CERTIFICATE

It is certified that the work contained in the dissertation entitled LDPC Codes: Bounds on the Rate for FSMCs and some results on Minimal Stopping Sets by **Pulkit Grover** has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

(Ajit Kumar Chaturvedi)
Associate Professor,
Department of Electrical Engineering,
Indian Institute of Technology,
Kanpur-208016.

July 2005

Acknowledgements

I wish to acknowledge the support of my advisor, Dr AK Chaturvedi, throughout my stay here, more so in the last three years. His keen insights, shared enthusiasm for results and flexibility with regards to the problems I addressed have kept me motivated and active.

Thanks also to the faculty of Electrical Engineering, IITK, particularly Dr RK Bansal and Dr Govind Sharma, with whose courses began my interests in probability. To the faculty of Mathematics department, IITK, who have helped me enhance my interests in algebra and probability. To Dr Amin Shokrollahi and Dr Rüdiger Urbanke at EPFL for enlightening discussions on LDPC codes and beyond in the summers of 2004.

Thanks also to my friends: Rajesh*ji*, Arya and Sahu*ji* for stimulating discussions, Mayank for pointing out that a sequence in the dissertation converges eventually almost-surely, and not almost-surely, and the rest of information systems group for a lot more than academics.

This dissertation marks the end of six years of my stay at IITK. The place is made special by the intellectually adept and stimulating people here, and I am grateful for the way in which their presence has helped me define myself. The myriad of possibilities here beckon one to essay different things, and find one's own route. In particular, the flights at the gliding club, the tennis matches, the adventure sports activities and spic-macay events have given me moments of peak that can seldom be attained elsewhere.

And finally, gratitude to my family for backing me throughout, despite my apparently experimental decisions, and for patiently waiting for me to complete my dissertation.

Abstract

Since their re-discovery in mid-90's, Low-Density Parity-Check (LDPC) codes have been shown to be capacity approaching for a large class of memoryless channels. Recently their performance has been analysed over certain Finite State Markov Channels (FSMCs), and has been found to be encouraging. In this dissertation, we present upper bounds on the rate of LDPC codes for reliable communication over FSMCs. We consider the class of FSMCs in which the channel behaves as a Binary Symmetric Channel (BSC) in each state. A simple upper bound for all non-inverting FSMCs is first derived. A tighter bound is then presented for the case of Gilbert-Elliott channels. Tighter bounds are also derived for FSMCs which behave as a BSC in each state. However, the latter bounds hold only *almost surely* for *randomly constructed* sequence of LDPC codes. Finally, we extend these bounds to arbitrary symmetric FSMCs. To establish the utility of of random construction of LDPC codes, we prove the concentration theorem for Belief Propagation (BP) decoding over Markov channels. The derivation of these bounds is a generalization of bounds given by Gallager for BSCs. Since these bounds are derived for optimal Maximum-Likelihood decoding, they also hold for BP decoding.

These bounds prove that if a sequence of codes has a constant average number of 1's in corresponding parity check matrices, the sequence can not achieve capacity of any FSMC. Furthermore, using the derivations of bounds on the rate, we also derive lower bounds on density of parity check matrices for given performance over FSMCs.

We also introduce the concept of *minimal stopping sets* in the decoding of LDPC codes over Binary Erasure Channel (BEC). The significance of minimal stopping sets is explained, and bounds on their number are derived for LDPC codes with $\lambda_2 = 0$.

Contents

Chapter 1

Introduction

Shannon, in his 1948 paper "A Mathematical Theory of Communication" [Sha48] wrote:
"The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point."

Consider Fig 1. The source produces a message which is to be reliably reproduced at the receiving end. The channel corrupts the transmitted symbols, inducing errors in the received message. To correct these errors, redundant symbols dependent on the message are introduced at the transmitting end. Using appropriate redundancy, some of the errors induced by the channel can be corrected. This is the essence of *error correction coding*.

The addition of the redundant symbols is done by the *encoder* at the transmitting end. The inverse operation of the encoder, and the correction of errors, is done at the *decoder*, which is at the receiver. The mapping of information symbols to symbols at the output of the encoder is called the *error correcting code*, or simply, the code. If the encoder maps k information symbols to $n(> k)$ output symbols, the n output symbols constitute a *codeword*, and n is said to be the *codelength* or the *blocklength*. Collection of all the codewords, along with the mapping from information symbols to the codewords, defines the code. The *rate* of the code is defined as the ratio of the number of information bits transmitted across the

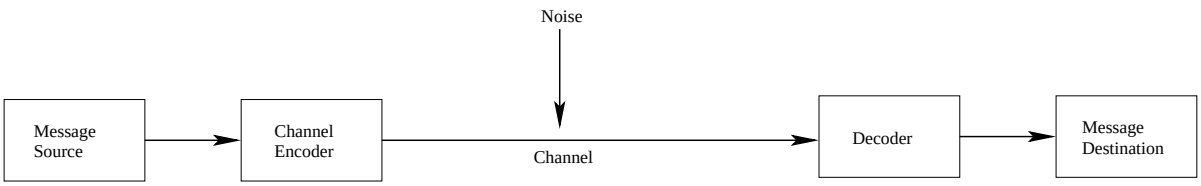

Figure 1.1: A communication system

channel to the number of bits used to transmit that information, which is $\frac{k}{n}$.

In [Sha48], Shannon proved that for transmission rates below a certain constant, known as *channel capacity*, C, one can design codes which operate at as low probability of error as desired. That is for any rate $R < C$, we can design coding techniques such that as the codelength converges to infinity, the probability of error converges to 0.

However, the proof in [Sha48] was non-constructive: he showed that there exists a sequence of codes which *achieves* capacity. That is, there exists a sequence of codes $\{\mathcal{C}_n\}_{n=1}^{\infty}$ of rates R_n and probabilities of error P_e^n such that $R_n \to C$ and $P_e^n \to 0$. But no explicit construction of such a sequence was given. Instead it was shown that a sequence of randomly chosen codes would achieve capacity with a high probability. However, encoding and decoding of such a random codes is a computationally expensive exercise. For the ensuing 45 years after the publication of Shannon's results [Sha48], coding theorists designed highly structured techniques for error correction with small probabilities of errors for small blocklengths. The strong code structure is helpful in realizing fast encoding and decoding algorithms, as well as minimizing the probability of error. However, these code constructions perform poorly at large blocklengths. Hence the goal of capacity achieving codes remained elusive.

In 1993, Berrou et al [BGT93] presented the *Turbo Codes*, a class of codes which came close to channel capacity for Additive White Gaussian Noise (AWGN) channels with fast decoding algorithms. The discovery of Turbo codes resulted in a flurry of results dealing with analysis of these codes, improvement in their performance, and further exploration of similar codes. The denouement of this exploration was the discovery of a whole new class of codes, including the rediscovery of Low-Density Parity-Check (LDPC) codes [Mac99], which exhibit superior performance [RSU01][LMSS01] and have fast encoding [RU01a] and decoding [RU01b] algorithms.

LDPC Codes were invented by Gallager in his PhD dissertation [Gal60]. Due to impediments of limited computational speeds, their competency in approaching capacity could not be appreciated and they were almost in oblivion until their rediscovery. An LDPC code is represented by an $m \times n$ ($m < n$) *parity-check* binary matrix $\mathbf{H}$. The null space of the parity check matrix is the set of valid codewords. That is, $\mathbf{x}$ is a valid codeword iff $\mathbf{Hx^T} = 0$. LDPC codes have sparse parity check matrices, which means that the number of 1's in the

matrix is much smaller than the number of 0's. Gallager also proposed a sub-optimal iterative decoding of LDPC codes, which has low-complexity. This decoding has been generalized to belief-propagation decoding [RU01b] which is commonly used today.

LDPC codes have been shown to be capacity approaching over various memoryless channels [RU01b]. They outperform Turbo codes over various channels [RU01b], including AWGN channels [CJRU01]. Recently, their performance has been analyzed over some channels with memory [Eck04][EKP03], and has been found to be promising. The authors in [Eck04][EKP03] also discuss the design and decoding issues for these channels.

In [Gal60, pp 37–38] Gallager gave information theoretic bounds on the rate that LDPC codes can achieve. The problem is defined as follows:

Consider the ensemble of LDPC codes for which there are d 1's in each row of the corresponding $\mathbf{H}$ matrices. *What is the maximum rate that we can achieve using these codes such that the probability of error converges to 0?* Evidently, this rate is less than the channel capacity. Under the restriction of fixed number of 1's in each row of the parity check matrix, can we find tighter bounds on the rate?

Gallager gave such upper bounds on the rate for Binary Symmetric Channels (BSCs). The bounds also yield results on the number of 1's required in the rows of $\mathbf{H}$ in order to achieve desired performance in asymptotic sense, as well as for finite lengths [SU03]. In this dissertation, we generalize the bounds found by Gallager to a certain class of channels with memory. The channels we consider are Finite State Markov Channels (FSMCs). The bounds are derived under the assumption of optimal Maximum-Likelihood (ML) decoding at the receiver. Since BP decoding is sub-optimal, the bounds continue to hold for performance of BP decoded LDPC codes, and so also for any improved decoding thereof.

We also analyze the decoding of LDPC codes over Binary Erasure Channels (BECs). In this context, we define *minimal stopping sets* and elucidate their significance. We also find bounds on the number of minimal stopping sets for a given ensemble of LDPC codes having no columns with two 1's in their corresponding $\mathbf{H}$ matrices. These bounds could be useful in finding complexity of algorithms which improve performance of LDPC codes.

1.1 Organization

In chapter 2 we introduce LDPC codes and Finite State Markov channels. We proceed to give the derivations and the expressions for the bounds on the rate derived by Gallager [Gal60], and the generalization of these bounds to memoryless symmetric channels [BKLM02]. We also introduce BP decoding of LDPC codes over BECs.

In chapter 3 we derive bounds on the rate of LDPC codes over a class of FSMCs, which we refer to as *simple* FSMCs. The derivation of these bounds is then used to find lower bounds on *density* of $\mathbf{H}$ for given performance over FSMCs. We then generalize the bounds on rate to general FSMCs. The bounds on general FSMCs may, however, be loose.

In chapter 4 we elucidate the importance of stopping sets, and draw conclusions pertaining to the ability of LDPC codes to achieve capacity. Next we define minimal stopping sets for a given LDPC code. We explain their significance in relation to decoding of LDPC codes over BECs. We then find bounds on the number of minimal stopping sets for a given ensemble of LDPC codes having no columns of weight 2.

We conclude in chapter 5.

Chapter 2

LDPC Codes and Finite State Markov Channels

2.1 LDPC Codes

Low-Density Parity-Check (LDPC) codes are linear codes for which the parity-check matrix $\mathbf{H}$ is very sparse, that is, the number of 1's in the matrix is much smaller than the number of 0's. Any vector $\mathbf{x}$ is a valid codeword if $\mathbf{Hx^T} = \mathbf{0}$. Thus rows of $\mathbf{H}$ represent the parity checks, that is, a constraint that some of the elements of $\mathbf{x}$ (viz. those which correspond to locations in the row where 1's occur) have to satisfy. Any valid codeword has to satisfy all the parity checks.

Alternatively, LDPC codes are represented by means of *Tanner graphs* [Tan81]. These are bipartite graphs, with nodes on left denoting the columns and the ones on the right denoting the rows of $\mathbf{H}$. An edge connects the left *variable* nodes to the right *check* nodes if there is a 1 in the corresponding location in the $\mathbf{H}$ matrix. An example Tanner graph and its corresponding parity check equations is shown in Fig 2.1. The *degree* of a node is the number of edges connected to that node. Hence the degree of a variable node is the number of 1's in the corresponding column in $\mathbf{H}$, also called as *column weight* for that column. Similarly, degree of a check node is the *row weight* of the corresponding row in $\mathbf{H}$. The set of variable nodes is denoted by $\mathcal{V}$.

Gallager in [Gal60] introduced codes which are now called *regular* LDPC codes[RU01b]. These codes have fixed c 1's in each row, and d 1's in each column of the parity check matrix.

The idea of regular LDPC codes was extended to *irregular* LDPC codes in [RSU01][LMSS01],

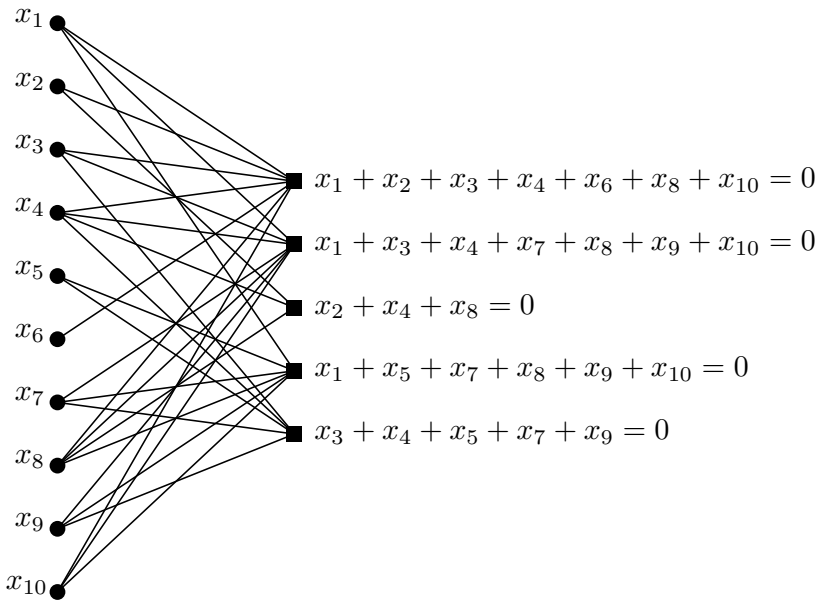

Figure 2.1: Tanner graph and corresponding parity check equations

where it was also shown that irregular LDPC codes exhibit better performance. An irregular LDPC code is defined using left and right degree polynomials, $\lambda(x)$ and $\rho(x)$. For $\lambda(x) = \sum_{i\geq 2} \lambda_i x^{i-1}$, λ_i is the fraction of *edges* of left degree i. $\rho(x)$ is similarly defined for distribution of edges of right degree i. Alternatively, they can be defined using distributions of degrees of variable and check *nodes*. It turns out (see [RU01b]) that using distributions of edges is mathematically more convenient to work with when analyzing the decoding of LDPC codes. However, in our results, we also need the distributions of check node degrees. Hence we define similar polynomial $\omega(x)$ for left node degree distribution, that is for row weight distribution. Thus ω_i denotes the fraction of variable nodes of degree i. Note that $\rho(x)$ and $\omega(x)$ are related by $\omega_d = \frac{\frac{\rho_d}{d}}{\sum_i \frac{\rho_i}{i}}$.

Irregular LDPC codes have been shown to be capacity-achieving for Binary Erasure Channels (BEC), and capacity-approaching for a large number of MBIOS channels, including the Binary Symmetric Channel (BSC) and the Additive White Gaussian Noise (AWGN) channel. Recently, their performance has been analyzed over a class of channels with memory, known as Markov channels, and more specifically on Gilbert-Elliott (GE) channel [Eck04][EKP03], and has been found to be promising.

2.1.1 Socket construction

Socket construction of LDPC codes was introduced in [RU01b]. Here we describe the construction for regular codes. An ensemble $C^n(c,d)$ of (c,d) regular LDPC codes of length n is defined as follows. Assign to each node c or d sockets, in accordance with its degree. Thus there are nc variable node sockets and $md = nc$ check node sockets. Enumerate the variable node sockets from 1 to nc. Arbitrarily enumerate the check node sockets from 1 to nc. This corresponds to generating a permutation of the set $\{1, 2, \ldots, nc\}$ and assigning values to the check nodes according to the permutation. The corresponding bipartite graph is then defined by identifying edges with pairs of sockets and letting the set of such pairs be $\{(i, \pi(i)), i = 1, 2, \ldots, nc\}$. In case of multiple edges between two nodes, an edge is assigned if the number of such edges is odd. The collection of codes generated by these permutations is called $C^n(c,d)$. Since the choice of labeling is arbitrary, this construction induces a uniform distribution on the set $C^n(c,d)$. Similar construction is performed for irregular LDPC codes.

