Ordered Subsets Message-Passing

Joseph A. O'Sullivan, Naveen Singla*

August 31, 2005

Abstract

A novel message-passing algorithm is proposed for decoding on graphs having short cycles. The algorithm, termed the "ordered subsets message-passing" (OSMP) algorithm, uses the idea of partitioning the measured data into subsets to break short cycles. This idea is commonly employed in imaging for accelerating image reconstruction algorithms. The OSMP algorithm is then applied for joint equalization and decoding for information storage systems having two-dimensional (2-D) intersymbol interference (ISI) during readback. Simulation results are provided to show that the OSMP algorithm brings about an improvement in performance over the case when the measured data are not partitioned into subsets. It is also shown that the performance of the OSMP algorithm for the 2-D ISI channel concentrates around the behavior of the decoder when the input is a sequence of equally likely independent and identically distributed random variables. From this it can be concluded that the probability of decoding error can be made arbitrarily small for long enough block lengths and noise variance lower than a threshold.

1 Introduction

Message-passing algorithms have generated a lot of interest recently due to the fact that a wide variety of algorithms developed for different applications can be derived as specific instances of message-passing algorithms [1], [7], [10]. For channel coding, message-passing algorithms achieve near-capacity performance when used for iterative decoding. The sum-product algorithm [7] for the decoding of low-density parity-check (LDPC) codes [9] and the forward/backward algorithm [2] for the decoding of turbo codes [3] are message-passing algorithms which get to within a fraction of a dB of the Shannon capacity on a wide variety of channels. When applied to decoding, message-passing algorithms compute posterior probabilities of the transmitted data given the measured data by passing "messages" on a graph. This graph represents a behavioral or probabilistic (or both) modeling of the underlying system [7]. The message-passing algorithm does not converge to the exact posteriors if cycles are present in the graph, resulting in a loss of performance.

In this paper we describe a novel message-passing algorithm for decoding on graphs with cycles. The algorithm, termed the "ordered subsets message-passing" (OSMP) algorithm, partitions the observed data into subsets and alters the message-passing schedule to break the short cycles in the graph. The idea of partitioning the measured data is commonly employed in imaging [4], [5] to increase the rate of convergence of reconstruction algorithms. The OSMP algorithm is applied for joint equalization and decoding for information storage systems having two-dimensional (2-D) intersymbol interference (ISI) during readback. We show by simulations that the OSMP algorithm outperforms the full graph algorithm [12] which is the unordered counterpart of the OSMP algorithm. A concentration result for the OSMP algorithm applied to a 2-D ISI channel is also shown to hold. This concentration result is an extension of the result proved in [6], for the case of 1-D ISI channels, to 2-D ISI channels.

We begin the paper in Section 2 by describing the OSMP algorithm in a general setting. In Section 3 the OSMP is applied for joint equalization and decoding for channels with 2-D ISI. The concentration result is also proved in Section 3. Some concluding remarks are given in Section 4.

2 Ordered Subsets Message-Passing

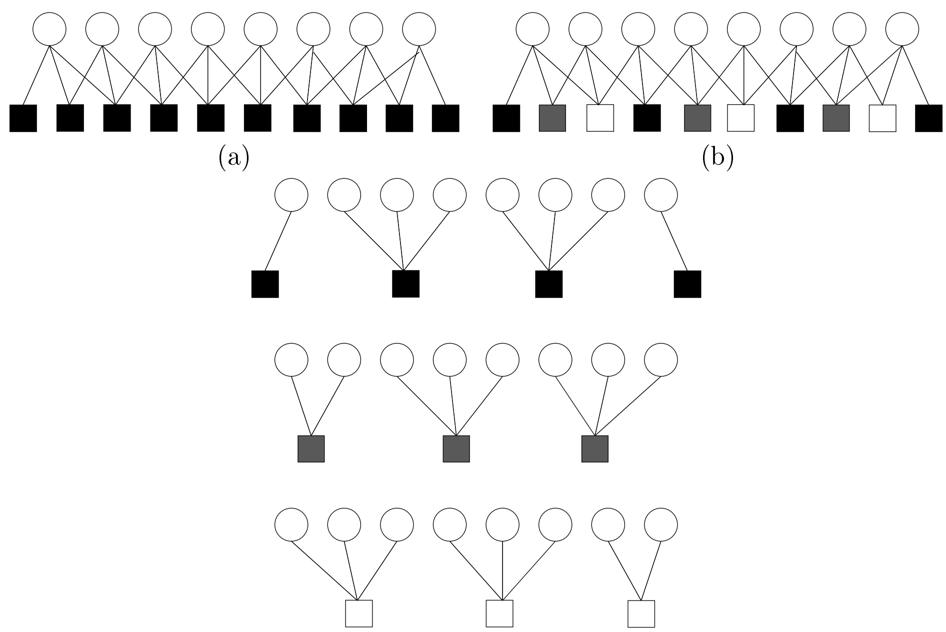
The idea of ordered subsets was originally proposed for speeding up algorithms, such as the Expectation Maximization (EM) algorithm, for image reconstruction [4]. This is done by partioning the measured

