OPTIMAL FINITE ALPHABET SOURCES OVER PARTIAL RESPONSE CHANNELS

A Thesis

by

DEEPAK KUMAR

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

August 2003

Major Subject: Electrical Engineering

OPTIMAL FINITE ALPHABET SOURCES

OVER PARTIAL RESPONSE CHANNELS

A Thesis

by

DEEPAK KUMAR

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Approved as to style and content by:

Scott L. Miller (Co-Chair of Committee)	Krishna R. Narayanan (Co-Chair of Committee)
Jianxin Zhou (Member)	Shankar P. Bhattacharyya (Member)
Chanan Singh (Head of Department)	

August 2003

Major Subject: Electrical Engineering

ABSTRACT

Optimal Finite Alphabet Sources

Over Partial Response Channels. (August 2003)

Deepak Kumar, B. Tech., Indian Institute of Technology, Bombay

Co–Chairs of Advisory Committee: Dr. Scott L. Miller
Dr. Krishna R. Narayanan

We present a serially concatenated coding scheme for partial response channels. The encoder consists of an outer irregular LDPC code and an inner matched spectrum trellis code. These codes are shown to offer considerable improvement over the i.i.d. capacity (> 1 dB) of the channel for low rates (approximately 0.1 bits per channel use). We also present a qualitative argument on the optimality of these codes for low rates. We also formulate a performance index for such codes to predict their performance for low rates. The results have been verified via simulations for the $(1-D)/\sqrt{2}$ and the $(1-D+0.8D^2)/\sqrt{2.64}$ channels. The structure of the encoding/decoding scheme is considerably simpler than the existing scheme to maximize the information rate of encoders over partial response channels.

To my Mom

ACKNOWLEDGMENTS

While working on this thesis, I was immensely helped by Dr. Aleksandar Kavčić and Dr. Xiao Ma (at the Division of Engineering and Applied Sciences, Harvard University). I had a helpful and motivating company of my friends associated with the Wireless Communications Lab. Vivek, Doan, Jun, Pradeep, Janath, Hari and Nitin helped me a lot. I would like to thank Parul for proof reading the report. I would also like to take this opportunity to thank Prof. Mary Saslow, Dr. Scott L. Miller, Dr. Jianxin Zhou, Dr. Erchin Serpedin, Dr. Carl Gagliardi and Dr. Chia-Ren Hu for making learning a wonderful experience at Texas A&M.

Last but not the least I would like to thank my advisors, Dr. Scott L. Miller and Dr. Krishna R. Naryanan for giving me the opportunity to work on this interesting topic. I really appreciate them being so cool, helpful, motivating and professional.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

FIGURE Page

CHAPTER I

INTRODUCTION

Nearly half a century ago, Shannon developed the concept of channel capacity. In his seminal papers he proved the channel coding theorem and its converse. The essence of the channel coding theorem is that reliable communication (with probability of error tending to zero) is possible only if the rate is below the channel capacity. Shannon also provided a non constructive proof for the existence of code book(s) which provide reliable communication for any rate less than the channel capacity. Even though, Shannon's work guaranteed the existence of capacity achieving code book(s), finding them for any given channel has been a very difficult problem attempted by many researchers.

Suppose we wish to transmit data reliably over a given channel. We would like to do so at rates close to (but less than) the capacity. For this we would like to find the capacity of the channel and also an efficient coding/decoding scheme. In this thesis, we consider a class of channels called the ISI (Inter Symbol Interference) or the partial response channels. Many of the channels in practical communication systems and the magnetic recording systems are modeled as partial response channels.

A. Problem Definition

1. Notation

We represent a vector of random variables $[X_i, X_{i+1}, \cdots, X_j]^T$ by X_i^j. Its realization is represented by x_i^j. Vectors and matrices are represented by bold face letters. Unless otherwise mentioned, $P(.)$ and $f(.)$ will represent probability of a realization of a

---

The journal model is *IEEE Transactions on Automatic Control.*

discrete random variable and the probability density function of a continuous random variable respectively.

2. Source/Channel Model

Partial response channels belong to the class of time-invariant indecomposable finite state channels (FSC) as described in [1]. The partial response channels are modeled as a discrete time linear filter $g(D)$. The front end receiver noise is modeled by additive white Gaussian noise (Figure 1).

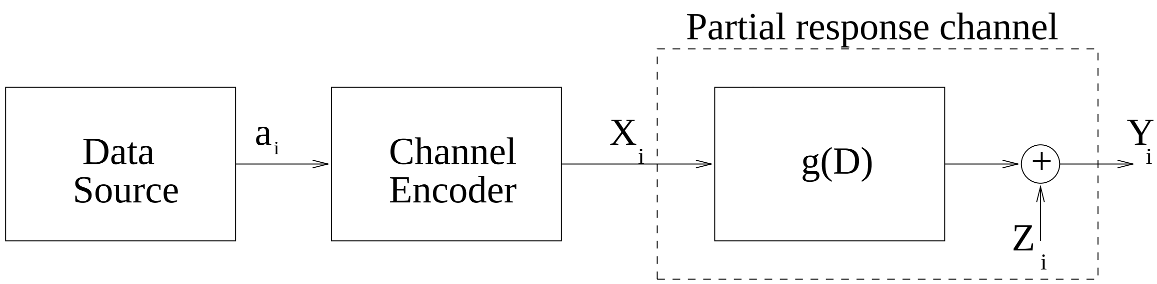

Fig. 1. Source/channel model.

