Joint Iterative Decoding of LDPC Codes for Channels with Memory and Erasure Noise

Henry D. Pfister and Paul H. Siegel

***Abstract*—This paper investigates the joint iterative decoding of low-density parity-check (LDPC) codes and channels with memory. Sequences of irregular LDPC codes are presented that achieve, under joint iterative decoding, the symmetric information rate of a class of channels with memory and erasure noise. This gives proof, for the first time, that joint iterative decoding can be information rate lossless with respect to maximum-likelihood decoding. These results build on previous capacity-achieving code constructions for the binary erasure channel. A two state intersymbol-interference channel with erasure noise, known as the dicode erasure channel, is used as a concrete example throughout the paper.**

***Index Terms*—joint iterative decoding, intersymbol interference, low-density parity-check codes, dicode erasure channel, symmetric information rate, capacity-achieving codes.**

I. 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], [2]. This was followed by the work of Richardson, Shokrollahi, and Urbanke [3], which showed that sequences of iteratively decoded LDPC codes may also approach the channel capacity of the binary symmetric channel and the binary-input additive white Gaussian noise (AWGN) channel. Since then, density evolution (DE) [4] has been used to optimize irregular LDPC codes for a variety of memoryless channels (e.g., [5]), and the results suggest, for each channel, that sequences of iteratively decoded LDPC codes can indeed achieve the channel capacity. In fact, the discovery of a binary-input memoryless output-symmetric (BMS) channel whose capacity cannot be approached by LDPC codes would be more surprising than a proof that iteratively decoded LDPC codes can achieve the capacity of BMS channels.

The idea of decoding a code transmitted over a channel with memory via iteration was first introduced by Douillard, *et al.* in the context of turbo codes and is known as *turbo equalization* [6]. Turbo equalization is applied to magnetic recording channels in [7] with a particular emphasis on the effect of channel precoding. Turbo equalization can also be extended to LDPC codes by constructing one large graph which represents the constraints of both the channel and the code. This idea is also referred to as *joint iterative decoding*, and was investigated for partial-response channels by Kurkoski, Siegel, and Wolf in [8]. Tüchler, Koetter, and Singer also considered a variety of channel detectors for turbo equalization in [9]. While this paper focuses mainly on channels with intersymbol interference (ISI), a number of nice results [10], [11], [12] have also been obtained for finite-state fading channels such as the Gilbert-Elliot channel [13].

Until recently, it was difficult to compare the performance of turbo equalization with channel capacity because the binary-input capacity of channels with memory could only be loosely bounded. Recently, new methods were proposed to compute good estimates of the achievable information rates of finite-state (FS) channels [14], [15], [16], and a number of authors have already designed LDPC based coding schemes which approach the achievable information rates of these channels [16], [17], [18], [19], [20], [21], [22]. As is the case with DE for general BMS channels, the evaluation of code thresholds and the optimization of these thresholds is done numerically. For FS channels, analysis of joint decoding is particularly complex because the BCJR algorithm [23] is used to decode the channel.

Since the capacity of a channel with memory is generally not achievable via equiprobable signaling, one can instead aim for the symmetric information rate (SIR) of the channel [24]. The SIR, also known as independent uniform capacity $C_{i.u.d.}$, is defined as the maximum information rate achievable via random coding with independent equiprobable input symbols. This rate is a popular benchmark because it is achievable with random linear codes.

In this paper, we introduce the concept of generalized erasure channels (GECs). For these channels, we show that DE can be done analytically for the joint iterative decoding of irregular LDPC codes and the channel. This allows us to construct sequences of LDPC degree distributions that achieve the SIR using joint iterative decoding. This gives proof, for the first time, that joint iterative decoding of the code and channel can, for a carefully chosen irregular LDPC code, be information rate lossless with respect to maximum-likelihood decoding of the code and channel. The best example of a GEC is the dicode erasure channel (DEC), which is simply a binary-input channel with a linear response of $1 - D$ and erasure noise. This work was initiated in [25, p. 113][26] but stalled for some time due to mathematical difficulties. Recently, these difficulties were overcome using more sophisticated mathematical methods and this paper presents this new approach.

Manuscript received April 13, 2007; revised October 31, 2007. This work was supported by the Center for Magnetic Recording Research at the University of California, San Diego and the National Science Foundation under grants CCR-0219582 and CCF-0514859. The material in this paper was presented in part at the 3rd International Symposium on Turbo Codes, Brest, France, 2003.

H. D. Pfister is with the Electrical and Computer Engineering Department at Texas A&M University, College Station, USA (e-mail: hpfister@tamu.edu).

P. H. Siegel is with the Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla CA 92093-0401 USA (e-mail: psiegel@ucsd.edu).

Digital Object Identifier 10.1109/JSAC.2008.080209.

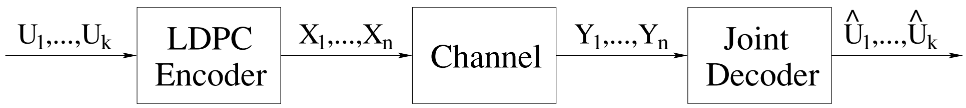


Fig. 1. Block diagram of the system.

It should be noted that many of the techniques in this paper are related to those used in the analysis of irregular repeat-accumulate (IRA) and accumulate-repeat-accumulate (ARA) codes for the BEC [27], [28]. For example, all three papers construct degree distributions algebraically and must verify that a function has a power series expansion with non-negative coefficients. The main technical achievement in this paper is the generalization of these techniques and the application of them to channels with memory (rather than to codes with accumulators). In particular, Theorem 4 is a generalization of a technique used in [27, Appendix A] and Lemma 6 can be used to resolve [27, Conjecture 1] in the affirmative. The use of Polya's Criteria to prove the non-negativity of power series expansions was introduced in [28] and also plays a small role in this work (e.g., Theorem 5). On the other hand, a number of results (e.g., Theorem 1) have their roots in [25, p. 113][26] and were expanded in [27], [28].

In Section II, we introduce the basic components of the system including GECs, the DEC, the joint iterative decoder, and irregular LDPC codes. In Section III, we derive a single parameter recursion for the DE of the joint iterative decoder which enables us to give necessary and sufficient conditions for decoder convergence. These conditions are also used to construct and truncate algebraic degree distributions. In Section IV, sequences of degree distributions are defined and shown to achieve the SIR. In Section V, we show that it is possible to construct a wide variety of GECs. Finally, we offer some concluding remarks in Section VI.

II. System Model

A. Description

We note that random variables will be denoted using capital letters (e.g., U, X, Y) and sequences will be denoted using subscript/superscript notation (e.g., $X_i^j = X_i, X_{i+1}, \ldots, X_j$). The system we consider is fairly standard for the joint iterative decoding of an LDPC code and a channel with memory. Equiprobable information bits, $U_1^k \in \{0,1\}^k$, are encoded into an LDPC codeword, $X_1^n \in \{0,1\}^n$, which is observed through a GEC[1] as the output vector, $Y_1^n \in \mathcal{Y}^n$ where $\mathcal{Y}$ is the channel output alphabet. 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 Y_1^n 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. Fig. 1 shows the block diagram of the system, and Fig. 2 shows the Gallager-Tanner-Wiberg (GTW) graph of the joint iterative decoder. The standard sum-product algorithm [29], [30] is used to pass messages (e.g., log-likelihood ratios (LLRs)) around the decoding graph which describe the decoder's belief that a bit is 0 or 1. For any edge in the graph, let the random variable E be the correct value (given by the encoded value of the bit node attached to that edge). The LLR passed along that edge is

$$\log \frac{\Pr(E=0)}{\Pr(E=1)}.$$

This decoder uses the turbo equalization schedule (i.e., the full channel detector is run each iteration) because this allows the effect of the channel detector to be characterized compactly with a single function. Expanding the channel detector into its own factor graph would also allow the standard single-edge message-passing schedule to be used for joint decoding, but the analysis becomes significantly more complicated. In general, changing the message-passing schedule does not affect the DE threshold, but optimizing the schedule can achieve some complexity savings in practice.

Including a random scrambling vector in the channel makes the performance independent of the transmitted codeword and allows one to analyze the decoder using the all-zero codeword assumption.

B. The Generalized Erasure Channel

Density Evolution involves tracking the evolution of message distributions as messages are passed around the decoder. Since the messages passed around the GTW graph of the joint decoder are LLRs, we find that DE tracks the LLR distribution of the messages. Let L be a random variable representing a randomly chosen LLR at the output of the channel decoder. The complexity of DE depends on the support of the distribution of L.

Definition 1: A *symmetric erasure distribution* is a LLR distribution supported on the set $\{-\infty, 0, \infty\}$ which also satisfies $\Pr(L=-\infty) = \Pr(L=\infty)$. Such distributions are one dimensional, and are completely defined by the erasure probability $\Pr(L=0)$.

Definition 2: 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 optimal[2] APP channel detector (e.g., the distribution of the x_3 messages in Fig. 2) is a symmetric erasure distribution whenever the *a priori* LLR distribution (e.g., the distribution of the x_2 messages in Fig. 2) is a symmetric erasure distribution. Our closed form analysis of this system requires that all the densities involved in DE are symmetric erasure distributions. This allows DE of the joint iterative decoder to be represented by a single parameter recursion. Let $f(x)$ be a function

[1]The channel includes the addition and removal of a random scrambling vector to guarantee channel symmetry.

[2]A suboptimal detector with this property would also be compatible with the DE portion of the analysis.

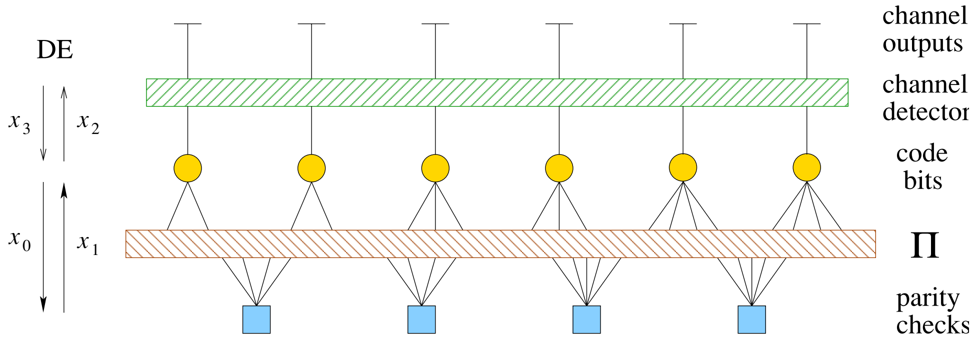


Fig. 2. Gallager-Tanner-Wiberg graph of the joint iterative decoder.

which maps the erasure probability, x, of the *a priori* LLR distribution to the erasure probability at the output of the channel detector. This function completely determines the DE properties of a GEC. Abusing terminology, we refer to $f(x)$ as the *extrinsic information transfer* (EXIT) function of the channel [31].

Of course, $f(x)$ is closely related to the mutual information transfer function, $T(I)$, used by the EXIT chart analysis of ten Brink [32]. This enables the use of a remarkable connection between the mutual information and the mutual information transfer function that was introduced by ten Brink in [33] and extended in [34]. Let the channel input vector X_1^n be i.i.d. and the channel output vector Y_1^n be defined by the length-n vector channel $\Pr\left(Y_1^n = y_1^n | X_1^n = x_1^n\right)$. Let Z_1^n be X_1^n observed through a BEC with erasure probability $1 - I$ for some $I \in [0, 1]$. The mutual information transfer function is defined by

$$T_n(I) \triangleq \frac{1}{n} \sum_{i=1}^{n} I(X_i; Y_1^n, Z_1^n \backslash Z_i),$$

where $Z_1^n \backslash Z_i$ denotes the random vector Z_1^n without Z_i. Its relationship to the mutual information is given by

$$\frac{1}{n} I(X_1^n; Y_1^n) = \int_0^1 T_n(I) dI. \qquad (1)$$

This result follows directly from [34, Thm. 1] if one changes their notation using $\{I_E \to T_n, I_A \to I, V \to X, Z \to A\}$ and replaces the variable of integration with $I_A \to \frac{I_A}{I_{A,max}}$.

For GECs, we assume further that each X_i is binary and equiprobable. Since the mutual information of a BEC with erasure probability x is $1 - x$, this implies that the channel EXIT function is $f_n(x) \triangleq 1 - T_n(1-x)$. Since any BEC with erasure probability $x' > x$ can be viewed as a cascade of two memoryless erasure channels[3] with erasure probabilities x and $1 - \frac{1-x'}{1-x}$, one can apply the Data Processing Inequality [35] to show that $T_n(\cdot)$ and $f_n(\cdot)$ are both non-decreasing functions.

Now, we consider the limit as n goes to infinity. If $T_n(I)$ converges pointwise to $T(I)$, then $f_n(x)$ converges pointwise to $f(x)$ and Lebesgue's dominated convergence theorem (e.g., $0 \le T_n(I) \le 1$) shows that both sides of (1) converge to $\int_0^1 T(I) dI$. When the input process is binary and equiprobable, the LHS of (1) equals the SIR (denoted I_s) and this gives

$$I_s = \int_0^1 T(I) dI = 1 - \int_0^1 f(x) dx. \qquad (2)$$

We mentioned previously that $f(x)$ completely characterizes the DE properties of a GEC, and now we see that it also determines the SIR of the channel. These two properties motivate us to treat all GECs with the same EXIT function as equivalent from a DE and information rate perspective.

C. The Dicode Erasure Channel

The dicode erasure channel (DEC) is a binary-input channel based on the dicode channel (i.e., a linear system with $H(z) = 1 - z^{-1}$ and Gaussian noise) used in magnetic recording [36]. The output of the linear system with $H(z) = 1 - z^{-1}$ and binary inputs (e.g., $+1, 0, -1$) is erased with probability ϵ and transmitted perfectly with probability $1 - \epsilon$. The precoded DEC (pDEC) is essentially the same, except that the input bits are differentially encoded prior to transmission. This modification simply changes the input labeling of the channel state diagram. The state diagram of the dicode channel is shown with and without precoding in Fig. 4.