*The authors are with the Department of Electrical Engineering, Washington University in St. Louis, MO 63130, USA (e-mail: jao@ee.wustl.edu).

data into subsets and using only one subset at every iteration for reconstruction ignoring all the other subsets. It was shown empirically by Hudson and Larkin [4] that the ordered subsets EM algorithm increases the rate of convergence by a factor proportional to the number of subsets.

In decoding, the idea of ordered subsets can be applied to break short cycles in the graph, typically a factor graph, on which the message-passing algorithm operates. The presence of short cycles is detrimental to the performance of the message-passing algorithm. In [7], Kschischang *et al.* propose some transformations to convert a factor graph with cycles into a cycle-free factor graph. The first method they propose is clustering, where nodes of the same kind (variable or function) are clustered together with appropriate modifications to the function nodes. The second method, stretching variable nodes, extends the region of influence of a particular variable node in the graph. By this stretching transformation it is possible to make certain edges or even variable nodes redundant and delete them from the graph thus breaking some cycles. Although both of these transformations can break short cycles, they do so at the cost of increased decoding complexity.

The ordered subsets approach to decoding is to partition the measured data into subsets. The message-passing algorithm then cycles through these subsets in a fixed order using only one subset at each step and ignoring the messages from all the other subsets. This partitioning into subsets, if done properly, can eliminate short cycles in the graph. For decoding applications the message-passing algorithm has a complexity which is linear in the code block length per iteration. Hence, for mutually exclusive (and exhaustive) subsets, the time taken for a single pass through all the subsets will be comparable to what is taken for one iteration on the unpartitioned graph. The idea is illustrated in Fig. 1 for a very small graph.



(a) (b) (c)

Figure 1: (a) Factor graph with many length four cycles. Squares represent variable nodes and circles represent function nodes. (b) Partitioned factor graph with variable nodes divided into three subsets (c) Message-passing schedule using one subset per iteration eliminates length four cycles.

The idea of partitioning the measured data is also used in the decoding of turbo codes. For a classical rate one-third parallel concatenated turbo code the codeword is made up of the systematic bits, y_0, and y_1, y_2, respectively obtained from the convolution code using y_0 and its permutation. During decoding of turbo codes the measurements are divided into two overlapping subsets, the first subset corresponds to the codeword bits $[y_0, y_1]$ and the second one to $[y_0, y_2]$. Message-passing (forward/backward algorithm) is done over the portion of the graph representing the first subset. This is followed by message-passing on the other subset using extrinsic information provided by the first subset decoding. The cycle then continues with message-passing on the first subset using extrinsic information passed by the second subset decoding. Hence decoding of turbo codes is a special instance of OSMP where the measured data is partitioned into overlapping subsets. The overlapping data are the measurements of the systematic bits, y_0. For an illustration of the step-by-step decoding of turbo codes using factor graphs see [8].

3 Application of OSMP for Joint Equalization and Decoding

With existing data storage technologies challenging physical limits, there is a need for new and more improved technologies to keep up with the demands of data storage. A 2-D paradigm for storage offers a promising direction. Research in the magnetic storage industry is already being conducted for 2-D patterned media. Together with advances in read and write technology, this promises revolutionary changes in data storage in the near future. One aspect of data storage systems which has been extensively studied for one-dimensional (1-D) storage is the ISI. The problem of ISI for a 2-D medium is more complicated since the ISI is also 2-D and algorithms commonly used to mitigate ISI in one dimension, like the BCJR and Viterbi algorithms, have no direct generalization to two dimensions. Although much work has been done on equalization for 2-D ISI channels, relatively little work has been done for combined equalization and decoding for such channels. Some joint equalization and decoding schemes for 2-D ISI channels are discussed in [12], [13] which employ the turbo-principle or message-passing. The ability of the OSMP algorithm to break cycles motivates its use for 2-D ISI channels.