We consider channels with binary input $X_i \in \{+1, -1\}$, and output Y_i given by,

$$Y_i = \sum_{k=0}^{m} g_k X_{i-k} + Z_i, \tag{1.1}$$

where $g_0, g_1, \cdots, g_m$ are the channel taps and Z_i is a zero mean white Gaussian noise process. The variance of the noise is given by $E[Z_i^2] = \sigma^2$. The channel is often represented as a transfer polynomial in terms of the delay element D,

$$g(D) = \sum_{k=0}^{m} g_k D^k. \tag{1.2}$$

The channel encoder maps the binary bit sequence $\{a_i\}$, to the sequence $\{X_i\}$ which is sent over the channel. The channel encoder introduces certain redundancy (and hence has a rate loss) in the binary data sequence before transmitting it over the channel.

3. Problem Statement

In this thesis, we discuss the design of the channel encoder (Figure 1) so that we are able to transmit error free information at the maximum rate possible. The maximum such rate possible is upper bounded by the channel capacity (C) in bits per channel use.

$$C = \lim_{N\to\infty} \sup_{P_{X_1^N}} \frac{1}{N} I(X_1^N; Y_1^N). \tag{1.3}$$

Estimating the capacity and finding the capacity achieving codebook(s) for partial response channels have been open problems for a long time. While some research has been focussed on designing codes which either match the channel spectrum [2, 3] or maximize the free distance between coded sequences at the output of the channel [3, 4], the problem of maximizing the information rate over partial response channels has received much attention from the coding community only recently.

In its generality, the problem of designing capacity achieving encoders is very difficult. Thus we will be imposing certain simplifying constraints (usually on the encoder structure) and look for implementable solutions to the above problem. We will discuss more about these constraints in Chapters IV and V.

B. Structure

Chapter II reviews certain theorems/algorithms which are extensively used throughout the thesis. In Chapter III we present a survey of the various schemes to find/estimate the capacity of a channel. We also discuss certain bounds to the channel capacity. In Chapter IV, we review the design of encoders which tend to maximize the information rate over a channel. However this leads to very complex (practically unimplementable) encoders. Thus in Chapter V we propose certain simple encoders (based on matching channel spectrum). We demonstrate their optimality using simulation results and also present a qualitative argument on their asymptotic optimality (for low information rates). In Chapter VI, we summarize the contributions of this thesis.

CHAPTER II

TECHNICAL BACKGROUND

A. The Shannon-McMillan-Breiman Theorem

The information rate over a partial response channel (Figure 1) is defined as

$$\begin{aligned} I(X;Y) &= \lim_{N\to\infty} \frac{1}{N} I(X_1^N;Y_1^N) \\ &= \lim_{N\to\infty} \frac{1}{N}[h(Y_1^N) - h(Y_1^N|X_1^N)] \\ &= \lim_{N\to\infty} \frac{1}{N} h(Y_1^N) - h(Z) \end{aligned} \tag{2.1}$$

For a white Gaussian noise process Z, with variance σ^2, the differential entropy $h(Z)$ is $\frac{1}{2}\log_2(2\pi e\sigma^2)$. Thus we can get an estimate of the information rate over a channel if we have an estimate of $\lim_{N\to\infty}\frac{1}{N}h(Y_1^N)$. The Shannon-McMillan-Breiman theorem discussed in section 15.7 of [5], helps us do this for stationary ergodic sources over partial response channels.

The Shannon-McMillan-Breiman Theorem A.1 *If H is the entropy rate of a finite-valued stationary ergodic process $\{Y_1^N\}$, then*

$$-\frac{1}{N}\log_2 P(y_0, y_1, \cdots, y_{N-1}) \to H, \qquad \textit{with probability 1.} \tag{2.2}$$

A stronger version of the Shannon-McMillan-Breiman theorem is its extension for continuous valued random variables Y_1^N. In this case

$$\lim_{N\to\infty} -\frac{1}{N}\log_2 f(y_1^N) = h(Y) \qquad \text{almost surely,} \tag{2.3}$$

where $f(Y_1^N)$ is the probability density function of Y_1^N.

The essence of the Shannon-McMillan-Breiman Theorem is that we can estimate $h(Y)$ by finding the probability of occurrence of an arbitrarily long sequence. We will use this fact for estimating the information rate over the partial response channels.

B. BCJR Algorithm

The Bahl Jelinek Cocke and Raviv (BCJR) algorithm [6] is used to estimate the a posteriori probabilities of the states and transitions (and hence the data bits) of a Markov source observed through a discrete memoryless channel. We will use slight modifications of the original BCJR algorithm to:

1. Calculate $f(y_1^N)$, and hence to get an estimate of the information rate over a channel.

2. Calculate the log likelihood ratios of the bits transmitted

$$L(a_t) = \log \frac{P(a_t = 1|y_1^N)}{P(a_t = 0|y_1^N)}. \tag{2.4}$$

   The log likelihood ratios will be used for LDPC code design in Section D of Chapter V.

3. Calculate the a posteriori probabilities of state transition. These will be used in Kavčić's Algorithm described in the next section.

We will now describe a modified (normalized) version of BCJR algorithm. We will skip the intermediate steps wherever they directly follow from arguments similar to those presented in [6].