The simplicity of the DEC allows the BCJR algorithm for the channel to be analyzed in closed form. The method is similar to the exact analysis of turbo codes on the BEC [37], and the result shows that the DEC is indeed a GEC. Leaving the details to Appendix A, we give EXIT functions 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 EXIT function of the channel detector is

$$f_{DEC(\epsilon)}(x) = \frac{4\epsilon^2}{(2 - x(1-\epsilon))^2}. \qquad (3)$$

Using a precoder changes this function to

$$f_{pDEC(\epsilon)}(x) = \frac{4\epsilon^2 x \left(1 - \epsilon(1-x)\right)}{\left(1 - \epsilon(1-2x)\right)^2}. \qquad (4)$$

One can also compute the SIR of the DEC by analyzing only the forward recursion of the BCJR algorithm [25, p. 144]. This results in

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

and it is easy to verify that applying (2) to both (3) and (4) also gives the same result. Fig. 3 provides a few examples of these channel EXIT functions.

[3] The first is a BEC, while the second is a ternary input erasure channel which forwards the erasures.

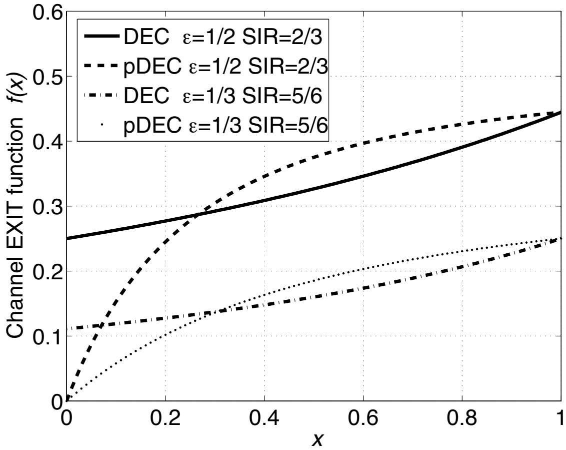


Fig. 3. The channel EXIT functions of the DEC and pDEC as a function of x for two different channel erasure probabilities.

D. Irregular LDPC Codes

Irregular LDPC codes are a generalization of Gallager's LDPC codes [38] that have been shown to perform remarkably well under iterative decoding [3]. They are probably best understood by considering their graphical representation as a bipartite graph, which is shown in the lower part of Fig. 2. Iterative decoding is performed by passing messages along the edges of this graph, and the evolution of these messages can be tracked using DE. In general, when we speak of an LDPC code we are referring to the ensemble of codes formed by picking a random bipartite graph with the proper degree structure. For a more complete description of LDPC codes, one should consult the book by Richardson and Urbanke [31].

For asymmetric memoryless channels, the standard DE assumption of channel symmetry clearly does not hold. This means that DE, without any modifications, can only be applied to one codeword at a time. In [39], a conceptual device named the i.i.d. channel adapter is introduced to overcome this problem. The basic idea is that the channel itself is modified to automatically add a random scrambling pattern to the input sequence and then automatically remove it from the output sequence. Using this, the decoder's output statistics become independent of the input sequence and the all-zero input can be used to assess performance. The concentration theorem follows as a simple corollary when applied to the super-channel formed by the channel and adapter.

For channels with memory, one must rely on the more sophisticated analysis of [19]. This approach chooses a random coset of the LDPC code (by randomizing the odd/even parity of each check) and provides a DE analysis and a concentration theorem for a windowed channel detector[4]. The same result holds for an i.i.d. channel adapter with a windowed decoder. Still, it is not entirely clear that these two methods are identical. For example, if the channel input sequence is fixed to the all-zero sequence, then the first method averages over all possible 2^n transmitted sequences while the second method only averages over 2^{n-k} transmitted sequences. When the channel input sequence is chosen uniformly, however, both approaches average over all 2^n transmitted sequences.

[4]Windowed channel detectors are discussed more thoroughly in Section III-A.

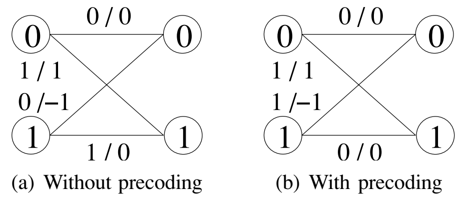


Fig. 4. State diagrams for the noiseless dicode channel with and without precoding. The edges are labeled by the input/output pair.

In this paper, we assume that an i.i.d. channel adapter is always used on the channel. This guarantees that the LLR distribution, provided by the channel detector, of *channel input bits* is symmetric. In contrast, without an i.i.d. channel adapter, the channel detector for the precoded DEC produces more reliable information when 1's are transmitted. The same effect can probably be achieved with the coset method of [19], but similar statements seem to require a joint analysis of the code and channel.

The degree distribution of an irregular LDPC code can be viewed either from the edge or node perspective, and this work is simplified by using both perspectives. Let $\lambda(x)$ be a polynomial defined by $\lambda(x) = \sum_{\nu\geq 1} \lambda_\nu x^{\nu-1}$, where λ_ν is the fraction of edges attached to a bit node of degree ν. Likewise, let $\rho(x)$ be a polynomial defined by $\rho(x) = \sum_{\nu\geq 1} \rho_\nu x^{\nu-1}$, where ρ_ν is the fraction of edges attached to a check node of degree ν. We refer to $\lambda(x)$ and $\rho(x)$ as the bit and check degree distribution from the edge perspective. Let $L(x)$ be a polynomial defined by $L(x) = \sum_{\nu\geq 1} L_\nu x^\nu$, where L_ν is the fraction of bit nodes with degree ν. We refer to $L(x)$ as the bit degree distribution from the node perspective. The coefficients of all these polynomials represent a fraction of some whole, and that means that $\lambda(1) = \rho(1) = L(1) = 1$. Since $L(x)$ has no degree zero bits, we also have $L(0) = 0$. Finally, we note that the possibility of bit and check nodes with degree 1 was included intentionally, and we cannot assume that $\lambda(0) = 0$ or $\rho(0) = 0$.

One can transform a degree distribution from the node perspective to the edge perspective by noting that each node of degree ν contributes ν edges to the edge perspective. Counting from the edge perspective and normalizing gives

$$\lambda(x) = \frac{\sum_{\nu\geq 1} L_\nu \nu x^{\nu-1}}{\sum_{\nu\geq 1} L_\nu \nu} = \frac{L'(x)}{l}, \tag{5}$$

where $l = \sum_{\nu\geq 1} L_\nu \nu = L'(1)$ is the average bit degree. One can also switch from the edge to node perspective by integrating both sides of (5) from zero to one. This gives the formula $l = 1/\int_0^1 \lambda(t)dt$. Since the same rules hold for check degree distributions, this allows us to define the average check degree $r = 1/\int_0^1 \rho(t)dt$. Finally, we note that the design rate of an irregular LDPC code is given by

$$R = 1 - \frac{l}{r}. \tag{6}$$

Iterative decoding of irregular LDPC codes on the BEC, with erasure probability ϵ, was introduced by Luby *et al.* in [1]

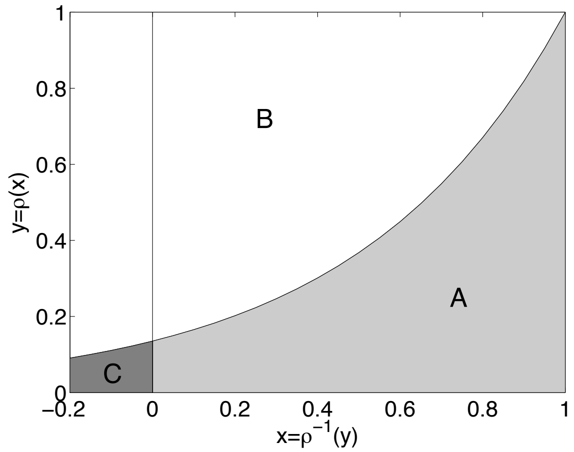


Fig. 5. Schematic showing $\rho(x)$ and $\rho^{-1}(x)$ when $\rho(0) > 0$.

and refined in [40]. The recursion for the erasure probability out of the bit nodes is given by

$$x_{i+1} = \epsilon\lambda\left(1 - \rho(1 - x_i)\right). \tag{7}$$

The recursion is guaranteed to converge to zero as long as $\epsilon\lambda\left(1 - \rho(1 - x)\right) < x$ for all $x \in (0, 1]$.

In many cases, we will also deal with the inverse function $\rho^{-1}(x)$. Since $\rho(x)$ has a power series expansion with positive coefficients, it is strictly increasing on $[0, 1]$ and $\rho^{-1}(x)$ exists and maps $[\rho(0), 1]$ to $[0, 1]$. If $\rho(0) > 0$, then we will also need $\rho(x)$ to be nicely behaved for $x < 0$. Therefore, we require the following condition in general.

Condition 1: The function $\rho(x)$ has a single-valued inverse $\rho^{-1}(x)$ for $x \in (0, 1]$ and is strictly increasing on that domain[5].

Lemma 1: Under Condition 1, we find that

$$\int_0^1 \left(1 - \rho^{-1}(1 - x)\right) dx = \frac{1}{r} + \int_{\rho^{-1}(0)}^0 \rho(x)dx \geq \frac{1}{r}, \tag{8}$$

with equality iff $\rho(0) = 0$.

Proof: From a geometric point of view, $\rho(x)$ divides the unit square into two pieces (denoted A and B in Fig. 5) and therefore

$$\int_0^1 \rho(x)dx + \int_{\rho(0)}^1 \rho^{-1}(x)dx = 1. \tag{9}$$

To show (8), we write

$$\begin{aligned}1 - \int_0^1 \rho^{-1}(1 - x)dx &= 1 - \int_0^1 \rho^{-1}(x)dx \\ &= \int_0^1 \rho(x)dx - \int_0^{\rho(0)} \rho^{-1}(x)dx \\ &= \int_0^1 \rho(x)dx + \int_{\rho^{-1}(0)}^0 \rho(x)dx,\end{aligned}$$

where $-\int_0^{\rho(0)} \rho^{-1}(x)dx = \int_{\rho^{-1}(0)}^0 \rho(x)dx$ because both equal the area of C in Fig. 5. This completes the proof. ■

[5]While this condition is not too restrictive (e.g., it allows check-regular and check-Poisson ensembles), it does exclude large classes of $\rho(x)$ functions. For example, $\rho^{-1}(0)$ does not exist for all even $\rho(x)$ functions (i.e., $\rho(x) = \rho(-x)$) satisfying $\rho(0) > 0$.

III. Analysis of Joint Iterative Decoding

A. Single Parameter Recursion

Now, we consider a turbo equalization system which performs one channel iteration for each LDPC code iteration. The EXIT 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 very similar to (7). The difference is that the effective channel erasure probability (i.e., x_3) should now decrease with each iteration.

We also assume that a small number of pilot bits (a fraction δ of the blocklength) are randomly interspersed in the data stream to help the channel decoder. These bits are known at the receiver and transmitted through the channel, but are not included in the code. Since this increases the number of channel uses without affecting the number of information bits, the effective code rate becomes

$$\widetilde{R} = \frac{k}{(1+\delta)n} = \frac{1}{1+\delta}R.$$

The main benefit is that the pilot bits reduce the erasure rate of messages passed to the channel detector by a factor of $\frac{1}{1+\delta}$. This helps the decoder to get started in decoding and also allows several proofs to be simplified.

The joint decoding graph is shown in Fig. 2 and message-passing schedule and variables are shown on the left. Let $x_0^{(i)}$, $x_1^{(i)}$, $x_2^{(i)}$, and $x_3^{(i)}$ denote the erasure rate of messages passed during iteration i. The update equations are given by

$$\begin{aligned}x_0^{(i+1)} &= x_3^{(i)}\lambda\left(x_1^{(i)}\right) \\ x_1^{(i+1)} &= 1 - \rho\left(1 - x_0^{(i+1)}\right) \\ x_2^{(i+1)} &= L\left(x_1^{(i+1)}\right) \\ x_3^{(i+1)} &= f\left(\tfrac{1}{1+\delta}x_2^{(i+1)}\right).\end{aligned}$$

The first two equations simply describe LDPC decoding when the channel erasure parameter is $x_3^{(i)}$ instead of the fixed constant ϵ. The third equation describes the message passing from the code to the channel detector. It reflects the fundamental difference between the messages passed from the bit nodes to the check nodes and the messages passed from the bit nodes to the channel detector. The difference is due to the fact that a bit node sends one message for each check edge towards the check nodes and only 1 message to the channel detector. Consider a bit node of degree ν whose check edges carry erasure with probability x. The erasure probability is weighted by the fraction of nodes with degree ν and is given by $\sum_{\nu\geq 1} L_\nu x^\nu = L(x)$. The fourth equation takes the channel detector and pilot bits into account and simply maps side information from the code through the EXIT function $f(x)$. Tracking these update equations through one full iteration of the joint decoder gives

$$x_0^{(i+1)} = f\left(\tfrac{1}{1+\delta}L\left(1-\rho(1-x_0^{(i)})\right)\right)\lambda\left(1-\rho(1-x_0^{(i)})\right). \tag{10}$$

Since DE only analyzes the average behavior of iterative decoding, the use of DE is typically motivated by the fact that the actual behavior concentrates around the average with high probability [4][19]. One must be a careful, however, making this statement for the decoder in this paper because the implied decoding graph is not tree-like due to long-range effects in the channel detector. This technical problem can be avoided by analyzing a windowed channel detector (see [19] and [31, Sec. 6.4]) which returns the LLR

$$\ell_i^{(w)} = \log \frac{\Pr\left(X_i = 0 | Y_{i-w}^{i+w}, W_{i-w}^{i+w}\right)}{\Pr\left(X_i = 1 | Y_{i-w}^{i+w}, W_{i-w}^{i+w}\right)},$$

where Y_{i-w}^{i+w} is the windowed channel output vector and W_{i-w}^{i+w} is the windowed *a priori* information (e.g., the edges represented by x_2 in Fig. 2). Edge effects are handled by truncating the window to those Y_j and W_j with $1 \leq j \leq n$. Let $f^{(w)}(x)$ be the channel EXIT function (i.e., erasure probability) for a channel detector restricted to window size w. Since a decoder with larger w has more information, it follows that $w > w'$ implies the LLR distribution of $\ell_i^{(w')}$ is degraded with respect to that of $\ell_i^{(w)}$. This implies that $f^{(w)}(x)$ is non-increasing in w and and that, for any $\zeta > 0$, there is a $w < \infty$ such that $f^{(w)}(x) \leq f(x) + \zeta$. Therefore, applying the concentration theorem for joint decoding with a finite window-size channel detector [19] has negligible loss with respect to the DE analysis for $f(x)$.