We denote a parity check matrix of a code of rate R and blocklength n by $\mathbf{H}$. Let the number of rows of $\mathbf{H}$ be m. If the rows of $\mathbf{H}$ are linearly independent, then for regular codes, $R = 1 - \frac{m}{n} = 1 - \frac{c}{d}$. However, in general, the rows needn't be linearly independent, and hence the actual rate of the code would be greater than this value, which we call the *design rate* R_d. Thus $R_d = 1 - \frac{c}{d}$ for regular codes and $R_d = 1 - \frac{\int_0^1 \rho(x)dx}{\int_0^1 \lambda(x)dx}$ for irregular codes. Since all the rows of $\mathbf{H}$ need not be linearly independent, $R \geq R_d$.

2.1.2 Notation and definitions

A vector comprising of first n elements of a sequence of random variables is denoted in bold letters with superscript n. For example, the channel input vector is denoted by $\mathbf{X}^n := \{X_1, X_2, \ldots, X_n\}$. Similarly, the channel output vector is denoted by $\mathbf{Y}^n$ and the error vector by $\mathbf{Z}^n$. k^{th} element of any vector $\mathbf{X}^n$ is denoted by X_k.

Throughout this dissertation, $\log(\cdot)$ denotes logarithm to the base 2, and $\ln(\cdot)$ denotes logarithm to the base e. $H(\cdot)$ denotes the binary entropy function. We use the same notation $H(\cdot)$ for discrete or continuous random variables, or random vectors. The entropy of a

continuous random variable X with distribution function $f_X(x)$ is given by

$$H(X) = -\int_{-\infty}^{\infty} f_X(x)\log(f_X(x))dx \tag{2.1}$$

We now introduce some definitions which are useful in understanding the subsequent contents.

Definition 2.1 (Syndrome Vector). *The syndrome vector comprises of the results of the m parity check equations when applied to $\mathbf{Y}^n$, the received vector. It is denoted by $\mathbf{S}^{n(1-R)}$. i^{th} syndrome element is denoted by S_i.*

Definition 2.2 (Density). *For a binary linear code $\mathcal{C}$ with a parity check matrix $\mathbf{H}$, the density of $\mathbf{H}$, denoted by $\Delta(\mathbf{H})$, is defined as the ratio of the total number of 1's in a parity-check matrix to the code dimension (see [SU03]).*

The design density, Δ^d, is defined in an analogous manner to the design rate R_d. Note that $\Delta_d \leq \Delta$.

Definition 2.3 (Gap). *In a row of the parity check matrix $\mathbf{H}$, suppose two 1's, which are separated by a string of 0's, occur at locations n_1 and n_2. Then the gap between the two 1's is defined as $|n_2 - n_1|$.*

Definition 2.4 (Reliable Communication). *A sequence of codes $\{\mathcal{C}_n\}_{n=1}^{\infty}$ of corresponding bit error probabilities P_b^n is said to communicate reliably if $P_b^n \overset{n\to\infty}{\longrightarrow} 0$. Similarly, reliable communication is also defined in block error probability sense. [BKLM02]*

Definition 2.5 (Binary Symmetric Channel). *A binary symmetric channel is a memoryless binary-input binary-output channel, with $Pr(Y = 1|X = 0) = Pr(Y = 0|X = 1) = \eta$.*

Definition 2.6 (Binary Erasure Channel). *A binary Erasure channel is a memoryless channel with input alphabet $\mathcal{X} = \{0, 1\}$ and output alphabet $\mathcal{Y} = \{0, 1, E\}$, where E denotes an erasure, that is, the receiver is unable to detect the value of input bit. The conditional probability of output given the input is*

$$p_{Y|X}(y|x) = \begin{cases} p & if \quad y = E \\ 1-p & if \quad y = x \\ 0 & otherwise \end{cases} \tag{2.2}$$

Definition 2.7 (Random Construction of LDPC Codes). *In random construction of LDPC codes, a code is chosen randomly from $C^n(\lambda, \rho)$ with uniform probability.*

Note that this random construction induces a uniform probability distribution on codes of given (λ, ρ) for a given length. This further induces a probability distribution on the product space consisting of sequences of codes of different lengths for a given (λ, ρ).

2.2 Decoding of LDPC codes

A major advantage of LDPC codes is the low-complexity Belief Propagation (BP) decoding introduced in [RU01b]. BP decoding, though sub-optimal, is of linear time decoding complexity (in blocklength), as compared to exponential time decoding complexity of random codes, or quadratic time decoding complexity of some well known codes, for example the Reed-Solomon (RS) codes. At the large lengths needed to approach capacity, even quadratic time complexity is quite expensive.

In this dissertation, we require only the BP decoding over the BEC. BP decoding assumes a particularly simple form over the BEC. The decoding proceeds as follows:

First the variable nodes corresponding to unerased data bit values are marked, the received values corresponding to these nodes are added to the the check node they are connected to. This value is assigned to the check node. Then these nodes and the edges connected to these nodes are erased. In the next step, all check nodes of reduced degree 1 are marked, and the connected variable node is given the assigned value of the check node. Now the two-step procedure is repeated in the next iteration.

The reader is referred to [RU01b] for detailed description of BP and related algorithms.

2.2.1 Stopping sets

Definition 2.8 (Stopping Sets). *A stopping set $\mathcal{S}$ is a subset of $\mathcal{V}$, the set of variable nodes, such that all neighbors of $\mathcal{S}$ are connected to $\mathcal{S}$ at least twice. [DPT+02]*

Observe that, trivially, the null set is also a stopping set. Also, if $\mathcal{S}_1$ and $\mathcal{S}_2$ are two stopping sets, then evidently, $\mathcal{S}_1 \cup \mathcal{S}_2$ is also a stopping set.

Lemma 2.1 (Combinatorial Characterization of Iterative Decoder Performance). *[DPT+02] For a given LDPC code $\mathcal{C}$ transmitted over the erasure channel, we decode iteratively (using BP algorithm) until the codeword has been recovered or the decoder fails to progress further. Let $\mathcal{E}$ denote the subset of set of variable nodes which is erased by the channel. Then the set of erasures which remain when the decoder stops is equal to the unique maximal stopping set of $\mathcal{E}$.*

Proof See [DPT+02]. □

The distribution of stopping sets is used to do the finite length analysis of LDPC codes over erasure channel in [DPT+02], and to find asymptotic block error probability for a given distribution in [OVZ05].

Using stopping set distributions for irregular ensembles, Orlitsky et al (in [OVZ05, Theorem 13]) discovered the following interesting result

Theorem 2.2. *For ensemble $C^n(\lambda, \rho)$, if $\lambda'(0)\rho'(1) < 1$, then there exists a constant α^* such that $\forall\ \alpha < \alpha^*$*

$$\lim_{n\to\infty} Pr(s^* \leq \alpha n) = 1 - \sqrt{1 - \lambda'(0)\rho'(1)} \tag{2.3}$$

where s^* is the size of the smallest non-empty stopping set.

In particular, for codes with no variable nodes of degree 2 (and hence, $\lambda_2 = 0$), with probability converging to 1, the size of smallest stopping set grows linearly in n.

2.3 Finite State Markov Channels

In a *Finite State Markov Channel (FSMC)*, the channel behavior at any given time instant is dependent only on the state of the channel at that time. The state determines the output probability distribution. The underlying state-space has a Markov chain structure, with finitely many states.

Our emphasis is on *simple* FSMCs, i.e. FSMCs for which the channel behaves as a BSC in each state.

We assume that the Markov chain has a unique steady state distribution, for which the Markov chain should be irreducible and aperiodic [Ros02]. C_m denotes the capacity of an FSMC, where m is the number of states in the FSMC, each of which are denoted by integers $s = 1, 2, \ldots, m$. η_i and γ_i denote the crossover probability and the steady state probability, respectively, of the i^{th} state. The FSMC is said to be *non-inverting* if $\eta_i < 0.5\ \forall\ i \in \{1, 2, \ldots, m\}$.

Throughout this dissertation, we consider only symmetric FSMCs, that is, FSMCs which are symmetric in each state. In i^{th} state, the channel has output distribution given by $f_i(y)$ such that

$$f_i(y|X = 0) = f_i(-y|X = 1) \tag{2.4}$$

Evidently, simple FSMC is a special case case of symmetric FSMCs.

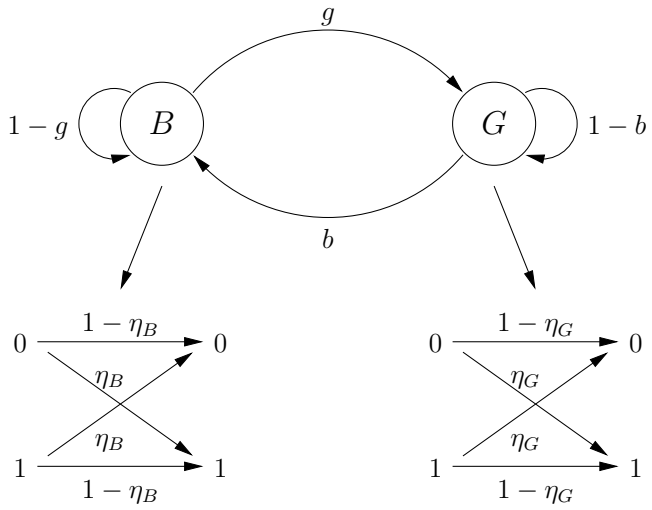

Figure 2.2: Gilbert-Elliott (GE) channel

2.3.1 Gilbert-Elliott Channels

Gilbert-Elliott (GE) channels are a special class of FSMCs. They are two-state simple FSMCs (Fig 2.3.1). The states are called "good" state and "bad" state. The "good" ("bad") state is denoted by G (B), transition probability from G to B (B to G) by b (g), the corresponding crossover probability in the state G (B) by η_G (η_B), where $\eta_G < \eta_B$. If $\eta_G < \eta_B < 0.5$, the channel is said to be *non-inverting*, and if $g + b < 1$, the channel is said to be *non-oscillatory*. We assume that the GE channel is non-inverting and non-oscillatory. The steady state probability of G (B) is denoted by $\gamma_G(\gamma_B)$. Also, C_{GE} denotes the capacity of a GE channel. Much of the notations and definitions here are borrowed from [MBD89] in which the capacity of GE channels was also derived.

2.4 Gallager's bound on the rate for the BSC

LDPC codes were introduced by Gallager in [Gal60]. In the same work, he gave bounds on the rate of regular LDPC codes for reliable communication over a BSC. The bound Gallager

found is

$$R < 1 - \frac{H(\eta)}{H(\eta_d)} \tag{2.5}$$

where $\eta_d = \frac{1+(1-2\eta)^d}{2}$. Observe that as $d \to \infty$, $\eta_d \to 0.5$ and hence $H(\eta_d) \to 1$. Also, the capacity of BSC with crossover parameter η is given by $C_{BSC}(\eta) = 1 - H(\eta)$. Thus LDPC codes for fixed d are bounded below capacity. Also, for a sequence of LDPC codes to achieve capacity of a BSC, d must converge to infinity.

It is tempting to conclude that LDPC codes can not achieve capacity, since as $d \to \infty$, the matrix $\mathbf{H}$, seemingly, would no longer be low density. However, we can have d increase slowly with respect to the blocklength such that the ratio of number of 1's in each row to the block-length would converge to 0. Indeed, it was proved by MacKay in [Mac99] that LDPC codes are capacity achieving for any stationary ergodic channel (under ML decoding). Hence there exists a sequence of LDPC codes which is capacity achieving under ML decoding, and the number of 1's in each row converges to infinity.

2.5 Generalization of Gallager's Bound: The Bound of Burshtein et al

The bound was generalized by Burshtein et al in [BKLM02] to Memoryless Binary Input Output Symmetric (MBIOS) channels and irregular LDPC codes. By an MBIOS channel we mean a channel with binary input $X \in \{0, 1\}$ and output symbol $Y \in \mathcal{R}$ such that

$$P(Y = y|X = 1) = P(Y = -y|X = 0) = f(y) \tag{2.6}$$

The crossover probability η of an MBIOS channel is defined as

$$\eta = \frac{1}{2} \int_{-\infty}^{\infty} min(f(y), f(-y))dy \tag{2.7}$$

Observe that in case of the BSC this reduces to the usual crossover probability.
For regular codes, the bound found in [BKLM02] is

$$R \leq 1 - \frac{1 - C}{H(\eta_d)} \tag{2.8}$$

where $\eta_d = \frac{1+(1-2\eta)^d}{2}$, and C is the capacity of the MBIOS channel.
For irregular codes, the bound they find is

$$R \leq 1 - \frac{1-C}{\sum_d \omega_d H(\eta_d)} \tag{2.9}$$

Using Jensen's inequality, the bound in (2.9) can be relaxed to

$$R \leq 1 - \frac{1-C}{H(\eta_{\bar{d}})} \tag{2.10}$$

where $\bar{d} = \sum_d \omega_d d$ is the average check node degree.
This bound is the same as bound for regular codes with the row weight for regular codes replaced by average row weight. From the same arguments as in section 2.4, the average row weight must converge to infinity for any sequence of codes to be capacity achieving on any MBIOS channel.

2.6 The Derivation of the Bounds on the Rate

The bounds on the rate in section 2.4 and section 2.5 are derived in a similar 2-step process. We first give Gallager's proof for BSC [Gal60], and then the generalization for MBIOS channels by Burshtein et al [BKLM02]. In the first part, the problem is reduced to finding an upper bound on the entropy of the syndrome vector $\mathbf{S}^{n(1-R)}$. In the second part, the upper bound on syndrome entropy is obtained.

2.6.1 Derivation of Gallager's bound for BSCs

Assume that there are d 1's in each row of $\mathbf{H}$. The crossover probability of the BSC is denoted by η.

The underlying idea is same as that used in proof of converse channel coding theorems: Fano's inequality. Fano's inequality [Gal68] states

$$P_b log_2(M-1) + H(P_b) \geq \frac{1}{n} H(\mathbf{X}^n|\mathbf{Y}^n) \tag{2.11}$$

where P_b is the probability of symbol error. For binary case, $M = 2$, so the equation reduces

to

$$H(P_b) \geq \frac{1}{n} H(\mathbf{X}^n|\mathbf{Y}^n) \tag{2.12}$$

Thus, if $\frac{1}{n}H(\mathbf{X}^n|\mathbf{Y}^n)$ is strictly positive, then so is P_b, the probability of bit error. It follows that for reliable communication, $\frac{1}{n}H(\mathbf{X}^n|\mathbf{Y}^n) \to 0$. To find upper bounds on the rate, we find the threshold rate above which $\frac{1}{n}H(\mathbf{X}^n|\mathbf{Y}^n) > \epsilon > 0$.

For block error probability P_B, the symbol alphabet size is 2^n. Using (2.11)

$$RP_B + \frac{H(P_B)}{n} \geq H(\mathbf{X}^n|\mathbf{Y}^n) \tag{2.13}$$

If $P_b \nrightarrow 0$, then $P_B \nrightarrow 0$, and hence the same upper bounds on rate apply to reliable communication with respect to P_B as well. (2.13) shows that the bound can not be improved upon even if we apply Fano's inequality on the whole block.

2.6.1.1 Reducing the problem to bounding the entropy of individual syndromes

We know that

$$\frac{1}{n} I(\mathbf{X}^n; \mathbf{Y}^n) = \frac{1}{n} H(\mathbf{X}^n) - \frac{1}{n} H(\mathbf{X}^n|\mathbf{Y}^n) = \frac{1}{n} H(\mathbf{Y}^n) - \frac{1}{n} H(\mathbf{Y}^n|\mathbf{X}^n)$$

Thus,

$$\frac{1}{n} H(\mathbf{X}^n|\mathbf{Y}^n) = \frac{1}{n} H(\mathbf{X}^n) - \frac{1}{n} H(\mathbf{Y}^n) + \frac{1}{n} H(\mathbf{Y}^n|\mathbf{X}^n) \tag{2.14}$$

For any code, $\mathbf{X}^n$ is the encoded data which is in 1-1 mapping with the information symbols. The information symbols are uniformly distributed over the 2^{nR} values. Thus, $\mathbf{X}^n$ takes any value from the 2^{nR} codewords with uniform probability distribution. Therefore

$$H(\mathbf{X}^n) = nR \tag{2.15}$$

Notice that

$$H(\mathbf{Y}^n|\mathbf{X}^n) = H(\mathbf{Y}^n + \mathbf{X}^n|\mathbf{X}^n) = H(\mathbf{Z}^n|\mathbf{X}^n) = H(\mathbf{Z}^\mathbf{n}) = nH(\eta) \tag{2.16}$$

where the last equality follows from the fact that errors are independent of each other, since the channel is memoryless.