Let us assume that we have a finite state Markov source transmitting symbols over an AWGN channel. Let a_1^N be the input data sequence to the source. Let X_1^N be the output of the source and y_1^N, the output of the channel (which we observe). In general a_i can be a binary k-tuple and X_i a binary n-tuple. Thus the encoder will be a rate k/n encoder. Let us assume that the data symbol a_t corresponds to the transition from state S_{t-1} to S_t. In what follows, the variables m and m' will be used to index the states of a Markov source.

Let us define the quantities:

$$\lambda_t(m) = P(S_t = m | y_1^N) \tag{2.5}$$

$$\sigma_t(m', m) = P(S_{t-1} = m', S_t = m | y_1^N) \tag{2.6}$$

The BCJR algorithm just defines the iterative equations to compute the above mentioned probabilities. Before we describe the iterative relations, let us describe how we are going to use the above quantities to compute the log likelihood ratios for the case when a_t are binary.

$$L(a_t) = \log \frac{\sum_{m',m:a_t=1} \sigma_t(m', m)}{\sum_{m',m:a_t=0} \sigma_t(m', m)}. \tag{2.7}$$

We define a few more quantities.

$$\alpha_t(m) = P(S_t = m | y_1^t). \tag{2.8}$$

$$\beta_t(m) = \frac{f(y_{t+1}^N)}{f(y_t^N | y_1^{t-1})}. \tag{2.9}$$

$$\gamma_t(m', m) = f(S_t = m, y_t | S_{t-1} = m'). \tag{2.10}$$

The quantities defined above can be used to calculate $\sigma_t(m', m)$ and $\lambda_t(m)$ by the following equations.

$$\begin{aligned} \sigma_t(m', m) &= \alpha_{t-1}(m')\gamma_t(m', m)\beta_t(m). \\ \lambda_t(m) &= \sum_{m'} \sigma_t(m', m). \end{aligned} \tag{2.11}$$

The quantity $\gamma_t(m', m)$ can be computed by

$$\gamma_t(m', m) = \sum_i P(a_t = i | S_t = m, S_{t-1} = m') f(y_t | S_t = m, S_{t-1} = m') P(S_t = m | S_{t-1} = m'). \tag{2.12}$$

The first term in the above summation is a 1 or a 0. The second term is given by

$$f(y_t | x_t) = \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{||y_t - x_t||^2}{2\sigma^2}}, \tag{2.13}$$

where X_t is the symbol transmitted by the source when a transition from state m' to m took place.

If we encode with the constraints that our Markov source should start and end at the state 0, then we have the following initializations for α and β respectively.

$$\alpha_0(m) = \begin{cases} 1 & m = 0 \\ 0 & m \neq 0 \end{cases}$$

$$\beta_N(m) = \begin{cases} 1 & m = 0 \\ 0 & m \neq 0 \end{cases} \tag{2.14}$$

The *backward recursion* used to compute β is given by

$$\beta_t(m) = \frac{\sum_{m'} \beta_{t+1}(m') \gamma_{t+1}(m, m')}{\sum_{m',m} \alpha_t(m') \gamma_{t+1}(m', m)}. \tag{2.15}$$

The *forward recursion* used to compute α is given by

$$\alpha_t(m) = \frac{\sum_{m'} \alpha_{t-1}(m') \gamma_t(m', m)}{\sum_{m',m} \alpha_{t-1}(m') \gamma_t(m', m)}. \tag{2.16}$$

By equations 2.5-2.16, we have completely described the BCJR algorithm. However, before we move ahead, let us briefly review how we can estimate $h(Y)$ from just the forward recursion equation 2.16 of the BCJR algorithm and the Shannon-McMillan-

Breiman theorem.

$$\begin{aligned} h(Y) &= \lim_{N\to\infty} -\frac{1}{N}\log_2 f(y_1^N) \\ &= \lim_{N\to\infty} -\frac{1}{N}\sum_{i=1}^{N}\log_2 f(y_i|y_1^{i-1}) \\ &= \lim_{N\to\infty} -\frac{1}{N}\sum_{i=1}^{N}\log_2(\sum_{m',m}\alpha_{i-1}(m')\gamma_i(m',m)) \end{aligned} \tag{2.17}$$

Note that the term inside the logarithm is just the normalizing factor in equation 2.16.

C. Kavčić's Algorithm

Kavčić [7] conjectures an algorithm that optimizes the transition probabilities of a Markov source over a memoryless channel to maximize the information rate. Even though the algorithm is not proven it is verified to give desired results for low order Markov sources. The algorithm reduces to an expectation maximization version of the Arimoto-Blahut algorithm [8] in the special case when the input Markov source is memoryless. The algorithm also reduces to the special case of maximizing the entropy of the Markov source (Section 8.3 [9]) when the channel is noiseless. We describe the algorithm here without going in detail over the intermediate steps.

Consider a Markov source with M states. The source is described by a probability transition matrix $\mathbf{P}$.

$$P_{ij} = P(S_{t+1} = j|S_t = i). \tag{2.18}$$

We can derive the steady state probabilities μ_i's for the source by solving $\mathbf{P}^T\mu = \mu$. Also, the symbols which are transmitted on transition between respective states need to be specified. The Markov source may also obey certain constraints like transition between some states is not allowed. Such constraints frequently arise for run length limited sequences [10]. The corresponding entry in the probability transition matrix is 0. The allowable state transitions are often represented by a set τ of state pairs.

$$\tau = \{(i, j) : P_{ij} \neq 0\}. \tag{2.19}$$

An example of a 2 state Markov source with transition probability matrix

$$P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}, \tag{2.20}$$

is given in Figure 2. Here $0 \leq p, q \leq 1$. The mutual information rate between the