B. Conditions for Convergence

Using the recursion (10), we can derive a necessary and sufficient condition for the erasure probability to converge to zero. This condition is typically written as a basic condition which must hold for $x \in (0,1]$ and a stability condition which simplifies the analysis at $x = 0$. The basic condition implies there are no fixed points in the iteration for $x \in (0,1]$ and is given by

$$f\left(\tfrac{1}{1+\delta} L\left(1-\rho(1-x)\right)\right) \lambda\left(1-\rho(1-x)\right) < x. \qquad (11)$$

Verifying this condition for small x is simplified by requiring that $x = 0$ is a stable fixed point of the recursion. This is equivalent to evaluating the derivative of (11) at $x = 0$, which gives the stability condition[6]

$$\left(\tfrac{1}{1+\delta} l\, \lambda^2(0) f'(0) + \lambda'(0) f(0)\right) \rho'(1) < 1. \qquad (12)$$

In the remainder of the paper, we will assume that the channel EXIT function $f(x)$ either satisfies $f(0) > 0$ or its one-sided derivative at 0 satisfies $f'(0^+) > 0$.

Lemma 2 (Rate Gap): The DE iteration converges to zero if and only if

$$f(x) < \frac{1-\rho^{-1}\left(1-L^{-1}\left((1+\delta)x\right)\right)}{\lambda\left(L^{-1}\left((1+\delta)x\right)\right)} \qquad (13)$$

for $x \in (0, x^*]$ where $x^* = \frac{1}{1+\delta} L\left(1-\rho(0)\right)$. For any channel and code satisfying (13), let the non-negative slack be

$$s(x) \triangleq \frac{1-\rho^{-1}\left(1-L^{-1}\left((1+\delta)x\right)\right)}{\lambda\left(L^{-1}\left((1+\delta)x\right)\right)} - f(x). \qquad (14)$$

[6]One may derive (12) from (11) using the facts that $\rho(1) = 1$, $L(0) = 0$, and $L'(x) = l\,\lambda(x)$.

In this case, the *rate gap* between the channel SIR and the effective code rate $\widetilde{R} = \frac{1}{1+\delta} R$ is given by $\Delta \triangleq I_s - \widetilde{R} \geq \int_0^{x^*} s(x)dx$, with equality if $x^* = 1$.

Proof: Using the fact that $L(x), \rho(x)$ are generating functions of non-negative integer random variables, it follows that $g(x) = \frac{1}{1+\delta} L\left(1-\rho(1-x)\right)$ is increasing for $x \in [0,1]$. This implies that $g^{-1}(x)$ exists and maps $[0, x^*]$ to $[0,1]$. Using this, we derive (13) from (11) by dividing both sides by $\lambda(1-\rho(1-x))$ and substituting $x \to g^{-1}(x)$. Since this derivation is reversible, we find that the conditions (13) and (11) are actually equivalent.

Now, we can integrate (14) by treating the two terms separately. Using (2) and the fact that $f(x) \leq 1$ on $[0,1]$, it is easy to verify that the second term satisfies

$$\int_0^{x^*} f(x)dx \geq \int_0^1 f(x)dx - (1-x^*) = x^* - I_s, \qquad (15)$$

with equality if $x^* = 1$. The first term requires the variable change $x = \frac{1}{1+\delta} L(y)$ so that $y^* = 1-\rho(0)$, and gives

$$\begin{aligned}\int_0^{y^*} \frac{1-\rho^{-1}(1-y)}{(1+\delta)\lambda(y)} L'(y)dy &= l \int_0^{y^*} \frac{1-\rho^{-1}(1-y)}{1+\delta} dy \\ &= l \int_{\rho(0)}^{1} \frac{1-\rho^{-1}(t)}{1+\delta} dt \\ &= \frac{l}{1+\delta}\left(\frac{1}{r} - \rho(0)\right), \qquad (16)\end{aligned}$$

where the last step uses (9). Using the convexity of $L(x)$, we find that $L\left(1-\rho(0)\right) \geq 1 - L'(1)\rho(0)$ and this implies that $x^* \geq \frac{1-l\,\rho(0)}{1+\delta}$. Putting the two terms, (15) and (16), together and using this bound gives

$$\begin{aligned}\int_0^{x^*} s(x)dx &\leq \left(\frac{1}{1+\delta}\frac{l}{r} - \frac{l\,\rho(0)}{1+\delta}\right) - (x^* - I_s) \\ &\leq \left(\frac{1}{1+\delta}\frac{l}{r} - \frac{l\,\rho(0)}{1+\delta}\right) - \left(\frac{1-l\,\rho(0)}{1+\delta} - I_s\right) \\ &\leq I_s - \frac{1}{1+\delta}\left(1 - \frac{l}{r}\right),\end{aligned}$$

with equality if $x^* = 1$. Using (6) and the fact that the effective code rate is $\frac{1}{1+\delta} R$ completes the proof. ■

Example 1: Applying Lemma 2 to the (3,6)-regular LDPC code (i.e., $\lambda(x) = x^2$, $\rho(x) = x^5$, and $\delta = 0$) gives

$$f(x) < x^{-2/3}\left(1 - \left(1 - x^{1/3}\right)^{1/5}\right).$$

This function, shown in Fig. 6, is not monotonically increasing and therefore *cannot* be the EXIT function of any real channel. Still any channel function which lies strictly below this curve will satisfy (11) with $\delta = 0$. We can also test this formula with the (2,4)-regular LDPC code (i.e., $\lambda(x) = x$, $\rho(x) = x^3$, and $\delta = 0$) and this gives

$$f(x) < x^{-1/2}\left(1 - \left(1 - x^{1/2}\right)^{1/3}\right). \qquad (17)$$

This function, shown in Fig. 6, is monotonically increasing and therefore *could* be the EXIT function of a real channel. In fact, the (2,4)-regular LDPC code is capacity-achieving for any channel (if it exists) whose EXIT function satisfies (17) with equality.

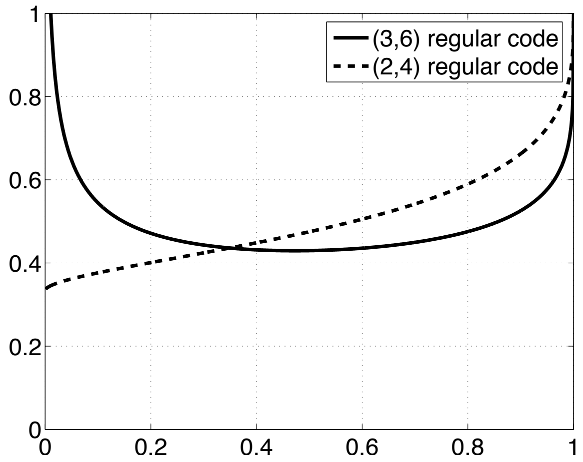


Fig. 6. Upper bounds on $f(x)$ given by Lemma 2.

Lemma 3 (Stability Conditions): The general stability condition (12) has two special cases: (i) If the channel EXIT function satisfies $f(0) > 0$, then the code cannot have degree 1 bits and the stability condition is given by

$$\lambda_2 f(0)\rho'(1) < 1. \tag{18}$$

(ii) If the channel EXIT function is differentiable at zero and satisfies $f(0) = 0$, then the stability condition depends on the fraction of degree 1 bits and is given by

$$\tfrac{1}{1+\delta} l\, \lambda_1^2 f'(0)\rho'(1) < 1.$$

Proof: (i) If the code has degree 1 bits, then $\lambda(0) > 0$. In this case, $f(0)\lambda(0) > 0$ and (11) cannot hold in the neighborhood of $x = 0$. Assuming $\lambda(0) = 0$ allows (12) to simplified to the given form. (ii) If $f(0) = 0$, then (12) simplifies to the given form. ■

C. Solving for $L(x)$

In this section, we solve explicitly for an $L(x)$, for some $f(x)$ and $\rho(x)$, which satisfies (11) with equality for[7] $\delta = 0$. The main idea is that $\lambda(x) = L'(x)/l$ transforms

$$f\left(L\left(1-\rho(1-x)\right)\right)\lambda\left(1-\rho(1-x)\right) = x \tag{19}$$

into a differential equation. The following definitions will be used throughout the remainder of the paper.

Definition 3: The integral of the channel EXIT function is denoted

$$F(x) \triangleq \int_0^x f(t)dt.$$

On $[0,1]$, this integral is well-defined because $f(x)$ is non-decreasing and the result $F(x)$ is strictly increasing because either $f(0) > 0$ or $f'(0) > 0$. For a check degree distribution $\rho(x)$, the matching bit degree distribution [41] (in the memoryless case) is denoted

$$q(x) \triangleq 1 - \rho^{-1}(1-x),$$

and its integral is denoted

$$Q(x) \triangleq \int_0^x q(t)dt.$$

Under Condition 1, this integral is also well-defined on $[0,1]$ because $q(x)$ is non-decreasing.

Theorem 1: Using Condition 1 and the terms defined in Definition 3, we find that

$$L(x) = F^{-1}\left(F(1)\frac{Q(x)}{Q(1)}\right) \tag{20}$$

is a solution of (19) on $[0,1]$ which satisfies $L(0) = 0$ and $L(1) = 1$. This solution also satisfies

$$l = \frac{F(1)}{Q(1)} \tag{21}$$

$$L'(1) = \frac{F(1)q(1)}{f(1)Q(1)}. \tag{22}$$

Finally, the $L(x)$ solution is unique if $f(x)$ and $\rho(x)$ are Lipschitz continuous on $[0,1]$.

Proof: Under Condition 1, we find that $q(x)$ is a one-to-one mapping of $[0,1]$ to $[0,a]$ for some $a \geq 0$. So, we start with (19), change variables $x \to q(x)$, and substitute $\lambda(x) = L'(x)/l$ to get

$$f\left(L(x)\right)\frac{L'(x)}{l} = q(x).$$

Since $f(x)$ and $q(x)$ are both non-decreasing, they are both integrable. Therefore, multiplying both sides by l and integrating gives

$$F\left(L(x)\right) = l\,Q(x) + C. \tag{23}$$

Requiring that $L(0) = 0$ and $L(1) = 1$ shows that $C = 0$ and $l = F(1)/Q(1)$.

Since $F(x)$ is a strictly-increasing mapping from $[0,1]$ to $[0, F(1)]$, we apply the unique inverse to both sides of (23) to get (20). If $f(x)$ and $\rho(x)$ are Lipschitz continuous, then Picard's Theorem [42, p. 734] shows that $L(x)$ is the unique solution to the 1st order differential equation defined by (19), $L(0) = 0$, and $L(1) = 1$. ■

D. Truncating the Degree Distribution

In this section, we use Theorem 1 to construct codes whose rates are close to the SIR of the channel. Since the algebraic construction of SIR-achieving codes generally requires truncating the power series expansion of a function, it is important that the truncated power series represent the original function well. Therefore, rigorous treatment is naturally based on complex analysis and the theory of analytic functions (i.e., functions well represented by power series expansions). The following definitions will be needed for the remainder of this paper.

Definition 4 (Functions and Power Series): Let $\mathcal{A}$ be the set of functions analytic on the open unit disc $D = \{x \in \mathbb{C} \,|\, |x| < 1\}$. Each function $a \in \mathcal{A}$ has a well-defined power series expansion about zero which converges everywhere in D. Henceforth, we will refer to this expansion as the *power series* of a. Let $\overline{\mathcal{A}} \subset \mathcal{A}$ be defined as the $a \in \mathcal{A}$ such that the power series also converges on the boundary ∂D

[7]The main purpose of δ is to simplify the truncation of the degree distribution in Theorem 2.

(i.e., converges on the entire closed disc $\overline{D}$). Let $\mathcal{P} \subset \overline{\mathcal{A}}$ be $a \in \overline{\mathcal{A}}$ such that the power series coefficients are non-negative. For finite sums, we define the truncated power series of $a \in \mathcal{A}$ for $m \in \mathbb{R}^+$ to be

$$a(x;m) \triangleq \sum_{i=0}^{\lfloor m \rfloor} a_i x^i + (m - \lfloor m \rfloor) a_{\lceil m \rceil} x^{\lceil m \rceil}. \qquad (24)$$

This notation is also used for the derivative a' and it follows that

$$\frac{\partial}{\partial x} a(x;m) = a'(x;m-1).$$

Since the goal is to use the series expansion of $L(x)$ from (20) as a bit degree distribution, we require at a minimum that $L \in \mathcal{P}$. Even under this condition, there is still some work to be done. The main problem is that the design rate of the code does not equal the SIR $I_s = 1 - F(1)$. For example, if $\rho(0) = 0$, then (6), (8), and (22) show that

$$\begin{aligned} R &= 1 - \frac{L'(1)}{r} \\ &= 1 - \frac{F(1) q(1)}{r\, f(1) Q(1)} \\ &= 1 - \frac{F(1)}{f(1)} \end{aligned} \qquad (25)$$

because $q(1) = 1$ and $Q(1) = \frac{1}{r}$. If $f(1) < 1$, then the code rate can be increased to I_s by truncating and scaling $L(x)$.

Theorem 2: Let $f(x)$ and $\rho(x)$ satisfy the conditions of Theorem 1 and assume (20) produces an $L \in \mathcal{P}$. A degree distribution $\widetilde{L}$, which achieves an effective rate of $\widetilde{R} \geq I_s - \frac{\delta}{1+\delta}$ and satisfies (11), is given by

$$\widetilde{L}(x) = \frac{L(x;M)}{L(1;M)}$$

where M is the largest positive root of $L'(1; M-1) = \frac{F(1)}{Q(1)}$ and the pilot fraction δ is chosen such that $\frac{1}{1+\delta} = L(1;M)$. In addition, if $f(1) = 1$ and $\rho(0) = 0$, then $M = \infty$, $L(1;M) = 1$, and any $\delta > 0$ leads to successful decoding.

Proof: First, we note that if $f(1) = 1$ and $\rho(0) = 0$, then (25) shows the design rate is $I_s = 1 - F(1)$ and no truncation is required. Decoding will be successful for any pilot fraction $\delta > 0$ because any non-decreasing $f \in \mathcal{A}$ with $f(1) = 1$ is either constant (e.g., $f(x) = 1$ is degenerate with $I_s = 0$) or strictly increasing. Therefore, any $\delta > 0$ is sufficient to satisfy (11).