It now suffices to upper bound the entropy $H(\mathbf{Y}^n)$ in (2.14).

Gallager's clever observation was this: the information content of $\mathbf{Y}^n$ is same as information content of ${\mathbf{Y_1}}^{nR}$, which are the received bits at any nR linearly independent locations in the code, and $\mathbf{S}^{n(1-R)}$, the syndrome vector[1]. Therefore,

$$\begin{aligned} H(\mathbf{Y}^n) = H({\mathbf{Y_1}}^{nR}; \mathbf{S}^{n(1-R)}) \quad &= H({\mathbf{Y_1}}^{nR}) + H(\mathbf{S}^{n(1-R)} | {\mathbf{Y_1}}^{nR}) \\ &\leq H({\mathbf{Y_1}}^{nR}) + H(\mathbf{S}^{n(1-R)}) \end{aligned} \tag{2.17}$$

where the last inequality follows from the fact that conditioning reduces entropy. Now ${\mathbf{Y_1}}^{nR} = {\mathbf{X_1}}^{nR} + {\mathbf{Z_1}}^{nR}$, where ${\mathbf{X_1}}^{nR}$ and ${\mathbf{Z_1}}^{nR}$ are the vectors corresponding to characters at independent locations in the transmitted codeword and the error vector respectively. Since ${\mathbf{X_1}}^{nR}$ is the vector corresponding to nR independent positions in the transmitted word, it specifies a codeword uniquely, and hence the distribution of ${\mathbf{X_1}}^{nR}$ is uniform over its possible 2^{nR} values.

Since ${\mathbf{X_1}}^{nR}$ has a uniform distribution over all its possible 2^{nR} values, and ${\mathbf{Z_1}}^{nR}$ is independent of ${\mathbf{X_1}}^{nR}$, ${\mathbf{Y_1}}^{nR} = {\mathbf{X_1}}^{nR} + {\mathbf{Z_1}}^{nR}$ also has a uniform distribution over all its possible 2^{nR} values. Thus

$$H({\mathbf{Y_1}}^{nR}) = nR \tag{2.18}$$

Now it is sufficient to upper bound $H(\mathbf{S}^{n(1-R)})$, the entropy of the syndrome vector. In general, the syndromes are not independent of each other. From the chain rule, and from the fact that conditioning reduces entropy, we get

$$H(\mathbf{S}^{n(1-R)}) \leq \sum_{i=1}^{n(1-R)} H(S_i) \tag{2.19}$$

The problem now reduces to finding entropy of individual syndromes.

2.6.1.2 Finding entropy of individual syndromes

In the case of BSC, this part is simple. Since there are d 1's in each row, probability that a particular parity check is satisfied is the same as the probability that even number of errors

[1]Given $\mathbf{Y}^n$, both ${\mathbf{Y_1}}^{nR}$ and $\mathbf{S}^{n(1-R)}$ can decidedly be obtained. Given ${\mathbf{Y_1}}^{nR}$ and $\mathbf{S}^{n(1-R)}$, finding $\mathbf{Y}^n$ is same as finding values of $\mathbf{Y}^n$ on locations other than those corresponding to ${\mathbf{Y_1}}^{n(1-R)}$. The problem reduces to solving a system of $n(1-R)$ linear equations with $n(1-R)$ variables, which has a unique solution. Note that R is the actual rate of the code, and not the design rate.

have occurred in the check set, which is

$$\sum_{i\ even} \binom{d}{i} \eta^i (1-\eta)^{d-i} = \frac{1+(1-2\eta)^d}{2} = \eta_d \tag{2.20}$$

Thus,

$$H(S_i) = H(\eta_d) \tag{2.21}$$

Assume that $R = 1 - \frac{H(\eta)-\delta}{H(\eta_d)}$. Using (2.21), (2.19), (2.18), (2.17), (2.15) and putting appropriate values in (2.14), using Fano's inequality (2.12), we get a lower bound of δ on the probability of error. Hence for reliable communication, we get the bound on the rate as

$$R \leq 1 - \frac{H(\eta)}{H(\eta_d)} \tag{2.22}$$

2.6.2 Derivation of Burshtein et al's bound for general MBIOS channels

Following very much in Gallager's footsteps, Burshtein et al generalized the bounds on the rate over BSC to arbitrary Memoryless Binary Input Output Symmetric (MBIOS) channels. Consider an MBIOS channel as in section 2.5. We give the derivation of the bound in (2.8) ([BKLM02]). As in section 2.6.1 we use Fano's inequality (2.12) and find a lower bound to $\frac{1}{n}H(\mathbf{X}^n|\mathbf{Y}^n)$. Since the channel is memoryless,

$$H(\mathbf{Y}^n|\mathbf{X}^n) = \sum_{l=1}^{n} H(Y_l|Y_1,\ldots,Y_{l-1},\mathbf{X}^n) = nH(Y|X) \tag{2.23}$$

Thus to lower bound $H(\mathbf{X}^n|\mathbf{Y}^n)$ in (2.12), it is again sufficient to upper bound $H(\mathbf{Y}^n)$. For this, define binary random variable ξ_l, $l = 1,\ldots,n$ as follows

$$\begin{aligned} P(\xi_l = 1|f(Y_l) > f(-Y_l)) &= 1 \\ P(\xi_l = 1|f(Y_l) < f(-Y_l)) &= 0 \\ P(\xi_l = 1|f(Y_l) = f(-Y_l)) &= 0.5 \end{aligned} \tag{2.24}$$

That is, ξ_l is result of performing hard decision on Y_l. Note that the case $f(Y_l) = f(-Y_l)$ is not merely a technical point. For example, BEC has $f(Y_l) = f(-Y_l)$ at $Y_l = 0$.

Thus we get $Pr(\xi = 0) = \frac{1}{2}$. Now for any Y_l and ξ_l,

$$H(Y_l|\xi_l) = H(Y_l) - I(Y_l;\xi_l) = H(Y_l) - H(\xi_l) + H(\xi_l|Y_l) = H(Y_l) - 1 + H(\xi_l|Y_l) \quad (2.25)$$

Also,

$$H(\mathbf{Y}^n) = H(\xi^n) + H(\mathbf{Y}^n|\xi^n) - H(\xi^n|\mathbf{Y}^n) = H(\xi^n) + H(\mathbf{Y}^n|\xi^n) - nH(Z_l|Y_l) \quad (2.26)$$

Now

$$H(\mathbf{Y}^n|\xi^n) = \sum_{l=1}^{n} H(Y_l|Y_1,\ldots,Y_{l-1},\xi^n) \leq nH(Y_l|Z_l) \quad (2.27)$$

since conditioning reduces entropy. Using equations (2.26),(2.27) and (2.25)

$$H(\mathbf{Y}^n) \leq H(\xi^n) + nH(Y) - n \quad (2.28)$$

Now it is sufficient to upper bound $H(\xi^n)$. Once again,

$$H(\xi^n) \leq nR + H(\mathbf{S}^{n(1-R)}) \leq nR + n(1-R)H(S_l) \quad (2.29)$$

And using the derivation in section 2.6.1, we get the upper bound of (2.8).

Chapter 3

Bounds on Rate for Finite State Markov Channels

In this chapter we generalize the bounds on the rate derived in [Gal60] to the class of simple Finite State Markov Channels under consideration. We also look at a possible generalization to general symmetric FSMCs. As before, we first reduce the problem to finding or bounding the entropy of individual syndromes. The derivation in this first part is similar to that in [Gal60]. However, the derivation in the second part is considerably different from the derivation in [Gal60].

3.1 Upper bounds on the rate: Reducing the problem

3.1.1 Simple FSMCs

Consider a linear code $\mathcal{C}_n$ of blocklength n and parity check matrix $\mathbf{H}$. Using Fano's inequality (2.12), if $\frac{1}{n}H(\mathbf{X}^n|\mathbf{Y}^n)$ is strictly positive, then so is P_b, the probability of bit error. For ease of exposition, we first derive the bound for regular codes.

Suppose that $\mathcal{C}_n$ is transmitted over an m-state simple FSMC. Suppose all rows of parity check matrices of C_n have a constant weight d. We derive bounds on the rate for such a sequence of codes for reliable communication.

As before,

$$\frac{1}{n}H(\mathbf{X}^n|\mathbf{Y}^n) = \frac{1}{n}H(\mathbf{X}^n) - \frac{1}{n}H(\mathbf{Y}^n) + \frac{1}{n}H(\mathbf{Y}^n|\mathbf{X}^n) \tag{3.1}$$

Also,

$$H(\mathbf{X}^n) = nR \tag{3.2}$$

As in memoryless case,

$$H(\mathbf{Y}^n|\mathbf{X}^n) = H(\mathbf{Y}^n + \mathbf{X}^n|\mathbf{X}^n) = H(\mathbf{Z}^n|\mathbf{X}^n) = H(\mathbf{Z^n}) \tag{3.3}$$

where the last equality follows from the fact that for an FSMC, the errors are independent of the input sequence.

Also, it can be inferred from [GV96, Lemma 4.2] that the sequence $\{H(Z_i|\mathbf{Z}^{i-1})\}_{i=1}^{\infty}$ is monotonically decreasing[1]. Therefore, $\frac{1}{n}H(\mathbf{Z}^n) \geq \lim_{i\to\infty} H(Z_i|\mathbf{Z}^{i-1})$. Thus,

$$\frac{1}{n}H(\mathbf{Y}^n|\mathbf{X}^n) \geq \lim_{i\to\infty} H(Z_i|\mathbf{Z}^{i-1}) \tag{3.4}$$

We need the following Lemma in the sequel.

Lemma 3.1. *The capacity of a simple FSMC is:*

$$C_m = 1 - \lim_{i\to\infty} H(Z_i|\mathbf{Z}^{i-1}) \tag{3.5}$$

Proof The derivation is similar to the derivation of channel capacity for GE channels in [MBD89].

$$C_m = \lim_{n\to\infty} \frac{1}{n} \max_{p(\mathbf{X}^n)} I(\mathbf{X}^n; \mathbf{Y}^n) = \lim_{n\to\infty} \frac{1}{n} \max_{p(\mathbf{X}^n)} \big(H(\mathbf{Y}^n) - H(\mathbf{Y}^n|\mathbf{X}^n)\big) = 1 - \lim_{n\to\infty} \frac{1}{n} H(\mathbf{Z}^n) \tag{3.6}$$

The last equality follows because $H(\mathbf{Y}^n)$ achieves its maximum value n when $\mathbf{X}^n$ is uniformly distributed over the possible 2^n values, and the distribution of $H(\mathbf{Z}^n)$ does not depend on distribution of $\mathbf{X}^n$. Also, as proved in [GV96]

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(Z_i|\mathbf{Z}^{i-1}) = \lim_{i\to\infty} H(Z_i|\mathbf{Z}^{i-1}) \tag{3.7}$$

Thus, from the chain rule and (3.6),

$$C_m = 1 - \lim_{i\to\infty} H(Z_i|\mathbf{Z}^{i-1}) \tag{3.8}$$

Using (3.2), (3.4) and Lemma 3.1, we get

$$\frac{1}{n}H(\mathbf{X}^n|\mathbf{Y}^n) \geq R - \frac{1}{n}H(\mathbf{Y}^n) + 1 - C_m \tag{3.9}$$

[1] $H(Z_i|\mathbf{Z}^{i-1}) = H(Z_{i+1}|Z_2, Z_3 \ldots, Z_i) \geq H(Z_{i+1}|\mathbf{Z}^i)$

Since we want to lower bound $\frac{1}{n}H(\mathbf{X}^n|\mathbf{Y}^n)$, we now find an upper bound on $\frac{1}{n}H(\mathbf{Y}^n)$. As in Gallager's proof, the information content of $\mathbf{Y}^n$ is same as information content of ${\mathbf{Y_1}}^{nR}$ and $\mathbf{S}^{n(1-R)}$, the syndrome vector. Therefore,

$$\begin{aligned} H(\mathbf{Y}^n) = H({\mathbf{Y_1}}^{nR}; \mathbf{S}^{n(1-R)}) &= H({\mathbf{Y_1}}^{nR}) + H(\mathbf{S}^{n(1-R)}|{\mathbf{Y_1}}^{nR}) \\ &\leq H({\mathbf{Y_1}}^{nR}) + H(\mathbf{S}^{n(1-R)}) \end{aligned} \tag{3.10}$$

where the last inequality follows from the fact that conditioning reduces entropy. Now ${\mathbf{Y_1}}^{nR} = {\mathbf{X_1}}^{nR} + {\mathbf{Z_1}}^{nR}$, where ${\mathbf{X_1}}^{nR}$ and ${\mathbf{Z_1}}^{nR}$ are the vectors corresponding to characters at independent locations in the transmitted codeword and the error vector respectively. Again, the distribution of ${\mathbf{X_1}}^{nR}$ is uniform over its possible 2^{nR} values and hence ${\mathbf{Y_1}}^{nR} = {\mathbf{X_1}}^{nR} + {\mathbf{Z_1}}^{nR}$ also has a uniform distribution over all its possible 2^{nR} values. Thus

$$H({\mathbf{Y_1}}^{nR}) = nR \tag{3.11}$$

Now it is sufficient to upper bound $H(\mathbf{S}^{n(1-R)})$, the entropy of the syndrome vector. In general, the syndromes are not independent of each other. From the chain rule, and from the fact that conditioning reduces entropy, we get

$$H(\mathbf{S}^{n(1-R)}) \leq \sum_{i=1}^{n(1-R)} H(S_i) \tag{3.12}$$

The problem now reduces to bounding entropy of individual syndromes.

3.2 Upper bounds on the rate

In this section we give the bounds on the rate for different channels by finding bounds on the entropy of the syndromes.

Consider a parity check equation, corresponding to a row of the parity check matrix $\mathbf{H}$. Let the places at which 1's occur in the equation be denoted by $n_1, n_2, \ldots n_d$, and the corresponding output random variables be denoted by $Y_{n_1}, Y_{n_2}, \ldots, Y_{n_d}$. Let $S = \sum_{i=1}^{d} Y_{n_i}$, where addition is over $GF(2)$. The entropy of a single syndrome is given by $H(S)$.

Since the input codeword to the channel $\mathbf{X}^n$ satisfies the parity check equations,

$\sum_{i=1}^{d} X_{n_i} = 0$. Hence

$$S = \sum_{i=1}^{d} Y_{n_i} = \sum_{i=1}^{d} (Y_{n_i} + X_{n_i}) = \sum_{i=1}^{d} Z_{n_i} \tag{3.13}$$

That is, a particular parity check is satisfied if there are even number of errors in locations corresponding to $\{Z_{n_i}\}_{i=1}^{d}$.

Since the state space is Markov, determining $Pr(S = 0)$ exactly is not possible without knowing the exact positions of 1's. Even if exact positions of 1's are known, the procedure to find $Pr(S = 0)$ would be tedious in general. So we develop some methods to bound this probability. We now proceed to derive bounds on the rate for different channels by finding bounds on $Pr(S = 0)$.

3.2.1 A simple upper bound for simple FSMCs

We first present a simple upper bound on the rate of LDPC codes that holds for all non-inverting simple FSMCs. We need the following Lemma.

Lemma 3.2. *For an m-state simple FSMC*

$$Pr(S = 0) = \frac{1}{2} + \frac{1}{2} \sum_{r_1, r_2, \ldots, r_m} \Pi_{i=1}^{m} (1 - 2\eta_i)^{r_i} Pr(r_1, r_2, \ldots, r_{m-1}) \tag{3.14}$$

where $Pr(r_1, r_2, \ldots, r_m)$ denotes the probability of making r_i visits to state i in d steps, that is, $\sum_{i=1}^{m} r_i = d$.