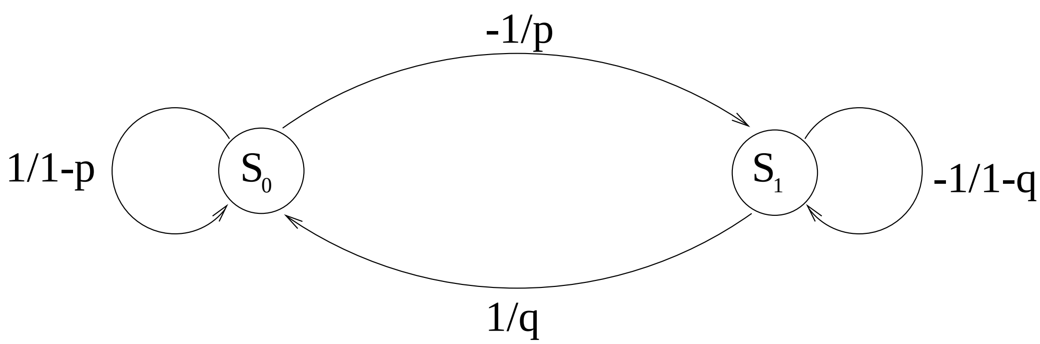

Fig. 2. An example of a 2 state Markov source.

output of the Markov source and the output of the channel can be shown to be

$$I(X;Y) = \sum_{i,j:(i,j)\in\tau} \mu_i P_{ij}[\log \frac{1}{P_{ij}} + T_{ij}], \tag{2.21}$$

where T_{ij} is defined as

$$T_{ij} = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} E[\log \frac{\sigma_t(i,j)^{\frac{\sigma_t(i,j)}{\mu_i P_{ij}}}}{\lambda_{t-1}(i)^{\frac{\lambda_{t-1}(i)}{\mu_i}}}]. \tag{2.22}$$

Note that the terms involved in the above equation are nothing but the a posteriori probabilities given by the BCJR algorithm (equations 2.5 and 2.6). By the law of large numbers, we can accurately estimate T_{ij}, if we perform the BCJR algorithm for sufficiently long trellis realizations.

Now that we have sufficient background, we formally state Kavčić's algorithm for optimizing transition probabilities of a Markov source.

1. *Initialize:* Choose any arbitrary transition probability matrix $\mathbf{P}$ satisfying $P_{ij} = 0$ if $(i, j) \notin \tau$.

2. *Expectation:* For the source defined by P, simulate a long realization over the channel. Use the BCJR algorithm and equation 2.22 to estimate $[T_{ij}]$.

3. *Maximization:* With $[T_{ij}]$ fixed, calculate $\mathbf{P}$ such that

$$[P_{ij}] = \arg \max_{[P_{ij}]} \sum_{i,j:(i,j)\in\tau} \mu_i P_{ij}[\log \frac{1}{P_{ij}} + T_{ij}]. \tag{2.23}$$

4. *Repeat:* Repeat the expectation and maximization steps until convergence.

The optimal probability transition matrix which maximizes the argument given in equation 2.23 is given by

$$P_{ij} = \frac{b_j}{b_i} \frac{A_{ij}}{W_{\max}}, \tag{2.24}$$

where the matrix $\mathbf{A}$ is defined as

$$A_{ij} = \begin{cases} 2^{T_{ij}} & \text{if } (i, j) \in \tau \\ 0 & \text{otherwise,} \end{cases} \tag{2.25}$$

and the vector $\mathbf{b}$ is the eigenvector of $\mathbf{A}$ corresponding to the maximal real eigenvalue $W_{\max}$. The proof of equation 2.24 requires some non trivial algebra using Lagrange's Multipliers [11]. We present the proof in Appendix A.

Before we move ahead, it might be useful to discuss the importance of Kavčić's algorithm. Following arguments similar to those presented in the proof of Theorem

2.7.4 of [5], we can show that the entropy of a Markov Source is a concave function of the joint probabilities of state transition [1]. The entropy is nothing but the information rate over the channel when the noise variance tends to 0. Thus for sufficiently high signal to noise ratios (SNR), we expect the information rate to have a global maximum as a function of the transition probabilities. Naturally, we would like to find this global maximum and try to "mimic" the performance of such a source.

D. Simulation Results

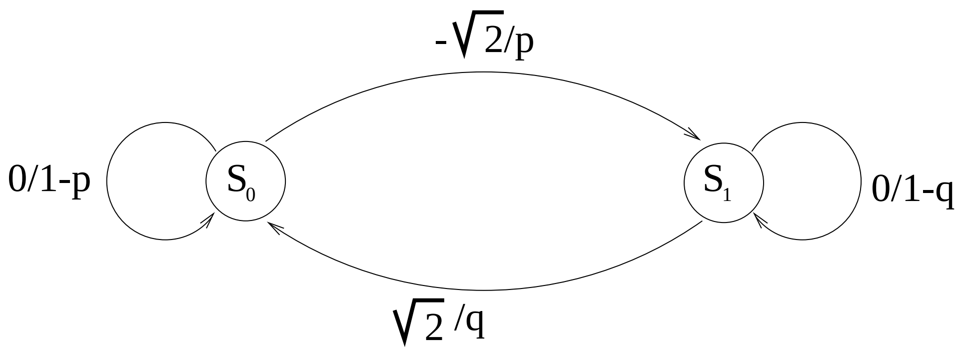

Fig. 3. Another example of a 2 state Markov source.