Next, we assume that $f(1) < 1$. In this case, $L'(1; m-1)$ is a continuous and non-decreasing map (in m) with $L'(1;0) = 0$ where (22) shows that[8] $\lim_{m\to\infty} L'(1; m-1) \geq \frac{F(1)q(1)}{Q(1)f(1)}$. Since $q(1) = 1 - \rho^{-1}(0) \geq 1$, we find that M exists and is finite. Furthermore, $L \in \mathcal{P}$ and $\lim_{m\to\infty} L'(1; m-1) > F(1)/Q(1)$ implies that the truncation of $L(x)$ to $\lceil M \rceil$ terms must remove positive terms. This allows us to write

$$\begin{aligned} f\left(\frac{1}{1+\delta}\widetilde{L}(x)\right) &= f\left(L(1;M)\widetilde{L}(x)\right) \\ &= f\left(L(x;M)\right) \\ &< f\left(L(x)\right) \end{aligned}$$

[8]Equality occurs unless the power series of $L'(x)$ diverges at $x = 1$.

and

$$\begin{aligned} \widetilde{\lambda}(x) &= \frac{\widetilde{L}'(x)}{\widetilde{L}'(1)} \\ &= \frac{L'(x; M-1)}{L'(1; M-1)} \\ &= \frac{L'(x; M-1)}{F(1)/Q(1)} \\ &< \frac{L'(x)}{F(1)/Q(1)} \\ &= \lambda(x) \end{aligned}$$

for $x \in (0,1]$, where the strict inequalities follow from the truncation of positive terms. Combining these two expressions shows that

$$\begin{aligned} f\left(\frac{1}{1+\delta}\widetilde{L}(x)\right)\widetilde{\lambda}(x) &< f\left(L(x)\right)\lambda(x) \\ &= 1 - \rho^{-1}(1-x), \end{aligned}$$

and proves (11) for the truncated distribution $\widetilde{L}(x)$. Finally, we compute the effective code rate with

$$\begin{aligned} \widetilde{R} &= \frac{1}{1+\delta}\left(1 - \frac{\widetilde{L}'(1)}{r}\right) \\ &= \frac{1}{1+\delta}\left(1 - \frac{(1+\delta)F(1)}{r\,Q(1)}\right) \\ &= 1 - \frac{\delta}{1+\delta} - F(1)\frac{1}{rQ(1)}. \end{aligned}$$

Using (8), we see that $Q(1) \geq \frac{1}{r}$ and $\widetilde{R} \geq I_s - \frac{\delta}{1+\delta}$. ■

Remark 1: We note that, by itself, Theorem 2 provides no guarantee that the achievable rate is close to I_s (i.e., that δ is small). In the next section, we will design sequences of codes which achieve this goal.

IV. Achieving the Symmetric Information Rate

A. Sequences of Codes

Now, we consider sequences of irregular LDPC code ensembles which can be used to communicate reliably at rates arbitrarily close to the SIR. The basic idea is to start with a sequence of check degree distributions and apply Theorems 1 and 2. This process is significantly simplified by choosing the check-Poisson LDPC ensemble (i.e., $\rho(x) = e^{\alpha(x-1)}$) because the entire sequence can be described in terms of a single bit degree distribution $L(x)$.

Theorem 3: Consider the check-Poisson LDPC codes implied by $\rho(x) = e^{\alpha(x-1)}$. In this case, Theorem 1 gives

$$L(x) = F^{-1}\left(F(1)\left(x + (1-x)\log(1-x)\right)\right), \qquad (26)$$

which is independent of α. If $L \in \mathcal{P}$, then we let $\widetilde{R}_\alpha$ be the effective rate achieved by the construction of Theorem 2. In this case, we can achieve the SIR by increasing α because

$$\lim_{\alpha\to\infty} \widetilde{R}_\alpha = I_s.$$

Proof: From $\rho(x)$, we find that $Q(x) = \frac{1}{\alpha}\left(x + (1-x)\log(1-x)\right)$. Applying Theorem 1, we see that α cancels out of $L(x)$. For any α, we can apply Theorem 2 to construct a code which satisfies (11). Since $L_i \leq 1$, we see that $L_i i \leq i$ and $L'(1;M-1) \leq \sum_{i=1}^{\lceil M \rceil} i \leq M^2$. This implies that M, which is defined by $L'(1;M-1) = F(1)/Q(1) = \alpha F(1)$, must satisfy $M \geq \sqrt{\alpha F(1)}$ so that $M \to \infty$ as $\alpha \to \infty$. The fact that $L \in \mathcal{P}$ also implies that $L(1;M) \to L(1) = 1$ as $M \to \infty$. Since δ is defined by $\frac{1}{1+\delta} = L(1;M)$, we see that $L(1;M) \to 1$ implies $\delta \to 0$. Therefore, the gap to the SIR $\frac{\delta}{1+\delta}$ vanishes as $\alpha \to \infty$. ■

The following definition sets the stage for a number of results that will be combined to prove the non-negativity, for a particular channel and check degree distribution, of the power series expansion of the degree distribution $L(x)$.

Definition 5: The G-function of an channel EXIT function is defined to be

$$G(x) \triangleq F^{-1}\left(F(1)x\right).$$

Computing the G-function of a channel is the first step towards constructing a SIR-achieving degree sequence for that channel. This function represents the transformation of the degree distribution, relative to a memoryless channel, required to achieve the SIR. For simplicity, we also define

$$\widetilde{Q}(x) \triangleq \frac{Q(x)}{Q(1)} \tag{27}$$

so that $L = G \circ \widetilde{Q}$ is the functional composition of $\widetilde{Q}$ followed by G.

Applying Theorem 3 requires that $L \in \mathcal{P}$, and the following describes two methods of proving $L \in \mathcal{P}$. The first is proved in Theorem 4 and is based on combining the power series of G and $\widetilde{Q}$ analytically. The second is proved in Theorem 5 and requires numerical verification of the convexity of a particular function to show that $L \in \mathcal{P}$.

The following Lemma gives a sufficient condition for $G \circ \widetilde{Q} \in \overline{\mathcal{A}}$ which depends on G and $\widetilde{Q}$ separately.

Lemma 4: If $G \in \overline{\mathcal{A}}$ and $\widetilde{Q} \in \mathcal{P}$, then $G \circ \widetilde{Q} \in \overline{\mathcal{A}}$.

Proof: We start by writing $L = G \circ \widetilde{Q}$ as the composition of two infinite sums

$$(G \circ \widetilde{Q})(x) = \sum_{j=0}^{\infty} G_j \left(\sum_{i=0}^{\infty} \widetilde{Q}_i x^i \right)^j .$$

The composed sum converges uniformly on $\overline{D}$ because the inner sum converges uniformly on $\overline{D}$ to a value with magnitude at most one and the outer sum converges uniformly on $\overline{D}$ as well. Therefore, the terms of the composed sum can be sorted by the degree of x without changing the answer. Sorting by degree exposes the power series of L and shows that it must also converge uniformly on $\overline{D}$. ■

The approach taken to prove $L \in \mathcal{P}$ is based on the approach taken in [27] for bit-regular IRA codes. If the $G \in \overline{\mathcal{A}}$ has an alternating power series which decays fast enough and Q satisfies a certain condition, then we can show $L \in \mathcal{P}$ by combining positive and negative terms.

Theorem 4: Let $G \in \overline{\mathcal{A}}$ have a power series which satisfies (i) the first non-zero term is positive with degree $n_0 \geq 1$, (ii) G_{n_0+k} has alternating sign for $k \in \mathbb{N}$, and (iii) $\exists \gamma > 0$ such that $G_{n_0+2k+1} \geq -\gamma G_{n_0+2k}$ for $k \in \mathbb{N}$. Let $\widetilde{Q} \in \mathcal{P}$ satisfy (iv) $\widetilde{Q} - z\widetilde{Q}^2 \in \mathcal{P}$ for $z \in [0,\gamma]$. In this case, we find that $G \circ \widetilde{Q} \in \mathcal{P}$.

Proof: Using Lemma 4, we see that $G \circ \widetilde{Q} \in \overline{\mathcal{A}}$. To show $G \circ \widetilde{Q} \in \mathcal{P}$, we must simply show that the power series of $G \circ \widetilde{Q}$ has non-negative coefficients. We start by noting that if $a, b \in \overline{\mathcal{A}}$ have non-negative power series, then both $a + b$ and $a \cdot b$ have non-negative power series. Expanding $G\left(\widetilde{Q}(x)\right) = \sum_{i=n_0}^{\infty} G_i \widetilde{Q}(x)^i$ gives

$$\sum_{k=0}^{\infty} \widetilde{Q}(x)^{2k+n_0-1} \left[G_{n_0+2k}\widetilde{Q}(x) + G_{n_0+2k+1}\widetilde{Q}(x)^2 \right].$$

Using (i) and (ii), we see that $G_{n_0+2k} \geq 0$ while $G_{n_0+2k+1} \leq 0$. Using (iii) and (iv), we see that the bracketed term has a non-negative power series. Since $\widetilde{Q}(x)^n$ has a non-negative power series for integer $n \geq 0$, we see that the entire expression is the sum of products of functions with non-negative power series. ■

Theorem 5 (Polya's Criteria): Let the function $a \in \mathcal{A}$ be well-defined and continuous on $\overline{D}$, and real on the subset $\overline{D} \cap \mathbb{R}$. If $h(x) \triangleq \mathrm{Re}\left[2a\left(e^{ix}\right)\right]$ is convex on $[0,\pi]$ and $a(0) \geq 0$, then $a \in \mathcal{P}$. Furthermore, if $\lim_{x\to 0} x\, a(1-x) = 0$, then the same holds even if a has a singularity at $x = 1$.

Proof: See [28, Appendix III]. ■

B. Check Degree Distributions

There are two families of check-degree distributions which are of primary interest. The first is the check-Poisson distribution defined by $\rho(x) = e^{\alpha(x-1)}$ and the associated $\widetilde{Q}_P$ is given by

$$\widetilde{Q}_P(x) = x + (1-x)\log(1-x). \tag{28}$$

The second is the check-regular distribution defined by $\rho(x) = x^{d-1}$ and its associated $\widetilde{Q}_R$ is given by

$$\widetilde{Q}_R(x) = x - (d-1)(1-x)\left(1 - (1-x)^{1/(d-1)}\right). \tag{29}$$

It turns out that these two families are actually asymptotically equivalent because $\widetilde{Q}_R$ converges to $\widetilde{Q}_P$ as $d \to \infty$. The following lemmas establish some properties of these $\widetilde{Q}$ functions which can be used to show $L \in \mathcal{P}$ for various channels.

Lemma 5: For $\widetilde{Q}_P$, we have (i) $\widetilde{Q}_P - z\widetilde{Q}_P^2 \in \mathcal{P}$ for $z \in \left[0, \frac{3}{10}\right]$ and (ii) $\sqrt{\widetilde{Q}_P} \in \mathcal{P}$.

Proof: For (i), see Appendix B-B. For (ii), we have a computer assisted proof based on applying Theorem 5 to the function

$$a(x) = \sqrt{\frac{x + (1-x)\log(1-x)}{x^2}}.$$

■

Lemma 6: For $\widetilde{Q}_R$, we have (i) $\widetilde{Q}_R - z\widetilde{Q}_R^2 \in \mathcal{P}$ for $z \in [0, \gamma_d]$ with

$$\gamma_d = \begin{cases} \frac{2d^2-3d+1}{6d^2+6d} & \text{if } 2 \leq d \leq 7 \\ \frac{6d^2-5d+1}{20d^2+20d} & \text{if } 7 \leq d < \infty \end{cases}. \tag{30}$$

Proof: See Appendix B-C. ■

Remark 2: Since $\gamma_d \nearrow \frac{3}{10}$, we expect that most code constructions will work equally well with the check-Poisson degree distribution (as $\alpha \to \infty$) and the check-regular distribution (as $d \to \infty$). This is not guaranteed, however, because it is easy to construct function sequences where $\lim_{d\to\infty} a_d \in \mathcal{P}$ does not imply that there exists any $d < \infty$ such that $a_d \in \mathcal{P}$.

C. Channel G-functions

The purpose of the following three sections is to present and discuss properties of channel G-functions for three different channel EXIT functions. The discussion focuses mainly on the conditions required by Theorem 4 to prove the non-negativity of L for that channel and a particular check degree distribution. In each case, the main result is the range of channel parameters for which we can prove $L \in \mathcal{P}$. There is also some discussion of channel parameters where numerical results can show conversely that $L \notin \mathcal{P}$.

1) The Dicode Erasure Channel: For the DEC, the G-function $G_{DEC(\epsilon)}(x)$ is computed in Appendix C-A and is given by

$$G_{DEC(\epsilon)}(x) = \frac{2x}{(1+\epsilon)+(1-\epsilon)x} = -\frac{2}{1-\epsilon}\sum_{n=1}^{\infty}\left(-\frac{1-\epsilon}{1+\epsilon}x\right)^n. \quad (31)$$

Applying Theorem 4, we see that if $\widetilde{Q} - \frac{1-\epsilon}{1+\epsilon}\widetilde{Q}^2 \in \mathcal{P}$ then $L \in \mathcal{P}$. Using Lemma 5 for the check-Poisson ensemble, we see that this requires $\frac{1-\epsilon}{1+\epsilon} < \frac{3}{10}$ or $\epsilon > 7/13$. In fact, this condition is tight and $\epsilon < 7/13$ implies that L has a negative power series coefficient.

2) The Precoded Dicode Erasure Channel: For the precoded DEC, the G-function $G_{pDEC(\epsilon)}(x)$ is computed in Appendix C-B and is given by

$$G_{pDEC(\epsilon)}(x) = \frac{\epsilon x}{1+\epsilon} + \sqrt{\frac{1-\epsilon}{1+\epsilon}x}\sqrt{1+\frac{\epsilon^2 x}{1-\epsilon^2}} \quad (32)$$
$$= \frac{\epsilon x}{1+\epsilon} + \sqrt{\frac{1-\epsilon}{1+\epsilon}x}\sum_{n=0}^{\infty}\binom{\frac{1}{2}}{n}\left(\frac{\epsilon^2 x}{1-\epsilon^2}\right)^n.$$

Since $\binom{1/2}{n+1} \geq -\binom{1/2}{n}$ for odd n, we can apply Theorem 4 to $G(x) = -1 + \sqrt{1+\frac{\epsilon^2 x}{1-\epsilon^2}}$. Therefore, if $\sqrt{\widetilde{Q}} \in \mathcal{P}$ and $\widetilde{Q} - \frac{\epsilon^2}{1-\epsilon^2}\widetilde{Q}^2 \in \mathcal{P}$ then $L \in \mathcal{P}$. For the check-Poisson ensemble, this requires $\epsilon < \sqrt{\frac{3}{13}}$ so that $\frac{\epsilon^2}{1-\epsilon^2} < \frac{3}{10}$. This condition, however, is not tight. Using Theorem 5, we have a computer assisted proof that $L \in \mathcal{P}$ for $\epsilon \leq \sqrt{\frac{1}{2}}$. Numerical results also prove that $\epsilon \geq 0.83$ implies $L \notin \mathcal{P}$ and suggest that $\epsilon < 0.82$ implies $L \in \mathcal{P}$.

