B. Optimization of the Algorithms

There exist some scenarios where both PS-SPA and SS-SPA, when implemented in the basic version described above, provide a poor performance. After deep investigations on this behavior, we can state that the most significant problem is the overestimation of the reliability of the propagated messages—this is a known issue of the SPA when applied to FGs with cycles [13]. A very simple way for overcoming this problem consists of adopting in (10) and (11) a value of $N_0$ larger than the actual one. The rationale of this trick is the following: since the problem is the overconfidence in the computed messages, we can make the algorithm less confident simply by describing the channel as if it added more noise than it really does. The effectiveness of this trick is proved by the simulation results reported in Section V.

Other analyses, carried out by means of EXIT charts as well as bit-error rate (BER) simulations, show that it is generally not convenient to execute more than one self-iteration neither for the PS-SPA nor for the SS-SPA, when employed in iterative detection/decoding schemes. Both algorithms indeed provide the best performance when, each time the SISO decoder provides them with intrinsic information, they execute just one self-iteration before feeding out the extrinsic information.

Hence, the possibility of executing more self-iterations will not be considered hereafter.

In iterative detection/decoding schemes, a further design option consists of setting all messages in the detection FG each time the PS-SPA and the SS-SPA are provided with intrinsic information, so that any memory of the previous iterations is removed. These algorithms will be referred to as PS-SPA-R and SS-SPA-R.

V. SIMULATION RESULTS

In this section, the performance of the proposed algorithms, implemented in the logarithmic domain [13], is assessed by means of computer simulations. All results reported in the following are related to binary phase-shift keying (BPSK) transmissions and real-valued sequences $\{g_i\}$. We consider two ISI channels, referred to as C1 and C2. Channel C1 is characterized by a shaping pulses $h(t) = h_{\text{LOR}}(t-t)h_{\text{LOR}}(t-T)$, $h_{\text{LOR}}(t)$ being a Lorentzian pulse with density $D = 3$ [19]—this model is commonly used for magnetic-storage systems. Channel C2 corresponds to faster-than-Nyquist signaling with roll-off factor $\alpha = 0.1$ and time-compression factor $\tau = 0.78$ [19]—this system provides a very attractive spectral efficiency by introducing intentional ISI [19]. Both channels are not characterized by a finite value of $L$ and were truncated by setting $L = 12$ the full-complexity FBA works on a 4096-state trellis.

Fig. 3 reports EXIT charts related to transmissions over channel C1 when the value of the SNR is equal to 3 dB.

![EXIT charts for different algorithms over channel C1, when a BPSK modulation is adopted and the value of the SNR is equal to 3 dB.](image)

We chose $L$ as the minimum integer such that $|g_i|/g_0 < 0.01$ for all values of $i$ larger than $L$ in [1].

The PS-SPA would be even faster, but we could not implement the message passing in parallel in our simulation environment. For this reason, an implementation over Field Programmable Gate Array (FPGA) is ongoing.

978-1-4244-2324-8/08/$25.00 © 2008 IEEE.

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2008 proceedings.
PS-SPA and the SS-SPA significantly improve with respect to when the assumed value of $N_0$ equals the actual one—in particular, $N_0$ is increased by 9 dB for the PS-SPA and by 5 dB for the SS-SPA. Moreover, we notice that the SS-SPA is noticeably more effective than the PS-SPA in this scenario, and fairly close to the performance of the optimal FBA.

Finally, in Fig. 4, we report the BER performance of various algorithms in a turbo equalization scheme [2], [3]. The results related to channel C1 and channel C2 are marked, respectively, by circles and squares. An LDPC code with codeword length of 50,000 bits and rate 1/2 is used in both cases. A detection instance is executed before each iteration of the SISO decoder, for a maximum of 50 iterations. The process also stops if, by checking the code syndrome, a valid codeword is found before the 50th iteration. No interleaver is used because of the random nature of the LDPC code. In all cases, the algorithms are implemented after a (coarse) optimization of the assumed values of $N_0$. We point out that the simulation results related to the full-complexity FBA, yet sufficient for estimating the performance loss due to the proposed algorithms, are incomplete, since it is nearly unfeasible to obtain reliable BER curves for turbo equalizers working on a 4096-state trellis. Let us notice the impressive performance of the proposed algorithms over channel C2. First, we remark that there is no need for considering more than 5 function nodes $\{H_i\}$, since the SS-SPA practically performs as the FBA when implemented with message resetting and $L_E = 5$. Then, we point out that the PS-SPA loses only few tenths of dB when implemented without message resetting and $L_E = 5$, and can be thus considered as the most convenient solution thanks to the low-latency properties discussed before. The proposed algorithms result just slightly less effective over channel C1. In particular, we notice that the SS-SPA loses about 0.7 dB with respect to the FBA, but, since the simulation was at least 100-time faster, this is still a very satisfactory result—to our knowledge, over this magnetic channel, the proposed solution is by far that providing the best performance/complexity tradeoff. Fig. 4 also shows the impact of the value of $L_E$ and message resetting in these scenarios. Interestingly, the SS-SPA performs better when message resetting is implemented, while the PS-SPA does when message resetting is not implemented.

Other simulation results, which are not shown here for a lack of space, confirm the effectiveness of the proposed algorithms even when non-binary modulations and different ISI channels are considered.

VI. CONCLUSIONS

We have presented a framework that leads to novel algorithms for SISO detection over ISI channels. The proposed schemes, which stem from a FG we have derived on the basis of the Ungerboeck approach, provide impressive complexity reductions with respect to the reference FBA algorithm. Moreover, we have presented a couple of optimization techniques that significantly improve the effectiveness of the proposed algorithms, such that they can be considered as the most convenient solutions, in terms of performance/complexity tradeoff, in various turbo equalization schemes.

REFERENCES


