Joint Iterative Decoding of LDPC Codes and Channels with Memory

Henry D. Pfister and Paul H. Siegel

Signal Transmission and Recording Group, University of California, San Diego
E-mail: {hpfister,psiegel}@ucsd.edu

Abstract: *This paper considers the joint iterative decoding of irregular low-density parity-check (LDPC) codes and channels with memory. It begins by introducing a new class of erasure channels with memory, known as generalized erasure channels. For these channels, a single parameter recursion for the density evolution of the joint iterative decoder is derived. This provides a necessary and sufficient condition for decoder convergence, and allows the algebraic construction of sequences of LDPC degree distributions. Under certain conditions, these sequences can achieve the symmetric information rate (SIR) of the channel using only iterative decoding. Example code sequences are given for two channels, and it is conjectured that they each achieve the respective SIR.*

Keywords: joint iterative decoding, erasure channel, capacity achieving, LDPC codes

1. INTRODUCTION

Sequences of irregular low-density parity-check (LDPC) codes that achieve the capacity of the binary erasure channel (BEC) under iterative decoding were first constructed by Luby, *et al.* in [1]. Their construction relies heavily on the fact that density evolution (DE) for the BEC has a simple closed form expression. The *joint iterative decoding* of a code and a channel with memory was first introduced by Douillard, *et al.* for turbo codes and is known as *turbo equalization* [2]. This approach can be generalized to LDPC codes by constructing a single large graph which includes both code and channel constraints [3].

In Section 2, we introduce the basic components of our system. First, we define a class of channels with memory, which we refer to as generalized erasure channels (GECs). Then, we describe the dicode erasure channel (DEC), which is a binary-input GEC with a linear response of $1 - D$ and erasure noise. In Section 3, we show that DE can be done analytically for the joint iterative decoding of a GEC and an irregular LDPC code. This allows us to algebraically construct sequences of irregular LDPC codes. Since we are using equiprobable signaling (i.e., linear codes), the maximum achievable information rate is the symmetric information rate (SIR) of the channel. In Section 4, we construct sequences of LDPC degree distributions which appear to achieve the SIR using iterative decoding.

2. BACKGROUND

2.1. System Model

The system we consider is fairly standard for the joint iterative decoding of an LDPC code and a channel with memory. Equiprobable information bits are encoded into an LDPC codeword, $\mathbf{X} = X_1, \ldots, X_n$, which is observed through a GEC as the output vector, $\mathbf{Y} = Y_1, \ldots, Y_n$. The decoder consists of an *a posteriori probability* (APP) detector matched to the channel and an LDPC decoder. The first half of decoding iteration i entails running the channel detector on $\mathbf{Y}$ using the *a priori* information from the LDPC code. The second half of decoding iteration i corresponds to executing one LDPC iteration using internal edge messages from the previous iteration and the channel detector output. Figure 1 shows the Gallager-Tanner-Wiberg (GTW) graph of the joint iterative decoder.

2.2. Generalized Erasure Channels

Since the messages passed around the GTW graph of the joint decoder are log-likelihood ratios (LLRs), DE involves tracking the evolution of the distribution of LLR messages passed around the decoder. Let L be a r.v. representing a randomly chosen LLR at the output of the channel decoder. If the distribution of L is supported on the set $\{-\infty, 0, \infty\}$ and $Pr(L = -\infty) = Pr(L = \infty)$, then we refer to it as a *symmetric erasure distribution*. Such distributions are one dimensional, and are completely defined by the erasure probability $Pr(L = 0)$. Our closed form analysis of this system requires that all the densities involved in DE are symmetric erasure distributions.

Definition. A *generalized erasure channel* (GEC) is any channel which satisfies the following condition for i.i.d. equiprobable inputs. The LLR distribution at the output of the channel detector is a symmetric erasure distribution whenever the *a priori* LLR distribution is a symmetric erasure distribution.