In this section we propose a message-passing algorithm for joint decoding and equalization for storage systems with 2-D ISI using the idea of ordered subsets. Simulation results show that this algorithm provides an improvement over the full graph message-passing algorithm [12]. It is also shown that the behavior of the decoder concentrates around the expected behavior. The expected behavior is defined as the behavior of the decoder when the channel input is a sequence of equally likely independent and identically distributed (i.i.d.) random variables. The proof is an extension of the proof in [6], for a 1-D ISI channel, to 2-D ISI channels.

This section is organized as follows: In subsection III-A, the channel model for the storage system and the full graph message-passing algorithm are described briefly. Subsection III-B describes the OSMP algorithm for this setting and shows simulation results to corroborate the fact that the OSMP algorithm does improve performance by eliminating short cycles. In subsection III-C we show that the concentration result holds and conclude that the probability of decoding error can be made arbitrarily small for large enough block lengths and when noise variance is below a threshold.

3.1 Channel Model and Full Graph Message-Passing

The recording system can be represented by the discrete time system shown in Fig. 2. Here $\mathbf{a}$ is the uncoded user data, the channel input, $\mathbf{X}$, is the encoded data, and $\mathbf{R}$ is the channel output. $\mathbf{X}$ is a matrix, typically rectangular, with elements $x(i,j) \in \{\pm 1\}$. The noise, $\mathbf{W}$, is assumed to be additive white Gaussian noise (AWGN). The channel output, $\mathbf{R}$, is a matrix with elements

$$r(i,j) = \sum_{k_1=0}^{L-1} \sum_{k_2=0}^{L-1} x(i-k_1, j-k_2) h(k_1,k_2) + w(i,j). \tag{1}$$

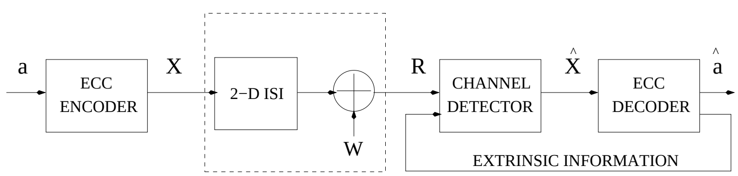


Figure 2: Discrete-time system.

In (1) L represents the number of elements over which the ISI extends in each dimension and $\mathbf{h}$ is the 2-D channel point spread function (PSF). For error correction we use LDPC coset codes [6] with the code graph chosen uniformly at random from the ensemble of regular graphs. For the purpose of illustration of our concepts and simulations we use the following PSF:

$$h = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 0.25 \end{pmatrix}.$$

The full graph algorithm is an *a-posteriori* probability (APP) based algorithm that is used for both equalization and decoding. This algorithm computes approximate APPs of the codeword bits given the observations by performing message-passing on a three-level graph of the LDPC code and the channel ISI. The upper two levels in this "full graph" represent the LDPC code bipartite graph and the lower two levels represent the channel ISI graph showing how the PSF connects the codeword bits to the measured data. The algorithm first iterates on the LDPC bipartite graph for a fixed number of iterations. If decoding fails then message-passing is performed on the three-level graph for a fixed number of iterations using the following schedule: Measured data nodes to variable nodes, variable nodes to the check nodes, check nodes to the variable nodes and finally variable nodes to the measured data nodes.

The performance of the full graph algorithm is shown in Fig. 3(a), where the LDPC code used is a blocklength 10000, regular (3,6) code. The decoding performance is compared to the performance of the LDPC code on a channel without ISI. The performance is degraded due to the presence of short cycles in the channel ISI graph as shown in Fig. 3(b). This motivates the use of alternative message-passing schedules to break these short cycles and improve performance.