3) The Simple Linear Channel: Since we can use the simple linear function $f(x) = ax+b$ as a first order approximation of any other channel EXIT function, we also consider this channel. The G-function $G_{lin}(x)$ is computed in Appendix C-C and is given by

$$G_{lin}(x) = \frac{b}{a}\left(\sqrt{1+\frac{2a}{b^2}\left(\frac{a}{2}+b\right)x} - 1\right) = \frac{b}{a}\sum_{n=1}^{\infty}\binom{\frac{1}{2}}{n}\left(\frac{2a}{b^2}\left(\frac{a}{2}+b\right)x\right)^n.$$

Since $\binom{1/2}{n+1} \geq -\binom{1/2}{n}$ for odd n, we can apply Theorem 4. This shows that, if $\widetilde{Q} \in \mathcal{P}$ and $\widetilde{Q} - \frac{2a}{b^2}\left(\frac{a}{2}+b\right)\widetilde{Q}^2 \in \mathcal{P}$, then $L \in \mathcal{P}$. For the check-Poisson ensemble, this means that $a < \left(\sqrt{\frac{13}{10}} - 1\right)b$ implies $\frac{2a}{b^2}\left(\frac{a}{2}+b\right) < \frac{3}{10}$ and $L \in \mathcal{P}$. Numerical results also prove that $a \geq 0.49\,b$ implies $L \notin \mathcal{P}$ and suggest that $a \leq 0.48\,b$ implies $L \in \mathcal{P}$. We also note that, if $b = 0$, then $G_{lin}(x) = \sqrt{x}$ and $L \in \mathcal{P}$ if and only if $\sqrt{\widetilde{Q}} \in \mathcal{P}$.

D. General Code Sequences

For general code sequences, it is necessary to show that each element in the sequence is in $\mathcal{P}$ and that the rate tends to the SIR. The basic idea is to start with a sequence of check degree distributions and apply Theorems 1 and 2. For a sequence of functions $\left\{a^{(k)}(x)\right\}$, we define the sequence of power series expansions by $a^{(k)}(x) = \sum_{i=0}^{\infty} a_i^{(k)} x^i$. The following lemma gives a condition sufficient to show the rate approaches the SIR.

Let $\left\{\rho^{(k)}(x)\right\}$ be a sequence of check degree distributions with $\rho^{(k)}(0) = 0$ and average degree $r_k = 1/\int_0^1 \rho^{(k)}(x)dx$. Let $\left\{L^{(k)}(x)\right\}$ be the associated sequence of bit degree distributions defined by Theorem 1. If $L^{(k)} \in \mathcal{P}$ for all $k \geq k_0$, then we can apply Theorem 2 to get the implied sequences $\{M_k\}$ and $\{\delta_k\}$. If $\delta_k \to 0$, then the sequence of codes achieves the SIR. The following Lemma provides a sufficient condition for a sequence of codes to achieve the SIR.

Lemma 7: If the sequence of derivatives $\left\{L^{(k)\,\prime}(x)\right\}$ satisfies $L^{(k)\,\prime} \in \mathcal{P}$ for all $k \geq k_0$, then $L^{(k)}(1; M_k) \geq 1 - \frac{r_k}{M_k}$ and $\frac{r_k}{M_k} \to 0$ implies that $\delta_k = O\left(\frac{r_k}{M_k}\right) \to 0$.

Proof: If $L^{(k)\,\prime} \in \mathcal{P}$, then we can use (8) and (22) to write

$$L^{(k)\,\prime}(1) = \sum_{i=1}^{\infty} L_i^{(k)} i = \frac{r_k F(1)}{f(1)} \leq r_k,$$

where the last step follows from $F(1) \leq f(1)$. Therefore, we have

$$1 - L^{(k)}(1; M_k) \leq \sum_{i=\lceil M_k \rceil}^{\infty} L_i^{(k)} \leq \frac{1}{M_k}\sum_{i=\lceil M_k \rceil}^{\infty} L_i^{(k)} i \leq \frac{r_k}{M_k}.$$

Since $\delta_k = \frac{1-L^{(k)}(1;M_k)}{L^{(k)}(1;M_k)}$, we find that $\frac{r_k}{M_k} \to 0$ implies $\delta_k = O\left(\frac{r_k}{M_k}\right) \to 0$. ■

Lemma 8: The truncation point M in Theorem 2 can be lower bounded by solving for M in the inequality

$$Q'\left(\frac{M}{M+1}\right) \geq \begin{cases} e^{-1} f(0) & \text{if } f(0) > 0 \\ e^{-1} f\left(F^{-1}\left(\frac{F(1)Q(\frac{1}{2})}{Q(1)}\right)\right) & \text{if } L_1 < \frac{F(1)}{Q(1)} \end{cases}.$$

Proof: Using the fact that $L \in \mathcal{P}$, we have the upper bound (for $x \in (0,1]$)

TABLE I
CHECK-POISSON RESULTS FOR THE DEC WITH $\epsilon = \frac{2}{3}$.

M_k	α_k	r_k	δ_k
10	4.73	4.77	0.0815
20	5.73	5.75	0.0408
40	6.76	6.77	0.0205
80	7.81	7.81	0.0103
160	8.87	8.87	0.0052

TABLE II
CHECK-REGULAR RESULTS FOR THE PRECODED DEC WITH $\epsilon = \frac{2}{3}$.

r_k	M_k	δ_k
3	4	0.0338
4	10	0.1100
5	27	0.0113
6	78	0.0039
7	223	0.0014
8	632	0.0005

$$\begin{aligned} L'(1; M-1) &\le \sum_{i=1}^{\lceil M \rceil} L_i i \\ &\le x^{-\lceil M \rceil + 1} \sum_{i=1}^{\lceil M \rceil} L_i i x^{i-1} \\ &\le x^{-M} L'(x). \end{aligned}$$

Since $\left(1 - \frac{1}{M+1}\right)^{-M} \le e$ (for $M \ge 1$) and $L'(1; M-1) = F(1)/Q(1)$, we can choose $x = 1 - \frac{1}{M+1}$ to get

$$\begin{aligned} \frac{F(1)}{Q(1)} &\le e\, L'\left(1 - \frac{1}{M+1}\right) \\ &= \frac{e\, F(1)}{Q(1)} \frac{Q'\left(1 - \frac{1}{M+1}\right)}{f\left(F^{-1}\left(F(1) Q\left(1 - \frac{1}{M+1}\right)/Q(1)\right)\right)}. \end{aligned}$$

To get the final result, we start by solving for $Q'\left(1 - \frac{1}{M+1}\right)$. If $f(0) > 0$, then the expression is simplified by using $f(x) \ge f(0)$. If $L_1 < F(1)/Q(1)$, then the expression is simplified by noting that truncation point satisfies $M \ge 1$ so that $Q\left(1 - \frac{1}{M+1}\right) \ge Q\left(\frac{1}{2}\right)$. ■

Remark 3: For check-Poisson and check-regular codes, applying Lemma 8 shows the truncation point M_k grows exponentially with the average right degree r_k for channels with $f(0) > 0$. For the check-Poisson sequence, we have $Q^{(k)\prime}(x) = -\frac{1}{\alpha_k} \ln(1-x)$ where the average right degree satisfies $\alpha_k \le r_k \le 2\alpha_k$ for $\alpha_k \ge 1$. Applying Lemma 8 for $\alpha_k \ge 1$ gives

$$M_k \ge e^{\alpha_k e^{-1} f(0)} - 1 \ge e^{r_k e^{-1} f(0)/2}. \tag{33}$$

For the check-regular sequence, we have $Q^{(k)\prime}(x) = 1 - (1-x)^{1/(r_k - 1)}$ and this gives

$$M_k \ge \left(\frac{1}{1 - e^{-1} f(0)}\right)^{r_k - 1}.$$

Therefore, check-regular sequences, which satisfy the conditions of Lemma 7, have rates approaching the SIR exponentially fast in the average right degree.

Example 2: A closed-form expression for $L(x)$ is obtained for check-Poisson codes on the DEC by combining (28) and (31) to get

$$L(x) = \frac{2\left(x + (1-x)\log(1-x)\right)}{(1+\epsilon) + (1-\epsilon)\left(x + (1-x)\log(1-x)\right)}.$$

Using the methods of [43], one can show that $L_M = O\left(M^{-2}\right)$ so that $1 - L(1; M) = O(M)$. Combining this with (33) implies the gap to capacity decays exponentially with the average check degree. A sequence of codes is easily constructed, for each k, by choosing an integer truncation point M_k and then calculating α_k and δ_k from $L'(1; M-1) = \alpha_k F(1)$ and $\frac{1}{1+\delta_k} = L(1; M_k)$. Table I gives numerical results for the sequence of codes (i.e., $M_k, \alpha_k, r_k, \delta_k$) when $\epsilon = \frac{2}{3}$.

Example 3: For the check-regular sequence $\rho^{(k)}(x) = x^{k-1}$ and the precoded DEC, we can write the expression for the sequence $L^{(k)}(x)$ by combining (29) and (32). Using the fact that $L_1^{(k)} = L^{(k)\prime}(0) = \sqrt{\frac{k}{10k-10}} < k\, F(1)$ for $k \ge 2$, one can combine Lemmas 7 and 8 to show that the gap to capacity decays exponentially with the check degree. Table II gives numerical results for $k = 3, \ldots, 8$ (i.e., r_k, M_k, δ_k) and $\epsilon = \frac{2}{3}$.

V. RESULTS FOR ARBITRARY GECS

A. The Existence of Arbitrary GECs

From what we have discussed so far, it is not clear that the set of GECs contains anything more than the DEC with and without precoding. Nothing in our analysis, however, prevents us from considering the much larger family of GECs implied by any non-decreasing $f(x)$ which maps the interval $[0, 1]$ into itself. Moreover, we believe that it is possible to construct, albeit somewhat artificially, a binary-input GEC for any such $f(x)$. This would mean that, in some sense, there is a GEC for every well-defined EXIT function.

The EXIT function of a GEC is defined as mapping from the *a priori* erasure probability of the channel decoder to the erasure probability of the extrinsic output. If the *a priori* messages from the code to the channel decoder are divided randomly into two groups of equal size, then the erasure probability in the two groups will be the same. Now, suppose that these groups of bits are sent through different GECs. In this case, the extrinsic messages from the first channel will have erasure probability $f_1(x)$ and the extrinsic messages from the second channel will have erasure probability $f_2(x)$. Since the two groups were chosen at random, the average erasure probability of all the extrinsic messages passed back to the code will be $\left(f_1(x) + f_2(x)\right)/2$. This idea of linearly combining channels was first introduced in the context of EXIT charts and doping [44]. It also extends naturally to an arbitrary weighted combinations of GECs.

Now, consider the performance of a length-$2k$ rate-$\frac{1}{2}$ binary linear code where each entry in the generator matrix is chosen uniformly from $\{0, 1\}$. It is well-known that this code will decode (with high probability) if the erasure rate is less than $\frac{1}{2}$ and fail to decode (with high probability as $k \to \infty$) if the erasure rate is greater than $\frac{1}{2}$. We would also like to to characterize the threshold of a code whose systematic bits are erased with probability x and the non-systematic bits are erased with probability y. To do this, we recall that a uniform random $c \times d$ binary matrix is full rank with probability $\psi_{c,d} = \prod_{j=0}^{c-1} \left(1 - 2^{j-d}\right)$ for $c \le d$ (e.g., [31, Prob. 3.20]). This implies that a submatrix consisting of any set of $k + \log_2 k$ columns will be full rank with probability $\psi_{k,k+\log_2 k} \to 1$ as $k \to \infty$. Therefore, a maximum likelihood decoder should be able to recover (by inverting the matrix associated with erased bits) all of the systematic bits as long as the average bit erasure rate, $\frac{x+y}{2}$, is less than $\frac{1}{2}$. If the average erasure rate is larger than $\frac{1}{2}$, then there will be, with probability 1, some systematic bits which cannot be recovered. Applying the area theorem shows that the limiting extrinsic erasure rate, at the output of an APP decoder for this code, must be given by function

$$U(x+y-1) = \begin{cases} 0 & \text{if } x+y<1 \\ 1 & \text{if } x+y>1 \end{cases},$$

where $U(x)$ is the unit step function and $U'(x)$ is the Dirac delta function. In this case, the value $U(0)$ is probably best defined as $\frac{1}{2}$ because the variation in the number of erasures is $O(\sqrt{k})$ while the transition width for decoding is $O(\log k)$. Therefore, with $x + y = 1$ decoding will succeed exactly half the time as $k \to \infty$.

This ensemble of codes can be treated as a GEC whose inputs are the systematic bits and whose outputs are the parity bits. In this case, the EXIT function of the GEC with parameter y and *a priori* erasure rate x is given by $U(x+y-1)$. If we choose the pdf of the channel erasure rate y to be

$$w(y) = f(0)U'(1-y) + f'(1-y) + (1 - f(1))\, U'(y),$$

then $\int_{0^-}^{1^+} w(y)dy = 1$ and the average EXIT function is given by $\int_{0^-}^{1^+} w(y)U(x+y-1)dy$. Using the substitution $z = 1-y$ and integrating by parts shows that

$$\int_{0^-}^{1^+} w(y)U(x+y-1)dy = f(x) + (1 - f(1))\, U(x-1),$$

which equals $f(x)$ for[9] $x \in [0,1)$. This means that any differentiable $f(x)$ can be approximated arbitrarily well by a linear combination of long systematic rate-$\frac{1}{2}$ codes. Furthermore, this construction of an arbitrary channel EXIT function allows the results of Sections III and IV to be applied (in a somewhat meaningful way) to any suitably analytic non-decreasing $f(x)$.