This allows DE of the joint iterative decoder to be represented by a single parameter recursion. Let

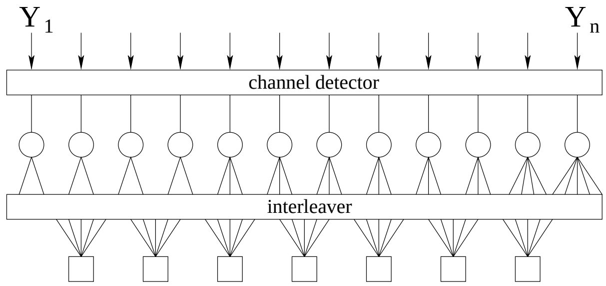

Figure 1: GTW graph of the joint iterative decoder.

$f(x)$ be a function which maps the erasure probability of the *a priori* LLR distribution, x, to the erasure probability at the output of the detector. The effect of the channel on the DE depends only on $f(x)$, which we refer to as the *erasure transfer function* (ETF) of the GEC. This function is very similar to the mutual information transfer function, $T(I)$, used by the EXIT chart analysis of ten Brink [4]. Since the mutual information of a BEC with erasure probability x is $1-x$, the mutual information transfer function and $f(x)$ are linked by the identity, $T(I) = 1 - f(1-I)$.

A remarkable connection between the channel SIR, I_s, and its mutual information transfer function was also introduced by ten Brink in [5]. This result,

$$\lim_{n\to\infty} \frac{1}{n} I(X_1,\ldots,X_n;Y_1,\ldots,Y_n) = \int_0^1 T(I)dI,$$

requires that $T(I)$ is computed using a symmetric erasure distribution as the *a priori* LLR distribution. Assuming the input process is i.i.d. and equiprobable makes the LHS equal the SIR, and using $T(I) = 1 - f(1-I)$ allows us to simplify this expression to

$$I_s = \int_0^1 T(I)dI = 1 - \int_0^1 f(x)dx. \tag{1}$$

Previously, we saw that $f(x)$ completely characterizes the DE properties of a GEC, and now we see that it can also be used to compute the SIR.

2.3. The Dicode Erasure Channel

The dicode erasure channel (DEC) is a binary-input channel based on the dicode channel used in magnetic recording. Essentially, the output of the standard dicode channel, $\{+1, 0, -1\}$, is erased with probability ϵ and transmitted perfectly with probability $1-\epsilon$. The precoded DEC is essentially the same, except that the input bits are differentially encoded prior to transmission. The state diagrams of these two channels are shown in [6, p. 161].

The simplicity of the DEC allows the BCJR algorithm for the channel to be analyzed in closed form. This analysis was done in [6, p. 179], and we simply state the ETFs for the DEC with and without precoding. If there is no precoding and the outputs of the DEC are erased with probability ϵ, then the ETF of the channel detector is

$$f(x) = \frac{4\epsilon^2}{(2 - x(1-\epsilon))^2}. \tag{2}$$

On the other hand, using a precoder changes this function to

$$f(x) = \frac{4\epsilon^2 x\,(1 - \epsilon(1-x))}{(1 - \epsilon(1-2x))^2}. \tag{3}$$

The SIR of the DEC was also computed in [6, p. 141], and is given by

$$I_s(\epsilon) = 1 - \frac{2\epsilon^2}{1+\epsilon}.$$

One can also get this expression for the SIR from either (2) or (3) by applying (1).

2.4. Irregular LDPC Codes

We assume that the reader is familiar with irregular LDPC codes and the standard polynomial description of their degree distributions (DDs). Accordingly, we use $\lambda(x)$ and $\rho(x)$ to denote the bit and check DDs from the edge perspective. This means that $\lambda(x) = \sum_{\nu\geq 1} \lambda_\nu x^{\nu-1}$, where λ_ν (or ρ_ν) denote the fraction of edges attached to a bit (or check) node of degree ν.

The DD of an irregular LDPC code can be viewed either from the edge or node perspective, and this paper is simplified by using both perspectives. We use $L(x)$ and $R(x)$ to denote the bit and check DDs from the node perspective. This means that $L(x) = \sum_{\nu\geq 1} L_\nu x^\nu$, where L_ν (or R_ν) denote the fraction of bit (or check) nodes with degree ν.