Let us consider the Markov source shown in Figure 3 transmitting symbols over the AWGN channel with variance $\sigma^2 = 1/2$. The information rate will depend on the values of p, q. Figure 4 shows the information rate as a function of p and q. The information rate was estimated using the forward recursion of the BCJR algorithm as discussed in equation 2.17. We observe that the information rate attains a maximum of around 0.68 at around $(p, q) = (0.55, 0.6)$. We also used Kavčić's Algorithm for

[1]There is a one to one relation between the conditional and joint state transition probabilities of a Markov source.

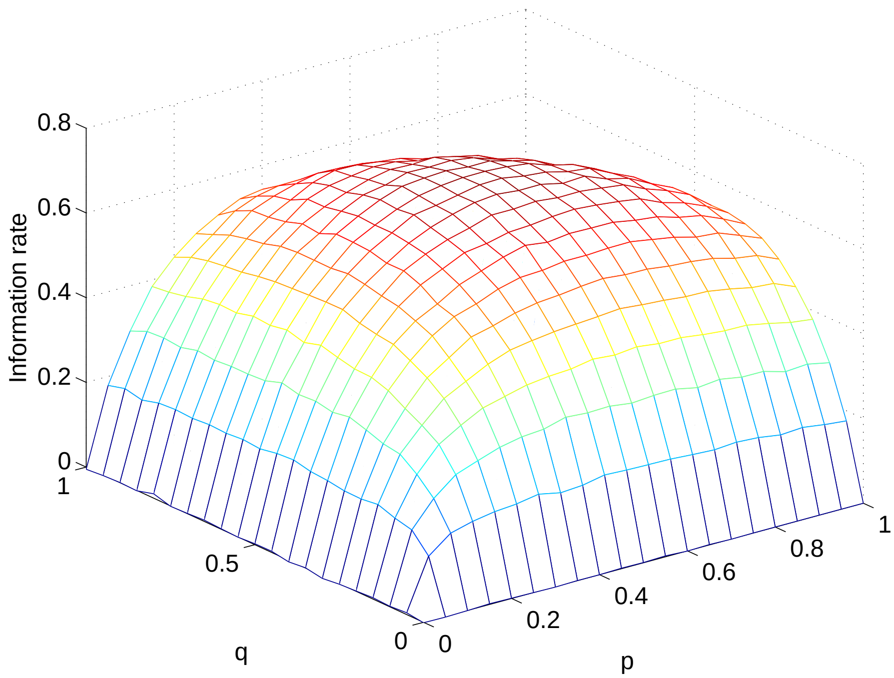

Fig. 4. Information rate of the source in Figure 3 as a function of p and q at a noise variance of $\sigma^2 = 1/2$.

the same setup and found that the optimal probability transition matrix is given by

$$P = \begin{bmatrix} 0.42 & 0.58 \\ 0.58 & 0.42 \end{bmatrix}. \tag{2.26}$$

The maximal value of information rate estimated is 0.676. Thus we find that both the results agree with each other and hence Kavčić's algorithm is verified to perform in this case.

The selection of the particular Markov source depicted in Figure 3 may seem arbitrary. But our motivation for selecting this source will become clear when we discuss the applications of Kavčić's algorithm for partial response channels in the next chapter.

CHAPTER III

CHANNEL CAPACITY

In this chapter we will be reviewing methods of estimating the capacity of partial response channels.

A. Lower Bounds

The lower bounds discussed in this section are derived from the idea that any achievable information rate by a Markov source on a partial response channel is a lower bound to the capacity. Before we proceed, it would be helpful to introduce the concept of a "*super source*."

Let us consider a Markov source at the input of a partial response channel. We can visualize the Markov source and the channel trellis together as a *super Markov source* over an AWGN channel. Note that the symbols at the output of the source will not remain binary. For example the super source formed by the Markov source in Figure 2 over the $(1-D)/\sqrt{2}$ channel is given by Figure 3.

1. Symmetric Information Rate

A scheme for estimating the uniform input information rate was proposed independently in [12] and [13]. The binary i.i.d. source can be visualized as a single state Markov source. This can be combined with the channel trellis to form a super source. The information rate of this super source over the AWGN channel can be estimated by using the forward recursion of BCJR algorithm (refer equation 2.17). This is often referred to as the *symmetric information rate* (SIR) or the *i.i.d. capacity* of the channel.

In [12], the authors demonstrate that interleaved random binary codes can achieve rates close to the SIR over partial response channels. In [14], the authors use the tools of density evolution to design *Low Density Parity Check* (LDPC) Codes which have thresholds very close to the SIR of the channel. Thus the SIR of a channel serves as a benchmark for the performance of codes over partial response channels. In Chapters IV and V we review/design codes which perform better than the SIR bound.

2. Extension of Kavčić's Algorithm for Partial Response Channels

Kavčić's algorithm tries to maximize the information rate of Markov sources over AWGN channel. Thus it can be used to optimize the transition probabilities of the super source to maximize its information rate. This will again serve as a lower bound to the channel capacity.