B. General Channels

In this section, we discuss some implications of this work on the joint iterative decoding of LDPC codes for general channels with memory. Of course, the design of capacity-achieving codes for general memoryless channels still remains an open problem. On one hand, this would appear to make the same problem for general channels with memory more difficult. On the other hand, allowing memory enlarges the space of possible channels and therefore it may be possible to solve the problem for a particularly "nice" channel.

[9]The problematic point $x=1$ is defined via $f(1) \triangleq \lim_{x \to 1^-} f(x)$.

We can also gain some insight into the stability condition for general channels with memory. Consider a general channel with memory which remains noisy even in the presence of perfect *a priori* information. This is analogous to a GEC with $f(0) > 0$. In this case, the stability condition is determined by the LLR density returned by the channel given perfect *a priori* information. For a GEC, this follows from the fact that (18) is the stability condition for an erasure channel with erasure probability $f(0)$. Let $P_0(x)$ be the LLR density at the extrinsic output of the channel decoder, for a general channel, when perfect *a priori* information is passed to the decoder. As long as $P_0(x)$ does not have perfect information itself (i.e., it is not equal to a delta function at infinity), then the stability condition is given by applying the memoryless channel condition from [3] to $P_0(x)$. This makes sense because, when the joint decoder is near convergence, the LLRs passed as *a priori* information to the channel decoder are nearly error free. A more rigorous analysis of this phenomenon is given in [25, p. 183].

VI. Concluding Remarks

In this paper, we considered the joint iterative decoding of irregular LDPC codes and channels with memory. We introduce a new class of erasure channels with memory, which we call generalized erasure channels (GECs). For these channels, we derive a single parameter recursion for density evolution of the joint iterative decoder. This allows us to state necessary and sufficient conditions for decoder convergence and to algebraically construct sequences of LDPC degree distributions which approach the symmetric information rate of the channel. This proves that the SIR is actually achievable via iterative decoding. This leaves two even bigger questions:

- Is it possible to construct degree distribution sequences which achieve the SIR for any GEC?
- Can this approach be extended to general channels using GEXIT functions?

Acknowledgement

The authors would like to thank Pascal Vontobel and the anonymous reviewers for very insightful comments which greatly improved the quality of this manuscript. They would also like to express their gratitude to the editors for organizing this special issue.

Appendix A
Analysis of the BCJR Decoder for the DEC

A. The Dicode Erasure Channel without Precoding

In this section, we compute the extrinsic erasure rate of a BCJR detector, for a DEC without precoding, as an explicit function of the channel erasure rate ϵ and the *a priori* erasure rate x. This is done by analyzing the forward recursion, the backward recursion, and the output stage separately. We note that expanding the decoder to consider *a priori* information is very similar to expanding the alphabet of our channel. Let W_t and Y_t be the *a priori* symbol and channel output received at time t, respectively. In this case, the decoder sees both the channel output $Y_t \in \{-, 0, +, e\}$ and an *a priori* symbol

$W_t \in \{0, 1, e\}$ (i.e., an observation of the channel input X_t through a BEC with erasure probability x).

Since the channel has only two states, the forward state probability vector $\underline{\alpha}^{(t)} = \left[\alpha_0^{(t)} \ \alpha_1^{(t)}\right]$ has only one degree of freedom and it suffices to consider the quantity $\alpha^{(t)} \triangleq \alpha_0^{(t)} = 1 - \alpha_1^{(t)} = \Pr(S_t = 0 | W_1^{t-1}, Y_1^{t-1})$. The real simplification, however, comes from the fact that the distribution of $\alpha^{(t)}$ has finite support when X is chosen uniformly from $\{0, 1\}$. Using this, we define the matrix $\left[\underline{\underline{A}}_t\right]_{ij} = \Pr(S_{t+1} = j, W_t, Y_t | S_t = i)$ and write the forward recursion as

$$\underline{\alpha}^{(t+1)} = \frac{\underline{\alpha}^{(t)} \underline{\underline{A}}_t}{\left\| \underline{\alpha}^{(t)} \underline{\underline{A}}_t \right\|_1}.$$

It is easy to verify that this recursion is identical to the simpler recursion,

$$\alpha^{(t+1)} = \begin{cases} \frac{1}{2} & \text{if } Y_t = e \text{ and } W_t = e \\ \alpha^{(t)} & \text{if } Y_t = 0 \text{ and } W_t = e \\ 0 & \text{if } Y_t = + \text{ or } W_t = 1 \\ 1 & \text{if } Y_t = - \text{ or } W_t = 0 \end{cases}.$$

Using the simple recursion, we see that, for all $t \geq \min\{i \geq 1 | Y_i \neq 0 \text{ or } W_i \neq e\}$, $\alpha^{(t)}$ will be confined to the finite set $\left\{0, \frac{1}{2}, 1\right\}$.

The symmetry of the input process actually allows us to consider even a smaller support set. The real difference between the three α values in the support set is whether the state is known perfectly or not. Accordingly, we define the known state set $\mathcal{K} \triangleq \{0, 1\}$ and the unknown state set $\mathcal{U} \triangleq \left\{\frac{1}{2}\right\}$. When $\alpha^{(t)} \in \mathcal{K}$, the state is known with absolute confidence, while $\alpha^{(t)} = \mathcal{U}$ corresponds to no prior knowledge. The symmetry of the input process implies that

$$\Pr(\alpha^{(t)} = 0 | \alpha^{(t)} \in \mathcal{K}) = \Pr(\alpha^{(t)} = 1 | \alpha^{(t)} \in \mathcal{K}) = \frac{1}{2}.$$

Therefore, the steady state probability of $\alpha^{(t)}$ can be computed using a reduced Markov chain with states $\mathcal{K}$ and $\mathcal{U}$. The reduced Markov chain transitions from state $\mathcal{K}$ to state $\mathcal{U}$ only if $W = e$ and $Y = e$, and this implies

$$\begin{aligned} \Pr(\alpha^{(t+1)} \in \mathcal{U} | \alpha^{(t)} \in \mathcal{K}) &= 1 - \Pr(\alpha^{(t+1)} \in \mathcal{K} | \alpha^{(t)} \in \mathcal{K}) \\ &= \epsilon x. \end{aligned}$$

The reduced Markov chain transitions from state $\mathcal{U}$ to state $\mathcal{U}$ only if $W = e$ and $Y \in \{e, 0\}$, and this implies

$$\begin{aligned} \Pr(\alpha^{(t+1)} \in \mathcal{U} | \alpha^{(t)} \in \mathcal{U}) &= 1 - \Pr(\alpha^{(t+1)} \in \mathcal{K} | \alpha^{(t)} \in \mathcal{U}) \\ &= x \left(\epsilon + \frac{1-\epsilon}{2}\right). \end{aligned}$$

Therefore, the steady state probability $p_\alpha(\cdot)$ is the unique normalized non-negative solution of

$$\begin{bmatrix} p_\alpha(\mathcal{K}) \\ p_\alpha(\mathcal{U}) \end{bmatrix}^T \begin{bmatrix} 1 - \epsilon x & \epsilon x \\ 1 - \frac{x(1+\epsilon)}{2} & \frac{x(1+\epsilon)}{2} \end{bmatrix} = \begin{bmatrix} p_\alpha(\mathcal{K}) \\ p_\alpha(\mathcal{U}) \end{bmatrix}^T,$$

given by $p_\alpha(\mathcal{U}) = 1 - p_\alpha(\mathcal{K}) = \frac{2\epsilon x}{2 - x(1+\epsilon) + 2\epsilon x}$.

The backward recursion is analyzed in an almost identical manner. Let the backward state probability vector be $\underline{\beta}^{(t)} = \left[\beta_0^{(t)} \ \beta_1^{(t)}\right]$. In this case, it suffices to consider the quantity $\beta^{(t)} \triangleq \beta_0^{(t)} = 1 - \beta_1^{(t)} = \Pr(S_t = 0 | W_t^n, Y_t^n)$. If we let $\left[\underline{\underline{B}}_t\right]_{ij} = \Pr(S_t = j, W_t, Y_t | S_{t+1} = i)$, then we can write the backward recursion as

$$\underline{\beta}^{(t-1)} = \frac{\underline{\beta}^{(t)} \underline{\underline{B}}_t}{\left\| \underline{\beta}^{(t)} \underline{\underline{B}}_t \right\|_1}.$$

Again, the recursion can be simplified and we have

$$\beta^{(t-1)} = \begin{cases} \frac{1}{2} & \text{if } Y_t = e \\ \beta^{(t)} & \text{if } Y_t = 0 \text{ and } W_t = e \\ 1 & \text{if } Y_t = + \text{ or } (Y_t = 0 \text{ and } W_t = 0) \\ 0 & \text{if } Y_t = - \text{ or } (Y_t = 0 \text{ and } W_t = 1) \end{cases}.$$

Using the simpler recursion, we see that, for all $t \geq \min\{i \geq 1 | Y_i \neq 0\}$, $\beta^{(t)}$ will be confined to the finite set $\left\{0, \frac{1}{2}, 1\right\}$.

Again, we use symmetry to reduce the Markov chain. The reduced Markov chain transitions from state $\mathcal{K}$ to the state $\mathcal{U}$ if $Y = e$, and this implies that

$$\begin{aligned} \Pr(\beta^{(t+1)} \in \mathcal{U} | \beta^{(t)} \in \mathcal{K}) &= 1 - \Pr(\beta^{(t+1)} \in \mathcal{K} | \beta^{(t)} \in \mathcal{K}) \\ &= \epsilon. \end{aligned}$$

The reduced Markov chain transitions from state $\mathcal{U}$ to state $\mathcal{U}$ if: (i) $Y = e$ or (ii) $W = e$ and $Y = 0$. This means that

$$\begin{aligned} \Pr(\beta^{(t+1)} \in \mathcal{U} | \beta^{(t)} \in \mathcal{U}) &= 1 - \Pr(\beta^{(t+1)} \in \mathcal{K} | \beta^{(t)} \in \mathcal{U}) \\ &= \epsilon + x\frac{1-\epsilon}{2}. \end{aligned}$$

Therefore, the steady state probability $p_\beta(\cdot)$ is the unique normalized non-negative solution of

$$\begin{bmatrix} p_\beta(\mathcal{K}) \\ p_\beta(\mathcal{U}) \end{bmatrix}^T \begin{bmatrix} 1 - \epsilon & \epsilon \\ 1 - \frac{2\epsilon + x(1-\epsilon)}{2} & \frac{2\epsilon + x(1-\epsilon)}{2} \end{bmatrix} = \begin{bmatrix} p_\beta(\mathcal{K}) \\ p_\beta(\mathcal{U}) \end{bmatrix}^T,$$

given by $p_\beta(\mathcal{U}) = 1 - p_\beta(\mathcal{K}) = \frac{2\epsilon}{(1-\epsilon)(2-x)+2\epsilon}$.

Now, we consider the output stage of the BCJR algorithm for the DEC without precoding. At any point in the trellis, there are now four distinct possibilities for forward/backward state knowledge: $(\mathcal{K}/\mathcal{K})$, $(\mathcal{K}/\mathcal{U})$, $(\mathcal{U}/\mathcal{K})$, and $(\mathcal{U}/\mathcal{U})$. At the extrinsic output of the decoder, the respective erasure probability conditioned on each possibility is: 0, ϵ, 0, and $(1+\epsilon)/2$. Therefore, the erasure probability of the extrinsic output of the BCJR is

$$\begin{aligned} P_e &= p_\beta(\mathcal{U}) \left(\epsilon\, p_\alpha(\mathcal{K}) + \frac{1+\epsilon}{2}\, p_\alpha(\mathcal{U})\right) \\ &= \frac{4\epsilon^2}{(2 - x(1-\epsilon))^2}. \end{aligned}$$

Decoding without *a priori* information is equivalent to choosing $x = 1$, and the corresponding expression simplifies to $4\epsilon^2/(1+\epsilon)^2$.

B. The Dicode Erasure Channel with Precoding

In this section, we compute the extrinsic erasure rate of the BCJR decoder, for the DEC using a $1/(1 \oplus D)$ precoder, as an explicit function of the channel erasure rate, ϵ, and the *a priori* erasure rate, x. This is done by analyzing the forward recursion, the backward recursion, and the output stage separately.

Our approach is the same as that of Section A-A, and this allows us to skip some details and give immediately the simplified forward recursion

$$\alpha^{(t+1)} = \begin{cases} \frac{1}{2} & \text{if } Y_t = e \text{ and } W_t = e \\ \alpha^{(t)} & \text{if } Y_t = 0 \text{ or } W_t = 0 \\ 1-\alpha^{(t)} & \text{if } Y_t = e \text{ and } W_t = 1 \\ 0 & \text{if } Y_t = + \\ 1 & \text{if } Y_t = - \end{cases}.$$

Using this, we see that, for all $t \geq \min\{i \geq 1 | Y_i \neq 0\}$, $\alpha^{(t)}$ will be confined to the finite set $\{0, \frac{1}{2}, 1\}$.

Again, the steady state probability of $\alpha^{(t)}$ can be computed using a reduced Markov chain with states $\mathcal{K}$ and $\mathcal{U}$. The reduced Markov chain transitions from state $\mathcal{K}$ to state $\mathcal{U}$ only if $W = e$ and $Y = e$, and this implies

$$\begin{aligned}\Pr(\alpha^{(t+1)} \in \mathcal{U} | \alpha^{(t)} \in \mathcal{K}) &= 1 - \Pr(\alpha^{(t+1)} \in \mathcal{K} | \alpha^{(t)} \in \mathcal{K}) \\ &= \epsilon x.\end{aligned}$$

The reduced Markov chain transitions from state $\mathcal{U}$ to state $\mathcal{K}$ only if $Y \in \{+, -\}$, and this implies

$$\begin{aligned}\Pr(\alpha^{(t+1)} \in \mathcal{K} | \alpha^{(t)} \in \mathcal{U}) &= 1 - \Pr(\alpha^{(t+1)} \in \mathcal{U} | \alpha^{(t)} \in \mathcal{U}) \\ &= \frac{1-\epsilon}{2}.\end{aligned}$$

Therefore, the steady state probability $p_\alpha(\cdot)$ is the unique normalized non-negative solution of

$$\begin{bmatrix} p_\alpha(\mathcal{K}) \\ p_\alpha(\mathcal{U}) \end{bmatrix}^T \begin{bmatrix} 1-\epsilon x & \epsilon x \\ \frac{1-\epsilon}{2} & \frac{1+\epsilon}{2} \end{bmatrix} = \begin{bmatrix} p_\alpha(\mathcal{K}) \\ p_\alpha(\mathcal{U}) \end{bmatrix}^T,$$

given by $p_\alpha(\mathcal{K}) = 1 - p_\alpha(\mathcal{U}) = \frac{1-\epsilon}{1-\epsilon+2\epsilon x}$.