Each coefficient represents a fraction of some whole, and this means that $\lambda(1) = \rho(1) = L(1) = R(1) = 1$. Since we do not allow degree 0 nodes, we also note that $L(0) = 0$ and $R(0) = 0$. The possibility of degree 1 nodes is allowed, however, and therefore one cannot assume that $\lambda(0) = 0$ or $\rho(0) = 0$.

The rate of an irregular LDPC code is given by $R = 1 - a_L/a_R$, where $a_L = L'(1) = 1/\int_0^1 \lambda(t)dt$ is the average bit degree and $a_R = R'(1) = 1/\int_0^1 \rho(t)dt$ is the average check degree. One can also switch from the bit to edge perspective by noting that each node of degree ν contributes ν edges to the edge perspective. For the bit nodes, this gives $\lambda(x) = L'(x)/L'(1) = L'(x)/a_L$, and a similar formula holds for the check nodes.

Iterative decoding of irregular LDPC codes on the BEC, with erasure probability δ, was introduced by Luby *et al.* in [1] and refined in [7]. These papers show that the recursion for the erasure probability of the bit-to-check messages is

$$x_{i+1} = \delta\lambda\left(1 - \rho(1 - x_i)\right). \tag{4}$$

3. JOINT ITERATIVE DECODING

3.1. Density Evolution Recursion

Now, we consider a turbo equalization system which performs one channel iteration for each LDPC code iteration. The function, $f(x)$, gives the fraction of erasures produced by the extrinsic output of the channel decoder when the *a priori* erasure rate is x. The update equation for this system is almost identical to (4). The main difference is that the parameter δ now changes with each iteration and is given by δ_i.

There is a fundamental difference between the bit-to-check messages and bit-to-channel messages. This difference is due to the fact that a degree ν bit node sends ν messages to the check nodes and only 1 message to the channel detector. If the erasure probability of all check-to-bit messages passed to a degree ν bit node is x, then the erasure probability of the bit-to-channel message is x^ν. Combining these two observations shows that the erasure probability of all bit-to-channel messages is given by $\sum_{\nu \geq 1} L_\nu x^\nu = L(x)$. Using (4), this means that the recursion for the erasure probability of the bit-to-check messages is given by

$$x_{i+1} = \delta_i \lambda \left(1 - \rho(1 - x_i)\right), \tag{5}$$

where $\delta_i = f\left(L\left(1 - \rho(1 - x_i)\right)\right)$.

3.2. Convergence Condition

Using the recursion (5), we can derive a necessary and sufficient condition for the erasure probability to converge to zero. This condition is written as a basic condition which must hold for $x \in (0, 1]$ and an auxiliary stability condition which simplifies the analysis at $x = 0$. The basic condition,

$$f\left(L\left(1 - \rho(1 - x)\right)\right) \lambda \left(1 - \rho(1 - x)\right) < x, \tag{6}$$

implies there are no fixed points in the iteration for $x \in (0, 1]$. Verifying this condition numerically for very small x can be difficult, so we require instead that $x = 0$ is a stable fixed point of the recursion. This is equivalent to evaluating the derivative of (6) at $x = 0$, which gives the stability condition

$$\lambda^2(0) f'(0) a_L \rho'(1) + \lambda'(0) f(0) \rho'(1) < 1. \tag{7}$$

Now, we can use (6) and (7) to say something about the code properties required by various channels: (i) if the channel has $f(0) > 0$, then the code cannot have any degree 1 bit nodes and the stability condition simplifies to $\lambda_2 f(0) \rho'(1) \leq 1$, (ii) if the channel has $f(0) = 0$, then some degree 1 bit nodes can be used and the stability condition simplifies to $\lambda_1^2 f'(0) a_L \rho'(1) < 1$, and (iii) if the channel has $f(1) = 1$, then the code must have degree 1 check nodes (which act as pilot bits) so that decoding progresses beyond $x = 1$.

3.3. Algebraic Degree Distributions

Armed with the convergence condition, we can now solve for DD pairs which satisfy (6) with equality. The basic idea is that the equation

$$f\left(L\left(1 - \rho(1 - x)\right)\right) \lambda \left(1 - \rho(1 - x)\right) = x$$

is actually a differential equation because $\lambda(x) = L'(x)/a_L$. This allows $\lambda(x)$ to be written in terms of $\rho(x)$ as

$$\lambda(x) = \frac{q(x)}{f\left(F^{-1}\left(a_L Q(x)\right)\right)}, \tag{8}$$

where $q(x) = 1 - \rho^{-1}(1 - x)$, $F(x) = \int_0^x f(t) dt$, and $Q(x) = \int_0^x q(t) dt$. The details of this derivation can be found in [6, p. 164]. The following theorem also shows why meeting (6) with equality is desirable.