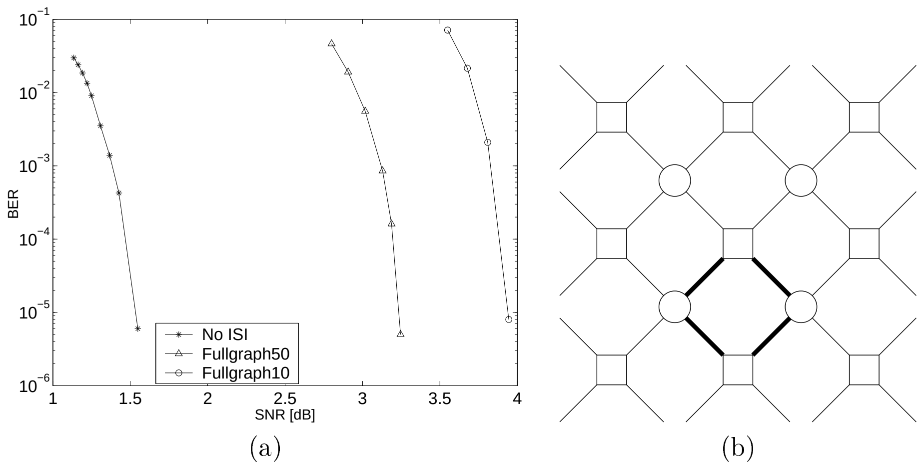


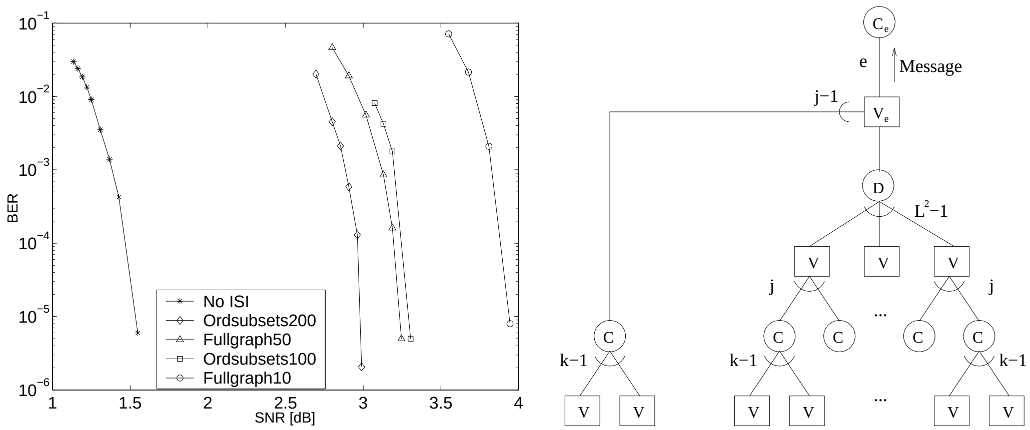
Figure 3: (a) Performance curves for the full graph message passing algorithm. The curve on the left is the performance of the LDPC code on an AWGN channel without ISI. The second and third curves from the left are the performance of the full graph algorithm for 50 and 10 iterations on the full graph. In each case 20 iterations are performed on the LDPC bipartite graph prior to message-passing on the three-level graph. (b) The channel ISI graph for the considered 2×2 PSF. The graph includes edges connecting variable nodes (squares) to the measured data nodes (circles). The 2-D ISI yields many length four cycles in this graph. One of these cycles is illustrated by the bold edges. The variable nodes are also connected to the check nodes (not shown) according to the parity-check matrix of the LDPC code.

3.2 OSMP for 2-D ISI Channels

In this subsection the OSMP algortihm is applied for the 2-D ISI channel. For a general $L \times L$ 2-D ISI we assign L^2 labels such that in any $L \times L$ block of the measured data, each entry has a unique label. Thus the measured data nodes are divided into L^2 mutually exclusive subsets. For example, the four measured data nodes shown in Fig. 3(b) are assigned distinct labels. As mentioned before, the message-passing on the channel ISI graph uses only one subset for each iteration and the algorithm then cycles through all the subsets in a fixed order. This process is continued until a pre-defined number of iterations is exhausted. The performance of this algorithm and that of the full graph algorithm are shown in Fig. 4(a). The LDPC code used is again a blocklength 10000 regular (3,6) code. From the curves it can be seen that the OSMP algorithm provides an improvement of about 0.3 dB at a bit error rate of 10^{-5} over the full graph algorithm.

3.3 Concentration Result

In [6] Kavčić *et al.* propose the use of LDPC coset codes for the ISI channel due to the fact that the channel has an input-dependent memory. The concentration result then is as follows: For information sequences generated uniformly at random, for almost all graphs and almost all cosets, the OSMP decoder behaves very close to the behavior of the decoder when the channel input is a sequence of equally likely i.i.d. binary random variables. The concentration result says that the number of errors made by the



(a) (b)

Figure 4: (a) Performance curves for the OSMP algorithm. The second and fourth curves from the left are the performance of the OSMP algorithm for 50 and 25 iterations on the partitioned three-level graph respectively. One iteration is completed after a single pass through all the subsets. The other three curves are the same as in Fig. 3(a). (b) The depth one message flow neighborhood for a regular (2,3) LDPC code and a 2×2 PSF. j and k are the variable and check node degrees respectively and L^2 is the number of variable nodes over which the 2-D ISI spreads for a general $L \times L$ PSF.

decoder are concentrated around the number of errors made by it when the inputs are equally likely i.i.d. binary random variables.