Proof See Appendix A. □

Since the channel is non-inverting in each state, $1 - 2\eta_i > 0 \ \forall i$. Hence $Pr(S = 0)$ in Lemma 3.2 can be lower bounded by

$$Pr(S = 0) \geq \frac{1}{2} + \frac{1}{2} \sum_{r_1, r_2, \ldots, r_m} (1 - 2\eta_m)^d Pr(r_1, r_2, \ldots, r_m) = \frac{1}{2} + \frac{(1 - 2\eta_m)^d}{2} \tag{3.15}$$

where, we assume without loss of generality that $\eta_m > \eta_i \ \ \forall \ i \neq m$. Define $p_{md} \stackrel{\Delta}{=} \frac{1}{2} + \frac{(1-2\eta_m)^d}{2}$. Thus $H(S) \leq H(p_{md})$, and from (3.12), $H(\mathbf{S}^{n(1-R)}) \leq n(1 - R)H(p_{md})$.

Suppose now that rate $R = 1 - \frac{1-C_m}{H(p_{md})} + \epsilon$, for some $\epsilon > 0$. Using (2.12), (3.9), (3.10), (3.11) and (3.12), we get

$$H(P_b) > \frac{1}{n} H(\mathbf{X}^n | \mathbf{Y}^n) \geq \epsilon H(p_{md}) \tag{3.16}$$

That is, reliable communication is not possible at this rate for any $\epsilon > 0$. Hence we get the following bound for regular codes

$$R \leq 1 - \frac{1 - C_m}{H(p_{md})} \tag{3.17}$$

Furthermore, (3.16) also gives us an error floor for rates exceeding this bound. For codes of any length, for rates exceeding the bound in (3.17) by ϵ, the probability of bit error is lower bounded by $H^{-1}(\epsilon H(p_{md}))$, for $H^{-1}(\epsilon H(p_{md})) < 0.5$.

For irregular codes, the expression for upper bound on $H(\mathbf{S}^{n(1-R)})$ changes to

$$H(\mathbf{S}^{n(1-R)}) \leq n(1-R) \sum_d \omega_d H(p_{md}) \tag{3.18}$$

Hence, the bound on the rate for irregular codes can similarly be found to be

$$R \leq 1 - \frac{1 - C_m}{\sum_d \omega_d H(p_{md})} \tag{3.19}$$

Similar error floors (as in (3.16)) can be obtained for irregular codes. Since all the bounds we give in this dissertation are of the same form, similar error floors can be derived for all of them.

Arguments in [SU03] concluded, using Jensen's inequality, that for memoryless channels, the expression of upper bound on the rate gives the result that for a sequence of codes to be capacity achieving, the density $\Delta(\mathbf{H})$ (see defn. 2.2) must converge to infinity. Since the expression for upper bound is same here with appropriate change in the expression for the capacity of a simple FSMC, we can conclude that even for simple FSMCs, for a sequence of LDPC codes to be capacity achieving, the density must converge to infinity.

In the next subsections, we tighten these bounds for the case of GE channels, and then for FSMCs.

3.2.2 Upper Bounds for non-inverting and non-oscillating GE channels

From Lemma 3.2, it can be seen that for non-inverting GE channels, $Pr(S = 0) > 0.5$, and hence, the entropy $H(S)$ increases with decrease in $Pr(S = 0)$. Thus, to upper bound $H(S)$, we lower bound $Pr(S = 0)$. To that end, we need the following Lemmas

Lemma 3.3. *For non-oscillating and non-inverting GE channels, $Pr(S = 0)$ decreases with increase in gap between any two 1's, keeping the gap between other 1's constant.*

Proof See Appendix B. □

Lemma 3.4. *For an m-state simple FSMC, if we keep increasing the gap between all the 1's, in the limit, $Pr(S = 0)$ converges to $Pr(S_{memless} = 0)$, where $S_{memless}$ is the random variable representing result of a parity check equation for a memoryless channel with error probability same as the average error probability of the FSMC in steady state.*

Proof Define k as the minimum gap between any two 1's in the row of $\mathbf{H}$ under consideration. First, we prove that as $k \to \infty$,

$$Pr(s_{n_1} = a_1, s_{n_2} = a_2, \ldots, s_{n_d} = a_d) \to \Pi_{i=1}^{d} \gamma_{a_i} \tag{3.20}$$

Notice that

$$\begin{aligned} &Pr(s_{n_1} = a_1, s_{n_2} = a_2, \ldots, s_{n_d} = a_d) \\ &\qquad = Pr(s_{n_1} = a_1) Pr(s_{n_2} = a_2 | s_{n_1} = a_1) \ldots Pr(s_{n_d} = a_d | s_{n_{d-1}} = a_{d-1}) \end{aligned} \tag{3.21}$$

As k increases, gap between each of n_i and n_{i-1} increases, and $Pr(s_{n_i} = a_i | s_{n_{i-1}} = a_{i-1}) \to \gamma_{a_i}$. This proves (3.20).

Now consider the following for b_i binary

$$\begin{aligned} &Pr(Z_{n_1} = b_1, Z_{n_2} = b_2, \ldots, Z_{n_d} = b_d) \\ &\quad = \sum_{a_i = 1,2,\ldots,m} Pr(Z_{n_1} = b_1, Z_{n_2} = b_2, \ldots, Z_{n_d} = b_d | s_{n_1} = a_1, s_{n_2} = a_2, \ldots, s_{n_d} = a_d) \\ &\qquad\qquad \times Pr(s_{n_1} = a_1, s_{n_2} = a_2, \ldots, s_{n_d} = a_d) \\ &\qquad \overset{k \to \infty}{\longrightarrow} \sum_{a_i = 1,2,\ldots,m; i = 1,2,\ldots,d} \Pi_{i=1}^{d} Pr(Z_{n_i} = b_i | s_{n_i} = a_i) \times Pr(s_{n_i} = a_i) \end{aligned} \tag{3.22}$$

Now,

$$\sum_{a_i=1,2,\ldots,m;i=1,2,\ldots,d} \Pi_{i=1}^{d} Pr(Z_{n_i} = b_i | s_{n_i} = a_i) \times Pr(s_{n_i} = a_i)$$

$$= \sum_{a_i=1,2,\ldots,m;i=1,2,\ldots,d} \Pi_{i=1}^{d} Pr(Z_{n_i} = b_i | s_{n_i} = a_i)\gamma_{a_i}$$

$$= \Pi_{i=1}^{d} \sum_{j=1}^{m} Pr(Z_{n_i} = b_i | s_{n_i} = j)\gamma_j = (1-q)^t q^{d-t} \quad (3.23)$$

where $q = \sum_{i=1}^{m} \eta_i \gamma_i$, and t is the number of $b_i's$ which are 1.
Using (3.22) and (3.23),

$$Pr(Z_{n_1} = b_1, Z_{n_2} = b_2, \ldots, Z_{n_d} = b_d) \overset{k\to\infty}{\longrightarrow} (1-q)^t q^{d-t} \quad (3.24)$$

Finally,

$$Pr(Z_{n_1} + Z_{n_2} + \ldots + Z_{n_d} = 0)$$

$$= \sum_{\text{even number of } b_i\text{'s are 1}} Pr(Z_{n_1} = b_1, Z_{n_2} = b_2, \ldots, Z_{n_d} = b_d) \to \frac{1 + (1-2q)^d}{2} \qquad \square \quad (3.25)$$

From Lemma 3.3 and Lemma 3.4, $H(S) \leq H(S_{memless}) = H(\frac{1+(1-2q)^d}{2})$, where $q \overset{\Delta}{=} \gamma_G \eta_G + \gamma_B \eta_B$ is the average probability of error in steady state of the FSMC.

Thus, similar to the derivation of (3.17), we get the following bound

$$R \leq 1 - \frac{1 - C_{GE}}{H(\overline{p_d})} \quad (3.26)$$

where C_{GE} is the capacity of a GE channel, and $\overline{p_d} \overset{\Delta}{=} \frac{1+(1-2q)^d}{2}$

Similar to (3.18), the bound on the rate for reliable communication for irregular codes is

$$R \leq 1 - \frac{1 - C_{GE}}{\sum_d \omega_d H(\overline{p_d})} \quad (3.27)$$

3.2.3 Tightening the bound for GE channels

In this section, we tighten the upper bound on the entropy of a syndrome, which can be used to tighten the bounds on the rate for GE channels.

Suppose in a row of a parity check matrix, the *maximum* gap between any two 1's is v,

that is, any two variables in that parity check equation are separated by a gap of no more than v. By Lemma 3.3, the entropy of the parity check increases as gap between 1's is increased. Hence the entropy of the given parity check will be lesser than or equal to the entropy of a parity check for which the gap between 1's is uniformly v. Note that in the latter case, the Markov chain relating Z_{n_i}'s is homogeneous, albeit the transition probabilities have changed.

It was proved by Pedler [Ped71] that for a homogeneous two state Markov chain, probability of visiting a particular state (say G) k times in d transitions is given by:
For $0 < k < d$

$$\begin{aligned} Pr(N_G = k) = \quad & (1-b)^k(1-g)^{d-k}F[-d+k,-k;1;\lambda] \\ & -\gamma_G l(1-b)^k(1-g)^{d-k-1}F[-d+k+1,-k;1;\lambda] \\ & -\gamma_B l(1-b)^{k-1}(1-g)^{d-k}F[-d+k,-k+1;1;\lambda] \end{aligned}$$

and

$$\begin{aligned} Pr(N_G = 0) &= (\gamma_G b + \gamma_B(1-g))(1-g)^{d-1} \\ Pr(N_G = d) &= (\gamma_G(1-b) + \gamma_B g)(1-b)^{d-1} \end{aligned}$$

where N_G is the random variable denoting number of times state G is visited, F is the hypergeometric function, $\gamma_G = \frac{g}{b+g}$ and $\gamma_B = \frac{b}{b+g}$ are the steady state probabilities of G and B respectively, $\lambda = \frac{gb}{(1-g)(1-b)}$, and $l = (1-g)(1-b) - gb$.

Now, using (3.68) of Appendix A, the entropy of a parity check can be determined if we know $Pr(N_G = k)$. For uniform gap v between 1's, the underlying state space for $Z'_{n_i}s$ is homogeneous Markov, and hence Pedler's result is applicable. Using v-step transition probabilities and (3.68), associated entropy can be calculated for different values of v. Fig 3.1 shows the variation of $H(S)$ with v for typical values of d, η_G, η_B, g and b. As v increases, we see that $H(S)$ converges to the entropy in the independent case, as expected.

Hence if v is known for each parity check, a tighter bound on entropy $H(S)$ can be obtained, and thus the bounds derived earlier for GE channels can be tightened. However, for fixed (λ, ρ), as we let $n \to \infty$, only in some specific constructions of LDPC codes would

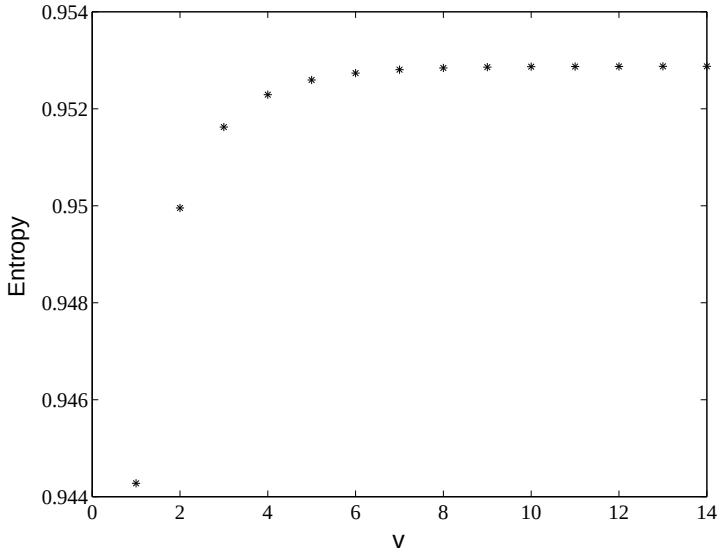

Figure 3.1: $H(S)$ versus v for $d = 10$, $\eta_G = 0.01$, $\eta_B = 0.1$, $g = 0.2$, $b = 0.3$.

the maximum gap between $1's$ remain constant. Hence, per se, the tightening of the bound in this section is not really useful. We revisit this point in section 3.3, where we bring out the utility of this result.

The tightening raises a natural question: can we conclude from this tightening that LDPC codes constructed in a manner that gap between 1's is bounded cannot achieve capacity? We show in Appendix A (eqn. (3.71)) that the proposed tightening does *not* lead to this conclusion.

3.2.4 An *almost-sure* bound for simple FSMCs

We first discuss the relevance of the concentration theorem for BP decoding of LDPC codes to the bound we give in this section. Then we derive *almost-sure* bounds for reliable communication over simple FSMCs.

Concentration theorem and upper bound on the rate

The bounds we give here hold *almost-surely* for ML decoding of a sequence of *randomly constructed* LDPC codes over such FSMCs. Since these bounds hold for ML decoding, they

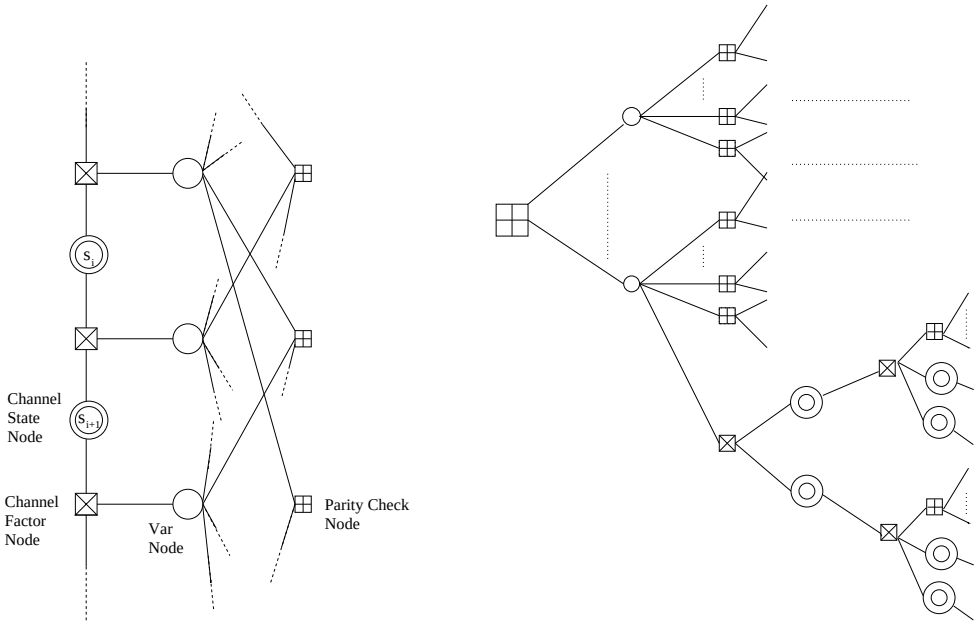

Figure 3.2: a) Factor graph for BP decoding of LDPC code over Markov channel b) Message passing neighborhood with a check node as root for scheduling in [Eck04]

also hold for BP decoding of randomly constructed LDPC codes. In what follows, we first elucidate why the concentration theorem for BP decoding of LDPC codes over FSMCs is needed for establishing the utility of random construction.

Density evolution, as discussed in [RU01b], is used to analyze the behavior of probability of error with number of iterations in BP decoding, assuming infinite block length. In the analysis, the evolution of the average probability density of messages passed (averaged over codes of given (λ, ρ), under tree assumption) is found.

The concentration theorem guarantees that for large lengths, almost all codes in ensemble $C^n(\lambda, \rho)$ have performance close to average performance under BP decoding. Provided that the graph is tree like, the behavior can be predicted using density evolution. Thus if density evolution of a degree distribution shows good performance, choosing a code randomly from ensemble of codes of given degree distribution would be a good construction for large block-lengths. Hence one can Analise the performance of random construction by density evolution, provided that the concentration theorem holds. Herein lies the importance of proving the concentration theorem.

The concentration theorem has been proved for BP decoding of LDPC codes over Memoryless channels [RU01b] and ISI channels [KMM03]. We prove it for the case of Markov channels. We formally state the concentration theorem theorem below.