The precoded case is also simplified by the fact that the state diagram of the precoded channel is such that time reversal is equivalent to negating the sign of the output. Therefore, the statistics of the forward and backward recursions are identical and $p_\beta(\mathcal{K}) = 1 - p_\beta(\mathcal{U}) = p_\alpha(\mathcal{K})$.

Now, we consider the output stage of the BCJR algorithm for the precoded DEC. At any point in the trellis, there are now four distinct possibilities for forward/backward state knowledge: $(\mathcal{K}/\mathcal{K})$, $(\mathcal{K}/\mathcal{U})$, $(\mathcal{U}/\mathcal{K})$, and $(\mathcal{U}/\mathcal{U})$. At the extrinsic output of the decoder, the respective erasure probability conditioned on each possibility is: 0, ϵ, ϵ, and ϵ. Therefore, the erasure probability of the extrinsic output of the BCJR is

$$\begin{aligned}P_e &= \epsilon\left(1 - p_\alpha(\mathcal{K}) p_\beta(\mathcal{K})\right) \\ &= \epsilon\left(1 - \frac{(1-\epsilon)^2}{(1-\epsilon+2\epsilon x)^2}\right) \\ &= \frac{4\epsilon^2 x(1-\epsilon(1-x))}{(1-\epsilon(1-2x))^2}.\end{aligned}$$

Again, decoding without *a priori* information is equivalent to choosing $x = 1$, and the corresponding expression simplifies to $4\epsilon^2/(1+\epsilon)^2$.

APPENDIX B
PROPERTIES OF $\widetilde{Q}$

A. General

In this section, we consider a property of the function $\widetilde{Q} \in \mathcal{P}$, defined in (27), for the check-Poisson and check-regular degree distribution. This property is crucial to proving that our algebraic construction of L is indeed a degree distribution (i.e., $L \in \mathcal{P}$). To simplify some expressions, we will use $[x^k]$ as an operator which returns the coefficient of x^k in the power series expansion of a function (e.g., $[x^k] L(x) = L_k$). In particular, we are interested in computing the largest γ such that $[x^k]\left(\widetilde{Q}(x) - \gamma \widetilde{Q}(x)^2\right) \geq 0$ for $k \geq 0$. Solving for γ implies the bound

$$g_k \triangleq \frac{[x^i]\left(\widetilde{Q}(x)^2\right)}{[x^i]\,\widetilde{Q}(x)} \leq \frac{1}{\gamma}$$

and, using the fact that $[x^i]\left(\widetilde{Q}(x)^2\right) = 0$ for $i < 4$, allows us to write

$$\gamma = \inf_{i \geq 4}\left(g_i^{-1}\right) = \left(\sup_{i\geq 4} g_i\right)^{-1}.$$

B. The Poisson Check Distribution

In this section, we prove that $\gamma = \frac{3}{10}$ for $\widetilde{Q}(x) = x + (1-x)\log(1-x)$. We start by computing the power series of $\widetilde{Q}$ and $\widetilde{Q}^2$. First, we compute the power series of $\widetilde{Q}$ with

$$\begin{aligned}\widetilde{Q}(x) &= \int_0^x \widetilde{Q}'(t)dt \\ &= -\int_0^x \log(1-t)dt \\ &= \sum_{i=1}^{\infty} \int_0^x \frac{1}{i} t^i dt \\ &= \sum_{i=1}^{\infty} \frac{1}{i(i+1)} x^{i+1} dt \\ &= \sum_{i=2}^{\infty} \frac{1}{i(i-1)} x^i. \end{aligned} \tag{37}$$

Next, we consider the quantity $\widetilde{Q}^2$ and write

$$\widetilde{Q}(x)^2 = x^2 + 2x(1-x)\log(1-x) + (1-x)^2\log^2(1-x).$$

The power series of $\log^2(1-x)$ can be computed with

$$\begin{aligned}\log^2(1-x) &= \int_0^x \left(\frac{d}{dt}\log^2(1-t)\right)dt \\ &= \int_0^x \frac{-2\log(1-t)}{1-t}dt \\ &= 2\int_0^x \left(\sum_{i=0}^{\infty} t^i\right)\left(\sum_{j=1}^{\infty}\frac{1}{j}t^j\right)dt \\ &= 2\int_0^x \sum_{i=0}^{\infty}\sum_{j=1}^{\infty}\frac{1}{j}t^{i+j}dt \\ &= 2\sum_{i=1}^{\infty} H_i \int_0^x t^i dt \\ &= \sum_{i=2}^{\infty}\frac{2H_{i-1}}{i}x^i, \end{aligned} \tag{38}$$

where $H_j = \sum_{i=1}^{j}(1/i)$ is the harmonic sum. Now, we can substitute the power series (37) and (38) into the formula for $\widetilde{Q}^2$. This gives (34), and we note that terms of degree less than 4 are dropped as needed.

Using (37) and (34), we can write

$$g_i = \frac{[x^i]\left(\widetilde{Q}(x)^2\right)}{[x^i]\,\widetilde{Q}(x)} = \left(\frac{4H_{i-3}+2i-6}{i(i-1)(i-2)}\right)\left(\frac{1}{i(i-1)}\right)^{-1} = \frac{4H_{i-3}+2i-6}{i-2}.$$

The forward difference of this sequence is given by

$$g_{i+1}-g_i = \frac{4H_{i-2}+2i-4}{i-1} - \frac{4H_{i-3}+2i-6}{i-2} = \frac{6-4H_{i-3}}{(i-1)(i-2)}.$$

We see that $g_{i+1}-g_i \le 0$ for $i \ge 5$ because $(i-1)(i-2) > 0$ and $6-4H_{i-3} \le 0$ in this range. Since g_i is non-increasing for $i \ge 5$, this means that we can reduce the range of the supremum and write

$$\gamma = \left(\sup_{4\le i\le 5} g_i\right)^{-1}.$$

Therefore, evaluating $g_4 = 3$ and $g_5 = 10/3$ shows that $\gamma = 3/10$.

C. The Regular Check Distribution

In this section, we prove that γ_d is given by (30) for $\widetilde{Q}(x) = x-(d-1)(1-x)\left(1-(1-x)^{1/(d-1)}\right)$. We start by defining $\alpha = \frac{1}{d-1}$ and computing the power series of $\widetilde{Q}$ and $\widetilde{Q}^2$. First, we compute the power series of $\widetilde{Q}$ using the binomial theorem to get

$$[x^k]\,\widetilde{Q}(x) = \begin{cases} \frac{1}{\alpha}\binom{\alpha+1}{k}(-1)^k & \text{if } k\ge 2 \\ 0 & \text{otherwise} \end{cases}.$$

Likewise, we can compute the power of $\widetilde{Q}^2$ by first squaring and then applying the binomial theorem. First, we square $\widetilde{Q}$ to get (35).

Next, we apply the binomial theorem (dropping all terms of degree less than 4) to get (36) for $k \ge 4$. This allows us to write the ratio sequence as

$$g_k(\alpha) = \frac{[x^k]\left(\widetilde{Q}(x)^2\right)}{[x^k]\widetilde{Q}(x)} = \frac{2}{\alpha}\left(\frac{k\alpha+\alpha+2}{k-\alpha-2}\right) + \frac{1}{\alpha}\frac{\binom{2\alpha+2}{k}}{\binom{\alpha+1}{k}} = \frac{2(k+1)}{k-\alpha-2} + \left(\frac{4}{\alpha(k-\alpha-2)} + \frac{1}{\alpha}\frac{\binom{2\alpha+2}{k}}{\binom{\alpha+1}{k}}\right) \triangleq \frac{2(k+1)}{k-\alpha-2} + h_k(\alpha).$$

The main difficulty with this expression lies in the ratio of binomial coefficients in $h_k(\alpha)$. To deal with this, we write

$$\frac{\binom{2\alpha+2}{k}}{\binom{\alpha+1}{k}} = \prod_{i=0}^{k-1}\frac{2\alpha+2-i}{\alpha+1-i} = \left(\prod_{i=0}^{4}\frac{2\alpha+2-i}{\alpha+1-i}\right)\left(\prod_{i=5}^{k-1}\frac{i-2\alpha-2}{i-\alpha-1}\right) = -\frac{8-32\alpha^2}{\alpha^2-5\alpha+6}\frac{\Gamma(4-\alpha)}{\Gamma(3-2\alpha)}\frac{\Gamma(k-2\alpha-2)}{\Gamma(k-\alpha-1)}.$$

Using the main result of [45], we find that

$$\frac{\Gamma(k-2\alpha-2)}{\Gamma(k-\alpha-1)} \ge \left(k-\frac{3}{2}\alpha-2\right)^{-1-\alpha} \ge (k-\alpha-2)^{-1-\alpha}$$

$$\begin{aligned}
\widetilde{Q}(x)^2 &= x^2 + 2x\left(-x+\sum_{i=2}^{\infty}\frac{1}{i(i-1)}x^i\right) + (1-x)^2\sum_{i=2}^{\infty}\frac{2H_{i-1}}{i}x^i \\
&= \sum_{i=4}^{\infty}\frac{2}{(i-1)(i-2)}x^i + \sum_{i=4}^{\infty}\frac{2H_{i-1}}{i}x^i - 2x\sum_{i=3}^{\infty}\frac{2H_{i-1}}{i}x^i + x^2\sum_{i=2}^{\infty}\frac{2H_{i-1}}{i}x^i \\
&= \sum_{i=4}^{\infty}\frac{2}{(i-1)(i-2)}x^i + \sum_{i=4}^{\infty}\frac{2H_{i-1}}{i}x^i - \sum_{i=4}^{\infty}\frac{4H_{i-2}}{i-1}x^i + \sum_{i=4}^{\infty}\frac{2H_{i-3}}{i-2}x^i \\
&= \sum_{i=4}^{\infty}\left(\frac{2}{(i-1)(i-2)} + \frac{2H_{i-1}}{i} - \frac{4H_{i-2}}{i-1} + \frac{2H_{i-3}}{i-2}\right)x^i \\
&= \sum_{i=4}^{\infty}\frac{4H_{i-3}+2i-6}{i(i-1)(i-2)}x^i.
\end{aligned} \tag{34}$$

$$\begin{aligned}
\widetilde{Q}(x)^2 &= \left(\frac{\alpha+1}{\alpha}\right)^2 x^2 - \frac{2(\alpha+1)}{\alpha^2}x\left(1-(1-x)^{\alpha+1}\right) + \frac{1}{\alpha^2}\left(1-(1-x)^{\alpha+1}\right)^2 \\
&= \left(\frac{\alpha+1}{\alpha}\right)^2 x^2 - \frac{2(\alpha+1)}{\alpha^2}x\left(1-(1-x)^{\alpha+1}\right) + \frac{1}{\alpha^2} - \frac{2}{\alpha^2}(1-x)^{\alpha+1} + \frac{1}{\alpha^2}(1-x)^{2\alpha+2} \\
&= \left(\frac{\alpha+1}{\alpha}\right)^2 x^2 - \left(\frac{2(\alpha+1)x-2}{\alpha^2}\right)\left(1-(1-x)^{\alpha+1}\right) + \frac{1}{\alpha^2}\left(1+(1-x)^{2\alpha+2}\right).
\end{aligned} \tag{35}$$

for $\alpha \in \left[0, \frac{1}{2}\right]$ and $k \geq 4$. This allows us to define $\overline{g}_k(\alpha) \triangleq \frac{2(k+1)}{k-\alpha-2} + \overline{h}_k(\alpha)$ and

$$\overline{h}_k(\alpha) \triangleq \frac{4}{\alpha(k-\alpha-2)}\left[1 - \frac{\left(2-8\alpha^2\right)\Gamma(4-\alpha)\left(k-\alpha-2\right)^{-\alpha}}{(\alpha^2-5\alpha+6)\,\Gamma(3-2\alpha)}\right]$$

so that $h_k(\alpha) \leq \overline{h}_k(\alpha)$ and $g_k(\alpha) \leq \overline{g}_k(\alpha)$. Taking the derivative of $\overline{h}_{z+\alpha+2}(\alpha)$ with respect to z shows that, for $\alpha \in \left[0, \frac{1}{2}\right]$, $\overline{h}_k(\alpha)$ is strictly decreasing in k if

$$k \geq 2 + \alpha + \left(\frac{(1+\alpha)\left(2-8\alpha^2\right)}{\alpha^2 - 5\alpha + 6}\frac{\Gamma(4-\alpha)}{\Gamma(3-2\alpha)}\right)^{1/\alpha}.$$

Taking the maximum of this expression over $\alpha \in \left[0, \frac{1}{2}\right]$ shows that $\overline{h}_k(\alpha)$ is strictly decreasing in k for $k \geq 14$. Taking the maximum of $\overline{h}_{14}(\alpha)$, for $\alpha \in \left[0, \frac{1}{2}\right]$, shows that $\overline{h}_k(\alpha) \leq \overline{h}_{14}\left(\frac{1}{2}\right) = \frac{16}{23}$ for $k \geq 14$. This implies that

$$\begin{aligned}\sup_{k\geq 14, \alpha\in\left[0,\frac{1}{2}\right]} g_k(\alpha) &\leq \sup_{k\geq 14, \alpha\in\left[0,\frac{1}{2}\right]} \overline{g}_k(\alpha) \\ &\leq \frac{16}{23} + \sup_{k\geq 14, \alpha\in\left[0,\frac{1}{2}\right]} \frac{2(k+1)}{k-\alpha-2} \\ &= \frac{76}{23}.\end{aligned}$$

Analyzing $\max_{4\leq k\leq 14} g_k(\alpha)$, for $\alpha \in \left[0, \frac{1}{2}\right]$, by computer shows that the maximum is always greater than $\frac{76}{23}$ and determined by either $g_4(\alpha)$ or $g_5(\alpha)$. A bit of algebra (via Mathematica) allows us to write

$$\sup_{k\geq 4} g_k(\alpha) = \begin{cases} \frac{20(\alpha+1)}{\alpha^2-5\alpha+6} & \text{if } 0 \leq \alpha \leq \frac{1}{7} \\ \frac{6(\alpha+1)}{\alpha^2-3\alpha+2} & \text{if } \frac{1}{7} \leq \alpha \leq \frac{1}{2} \end{cases}.$$