The proof of the concentration result for the OSMP algorithm applied to the 2-D ISI channel is the same as provided in [6] for a 1-D ISI channel. The proof uses the idea of message flow neighborhoods. Message flow neighborhoods are a generalization of directed neighborhoods, defined by Richardson and Urbanke [11] for bipartite graphs, to arbitrary graphs. The message flow neighborhood of depth l of an edge e is the subgraph that consists of the check node C_e, the variable node V_e, the edge e and all the nodes and edges that contribute to the computation of the message from V_e to C_e along the edge e. Fig. 4(b) shows a message flow neighborhood of depth one on the partitioned graph for a regular (2,3) LDPC code and a 2×2 PSF. The rest of the proof follows through exactly as in [6]. The only change is in the number of message flow neighborhoods. In this case there are $2^{N(l)}$ distinct message flow neighborhoods at a depth l, where

$$N(l) = \frac{(k-1)^l (j.(L^2-1)+j-1)^l - 1}{(k-1)(j.(L^2-1)+j-1) - 1}, \tag{2}$$

for a (j, k) regular LDPC code and an $L \times L$ PSF.

The conclusion of interest that can be drawn from this result is that there is at least one graph and one coset for which the decoding probability of error can be made arbitrarily small on an information sequence generated uniformly at random if the noise variance does not exceed a threshold and if the block length is long enough. The noise variance threshold is the supremum of all noise variances for which the error probability for i.i.d. binary inputs tends to zero for sufficiently large number of iterations of message-passing on the partitioned three-level graph.

4 Conclusions

A novel message-passing algorithm is proposed for decoding on graphs having short cycles. The algorithm uses the idea of ordered subsets from imaging to partition the measured data into subsets and for each iteration uses only one subset, ignoring all the other ones. The OSMP algorithm is then applied for joint equalization and decoding for channels with 2-D ISI and is shown to give an improvement in performance over its unordered counterpart by breaking the short cycles in the channel ISI graph. A concentration result for the OSMP is shown to hold following the proof in [6]. From the concentration result it can be concluded that there exists at least one graph and one coset for which the decoding error probability can be made arbitrarily small provided the block length is long enough and noise variance is lower than a threshold.

References

[1] S. M. Aji, and R. J. McEliece, "The generalized distributive law," *IEEE Trans. Inform. Theory*, vol. 46, pp. 325-343, Mar. 2000.

[2] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. 20, pp. 284-287, Sept. 1974.

[3] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes," in *Proc. IEEE Int. Conf. on Comm."*, Geneva, Switzerland, pp. 1064-1070, May 1993.

[4] H. M. Hudson, and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. Med. Imaging*, vol. 13, pp. 601-609, Dec. 1994.

[5] C. Kamphuis, and F. J. Beekman, "Accelerated iterative transmission CT reconstruction using an ordered subsets convex algorithm," *IEEE Trans. Med. Imaging*, vol. 17, pp. 1101-1105, Dec. 1998.

[6] A. Kavčić, X. Ma, and M. Mitzenmacher, "Binary intersymbol interference channels: Gallager codes, density evolution and code performance bounds," submitted to *IEEE Trans. Inform. Theory.*

[7] F. R. Kschischang, B. J. Frey, and H. A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, pp. 498-519, Feb. 2001.

[8] F. R. Kschischang, and B. J. Frey, "Iterative decoding of compound codes by probability propagation in graphical models," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 219-230, Feb. 1998.

[9] D. J. C. Mackay, "Good error-correcting codes based on very sparse matrices," *IEEE Trans. Inform. Theory*, vol. 45, pp. 399-431, Mar. 1999.

[10] R. J. McEliece, D. J. C. Mackay, and J.-F. Cheng, "Turbo decoding as an instance of Pearl's 'belief propagation' algorithm," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 140-152, Feb. 1998.

[11] T. J. Richardson, and R. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. Inform. Theory*, vol. 47, pp. 599-618, Feb. 2001.

[12] N. Singla, J. A. O'Sullivan, R. S. Indeck, and Y. Wu, "Iterative decoding and equalization for 2-D recording channels," *IEEE Trans. Magn.*, vol. 38, pp. 2328-2330, Sept. 2002.

[13] Y. Wu, J. A. O'Sullivan, R. S. Indeck, and N. Singla, "Iterative detection and decoding for separable two-dimensional intersymbol interference," submitted to *IEEE Trans. Magn.*, Jun. 2002.