Theorem 3.5 (Concentration Theorem for Markov Channels). *Let $E[Z]$ denote the expected number of incorrect messages in l^{th} iteration of BP decoding, where the expectation is over all the codes in $C^n(c,d)$ and the noise. Then the probability that the actual fraction of incorrect messages lies outside $(\frac{E[Z]}{nc} - \epsilon, \frac{E[Z]}{nc} + \epsilon)$ converges to 0 exponentially in n.*

Proof See Appendix C. □

We also need to prove that the neighborhood graph is a tree after a fixed l iterations with a probability that converges to 1 as $n \to \infty$. This is a simple extension to the corresponding proof for memoryless channels in [RU01b], and is proved in [Eck04].

3.2.4.1 Upper bound on the rate

We now derive an *almost-sure* upper bound on the rate of a sequence of LDPC codes for reliable communication over a simple FSMC. For each length n, we have an ensemble of codes $C^n(\lambda, \rho)$ and a uniform probability distribution over $C^n(\lambda, \rho)$. The probability space of sequence of codes is the product space of probability spaces corresponding to each n. The probability distribution over the product space is induced by uniform probability distribution over $C^n(\lambda, \rho)$. The bound is *almost-sure* in the sense that any sequence of codes $\{\mathcal{C}_n\}$ (with $\mathcal{C}_n \in C^n(\lambda, \rho)$) has to satisfy this bound with probability 1 (in the product space) if it communicates reliably.

Consider a parity-check equation. By Lemma 3.4, as k increases,

$$Pr(S = 0) \to Pr(S_{memless} = 0) = \frac{1 + (1 - 2q)^d}{2} \tag{3.28}$$

where $q = \sum_i \eta_i \gamma_i$ is the average probability of error in the steady state. Since binary entropy function is a continuous function of probability, $H(S) \to H(S_{memless})$ as k increases. In particular, we can choose k large enough such that $H(S) \leq H(S_{memless}) + \delta$ for any given $\delta > 0$.

Lemma 3.6. *For any $\epsilon > 0$, the fraction of codes which have at least ϵ fraction of rows in which at least two 1's are at a gap not exceeding k goes to 0 exponentially in n (for any fixed k, and for n large enough).*

Proof We prove the Lemma for regular codes.

First we look at the number of codes in the collection $C^n(c,d)$. The number of sockets in the random construction in [RU01b] is $n(1-R)d$. Hence the number of codes is equal to the number of permutations of set of $n(1-R)d$ different elements. Thus the total number of codes in $C^n(c,d)$ is $(n(1-R)d)!$. Using Stirling's approximation, i.e., $n! \approx n^n e^{-n}\sqrt{2\pi n}$, the total number of codes can be approximated as

$$(n(1-R)d)! \approx (n(1-R)d)^{n(1-R)d} e^{-n(1-R)d}\sqrt{2\pi n(1-R)d} \tag{3.29}$$

Now we find a bound on the number of codes which have at least an ϵ fraction of rows with gap between at least two 1's not exceeding k. Let τ_ϵ^k denote the set of such codes, and let $N_\epsilon^k = |\tau_\epsilon^k|$ denote the number of such codes.

Choose ϵ fraction of rows, where each row corresponds to a check node in the corresponding Tanner graph. This choice can be made in $\binom{n(1-R)}{\epsilon n(1-R)}$ ways. These are the rows for which at least two 1's have gap not exceeding k. For each of these rows, the first $d-1$ left sockets can be chosen in at most $(nd(1-R))^{d-1}$ ways. The last socket, which is at most a distance k from one of the 1's, can only be chosen in a constant (independent of n) number of ways, say c_1 ($c_1 \leq (k-1)c$). Here we do not put any restriction for the choice of the first $d-1$ sockets. Only in choice of the last socket is the constraint on gap used.

For the rest $(1-\epsilon)$ fraction of the rows, the number of ways in which variable node sockets can be chosen is at most $(nd(1-R))^d$. Thus,

$$N_\epsilon^k < \binom{n(1-R)}{\epsilon n(1-R)}((nd(1-R))^{d-1} \times (k-1)c)^{n(1-R)\epsilon}((nd(1-R))^d)^{n(1-R)(1-\epsilon)} \tag{3.30}$$

Using Stirling's approximation

$$\binom{n(1-R)}{\epsilon n(1-R)} \approx \frac{1}{(1-\epsilon)^{n(1-R)(1-\epsilon)}\epsilon^{n(1-R)\epsilon}\sqrt{2\pi n\epsilon(1-\epsilon)(1-R)}} \tag{3.31}$$

Using (3.29), (3.30) and (3.31), we can see that the probability of choosing a code from τ_ϵ^k converges to zero as

$$\frac{(nd(1-R))^{-n(1-R)\epsilon}((k-1)c)^{n(1-R)\epsilon}}{(1-\epsilon)^{n(1-R)(1-\epsilon)}\epsilon^{n(1-R)\epsilon}2\pi n(1-R)\sqrt{\epsilon(1-\epsilon)d}} = O\left(\frac{1}{ne^{n(\epsilon\ln(n)-c_2)}}\right) \tag{3.32}$$

for some constant c_2, and hence for n large enough, the decrease is exponential in n $\square$

The same proof works for irregular codes with bounded degrees, with the replacement of maximum left and right degrees in place of c, d. $\square$

From this Lemma, it follows that for large enough n (given k), all 1's are separated by a gap greater than k for at least $(1-\epsilon)$ fraction of rows with probability that goes to 1 exponentially in n. By our choice of k, it now follows that $H(S) > H(S_{memless}) + \delta$ for greater than ϵ fraction of rows with probability that converges to 0 exponentially in n. Since the probability decreases exponentially with n, an application of Borel-Cantelli Lemma shows that bound holds *eventually, almost-surely*[2]. That is, with probability 1, the event $H(S) > H(S_{memless}) + \delta$ for greater than ϵ fraction of rows will happen *only finitely often.*

From the above argument, we have an upper bound on the entropy of $(1-\epsilon)n(1-R)$ syndromes. For the entropy of rest of $\epsilon n(1-R)$ parity checks, we use the upper bound of 1, and arrive at the following bound on $H(\mathbf{S}^{n(1-R)})$

$$H(\mathbf{S}^{n(1-R)}) \leq n(1-R)(1-\epsilon)(H(S_{memless}) + \delta) + n(1-R)\epsilon \quad e.a.s.\ \forall\, \epsilon, \delta > 0 \tag{3.33}$$

$$\Rightarrow \frac{1}{n} H(\mathbf{S}^{n(1-R)}) \leq (1-R)H(S_{memless}) + \theta \quad e.a.s.\ \forall\, \theta > 0 \tag{3.34}$$

where $H(S_{memless}) = H(\frac{1+(1-2q)^d}{2})$, and *e.a.s.* denotes that the bound holds *eventually, almost surely.* Now, similar to the derivation in 3.2.1, for any sequence of regular LDPC codes that communicates reliably, the rate is bounded as follows

$$R \leq 1 - \frac{1-C_m}{H(\overline{p_d})} + \epsilon \quad \text{e.a.s. } \forall \epsilon > 0 \tag{3.35}$$

where $\overline{p_d} = \frac{1+(1-2q)^d}{2}$.

Define $R_0 \stackrel{\Delta}{=} 1 - \frac{1-C_m}{H(\overline{p_d})}$. For given c, d, the design rate $R_d = 1 - \frac{c}{d}$ and R_0 are fixed, and hence if $R_d > R_0$, the code rate $R(\geq R_d)$ would always (and hence, infinitely often) exceed $R_0 + \epsilon$ for $\epsilon = R_d - R_0$. Thus the above bound would be violated. Hence the same bound holds for $\epsilon = 0$ also[3] if we replace R by R_d. For the same reason, for fixed R_d and R_0, the

[2]The term *eventually, almost-surely* implies that there is a set of infinite sequences of probability 1 such that for each sequence of this set, the statement holds for all $n > n_0$, where n_0 may be dependent on the sequence

[3]Consider a sequence of random variables $\{X_n\}$. Suppose $X_n < K + \epsilon$ a.s for all $\epsilon > 0$. But this does

term *eventually* in (3.35) is redundant. Thus, the bound for regular codes of given (c, d) becomes

$$R_d \leq 1 - \frac{1 - C_m}{H(\overline{p_d})} \text{ a.s.} \tag{3.36}$$

Similarly for irregular codes of given (λ, ρ), the bound is

$$R_d \leq 1 - \frac{1 - C_m}{\sum_d \omega_d H(\overline{p_d})} \text{ a.s.} \tag{3.37}$$

There can exist sequences of codes of the given (λ, ρ) which defy the bound, but the set of such sequences is of probability 0 in the product space. Note that we do not need the channel to be non-inverting for the *almost-sure* bound to hold.

3.3 Lower bounds on parity-check density

The upper bounds in (3.17), (3.18), (3.26), (3.27), (3.36) and (3.37) are similar to expressions of upper bound on the rate derived in [BKLM02], with the capacity of general MBIOS channels replaced by C_m. In [SU03], lower bounds on parity check density of LDPC codes were derived for MBIOS channels using the same upper bounds. The bounds on density derived in [SU03] are of two types:

- *Type I:* For *any linear code* of given finite blocklength n and given probability of error when transmitted over a channel, using the lower bound on $\frac{1}{n}H(\mathbf{X}|\mathbf{Y})$, a lower bound on $\Delta(\mathbf{H})$ can be found (see [SU03, Thm 2.4]).

- *Type II:* For a sequence of linear codes $\{C_n\}$ which communicate reliably to achieve $1 - \epsilon$ of the capacity of an MBIOS channel, asymptotic lower bound on density of their parity check matrices can be found (see [SU03, Thm 2.1]).

Let Δ_m denote the density of parity-check matrix corresponding to $\mathcal{C}_m$. Similar to the derivation in [SU03], using the upper bound on the syndrome entropy derived in section 3.2.2, we can get the following *Type II* bound on asymptotic density of *any sequence of linear codes*

not guarantee that $X_n \leq K$ a.s. For example, take $X_n = 1 + \frac{1}{n}$ with probability 1. Then $X_n < 1 + \epsilon$ a.s. $\forall\, \epsilon > 0$. But evidently, $X_n \not\leq 1$ a.s. Hence we use this alternative argument to arrive at the bound in (3.36) and (3.37)

that achieves reliable communication at rate $(1-\epsilon)C_{GE}$ over GE channels

$$\underline{\lim}_{m\to\infty} \Delta_m > \frac{K_1 + K_2 \ln\frac{1}{\epsilon}}{1-\epsilon} \tag{3.38}$$

where $\underline{\lim}_{m\to\infty}$ denotes lim inf and

$$K_1 = \frac{(1-C_{GE})\ln\left(\frac{1-C_{GE}}{2\ln 2\, C_{GE}}\right)}{2C_{GE}\ln\left(\frac{1}{1-2q}\right)} \text{ and } K_2 = \frac{1-C_{GE}}{2C_{GE}\ln\left(\frac{1}{1-2q}\right)}$$

Using the expression for the upper bound on the syndrome entropy in section 3.2.2, we get the following *Type I* bound on density of parity check matrix of any linear code given its block length and performance over a GE channel

$$\Delta \geq \frac{1-(1-\epsilon)C_{GE}}{2(1-\epsilon)C_{GE}} \cdot \frac{\ln\left(\frac{1}{2\ln 2}\cdot\frac{1-C_{GE}+\epsilon C_{GE}}{\epsilon C_{GE}+H(P_b)}\right)}{\ln\left(\frac{1}{1-2q}\right)} \tag{3.39}$$

For non-inverting simple FSMCs, the bounds in (3.38) and (3.39) hold with q replaced by η_m and C_{GE} replaced by C_m.

Tighter *Type II* bounds for all simple FSMCs hold *almost-surely* for the design density Δ^d of a sequence of LDPC codes of given (λ, ρ). Similar to [SU03], we obtain the following bound

$$\Delta^d > \frac{K_1 + K_2 ln\frac{1}{\epsilon}}{1-\epsilon} \text{ a.s.} \tag{3.40}$$

We have only a probabilistic result for finite lengths (*Type I* bound). Based on the probabilistic bounds on the syndrome entropy for random construction derived in section 3.2.4, we can derive the following probabilistic bounds on Δ which hold for LDPC codes of given (λ, ρ)

$$\Delta \geq \frac{1-(1-\epsilon)C_m}{2(1-\epsilon)C_m} \cdot \frac{\ln\left(\frac{1}{2\ln 2}\cdot\frac{1-C_m+\epsilon C_m}{\epsilon C_m+H(P_e)+\theta}\right)}{\ln\left(\frac{1}{1-2q}\right)} \tag{3.41}$$

A lower bound on the probability with which this bound holds can be obtained from (3.32). This probability is dependent on choice of θ and n, and converges to 1 exponentially in n.

Since these bounds are derived in the same manner as the bounds on [SU03], we omit the derivations here. Similar bounds on the density can also be derived (using (2.13)) for a given block probability of error as shown in [SU03].

3.3.1 Tightening of lower bounds on density for GE channels using results in section 3.2.3

For using the tight bound for GE channels of section 3.2.3, we first prove the following Lemma.

Lemma 3.7. *For a non-inverting simple FSMC,* $Pr(S = 0) = Pr(\sum_{i=1}^{d} Z_{n_i} = 0)$ *decreases with increase in* d.

Proof The proof is by induction on d. Suppose we add another 1 to the parity check equation, which corresponds to adding another binary random variable Z to S. Now we have to consider $Pr(S + Z = 0)$.

First we note that

$$Pr(S + Z = 0) = Pr(S = 0; Z = 0) + Pr(S = 1; Z = 1) \tag{3.42}$$

Similar to the derivation of (3.79) in Appendix B, we can see that

$$Pr(S = 0; Z = 0) = Pr(S = 0) - Pr(Z = 1) + Pr(S = 1; Z = 1) \tag{3.43}$$

From (3.42) and (3.43), we get

$$\begin{aligned} Pr(S = 0) - Pr(S + Z = 0) &= Pr(Z = 1) - 2Pr(S = 1; Z = 1) \\ &= Pr(Z = 1) - 2Pr(Z = 1)Pr(S = 1|Z = 1) \end{aligned} \tag{3.44}$$

Since the channel is non-inverting, $Pr(S = 1|Z = 1) < 0.5$ (from Appendix A). Thus the quantity in (3.44) is positive, and the Lemma follows. □

We use this Lemma for the particular case of non-oscillating and non-inverting GE channels. Since $Pr(S = 0)$ decreases with increasing d (by Lemma 3.7), for regular codes, using the upper bound on syndrome entropy derived in 3.2.3, we can derive a tighter *Type I* lower bound on $\Delta(\mathbf{H})$ for GE channels.

Because there is no closed form expression for $P(S = 0)$ in 3.2.3, we could not prove the convexity of $H(S)$ as a function of d, and hence the lower bounds on $\Delta(\mathbf{H})$ do not extend directly to irregular codes. However, in light of Lemma 3.7, it is simple to see that the bounds continue to hold for maximum row weight, instead of average row weight.

Further, we can derive a necessary condition that an LDPC code of given row weights and given maximum gaps in the rows has to satisfy for given performance over GE channels.

Denote the probability of a parity check of d 1's at gaps uniformly v by p_d^v. Also, let ω_d^v denote the fraction of rows of weight d and maximum gap v. Then, using (2.12), (3.9), (3.10), (3.11) and (3.12) we get

$$H(P_e) \geq -(1-R)\sum_d \omega_d^v H(p_d^v) + 1 - C_{GE} \tag{3.45}$$

And hence, the code has to satisfy

$$\sum_d \omega_d^v H(p_d^v) \geq \frac{1 - C_{GE} - H(P_e)}{1-R} \tag{3.46}$$

3.4 Generalization to general symmetric FSMCs

For general symmetric FSMCs, we proceed in a manner similar to that of Burshtein et al discussed in section 2.6.2. Let $f_i(\cdot)$ be the corresponding output distribution for input $X = 1$ when the channel is in state i.