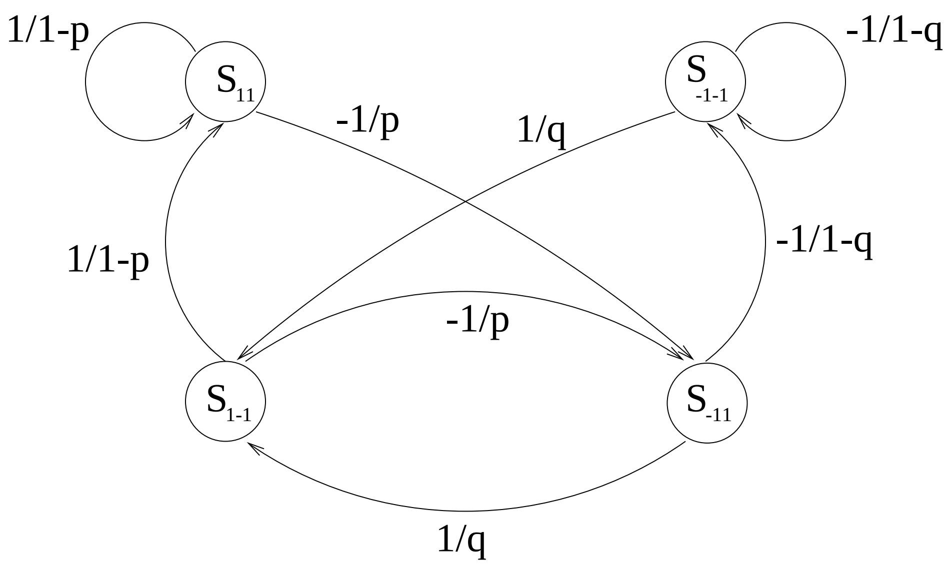

Fig. 5. Super source for Markov source in Figure 2 over $1-D+0.8D^2/\sqrt{2.64}$ channel. The noiseless output symbols are not shown for the sake of clarity.

However, Kavčić's algorithm requires that the transition probabilities of the source should be independent of each other. This may not be true always. For example the super source for the Markov source given in Figure 2 over the $(1 - D + 0.8D^2)/\sqrt{2.64}$ channel (Figure 5) has transition probabilities which are interdependent.

In most of the Markov sources, there are 2^m states, corresponding to the last m, binary bits transmitted. For such Markov sources, the super source will not contain any linearly dependent transition probabilities if the length of the channel is $\leq m+1$. A certain class of such Markov sources are called n^{th} *order channel extensions*, where n is a positive integer. The concept of channel extension is motivated from the concept of source extensions as defined in [10]. The n^{th} order channel extension of a channel has the same number of states as the channel trellis. However the branches correspond to the length n sequences. For example the source in Figure 2 is the 1^{st} order extension for any channel of length 2 (specifically the $(1 - D)/\sqrt{2}$ channel). The 3^{rd} order channel extension is shown in Figure 6.

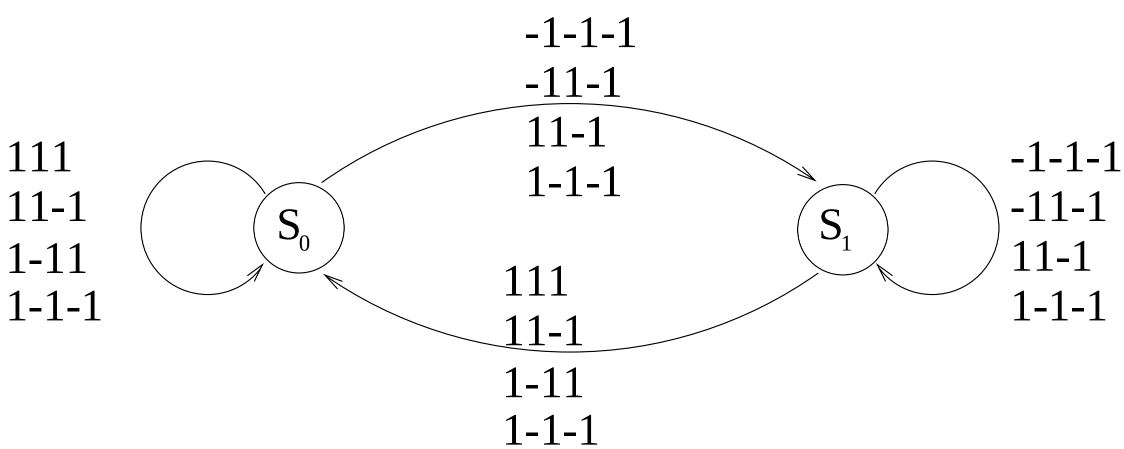

Fig. 6. 3^{rd} extension of a channel with length 2. Note that each branch in the figure represents 4 possible branches in the actual source.

Optimized higher order channel extensions are expected to perform better than the 1^{st} order channel extension. Thus, while the source in Figure 2 achieves an information rate of 0.676 over the $(1-D)/\sqrt{2}$ channel at $E_s/N_0 = 0$ dB, the source in Figure 6 achieves an information rate of about 0.71 bits per channel use.

However, one would not like to restrict oneself just to channel extensions while searching for the optimal source. Thus extension of Kavčić's algorithm, to optimize sources with linearly dependent transition probabilities is highly desirable. How to do so (by extending the proof of Appendix A), is still not clear.

B. Upper Bounds to Channel Capacity

1. Waterfilling Theorem

The Water Filling Theorem (Section 10.5 [5]) states that for continuous alphabet sources over partial response channels, the information rate is maximized if we distribute the input power (spectral density) as given by the waterfilling over inverse of channel frequency response. The result is intuitive in the sense that we are trying to "pump" in more information through the frequency band in which channel is good. An example of a water filling spectrum for the $(1-D)/\sqrt{2}$ channel is shown in Figure 7.