Theorem 1. *Consider any LDPC code ensemble, defined by the DDs $\lambda(x)$ and $\rho(x)$, which satisfies (6) for some GEC with ETF $f(x)$. The non-negative gap, Δ, between the rate of the LDPC code and the SIR of the channel, I_s, is given by*

$$\Delta = I_s - R = \int_0^1 g(x) dx,$$

where $g(x) = a_L q(x) - f\left(L(x)\right) L'(x)$ is non-negative.

Proof. This proof can be found in [6, p. 165]. □

3.4. Achieving the SIR

Now, we consider sequences of irregular LDPC code ensembles which can be used to communicate reliably at rates arbitrarily close to the SIR. The code sequence is defined by the sequence of algebraically generated DDs in a manner similar to [8]. The main difficulty that we will encounter is that the implied DDs generally have infinite support and may have negative components. We say that a DD is (i) *admissible* if its power series expansion about $x = 0$ has only non-negative coefficients and (ii) *realizable* if it is a polynomial (i.e., finite degree) whose coefficients sum to one. We say that a sequence of DDs is *SIR achieving* if, for any $\epsilon > 0$, there exists an k_0 such that, for all $k > k_0$, the kth DD is (i) realizable, (ii) satisfies (6), and (iii) has rate $R_k > I_s - \epsilon$.

In general, we construct SIR achieving sequences by starting with a sequence of realizable check DDs and defining algebraically a sequence of bit DDs, denoted $\widetilde{\lambda}(x)$. If each bit DD in this sequence is admissible with $\widetilde{\lambda}(1) > 1$, then we can form a sequence of realizable bit DDs, denoted $\lambda(x)$, by truncating the power series of each $\widetilde{\lambda}(x)$ so that it sums to one. Specifically, we generalize the notation of Section 2.4 and let $\lambda_i = \widetilde{\lambda}_i$ for $1 \leq i < N_k$, where N_k is the smallest integer such that $\sum_{i=1}^{N_k} \widetilde{\lambda}_i \geq 1$. The last term λ_{N_k} is then chosen so that $\lambda(1) = 1$.

$a_R^{(k)} = k$	$a_L^{(k)}$	R_k	Δ_k	N_k	α_k
4	1.595	0.6011	0.0655	4	0.16
5	1.903	0.6193	0.0473	7	0.069
6	2.102	0.6496	0.0170	9	0.048
7	2.411	0.6555	0.0111	19	0.025
8	2.718	0.6602	0.0064	33	0.014
9	3.030	0.6632	0.0034	56	0.0075
10	3.349	0.6651	0.0016	101	0.0042
11	3.677	0.6657	0.0009	184	0.0023

Table 1: Results for the precoded DEC with $\epsilon = 0.5$.

One problem with this method, which does not occur for the BEC [8], is that the truncation may cause the basic condition (6) to fail. To overcome this problem, we require the codes in sequence to satisfy the slightly stronger condition that

$$(1+\alpha_k) f\left(\widetilde{L}^{(k)}(x)\right)\widetilde{\lambda}^{(k)}(x) = q^{(k)}(x), \qquad (9)$$

where $\widetilde{L}(x)$ is defined implicitly via $\widetilde{\lambda}(x)$. This is the same as designing codes for a sequence of degraded channels given by $f^{(k)}(x) = (1+\alpha_k)f(x)$. Solving (9), in the same manner as (8), for the bit DD gives

$$\widetilde{\lambda}^{(k)}(x) = \frac{(1+\alpha_k)^{-1} q^{(k)}(x)}{f\left(F^{-1}\left(F(1) a_R^{(k)} Q^{(k)}(x)\right)\right)}. \qquad (10)$$

The details of this derivation can be found in [6, p. 166]. Each non-negative α_k is chosen so that (6) is satisfied for the original channel. This is not too difficult in practice because varying α_k only changes the truncation point for $\lambda^{(k)}(x)$.