We first derive relevant expression for capacity of general symmetric FSMC

Lemma 3.8. *The capacity of a symmetric FSMC is given by*

$$C_m = \lim_{l\to\infty} \underset{p(\mathbf{X}^l)}{max} H(Y_l|\mathbf{Y}^{l-1}) - \lim_{l\to\infty} H(Y_l|\mathbf{Y}^{l-1}, \mathbf{X}^l) \tag{3.47}$$

Proof

$$C_m = \underset{p(\mathbf{X}^n)}{max} \lim_{n\to\infty} \frac{1}{n} I(\mathbf{X}^n; \mathbf{Y}^n) = \underset{p(\mathbf{X}^n)}{max} \lim_{n\to\infty} \left(\frac{1}{n} H(\mathbf{Y}^n) - H(\mathbf{Y}^n|\mathbf{X}^n) \right) \tag{3.48}$$

By channel symmetry, $H(\mathbf{Y}^n|\mathbf{X}^n)$ is independent of input distribution.[4] The Lemma follows by application of chain rule. □

Let $p_c(\mathbf{X}^n)$ be the capacity achieving distribution. Despite the channel symmetry in each state, $p_c(\mathbf{X}^n)$ may not be the independent uniform distribution. Indeed, finding this input distribution for general FSMCs is an open problem.

Similar to derivation in section 2.6.1,

$$\frac{1}{n} H(\mathbf{X}^n|\mathbf{Y}^n) = R - \frac{1}{n} H(\mathbf{Y}^n) + \frac{1}{n} H(\mathbf{Y}^n|\mathbf{X}^n) \tag{3.49}$$

[4] Assuming $X \in \{1, -1\}$, $H(\mathbf{Y}^n|\mathbf{X}^n) = H(\mathbf{Y}^n \cdot \mathbf{X}^n|\mathbf{X}^n)$, where $\cdot$ is element-by-element product of the two vectors. Now $\mathbf{Y}^n \cdot \mathbf{X}^n$ is independent of $\mathbf{X}^n$.

Again,

$$\frac{1}{n}H(\mathbf{Y}^n|\mathbf{X}^n) \rightarrow \lim_{l\rightarrow\infty} H(Y_l|\mathbf{Y}^{n-1},\mathbf{X}^n) \tag{3.50}$$

Also, by channel symmetry, this quantity is independent of distribution of $\mathbf{X}^n$. Hence

$$\lim_{l\rightarrow\infty} H(Y_l|\mathbf{Y}^{l-1},\mathbf{X}^n) = \lim_{l\rightarrow\infty;\, p(\mathbf{X}^l)=p_c(\mathbf{X}^l)} H(Y_l|\mathbf{Y}^{l-1}) - C_m \tag{3.51}$$

where C_m is the capacity of the FSMC under consideration. Now,

$$H(Y_l|\mathbf{Y}^{l-1}) = H(Y_l|\xi_n,\mathbf{Y}^{l-1}) + H(\xi_l|\mathbf{Y}^l) - H(\xi_l|\mathbf{Y}^l) \tag{3.52}$$

We desire to find an upper bound on $H(\mathbf{Y}^n)$ similar to that in section 2.6.2. To that end, we define a binary random variable ξ as follows

$$\begin{aligned} Pr\big(\xi_l = 1|\sum_i \pi_i^l f_i(Y_l) > \sum_i \pi_i^l f_i(-Y_l)\big) &= 1 \\ Pr\big(\xi_l = 1|\sum_i \pi_i^l f_i(Y_l) < \sum_i \pi_i^l f_i(-Y_l)\big) &= 0 \\ Pr\big(\xi_l = 1|\sum_i \pi_i^l f_i(Y_l) = \sum_i \pi_i^l f_i(-Y_l)\big) &= 0.5 \end{aligned} \tag{3.53}$$

where $\pi_i^l = Pr(S_l = i|Y_1,\ldots,Y_{l-1})$. Now,

$$\frac{1}{n}H(\mathbf{Y}^n) = \frac{1}{n}H(\mathbf{Y}^n|\xi^n) - \frac{1}{n}H(\xi^n|\mathbf{Y}^n) + \frac{1}{n}H(\xi^n) \tag{3.54}$$

We tackle the first term in (3.54)

$$\begin{aligned} \frac{1}{n}H(\mathbf{Y}^n|\xi^n) \rightarrow \lim_{l\rightarrow\infty} H(Y_l|\xi^l,\mathbf{Y}^{l-1}) &= \lim_{l\rightarrow\infty} H(Y_l|\xi_l,\mathbf{Y}^{l-1}) \\ &= \lim_{l\rightarrow\infty}\big(H(Y_l,\xi_l|\mathbf{Y}^{l-1}) - H(\xi_l|\mathbf{Y}^{l-1})\big) \\ &= \lim_{l\rightarrow\infty}\big(H(\xi_l|\mathbf{Y}^l) + H(Y_l|\mathbf{Y}^{l-1}) - H(\xi_l|\mathbf{Y}^{l-1})\big) \end{aligned} \tag{3.55}$$

Now consider the second term in (3.54)

$$\frac{1}{n}H(\xi^n|\mathbf{Y}^n) \rightarrow \lim_{l\rightarrow\infty} H(\xi_l|\xi^{l-1},\mathbf{Y}^l) = \lim_{l\rightarrow\infty} H(\xi_l|\mathbf{Y}^l) \tag{3.56}$$

Using (3.54), (3.55) and (3.56)

$$\frac{1}{n}H(\mathbf{Y}^n) \to \lim_{l\to\infty}\left(H(Y_l|\mathbf{Y}^{l-1}) - H(\xi_l|\mathbf{Y}^{l-1}) + \frac{1}{l}H(\xi^l)\right) \tag{3.57}$$

Hence

$$\begin{aligned}\frac{1}{n}H(\mathbf{X}^n|\mathbf{Y}^n) \to & R - \lim_{l\to\infty} H(Y_l|\mathbf{Y}^{l-1}) + \lim_{l\to\infty} H(\xi_l|\mathbf{Y}^{l-1}) \\ & + \lim_{l\to\infty}\frac{1}{l}H(\xi^l) + \lim_{l\to\infty;\, p(\mathbf{X}^l)=p_c(\mathbf{X}^l)} H(Y_n|\mathbf{Y}^{n-1}) - C\end{aligned} \tag{3.58}$$

Conjecture 3.9. *The following holds for any symmetric FSMC*

$$\lim_{l\to\infty;\, p(\mathbf{X}^l)=p_c(\mathbf{X}^l)} H(Y_l|\mathbf{Y}^{l-1}) - 1 \geq \lim_{l\to\infty} H(Y_l|\mathbf{Y}^{l-1}) - \lim_{l\to\infty} H(\xi_l|\mathbf{Y}^{l-1}) \tag{3.59}$$

We prove the conjecture for the restricted case of MBIOS channels.

Proof for MBIOS channels:

For memoryless symmetric channels, $H(\mathbf{Y}^n)$, and hence $H(Y_l|\mathbf{Y}^{l-1})$ is maximized for uniform input distribution. Thus, for MBIOS channels

$$\lim_{l\to\infty;\, p(\mathbf{X}^l)=p_c(\mathbf{X}^l)} H(Y_l|\mathbf{Y}^{l-1}) - 1 = [H(Y_l) - H(\xi_l)]|_{p(X)=Unif\{0,1\}} \tag{3.60}$$

Therefore it suffices to show that the LHS term in (3.59) is maximized at uniform i.i.d. input distribution.

Assume that input distribution is $P(X=0)=\alpha,\ P(X=1)=1-\alpha$. We prove that the expression $H(Y_l) - H(\xi_l)$ is concave in α. Note that both $H(Y_l)$ and $H(\xi_l)$ are concave in α.

$$Y_l \sim f_Y(y) = \alpha f(y) + (1-\alpha)f(-y) \tag{3.61}$$

and hence,

$$H(Y) = -\int_{\infty}^{\infty}\left(\alpha f(y) + (1-\alpha)f(-y)\right)\log\left(\alpha f(y) + (1-\alpha)f(-y)\right)dy \tag{3.62}$$

Differentiating $H(Y)$ twice with respect to α

$$\begin{aligned}\frac{d^2H(Y)}{d\alpha^2} &= -\int_{-\infty}^{\infty}\frac{(f(y)-f(-y))^2}{\alpha f(y) + (1-\alpha)f(-y)}dy \\ &= -\left[\int_{\xi=1}\frac{(f(y)-f(-y))^2}{\alpha f(y) + (1-\alpha)f(-y)}dy + \int_{\xi=0}\frac{(f(y)-f(-y))^2}{\alpha f(y) + (1-\alpha)f(-y)}dy\right]\end{aligned} \tag{3.63}$$

Similarly, differentiating $H(\xi)$ twice with respect to α

$$\frac{d^2H(\xi)}{d\alpha^2} = -\left[\frac{\left(\int_{\xi=1} f(y)dy - \int_{\xi=1} f(-y)dy\right)^2}{\alpha\int_{\xi=1} + (1-\alpha)\int_{\xi=1} f(-y)dy} + \frac{\left(\int_{\xi=0} f(y)dy - \int_{\xi=0} f(-y)dy\right)^2}{\alpha\int_{\xi=0} + (1-\alpha)\int_{\xi=0} f(-y)dy}\right]$$

Consider the first terms in (3.63) and (3.64). Applying Cauchy-Schwartz inequality on $\frac{f(y)-f(-y)}{\sqrt{\alpha f(y)+(1-\alpha)f(-y)}}$ and $\sqrt{\alpha f(y) + (1-\alpha)f(-y)}$

$$\int_{\xi=1}\left(\frac{f(y)-f(-y)}{\sqrt{\alpha f(y)+(1-\alpha)f(-y)}}\right)^2 dy \times \int_{\xi=1}\left(\sqrt{\alpha f(y)+(1-\alpha)f(-y)}\right)^2 dy \geq \left(\int_{\xi=1}(f(y)-f(-y))dy\right)^2 \tag{3.64}$$

Using Cauchy-Schwartz inequality for the second terms as well

$$\frac{d^2(H(Y_l) - H(\xi_l))}{d\alpha^2} \leq 0 \tag{3.65}$$

and hence $H(Y_l) - H(\xi_l)$ is concave in α. Since it is also symmetric in α, the maximum is attained at $\alpha = 0.5$ □

If conjecture 3.9 holds, then the expression of lower bound on $\frac{1}{n}H(\mathbf{X}^n|\mathbf{Y}^n)$ reduces to that for simple FSMCs, and it is sufficient to bound $H(\xi^n)$.

Once the expression reduces to that of simple FSMCs, similar *almost-sure* bounds on the rate will hold. For 2-state channels, bounds which hold for all codes can be derived, as for the GE channels. Again, similar lower bounds on parity check density can also be derived.

Even if the conjecture in 3.9 does not hold, similar bounds on rate can be derived. However, these bounds would exceed capacity for large enough density. For example, for regular codes, the following bound holds

$$R \leq 1 - \frac{\lim_{l\to\infty;\, p(\mathbf{X}^n)=p_c(\mathbf{X}^n)} H(Y_n|\mathbf{Y}^{n-1}) - \lim_{l\to\infty} H(Y_l|\mathbf{Y}^{l-1}) + \lim_{l\to\infty} H(\xi_l|\mathbf{Y}^{l-1}) - C_m}{H(\overline{p_{md}})} \text{ a.s.} \tag{3.66}$$

where $\overline{p_d}$ is the average crossover probability of the channel.

Appendix A: Probability of a parity check equation being satisfied

We first prove Lemma 3.2 in a more general setting. Suppose we have some sequence $\{X_{n_i}\}_{i\geq 1}$ as input to a two state channel (which needn't even be Markov) which behaves as a BSC in each state. Let $\{Z_{n_i}\}_{i\geq 1}$ be the error sequence, and we want to calculate $Pr(S=0) = Pr(\sum_{i=1}^{d} Z_{n_i} = 0)$, where the addition is modulo two. Note that here we do not assume Markov modeling of the state space. (In Appendix B, we have a non-homogeneous Markov chain, and we use this result). We prove the following

$$Pr(\sum_{i=1}^{d} Z_{n_i} = 0) = \frac{1}{2} + \frac{1}{2} \sum_{r_1,r_2,\ldots,r_m} \Pi_{i=0}^{m} (1-2\eta_i)^{r_i} Pr(r_1, r_2, \ldots, r_m) \tag{3.67}$$

where $Pr(r_1, r_2, \ldots, r_m)$ denotes the probability of making r_i visits to state i in d steps, that is, $\sum_i r_i = d$.

For 2-state channels (in particular for GE channels), the above result reduces to

$$Pr(\sum_{i=1}^{d} Z_{n_i} = 0) = \frac{1}{2} + \frac{1}{2} \sum_{k=0}^{d} (1-2\eta_G)^k (1-2\eta_B)^{d-k} Pr(N_G = k) \tag{3.68}$$

where $Pr(N_G = k)$ is the probability of making k visits to the state G in the given d time instants.

Remark 3.1. *Note that in* (3.68)*, the probability is greater than* 0.5 *as long as* $\eta_i < 0.5 \; \forall \; i$ *(ensuring that the second term remains positive). This fact has been used frequently in the paper.*

Proof *Of Lemma 3.2* The proof is by induction on m. For $m = 1$, the case reduces to that for a BSC. As shown in [Gal60], the probability of even errors in d channel uses of a BSC is given by $\frac{1+(1+2\eta)^d}{2}$, where η is the crossover probability of the BSC. It is easy to verify that the expression given in (3.67) reduces to this.

We now assume that the result is true for $m-1$ states, and prove the result for m state systems.

$$\begin{aligned} Pr(\sum_{i=1}^{d} Z_{n_i} = 0) &= Pr\Big(\text{even errors in locations corresponding to } \sum_{i=1}^{d} Z_{n_i} = 0\Big) \\ &= Pr\Big(\text{even errors in locations corresponding to first } m-1 \text{ states;} \\ &\quad \text{even errors in locations corresponding to } m^{th} \text{ state}\Big) \\ &\quad + Pr\Big(\text{odd errors in locations corresponding to first } m-1 \text{ states;} \\ &\quad \text{odd errors in locations corresponding to } m^{th} \text{ state}\Big) \end{aligned} \tag{3.69}$$

We now investigate the first term in (3.69)

$$\begin{aligned} &Pr\Big(\text{even errors in locations corresponding to first } m-1 \text{ states;} \\ &\qquad \text{even errors in locations corresponding to } m^{th} \text{ state}\Big) \end{aligned}$$

$$\begin{aligned} &= \sum_{r_m=0}^{d} Pr\Big(\text{even errors in locations corresponding to first } m-1 \text{ states;} \\ &\quad \text{even errors in locations corresponding to } m^{th} \text{ state}|r_m\Big) Pr(r_m) \\ &= \sum_{r_m=0}^{d} Pr(\text{even errors in locations corresponding to first } m-1 \text{ states}|r_m) \\ &\quad \times Pr(\text{even errors in locations corresponding to } m^{th} \text{ state}|r_m) Pr(r_m) \\ &= \sum_{r_m=0}^{d} \frac{1}{2}\left(1 + \sum_{r_1,r_2,\ldots,r_{m-1}} \Pi_{i=1}^{m-1}(1-2\eta_i)^{r_i} Pr(r_1, r_2, \ldots, r_{m-1}|r_m)\right) \\ &\quad \times \frac{1+(1-2\eta_m)^{r_m}}{2} Pr(r_m) \\ &= \frac{1}{4}\Bigg(1 + (1-2\eta_m)^{r_m} + \sum_{r_1,r_2,\ldots,r_m} \Pi_{i=1}^{m-1}(1-2\eta_i)^{r_i} Pr(r_1, \ldots, r_m) \\ &\quad + \sum_{r_1,\ldots,r_m} \Pi_{i=1}^{m}(1-2\eta_i)^{r_i} Pr(r_1, \ldots, r_m)\Bigg) \end{aligned} \tag{3.70}$$

Doing similar analysis on second term of (3.69), and adding the result to (3.70), we arrive at

(3.67) (the middle two terms cancel out). □

Now we prove that for GE channels, as $d \to \infty$, $Pr(S = 0) \to \frac{1}{2}$. Note that

$$\frac{1}{2} \leq Pr(\sum_{i=1}^{d} Z_{n_i} = 0) = \frac{1}{2} + \frac{1}{2}\sum_{k=0}^{d}(1-2\eta_G)^k(1-2\eta_B)^{d-k}Pr(N_G = k) \tag{3.71}$$

$$\leq \frac{1}{2} + \frac{1}{2}\sum_{k=0}^{d}(1-2\eta_G)^d Pr(N_G = k) = \frac{1}{2} + \frac{1}{2}(1-2\eta_G)^d \overset{d\to\infty}{\longrightarrow} \frac{1}{2} \quad \square$$

This means that regardless of the gap between 1's, as $d \to \infty$, the bounds on the rate in 3.2.3 approach capacity. Note that the same proof works for FSMCs as well.