The conclusion of Lemma 6 is obtained by simplifying the expression for

$$\gamma_d = \left[\sup_{k\geq 4} g_k\left(\frac{1}{d-1}\right)\right]^{-1}.$$

APPENDIX C
THE DERIVATION OF CHANNEL G-FUNCTIONS

A. The DEC

Starting with (3), the definition of F gives

$$\begin{aligned}F_{DEC(\epsilon)}(x) &= \int_0^x \frac{4\epsilon^2}{(2-t(1-\epsilon))^2}dt \\ &= \frac{2x\epsilon^2}{2-(1-\epsilon)x}.\end{aligned}$$

Solving for the inverse shows that

$$F_{DEC(\epsilon)}^{-1}(x) = \frac{2x}{2\epsilon^2 + (1-\epsilon)x}.$$

Finally, computing the G-function results in

$$\begin{aligned}G_{DEC(\epsilon)}(x) &= F_{DEC(\epsilon)}^{-1}\left(F_{DEC(\epsilon)}(1)x\right) \\ &= \frac{\frac{4\epsilon^2}{1+\epsilon}x}{2\epsilon^2 + (1-\epsilon)\frac{2\epsilon^2}{1+\epsilon}x} \\ &= \frac{2x}{(1+\epsilon)+(1-\epsilon)x} \\ &= -\frac{2}{1-\epsilon}\sum_{n=1}^{\infty}\left(-\frac{1-\epsilon}{1+\epsilon}\right)^n x^n.\end{aligned}$$

B. The Precoded Dicode Erasure Channel

Starting with (4), the definition of F gives

$$\begin{aligned}F_{pDEC(\epsilon)}(x) &= \int_0^x \frac{4\epsilon^2 t\left(1-\epsilon(1-t)\right)}{\left(1-\epsilon(1-2t)\right)^2}dt \\ &= \frac{2\epsilon^2 x^2}{1-\epsilon(1-2x)}.\end{aligned}$$

Solving for the inverse shows that

$$\begin{aligned}F_{pDEC(\epsilon)}^{-1}(x) &= \frac{x + \sqrt{2(1-\epsilon)x + x^2}}{2\epsilon} \\ &= \frac{x + \sqrt{2(1-\epsilon)x}\sqrt{1+\frac{x}{2(1-\epsilon)}}}{2\epsilon}.\end{aligned}$$

Computing the G-function gives

$$\begin{aligned}G_{pDEC(\epsilon)}(x) &= F_{pDEC(\epsilon)}^{-1}\left(F_{pDEC(\epsilon)}(1)x\right) \\ &= \frac{\epsilon x}{1+\epsilon} + \sqrt{\frac{(1-\epsilon)x}{1+\epsilon}}\sqrt{1+\frac{\epsilon^2 x}{(1-\epsilon^2)}},\end{aligned}$$

which can be rewritten as

$$G_{pDEC(\epsilon)}(x) = \sqrt{\frac{1-\epsilon}{1+\epsilon}x}\sum_{n=0}^{\infty}\binom{\frac{1}{2}}{n}\left(\frac{\epsilon^2 x}{(1-\epsilon^2)}\right)^n + \frac{\epsilon x}{1+\epsilon}.$$

C. The Linear Channel $f(x) = ax + b$

The definition of F shows that

$$F_{lin}(x) = \frac{ax^2}{2} + bx$$

$$\begin{aligned}\left[x^k\right]\left(\widetilde{Q}(x)^2\right) &= \frac{2(\alpha+1)}{\alpha^2}\binom{\alpha+1}{k-1}(-1)^{k-1} - \frac{2}{\alpha^2}\binom{\alpha+1}{k}(-1)^k + \frac{1}{\alpha^2}\binom{2\alpha+2}{k}(-1)^k \\ &= \frac{2(\alpha+1)}{\alpha^2}\binom{\alpha+1}{k}\frac{k}{k-\alpha-2}(-1)^k - \frac{2}{\alpha^2}\binom{\alpha+1}{k}(-1)^k + \frac{1}{\alpha^2}\binom{2\alpha+2}{k}(-1)^k \\ &= \frac{2}{\alpha^2}\left(\frac{(\alpha+1)k}{(k-\alpha-2)} - 1\right)\binom{\alpha+1}{k}(-1)^k + \frac{1}{\alpha^2}\binom{2\alpha+2}{k}(-1)^k. \qquad (36)\end{aligned}$$

Using the quadratic formula, we see that the inverse is

$$\begin{aligned} F_{lin}^{-1}(x) &= \frac{-b+\sqrt{b^2+2ax}}{a} \\ &= \frac{b}{a}\left(\sqrt{1+\frac{2ax}{b^2}}-1\right) \\ &= \frac{b}{a}\sum_{n=1}^{\infty}\binom{\frac{1}{2}}{n}\left(\frac{2ax}{b^2}\right)^n. \end{aligned}$$

Computing the G-function gives

$$\begin{aligned} G_{lin}(x) &= F_{lin}^{-1}\left(F(1)x\right) \\ &= \frac{b}{a}\sum_{n=1}^{\infty}\binom{\frac{1}{2}}{n}\left(\frac{2a}{b^2}\left(\frac{a}{2}+b\right)x\right)^n. \end{aligned}$$

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] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman, "Efficient erasure correcting codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 569–584, Feb. 2001.

[3] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 619–637, Feb. 2001.

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

[5] J. Hou, P. H. Siegel, and L. B. Milstein, "Performance analysis and code optimization of low density parity-check codes on Rayleigh fading channels," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 924–934, May 2001.

[6] 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. – Oct. 1995.

[7] K. Narayanan, "Effect of precoding on the convergence of turbo equalization for partial response channels," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 686–698, April 2001.

[8] 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.

[9] M. Tüchler, R. Koetter, and A. Singer, "Turbo equalization: principles and new results," *IEEE Trans. Commun.*, vol. 50, pp. 754–767, May 2002.

[10] M. Mushkin and I. Bar-David, "Capacity and coding for Gilbert-Elliot channels," *IEEE Trans. Inform. Theory*, vol. 35, pp. 1277–1290, Nov. 1989.

[11] A. J. Goldsmith and P. P. Varaiya, "Capacity, mutual information, and coding for finite-state Markov channels," *IEEE Trans. Inform. Theory*, vol. 42, pp. 868–886, May 1996.

[12] A. W. Eckford, F. R. Kschischang, and S. Pasupathy, "Analysis of low-density parity-check codes for the Gilbert-Elliott channel," *IEEE Trans. Inform. Theory*, vol. 51, no. 11, pp. 3872–3889, 2005.

[13] E. N. Gilbert, "Capacity of a burst-noise channel," *The Bell Syst. Techn. J.*, vol. 39, pp. 1253–1265, Sept. 1960.

[14] D. Arnold and H. Loeliger, "On the information rate of binary-input channels with memory," in *Proc. IEEE Int. Conf. Commun.*, (Helsinki, Finland), pp. 2692–2695, June 2001.

[15] D. Arnold, H. A. Loeliger, P. O. Vontobel, A. Kavčić, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inform. Theory*, vol. 52, pp. 3498–3508, Aug. 2006.

[16] H. D. Pfister, J. B. Soriaga, and P. H. Siegel, "On the achievable information rates of finite state ISI channels," in *Proc. IEEE Global Telecom. Conf.*, (San Antonio, Texas, USA), pp. 2992–2996, Nov. 2001.

[17] J. Li, K. R. Narayanan, E. Kurtas, and C. N. Georghiades, "On the performance of high-rate TPC/SPC codes and LDPC codes over partial response channels," *IEEE Trans. Commun.*, vol. 50, pp. 723–734, May 2002.

[18] N. Varnica and A. Kavčić, "Optimized low-density parity-check codes for partial response channels," *IEEE Commun. Letters*, vol. 7, no. 4, pp. 168–170, 2003.

[19] A. Kavčić, X. Ma, and M. Mitzenmacher, "Binary intersymbol interference channels: Gallager codes, density evolution and code performance bounds," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1636–1652, July 2003.

[20] J. B. Soriaga and P. H. Siegel, "On near-capacity coding systems for partial-response channels," in *Proc. IEEE Int. Symp. Information Theory*, (Chicago, IL), p. 267, June 2004.

[21] A. Kavčić, X. Ma, and N. Varnica, "Matched information rate codes for partial response channels," *IEEE Trans. Inform. Theory*, vol. 51, pp. 973–989, Sept. 2005.

[22] J. B. Soriaga, H. D. Pfister, and P. H. Siegel, "Determining and approaching achievable rates of binary intersymbol interference channels using multistage decoding," *IEEE Trans. Inform. Theory*, vol. 53, April 2007.

[23] 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, March 1974.

[24] W. Hirt, *Capacity and Information Rates of Discrete-Time Channels with Memory*. PhD thesis, E.T.H., Zurich, Switzerland, 1988.

[25] 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.

[26] H. D. Pfister and P. H. Siegel, "Joint iterative decoding of LDPC codes and channles with memory," in *Proc. 3rd Int. Symp. on Turbo Codes & Related Topics*, (Brest, France), pp. 15–18, Sept. 2003.

[27] H. D. Pfister, I. Sason, and R. Urbanke, "Capacity-achieving ensembles for the binary erasure channel with bounded complexity," *IEEE Trans. Inform. Theory*, vol. 51, pp. 2352–2379, July 2005.

[28] H. D. Pfister and I. Sason, "Accumulate–repeat–accumulate codes: Capacity-achieving ensembles of systematic codes for the erasure channel with bounded complexity," *IEEE Trans. Inform. Theory*, vol. 53, pp. 2088–2115, June 2007.

[29] 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.

[30] S. M. Aji and R. J. McEliece, "The generalized distributive law," *IEEE Trans. Inform. Theory*, vol. 46, no. 2, pp. 325–343, 2000.

[31] T. J. Richardson and R. L. Urbanke, *Modern Coding Theory*. Cambridge, 2007.

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

[33] 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.

[34] A. Ashikhmin, G. Kramer, and S. ten Brink, "Extrinsic information transfer functions: model and erasure channel properties," *IEEE Trans. Inform. Theory*, vol. 50, pp. 2657–2674, Nov. 2004.

[35] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.

[36] K. A. S. Immink, P. H. Siegel, and J. K. Wolf, "Codes for digital recorders," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2260–2299, Oct. 1998.

[37] C. Méasson and R. Urbanke, "Asymptotic analysis of turbo codes over the binary erasure channel," in *Proc. 12th Joint Conf. on Comm. and Coding*, (Saas Fee, Switzerland), March 2002.

[38] R. G. Gallager, *Low-Density Parity-Check Codes*. Cambridge, MA, USA: The M.I.T. Press, 1963.

[39] J. Hou, P. H. Siegel, L. B. Milstein, and H. D. Pfister, "Capacity-approaching bandwidth-efficient coded modulation schemes based on low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2141–2155, Sept. 2003.

[40] 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.

[41] P. Oswald and M. A. Shokrollahi, "Capacity-achieving sequences for the erasure channel," *IEEE Trans. Inform. Theory*, vol. 48, pp. 3017–3028, Dec. 2002.

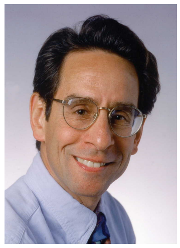
[42] M. Tenenbaum and H. Pollard, *Ordinary differential equations: an elementary textbook for students of mathematics, engineering, and the sciences*. Dover Publications, 1985.

[43] P. Flajolet and A. Odlyzko, "Singularity analysis of generating functions," *SIAM J. Disc. Math.*, vol. 3, no. 2, pp. 216–240, 1990.

[44] S. ten Brink, "Rate one-half code for approaching the Shannon limit by 0.1 dB," *Electronic Letters*, vol. 36, pp. 1293–1294, July 2000.

[45] C. L. Frenzen, "Error bounds for asymptotic expansions of the ratio of two gamma functions," *SIAM J. Math. Anal.*, vol. 18, pp. 890–896, May 1987.

Henry D. Pfister (S'99-M'03) received the B.S. degree in physics in 1995, the M.S. degree in electrical engineering in 2000, and the Ph.D. degree in electrical engineering in 2003, all from the University of California, San Diego. During 2003-2004, he worked as a Senior Engineer for Qualcomm, Inc. in San Diego. In 2005, he was a Postdoctoral Fellow in the School of Computer and Communication Sciences of the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland. He joined the faculty of the School of Engineering at Texas A\&M University, College Station in 2006, and he is currently an assistant professor in the Department of Electrical and Computer Engineering. His current research interests include information theory, channel coding, and iterative decoding with applications in wireless communications and data storage.



Paul H. Siegel (M'82-SM'90-F'97) received the S.B. and Ph.D. degrees in mathematics from the Massachusetts Institute of Technology (MIT), Cambridge, in 1975 and 1979, respectively.

He held a Chaim Weizmann Postdoctoral Fellowship at the Courant Institute, New York University. He was with the IBM Research Division in San Jose, CA, from 1980 to 1995. He joined the faculty of the School of Engineering at the University of California, San Diego in July 1995, where he is currently Professor of Electrical and Computer Engineering. He is affiliated with the California Institute of Telecommunications and Information Technology, the Center for Wireless Communications, and the Center for Magnetic Recording Research where he holds an endowed chair and currently serves as Director. His primary research interests lie in the areas of information theory and communications, particularly coding and modulation techniques, with applications to digital data storage and transmission.

Prof. Siegel was a member of the Board of Governors of the IEEE Information Theory Society from 1991 to 1996. He served as Co-Guest Editor of the May 1991 Special Issue on "Coding for Storage Devices" of the IEEE TRANSACTIONS ON INFORMATION THEORY. He served the same TRANSACTIONS as Associate Editor for Coding Techniques from 1992 to 1995, and as Editor-in-Chief from July 2001 to July 2004. He was also Co-Guest Editor of the May/September 2001 two-part issue on "The Turbo Principle: From Theory to Practice" of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. He was co-recipient, with R. Karabed, of the 1992 IEEE Information Theory Society Paper Award and shared the 1993 IEEE Communications Society Leonard G. Abraham Prize Paper Award with B. Marcus and J.K. Wolf. He holds several patents in the area of coding and detection, and was named a Master Inventor at IBM Research in 1994. He is a member of Phi Beta Kappa.