Theorem 2. *Let $\rho^{(k)}(x)$ be a sequence of realizable check DDs and let $\widetilde{\lambda}^{(k)}(x)$ be the sequence of bit DDs given by (10). Suppose that (i) the derivative of $f(x)$ is bounded on $[0,1]$ and $f(1) < 1$, (ii) each $\widetilde{\lambda}(x)$ given by (10) with $\alpha_k = 0$ is admissible, and (iii) the average check degree $a_R^{(k)}$ and maximum bit degree N_k satisfy $a_R^{(k)}/N_k \to 0$. In this case, there exists a α_k sequence such that the sequence of DDs defined above is SIR achieving.*

Proof. This proof can be found in [6, p. 168]. □

4. Results

In this section, we apply the strategy of Theorem 2 to a sequence of check distributions with a single non-zero coefficient. This type of check distribution is called regular, and the sequence is defined by $\rho^{(k)}(x) = x^{k-1}$. This allows us to rewrite (10) as

$$\widetilde{\lambda}^{(k)}(x) = \frac{(1+\alpha_k)^{-1}\left(1-(1-x)^{1/(k-1)}\right)}{f\left(F^{-1}\left(F(1)\frac{(k-1)(1-x)^{k/(k-1)}+kx}{k}\right)\right)}.$$

For the precoded DEC with $\epsilon = 0.5$, we constructed the check regular code sequence for $k = 4, \ldots, 11$. While we were unable to prove that the coefficients of each power series expansion are non-negative, we did verify this numerically for the first 200 coefficients. The results of this experiment are shown in Table 1. Since the choice of α_k in our construction guarantees that each code satisfies the convergence condition for the channel, all of the rates (R_k) and rate gaps (Δ_k) are valid. Although, we cannot prove that this sequence of codes satisfies the conditions of Theorem 2, we can still compare the results to the predictions of the theorem. For one, we find that N_k appears to be growing exponentially with k while Δ_k seems to be exponentially decaying.

For the DEC with $\epsilon = 0.85$ and no precoding, we also constructed the check regular code sequence. In this case, N_k grows so rapidly that we could only construct the codes with $k = 3, 4, 5$. This time, we verified numerically that the first 800 coefficients of each power series are non-negative. The results are shown in Table 2, and again we see exponential growth of N_k and decay of Δ_k.

$a_R^{(k)} = k$	$a_L^{(k)}$	R_k	Δ_k	N_k	α_k
3	2.370	0.2101	0.0088	14	0.00053
4	3.129	0.2177	0.0011	107	0.00018
5	3.906	0.2187	0.0002	757	0.00003

Table 2: Results for the DEC with $\epsilon = 0.85$.

REFERENCES

[1] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, D. A. Spielman, and V. Stemann, "Practical loss-resilient codes," in *Proc. of the 29th Annual ACM Symp. on Theory of Comp.*, pp. 150–159, 1997.

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

[3] B. M. Kurkoski, P. H. Siegel, and J. K. Wolf, "Joint message-passing decoding of LDPC codes and partial-response channels," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1410–1422, June 2002.

[4] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, pp. 1727–1737, Oct. 2001.

[5] S. ten Brink, "Exploiting the chain rule of mutual information for the design of iterative decoding schemes," in *Proc. 39th Annual Allerton Conf. on Commun., Control, and Comp.*, Oct. 2001.

[6] H. D. Pfister, *On the Capacity of Finite State Channels and the Analysis of Convolutional Accumulate-m Codes.* PhD thesis, University of California, San Diego, La Jolla, CA, USA, March 2003.

[7] M. G. Luby, M. Mitzenmacher, and M. A. Shokrollahi, "Analysis of random processes via and-or tree evaluation," in *SODA: ACM-SIAM Symposium on Discrete Algorithms*, pp. 364–373, Jan. 1998.

[8] M. A. Shokrollahi, "New sequences of linear time erasure codes approaching the channel capacity," in *Appl. Alg. in Eng. Commun. Comp.*, pp. 65–76, 1999.