Appendix B: Increase in gap increases syndrome entropy for non-inverting and non-oscillating GE channels

First we prove that

$$Pr(\sum_{i=1}^{d} Z_{n_i} = 0 | s_{n_d} = G) > Pr(\sum_{i=1}^{d} Z_{n_i} = 0 | s_{n_d} = B) \tag{3.72}$$

The proof is by induction on d. The result is trivially true for $d = 1$ (since $1 - \eta_G > 1 - \eta_B$).

Now, we prove the result for $d = k$ assuming that the result is true for $d = k - 1$. Let $t = n_k - n_{k-1}$ denote the gap between the k^{th} and $(k-1)^{th}$ 1's.

Define

$$\begin{aligned} b_t &\triangleq \frac{b - b(1-g-b)^t}{g+b} \\ g_t &\triangleq \frac{g - g(1-g-b)^t}{g+b} \end{aligned} \tag{3.73}$$

Using $t-$step transition probability matrix for a two state Markov chain

$$P^t = \begin{bmatrix} \frac{g+b(1-g-b)^t}{g+b} & \frac{b-b(1-g-b)^t}{g+b} \\ \frac{g-g(1-g-b)^t}{g+b} & \frac{b+g(1-g-b)^t}{g+b} \end{bmatrix} = \begin{bmatrix} 1-b_t & b_t \\ g_t & 1-g_t \end{bmatrix} \tag{3.74}$$

Where P is the single step transition probability matrix. Now, given s_{n_k}, Z_{n_k} is independent

of Z_{n_i} (for $i \neq k$). Thus,

$$\begin{aligned}
Pr(\sum_{i=1}^{k} Z_{n_i} = 0 | s_{n_k} = G) &= Pr(\sum_{i=1}^{k-1} Z_{n_i} = 0 | s_{n_k} = G) Pr(Z_{n_k} = 0 | s_{n_k} = G) \\
&\quad + Pr(\sum_{i=1}^{k-1} Z_{n_i} = 1 | s_{n_k} = G) Pr(Z_{n_k} = 1 | s_{n_k} = G) \\
&= \left[Pr(\sum_{i=1}^{k-1} Z_{n_i} = 0 | s_{n_{k-1}} = G)(1 - b_t) + Pr(\sum_{i=1}^{k-1} Z_{n_i} = 0 | s_{n_{k-1}} = B) b_t \right] (1 - \eta_G) \\
&\quad + \left[Pr(\sum_{i=1}^{k-1} Z_{n_i} = 1 | s_{n_{k-1}} = G)(1 - b_t) + Pr(\sum_{i=1}^{k-1} Z_{n_i} = 1 | s_{n_{k-1}} = B) b_t \right] \eta_G \quad (3.75)
\end{aligned}$$

Similarly, for conditioning on B we get:

$$\begin{aligned}
&Pr(\sum_{i=1}^{k} Z_{n_i} = 0 | s_{n_k} = B) \\
&\quad = \left[Pr(\sum_{i=1}^{k-1} Z_{n_i} = 0 | s_{n_{k-1}} = B)(1 - g_t) + Pr(\sum_{i=1}^{k-1} Z_{n_i} = 0 | s_{n_{k-1}} = G) g_t \right] (1 - \eta_B) \\
&\quad\quad + \left[Pr(\sum_{i=1}^{k-1} Z_{n_i} = 1 | s_{n_{k-1}} = B)(1 - g_t) + Pr(\sum_{i=1}^{k-1} Z_{n_i} = 1 | s_{n_{k-1}} = G) g_t \right] \eta_B \quad (3.76)
\end{aligned}$$

Now observe the terms in first square brackets in (3.75) and (3.76), which are expressions of probability of event $\sum_{i=1}^{k-1} Z_{n_i} = 0$ given s_{n_k} (G or B). Both the terms are of the form

$$f(a) = Pr(\sum_{i=1}^{k-1} Z_{n_i} = 0 | s_{n_{k-1}} = G) a + Pr(\sum_{i=1}^{k-1} Z_{n_i} = 0 | s_{n_{k-1}} = B)(1 - a)$$

where $a = 1 - b_t$ in (3.75) and $a = g_t$ in (3.76). Now,

$$\begin{aligned}
f(a) = a\big(&Pr(\sum_{i=1}^{k-1} Z_{n_i} = 0 | s_{n_{k-1}} = G) - Pr(\sum_{i=1}^{k-1} Z_{n_i} = 0 | s_{n_{k-1}} = B) \big) \\
&+ (1 - Pr(\sum_{i=1}^{k-1} Z_{n_i} = 0 | s_{n_{k-1}} = B))
\end{aligned}$$

which is an increasing function of a, since by induction assumption

$$Pr(\sum_{i=1}^{k-1} Z_{n_i} = 0 | s_{n_{k-1}} = G) > Pr(\sum_{i=1}^{k-1} Z_{n_i} = 0 | s_{n_{k-1}} = B) \quad (3.77)$$

We want to conclude that term under consideration is greater in (3.75). From (3.73), we can see that $1 - g_t > b_t$ for any non-oscillatory GE channel, and therefore, term a is greater in (3.75) and hence the term in first square brackets (which is under consideration here) is also greater in (3.75). Also it can be seen that in both (3.75) and (3.76), this term is greater than 0.5 (This follows since the channel is non-inverting, see Appendix A).

Now, the probability expressions in (3.75) and (3.76) can be written as $h(x, y) = xy + (1-x)(1-y)$, where, x takes the value of term in the first square brackets in each expression. Now, $h(x, y) = x(2y-1) - y + 1$, which is an increasing function of x if $y > 0.5$ and increasing function of y if $x > 0.5$. In our case, both $x > 0.5$ and $y > 0.5$, thus $h(x, y)$ is an increasing function of both x and y. It can also be seen that both x and y are larger in (3.75) and hence (3.72) follows. □

Increase in the gap between 1's increases entropy

Let $S = \sum_{i=1}^{d} Z_{n_i} = S_1 + S_2$, where $S_1 = \sum_{i=1}^{d_1} Z_{n_i}$and $S_2 = \sum_{i=d_1+1}^{d} Z_{n_i}$. We prove that if the gap $r \triangleq n_{d_1+1} - n_{d_1}$ increases, corresponding entropy $H(S)$ also increases. To prove this, it is sufficient to prove that $Pr(S_1 + S_2 = 0)$ decreases as r increases (since $Pr(S_1 + S_2 = 0) > 0.5$).

$$Pr(S_1 + S_2 = 0) = Pr(S_1 = 0; S_2 = 0) + Pr(S_1 = 1; S_2 = 1) \tag{3.78}$$

Note that

$$\begin{aligned} Pr(S_1 = 1; S_2 = 1) &= Pr(S_2 = 1) - Pr(S_1 = 0; S_2 = 1) \\ &= Pr(S_2 = 1) - Pr(S_1 = 0) + Pr(S_1 = 0; S_2 = 0) \end{aligned} \tag{3.79}$$

Since the terms $Pr(S_2 = 1)$ and $Pr(S_1 = 0)$ are independent of r, to prove $Pr(S = 0)$ decreases as r increases, it is sufficient to prove (from (3.78) and (3.79)) that $Pr(S_1 = 0; S_2 = 0)$ decreases as r increases.

We now prove that $Pr(S_1 = 0 | S_2 = 0)$ decreases as r increases (which suffices, as

$Pr(S_2 = 0)$ is independent of r).

$$\begin{aligned}
Pr(S_1 = 0|S_2 = 0) &= Pr(S_1 = 0|s_{n_{d_1+1}} = G)Pr(s_{n_{d_1+1}} = G|S_2 = 0) \\
&\quad + Pr(S_1 = 0|s_{n_{d_1+1}} = B)Pr(s_{n_{d_1+1}} = B|S_2 = 0) \\
&= \Big[Pr(S_1 = 0|s_{n_{d_1}} = G)Pr(s_{n_{d_1}} = G|s_{n_{d_1+1}} = G) \\
&\quad + Pr(S_1 = 0|s_{n_{d_1}} = B)Pr(s_{n_{d_1}} = B|s_{n_{d_1+1}} = G)\Big] \\
&\quad \times Pr(s_{n_{d_1+1}} = G|S_2 = 0) \\
&\quad + \Big[Pr(S_1 = 0|s_{n_{d_1}} = G)Pr(s_{n_{d_1}} = G|s_{n_{d_1+1}} = B) \\
&\quad + Pr(S_1 = 0|s_{n_{d_1}} = B)Pr(s_{n_{d_1}} = B|s_{n_{d_1+1}} = B)\Big] \\
&\quad \times Pr(s_{n_{d+1}} = B|S_2 = 0)
\end{aligned}$$

Thus,

$$\begin{aligned}
&Pr(S_1 = 0|S_2 = 0) \\
&= \left[Pr(S_1 = 0|s_{n_{d_1}} = G) \times \frac{g + b(1 - g - b)^r}{g + b} + Pr(S_1 = 0|s_{n_{d_1}} = B) \times \frac{b - b(1 - g - b)^r}{g + b}\right] \\
&\qquad \times Pr(s_{n_{d_1+1}} = G|S_2 = 0) \\
&+ \left[Pr(S_1 = 0|s_{n_{d_1}} = G) \times \frac{g - g(1 - g - b)^r}{g + b} + Pr(S_1 = 0|s_{n_{d_1}} = B) \times \frac{b + g(1 - g - b)^r}{g + b}\right] \\
&\qquad \times Pr(s_{n_{d_1+1}} = B|S_2 = 0) \\
&= C_1 + \frac{(1 - g - b)^r}{g + b} \times \Big[Pr(S_1 = 0|s_{n_{d_1}} = G) - Pr(S_1 = 0|s_{n_{d_1}} = B)\Big] \\
&\qquad \times \Big[bPr(s_{n_{d_1+1}} = G|S_2 = 0) - gPr(s_{n_{d+1}} = B|S_2 = 0)\Big]
\end{aligned}$$

where C_1 is a constant independent of r. It is easy to see that the two terms in product with $\frac{(1-g-b)^r}{g+b}$ are positive, and since channel is non-oscillatory, $(1-g-b) > 0$, so $Pr(S_1 = 0|S_2 = 0)$ decreases with increase in r. (To see that the second term in product with $\frac{(1-g-b)^r}{g+b}$ is positive,

notice that $Pr(s_{n_{d_1+1}} = G|S_2 = 0) = \frac{Pr(S_2=0|s_{n_{d_1+1}}=G)}{Pr(S_2=0)} \times \frac{g}{g+b}$, and a similar expression holds for $Pr(s_{n_{d_1+1}} = B|S_2 = 0)$. Now use (3.72)). Hence $Pr(S = 0)$ decreases with increase in r. Since $Pr(S = 0) > 0.5$, the entropy $H(S)$ increases as r increases.

Appendix C: Concentration Theorem for LDPC codes over Markov channels

In this section, we prove Theorem 1, the concentration theorem for BP decoding of LDPC codes over Markov channels. For a given graph of decoding LDPC codes over Markov channel (see Fig 3.2) there can be a number of possible message passing schedules. We prove the result for the scheduling proposed by Eckford et al in [Eck04][EKP03][5]. It is clear from the proof that the theorem would hold, albeit with different constants, for other scheduling choices as well.

The proof follows closely the proofs of concentration theorem given in [KMM03] (for ISI channels) and [RU01b] (for memoryless channels). [Eck04] indicated this for future work. In [KMM03], where concentration theorem is proved for ISI channels, the factor graph is similar to the factor graph for Markov channels. For Markov channels, unlike for the ISI channels, the noise is independent of the input. Thus the proof in [KMM03] emphasizes on the problem of input dependent noise sequence. Here we emphasize upon the application of Azuma's inequality for proving the concentration theorem for Markov channels.

We choose a code randomly from $C^n(c, d)$. Following [RU01b], we first construct a Doob edge-and-noise revealing martingale process. In the first nc steps of the process, one-by-one, the edges of the graph are revealed (in any order). In the last n steps, the received value

[5]The nodes are divided into set of factor nodes (parity checks and channel factors), and the set of variable nodes (symbol variables and channel states). In a full iteration, Sum product Algorithm (SPA) is calculated first at each factor node, and messages sent to appropriate variable nodes, and then SPA is calculated at each variable node, and messages sent to factor nodes. For further details, see [Eck04]

corresponding to each variable node is revealed. Let Z denote the number of messages in error at the end of l^{th} iteration. Denote a graph G with the received vector R as the pair (G,R). Define $N \stackrel{\Delta}{=} n(c+1)$. Define, for $i = 1, 2, \ldots, N$

$$Z_i(G,R) \stackrel{\Delta}{=} E[Z(G',R')|(G',R') =_i (G,R)] \tag{3.80}$$

where $(G',R') =_i (G,R)$ denotes all graphs (G',R') which are the same upto the i^{th} revealing. Then $Z_i's$ form a Doob's martingale process. Note that $Z_N = Z$ and $Z_0 = E[Z]$, where the expectation is over all graphs in the ensemble $C^n(c,d)$ and all values of R. Also note that total number of messages passed in LDPC subgraph is nc.

To apply Azuma's inequality (see appendix of [RU01b]), we first bound the number of neighborhoods in which an edge (connecting a variable node and a parity check node) can lie after some fixed l iterations, which is the same as the number of edges in a neighborhood after l iterations. Consider the graph in Fig 3.2 b). It is easy to see that for $d \geq 2$ the number of edges in the graph under consideration is less than the number of edges for a bipartite graph with variable node degree $c+1$, and check node degree d. The channel factor node results in a smaller increase in the number of edges in the graph than the increase if another variable node is added. An upper bound for a (c,d) graph for this number of edges is found in [RU01b], and is $2c^l d^l$. Hence the upper bound we get here is $2(c+1)^l d^l$.

As shown in [RU01b], it now follows that at most $8(c+1)^l d^l$ neighborhoods can be affected by exchange of end-points of two edges for the first nc steps. Similarly, for the last n steps, the change in channel output symbol can affect at most $2(c+1)^l d^l$ neighborhoods. Thus, as in [RU01b], we get the following

$$|Z_{i+1}(G,R) - Z_i(G,R)| \leq \alpha_i \leq 8(c+1)^l d^l \quad i = 0, 1, \ldots, N-1 \tag{3.81}$$

By an application of Azuma's inequality,

$$Pr\left(\frac{|Z - E[Z]|}{nc} > \epsilon\right) = Pr\left(|Z_N - Z_0| > nc\epsilon\right) \leq 2e^{-\frac{n^2c^2\epsilon^2}{2\sum_{i=0}^{N-1}\alpha_i^2}} = 2e^{-\frac{nc^2\epsilon^2}{128(c+1)^2(l-1)d^{2l}}} \tag{3.82}$$

where $\alpha_i = 8(c+1)^l d^l$. Thus this probability is bounded above exponentially in n (and hence converges to 0 as $n \to \infty$), which proves the concentration theorem.

The above proof is for regular LDPC codes, but the same proof goes through for irregular LDPC codes (with bounded maximum degrees) with maximum left and right degrees in place of c and d. It is evident from the proof that the concentration theorem holds, albeit with different constants, for other scheduling choices as well.

Chapter 4

Bounds on Stopping Sets

In this chapter we elucidate the importance of stopping sets in decoding of LDPC codes. We also introduce the concept of *minimal* stopping sets, explain their utility and give bounds on their number.

4.1 Conclusions from distributions of stopping sets

4.1.1 Capacity achieving sequences

Distributions of stopping sets are used to find the probability of error for finite length LDPC codes [DPT$^+$02][OVZ05] over the binary erasure channel. Interestingly, in [OVZ05], using the distributions of stopping sets, the authors conclude that block probability of error (P_B) does not converge to 0 for LDPC codes with $\lambda'(0)\rho'(1) > 0$. In fact, P_B converges to $1 - \sqrt{1 - \lambda'(0)\rho'(1)\epsilon}$. However, in all the known capacity achieving sequences of LDPC codes ([Sho00]), $\lambda'(0)\rho'(1) \to \frac{1}{\epsilon}$. Also, if the flatness condition of [Sho00] holds, then for a capacity achieving sequence of LDPC codes, $\lambda'(0)\rho'(1) \to \frac{1}{\epsilon}$, and hence P_B converges to 1.