The waterfilling capacity of the channel is given by

$$C = \int \frac{1}{2}\log_2\left(1 + \frac{(\nu - N(f))^+}{N(f)}\right) df, \tag{3.1}$$

where ν is such that $\int(\nu - N(f))^+ df$ is the total input power. The function $(.)^+$ is defined as

$$(x)^+ = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \tag{3.2}$$

Since the waterfilling theorem is valid for the case of continuous alphabet sources, min.$(1,$ waterfilling capacity$)$ will be an upper bound to the capacity of the channel.

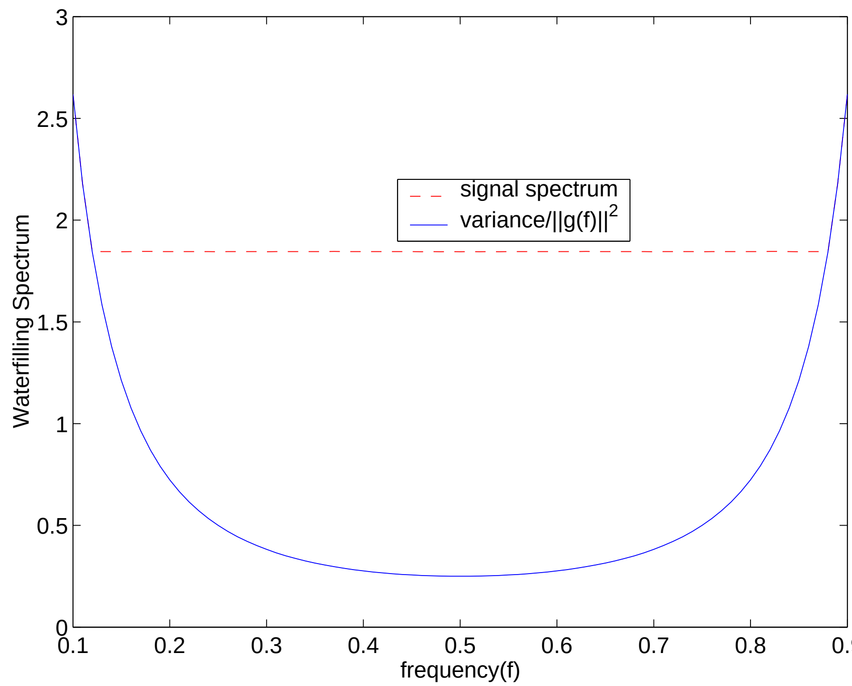

Fig. 7. Waterfilling spectrum over 1-D channel with $E_s/N_0 = 0$ dB.

2. Vonotobel-Arnold Upper Bound

In [15], the authors present a method to estimate an upper bound for the channel capacity of a partial response channel with binary input. The method is motivated by Problem 4.17 of [1]. Here, we discuss the generalization of the problem.

Consider a discrete source X with probability distribution $P_0(X)$, as an input to a discrete memoryless channel. Let Y be the discrete symbols observed at the output of the channel. Then

$$I(X;Y) \leq C \leq \max_{k} I(x = k; Y). \tag{3.3}$$

If the distribution $P_0(X)$ in the above equation is capacity achieving, then the inequalities turn into equalities. Thus we expect that as the probability distribution of the source tends towards a capacity achieving distribution, the bound to the channel capacity becomes tighter. The generalization of equation 3.3 to partial response channels is

$$\lim_{N\to\infty} \frac{1}{N} I(X_1^N; Y_1^N) \leq C \leq \lim_{N\to\infty} \max_{u_1^N} \frac{1}{N} I(x_1^N = u_1^N; Y_1^N). \tag{3.4}$$

Thus the problem of finding an upper bound to the capacity reduces to finding a "good" source and finding a sequence u_1^N, such that the information conveyed by that sequence is the maximum.

In their paper, [15], Vontobel et al. use Markov source optimized by Kavčić's Algorithm. Then they perform a trellis based search to find the "best" sequence. The algorithm is quite complex to implement and at low rates, the upper bound is very close to the waterfilling capacity.

CHAPTER IV

MATCHED INFORMATION RATE CODES

In [16], Kavčić, et al. design codes that achieve rates higher than the symmetric information rate of a partial response channel. The code consists of serially concatenated outer LDPC codes with an inner trellis code, as shown in Figure 8.

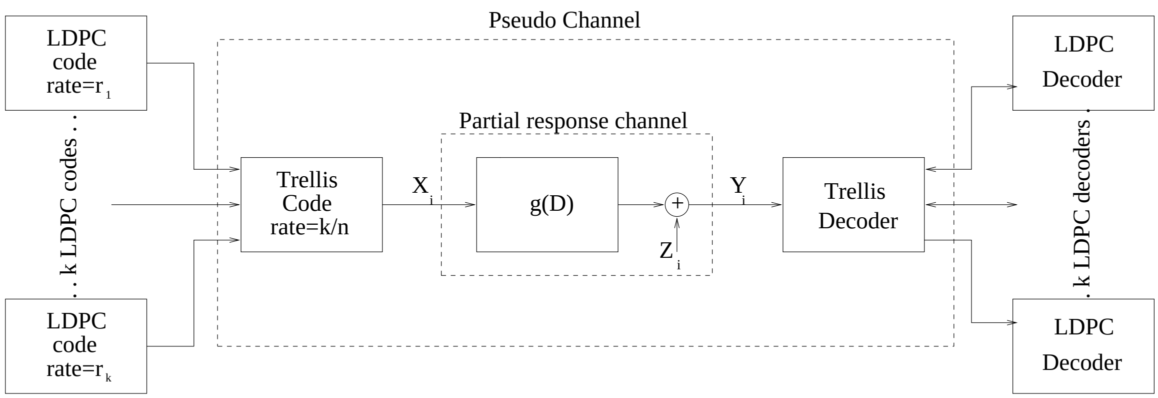

Fig. 8. Structure of the matched information rate code.