Thus for erasure channels, though we can construct sequences for which probability of bit error (P_b) converges to 0, and rate converges to capacity [Sho00], no sequences are known for which P_B converges to 0, as the rate converges to capacity. Shannon [Sha48], in his noisy

channel coding theorem, proved that we can construct capacity achieving sequences in which P_B converges to 0. Hence in Shannon sense, the goal of coding theory has not yet been achieved for *any* channel under linear time decoding.

In many applications, for example file transfer, it is important that P_B, and not just P_b, converge to 0. A single bit error may cause the file to crash, even though this may correspond to a low bit error probability.

The best known LDPC codes (in the sense of maximum rate for probability of error converging to 0), hence are not suitable for certain applications where P_B should converge to 0.

4.1.2 Utility of stopping set distributions for general channels

Stopping sets determine the performance of LDPC codes over erasure channels. Furthermore, even in the performance over general (non-erasure) channels, stopping sets play a vital role. For general channels, channel output values with poor observation reliability are analogous to erasures, and hence stopping sets are the parts of the graphs where lack of reliable information would lead to errors with a high probability. This concept was used to design LDPC codes with low error floors in [TJVW03], where the authors devised an algorithm to reduce the number of small stopping sets.

4.2 Minimal stopping sets

We define Minimal stopping sets as follows:

Definition 4.1 (Minimal Stopping Set). *A stopping set $\mathcal{S}(\neq \phi)$ is said to be a minimal stopping set if there does not exist any non-empty stopping set $\mathcal{S}_1 \subsetneqq \mathcal{S}$.*

The following Lemma gives the importance of minimal stopping sets in decoding of LDPC codes over BECs.

Lemma 4.1. *Decoding of an LDPC code over BEC is successful if and only if at least one bit is received in every minimal stopping set.*

Proof The proof is simple. Suppose the decoding is successful, then if some minimal stopping set has not received any bit, then (using Lemma 2.1) the decoder will not be able to find out the values in that stopping set, and we arrive at a contradiction.

Conversely, if the decoding is unsuccessful, it stops at a stopping set in which no bit has been received. If the stopping set is not minimal, it contains at least one minimal stopping set. But every minimal stopping set has received at least one bit. Hence we again arrive at a contradiction. □

Note that the above Lemma trivially holds for all stopping sets. The interesting part is that the Lemma also holds for minimal stopping sets, which are a small subset of all the stopping sets.

4.3 Bounds on the number of minimal stopping sets

Orlitsky et al [OVZ05] proved that for the smallest stopping set is linear in blocklength n if $\lambda_2 = 0$. They found a high probability lower bound $\alpha^* n$ ($\alpha^* > 0$) on the length of smallest stopping set.

A non-minimal stopping set can be formed by union of two stopping sets, or by adding to a stopping set a few variable nodes which do not form a stopping set in themselves, but form a stopping set when added to the existing stopping set.

If unions of all the minimal stopping sets were disjoint, the total number of stopping sets would be exponential in the number of minimal stopping sets. However, as the following theorem shows, this is not true for codes with $\lambda_2 = 0$, i.e. for codes which have smallest stopping set of size linear in n.

Theorem 4.2. *Let $K(n)$ denote the number of minimal stopping sets for an LDPC code $\mathcal{C}_n \in C^n(\lambda, \rho)$. Let $\alpha^* n$ be the size of the smallest non-empty minimal stopping set, and βn*

be the size of largest non-empty minimal stopping set. Then for n large enough

$$\left(R-\frac{1}{n}\right) ln\left(\frac{1}{1-\alpha^*}\right) \leq \frac{1}{n} ln\left(K(n)\right) \leq (1-\delta) ln\left(\frac{1}{1-\beta}\right) \tag{4.1}$$

That is, the number of minimal stopping sets is exponential in n . Since the total number of subsets of n size $\mathcal{V}$ is 2^n (which is only exponential, and not superexponential in n), the total number of stopping sets is only polynomial, and not exponential in the number of minimal stopping sets.

Proof We first prove the *left* inequality. The idea used is this: since we can decode successfully after there is 1 bit in each minimal stopping set, we must be able to accommodate nR bits for a rate R code without completion of decoding. That is, decoding should not stop for less than nR bits. If it does, then less than nR bits suffice to specify the codeword, in which case the rate is less than R, which is a contradiction.

Including repetitions, the number of bits $K(n)$ minimal stopping sets cover is at least $K(n) \times \alpha^* n$, since each minimal stopping set is of size at least $\alpha^* n$.

Therefore, the average number of minimal stopping sets each bit lies in, say γ

$$\gamma \geq \frac{K(n) \times \alpha^* n}{n} = \alpha^* K(n) \tag{4.2}$$

Now choose one bit which lies in at least γ minimal stopping sets. Thus the number of stopping sets left is at most $K(n) - \gamma = K(n)(1-\alpha^*)$. Finding the average again, and again choosing a bit which has above average minimal stopping sets, and continuing, we get that in r steps, the number of minimal stopping sets left is at most $\leq K(n)(1-\alpha^*)^r$.

For $r = nR - 1$, at least 1 stopping set must remain. Hence

$$\begin{aligned} K(n)(1-\alpha^*)^{nR-1} &> 1 \\ \Rightarrow \left(R-\frac{1}{n}\right) \ln\left(\frac{1}{1-\alpha^*}\right) &\leq \frac{1}{n} \ln\left(K(n)\right) \end{aligned} \tag{4.3}$$

which is the left inequality in 4.1.

Next we prove the *right* inequality in 4.1. The idea is used is the following: if δ is less than the threshold of iterative decoding over the BEC, then for n large enough, any set of $n(1-\delta)$ bit suffice with a high probability for successful decoding.

Choose any $n(1-\delta)$ bits. If $\mathcal{C} \in C^n(\lambda, \rho)$, then all permutations of $\mathcal{C}$ also belong to $C^n(\lambda, \rho)$. Hence any choice of k bits from the $\binom{n}{k}$ possible choices will leave the same number of stopping sets summed over all the codes.

Since maximum size of any minimal stopping set is less than βn, average number of stopping sets each bit lies in is no greater than $k\beta(nc)!$, since $(nc)!$ is the total number of codes in $C^n(\lambda, \rho)$. Note that this is the number of stopping sets over all codes in $C^n(\lambda, \rho)$

Similar to proof of left inequality, after r steps, at least $(na)!K(n)(1-\beta)^r$ minimum stopping sets are left. That is, at least $(nc)!K(n)(1-\beta)^r$ bits are left undecoded. If the average probability of block error is $E[P_B^n]$, an upper bound on the number of undecoded bits is $(nc)!E[P_B^n] \times n$ (assuming a worst case of all bits in error, in case of a block error) when $1-\delta$ fraction of the bits are received.

Thus,

$$(nc)!K(n)(1-\beta)^{n(1-\delta)} < (nc)!E[P_B^n] \times n \tag{4.4}$$

It is proved in [OVZ05, Theorem 16] that P_B^n goes to 0 polynomially in n, if there are no degree 2 nodes. Precisely,

$$E[P_B^n] = \Theta\left(\frac{1}{n^{\lceil \frac{l_m}{2} \rceil}}\right) \tag{4.5}$$

where l_m is the smallest variable node degree. Thus (4.4) becomes

$$K(n)(1-\beta)^{n(1-\delta)} < \frac{nc_1}{n^{\lceil \frac{l_m}{2} \rceil}} < 1 \tag{4.6}$$

Where c_1 is some constant depending on the block error probability. The last inequality holds for n large enough. Taking $\ln(\cdot)$ on both side, we get the desired inequality in (4.1).

It is likely that the number of minimal stopping sets is closer to the lower bound, because

we expect that among all the possible choices of nR bits, there would exist some choice for which successful decoding requires only a few more bits. The upper bound is based on a high probability analysis, and hence would be loose.

4.3.1 Significance of the number of minimal stopping sets

As shown in [TJVW03], to lower the error floor of LDPC codes, it is more efficient to reduce the number of small stopping sets rather than the number of small cycles in the code.

Algorithm in [TJVW03] decreases the number of stopping sets of size smaller than a fixed value, as shown in [RW04]. However, removing a stopping set of small size, which has another non-empty stopping set as a proper subset, would not reduce the block probability of error. A more efficient algorithm to reduce block error probability would only reduce the number of minimal stopping sets of size lesser than a fixed value. If we could design algorithms which reduce the number of minimal stopping sets, these bounds would give an estimate of their complexity.

Chapter 5

Conclusions

In this dissertation, we generalized the bounds on the rate of LDPC codes for reliable communication over BSC to FSMCs. We first derived a simple upper bound on the rate for reliable communication over all non-inverting simple FSMCs. Using this bound, we proved that for a sequence of LDPC codes to be capacity achieving over such a simple FSMC, the density $\Delta(\mathbf{H})$ must converge to infinity.

For non-inverting and non-oscillating GE channels, we obtained a tighter upper bound, and showed that this can be further tightened using the knowledge of maximum gap between 1's in the rows of $\mathbf{H}$.

The bound on the syndrome entropy derived for GE channels was used to derive lower bounds on the parity check density for given performance, which is a generalization of results of Sason and Urbanke [SU03]. The tighter upper bound, which uses the knowledge of maximum gap between 1's in the rows of $\mathbf{H}$, leads to tighter lower bounds on the density of regular codes. For irregular codes, this leads to a necessary condition on row weights that has to be satisfied for any irregular code for given performance over a GE channel. These bounds hold for *any* linear code.

For simple FSMCs, we also proved that the tight upper bound derived holds *almost-surely* for any sequence of randomly constructed codes of fixed (λ, ρ). Thus the bound is valid for BP decoding of randomly constructed LDPC codes, and to establish the utility of

this construction method, we proved the concentration theorem for LDPC codes over Markov channels.

We showed that the simple bound on the rate over simple FSMCs derived here can be used to derive *Type I* and *Type II* lower bounds on $\Delta(\mathbf{H})$. Using the derivation of *almost-sure* bound on the rate, the corresponding lower bound for finite length (*Type I*) holds with high probability, and lower bound for reliable communication on the design density (*Type II*) holds *almost-surely*.

We next showed how to extend these bounds to general FSMCs.

Further, we explained the significance of stopping sets and underlined the fact that in Shannon sense, we still do not have capacity achieving sequences for any non-trivial channel with linear time decoding. We also defined and examined minimal stopping sets. We found bounds on the number of minimal stopping sets for given minimum and maximum sizes of minimal stopping set for $\lambda_2 = 0$. The concept of minimal stopping sets can be utilized for constructing LDPC codes which give better performance of block error probabilities. The bounds on the number of minimal stopping sets would then yield the complexity of such constructions.

5.1 Scope for future work

We showed how to generalize the bounds on the rate to general FSMCs. However, the bounds we arrive at may not be the same as those for simple FSMCs. It would be interesting to examine if these bounds are the same as those for simple FSMCs. Also, it would be interesting to see if these bounds extend to other class of channels with memory, particularly ISI channels.

Algorithms for improvement in performance of LDPC codes based on increasing the length of minimal stopping sets would be more efficient than those based on increasing the length of small stopping sets. Therefore the design of of such algorithms is a problem worthy of investigation.

References

[BGT93] C Berrou, A Glavieux, and P Thitimajshima. Near Shannon Limit Error-Correcting Codes and Decoding: Turbo Codes. In *Proceedings of International Conference on Communications*, pages 1064–1070, 1993.

[BKLM02] D Burshtein, M Krivelevich, S Litsyn, and G Miller. Upper bounds on the rate of LDPC codes. *IEEE Transactions on Information Theory*, 48(9), September 2002.

[CJRU01] Sae-Young Chung, G. David Forney Jr., TJ Richardson, and Rudiger Urbanke. On the design of low-density parity-check codes within 0.0045 db of the shannon limit. *IEEE Communication Letters*, 5(2):58–60, Feb 2001.

[DPT+02] C Di, D Proietti, I Emre Telatar, TJ Richardson, and RL Urbanke. Finite-Length Analysis of Low-Density Parity-Check Codes on the Binary Erasure Channel. *IEEE Transactions on Information Theory*, 48(6):1570–1579, June 2002.

[Eck04] AW Eckford. *Low-Density Parity-Check Codes for Gilbert-Elliott and Markov-Modulated Channels*. PhD thesis, University of Toronto, 2004.

[EKP03] AW Eckford, FR Kschischang, and S Pasupathy. Analysis of Low-Density Parity-Check decoding over the Gilbert-Elliott channel. *submitted to IEEE Transactions on Information Theory*, 2003.

[Gal60] RG Gallager. *Low-Density Parity-Check Codes*. PhD thesis, MIT, Cambridge, 1960.

[Gal68] RG Gallager. *Information Theory and Reliable Communication*. Wiley, 1968.

[GV96] AJ Goldsmith and PP Varaiya. Capacity, mutual information, and coding for finite-state markov channels. *IEEE Transactions on Information Theory*, 42(3), May 1996.

[KMM03] A Kavcic, Xiao Ma, and M Mitzenmacher. Binary Intersymbol Interference Channels: Gallager Codes, Density Evolution, and Code Performance Bounds. *IEEE Transactions in Information Theory*, 49(7):1636–1652, July 2003.

[LMSS01] MG Luby, M Mitzenmacher, MA Shokrollahi, and DA Spielman. Improved Low-Density Parity-Check Codes Using Irregular Graphs. *IEEE Transactions on Information Theory*, 47(2):585–598, February 2001.

[Mac99] DJC MacKay. Good error-correcting codes based on very sparse matrices. *IEEE Transactions on Information Theory*, 45:399–431, March 1999.

[MBD89] M Mushkin and I Bar-David. Capacity and Coding for the Gilbert-Elliott Channels. *IEEE Transactions on Information Theory*, 35(6), November 1989.

[OVZ05] A Orlitsky, K Viswanathan, and J Zhang. Stopping Set Distribution of LDPC Code Ensembles. *IEEE Transactions on Information Theory*, 51(3):929–953, March 2005.

[Ped71] PJ Pedler. Occupation times for two state markov chains. *Journal of Applied Probability*, pages 381–390, 1971.

[Ros02] Sheldon M Ross. *Introduction to probability models*. Academic Press, 8 edition, 2002.

[RSU01] TJ Richardson, MA Shokrollahi, and RL Urbanke. Design of Capacity Approaching Irregular Low-Density Parity-Check Codes. *IEEE Transactions on Information Theory*, 47(2):619–637, February 2001.

[RU01a] TJ Richardson and RL Urbanke. Efficient encoding of low-density parity-check codes. *IEEE Transactions on Information Theory*, 47(2):638–656, 2001.

[RU01b] TJ Richardson and RL Urbanke. The Capacity of Low-Density Parity-Check Codes Under Message-Passing Decoding. *IEEE Transactions on Information Theory*, 47(2):599–618, Feb 2001.

[RW04] A Ramamoorthy and RD Wesel. Analysis of an algorithm for LDPC code construction. In *International Symposium on Information Theory (ISIT)*. IEEE, 2004.

[Sha48] CE Shannon. A Mathematical Theory of Communication. Technical report, Bell System Technical Journal, 1948.

[Sho00] Amin Shokrollahi. Capacity achieving sequences. In *Codes, Systems and Graphical Models*, number 123 in IMA Volumes in Mathematics and its Applications, pages 153–166, 2000.

[SU03] Igal Sason and Rudiger Urbanke. Parity-check density versus performance of binary linear block codes over memoryless symmetric channels. *IEEE Transactions on Information Theory*, 49(7):1611–1635, July 2003.

[Tan81] RM Tanner. A Recursive Approach to Low Complexity Codes. *IEEE Transactions on Information Theory*, 27(5):533–547, September 1981.

[TJVW03] T Tian, C Jones, J Villasenor, and RD Wesel. Construction of Irregular LDPC codes with low error floors. In *International Conference on Communications (ICC)*, Anchorage, Alaska, May 2003.