The inner trellis code tries to "mimic" the sub optimal Markov source derived from Kavčić's rate optimizing algorithm as discussed in Section A.2 of Chapter 3. With properly optimizing the outer LDPC codes, we can surpass the i.i.d. capacity of the channel. In this chapter we will discuss the results of [16].

A. Design of Inner Trellis Code

Let us consider the 3^{rd} order extension of the $1 - D$ channel. We optimize the transition probabilities for various SNR's, to maximize the information rate over the $(1 - D)/\sqrt{2}$ channel. The optimal source achieves an information rate of 0.5 at $E_s/N_0 = -2.7$ dB. The optimal transition probabilities of this source are listed in

Table I. Our aim is to "mimic" the performance of this source by a rate k/n trellis code, so as to achieve an overall rate of 0.5 at E_s/N_0 close to -2.7 dB.

A rate k/n trellis code is characterized by K states. Each state has 2^k branches coming out of it (corresponding to 2^k different possible inputs). Each branch is mapped to a binary n-tuple which is the output of the particular branch transition. Since we are trying to mimic the 3^{rd} extension of the channel we will choose $n = 3$. We would like the inner code rate k/n, to be greater than the target rate (= 0.5). Thus we choose $k = 2$.

We will split the $M(= 2)$ states of the 3^{rd} order extension into k_1 and k_2 states so that they minimize the Kullback Leibler Distance[1] [5] $D(\kappa||\mu)$, where κ is the probability distribution $(\frac{k_1}{K}, \cdots, \frac{k_M}{K})$, with the constraint that $\sum k_i \leq K$. The μ_i's are the steady state probabilities of the states of the Markov source.

Now, corresponding to each state i of the Markov Source we have $k_i 2^k$ branches in the encoder. We would like to choose a set of integers $\{n_{il}\}$ where $l = 0, 1, \ldots$ such that the transition probabilities of the branches emanating from state i are approximated by the fractions $\{\frac{n_{il}}{k_i 2^k}\}$. To do so we minimize the Kullback Leibler Distance between the transition probabilities of branches originating from state i and the above fractions. The problem of finding the approximate encoder is solved by solving the 2 optimization problems of minimizing the Kullback Leibler Distance. However the solution may not be unique. Out of all the possible set of solutions we choose the one which minimizes $\sum_i \kappa_i D(\{P_{il}\}||\{\frac{n_{il}}{k_i 2^k}\})$. The results for such an optimization are shown in Table I.

[1] The Kullback Leibler distance between 2 probability mass functions $p(.)$ and $q(.)$ is defined as $D(p||q) = \sum_x p(x) \log_2 \frac{p(x)}{q(x)}$. Furthermore, $D(p||q)$ is 0 iff $p = q$.

Table I. Optimal transition probabilities for the 3^{rd} order channel extension of $(1-D)/\sqrt{2}$ channel and their approximations used to design the trellis encoder of rate 2/3 with 10 states.

start state	end state	channel input	transition probability	integer approximation $\{n_{ij}\}$ $k_1 = 5, k_2 = 5, K = 10$	approximate probabilities $n_{ij}/k_1 2^k$
1	1	$1,1,1$	0.005	0	0
1	1	$-1,1,1$	0.146	3	0.15
1	1	$1,-1,1$	0.146	3	0.15
1	1	$-1,-1,1$	0.195	4	0.20
1	2	$1,1,-1$	0.066	1	0.05
1	2	$-1,1,-1$	0.231	5	0.25
1	2	$1,-1,-1$	0.145	3	0.15
1	2	$-1,-1,-1$	0.066	1	0.05
2	1	$1,1,1$	0.066	1	0.05
2	1	$-1,1,1$	0.145	3	0.15
2	1	$1,-1,1$	0.231	5	0.25
2	1	$-1,-1,1$	0.066	1	0.05
2	2	$1,1,-1$	0.195	4	0.20
2	2	$-1,1,-1$	0.146	3	0.15
2	2	$1,-1,-1$	0.146	3	0.15
2	2	$-1,-1,-1$	0.005	0	0

Thus all the 3-tuple outputs of the branches are derived by solving the optimization problem. However there are many ways by which the same branches may be connected between different states in the trellis encoder. To give an estimate of numbers, there are 5^{40} possible permutations of branches possible for the source with parameters represented in Table I. Which permutation do we choose? We randomly select a large number of possible permutations (say 10^4) and evaluate the i.i.d capacity of the trellis for that permutation. We select the permutation which gives the highest i.i.d. capacity.

B. Outer Code Design

The inner trellis code designed in the previous section requires that the input to it should be i.i.d. binary digits. Further we want the rate of the outer code to be 3/4. Ideally speaking these 2 conditions cannot be met simultaneously. However, LDPC codes with large block lengths can produce approximately i.i.d. sequences.

We can visualize the output of the outer code as an input to a *pseudo channel* (Figure 8), where the pseudo channel is the combination of inner trellis code, the partial response channel and the soft output trellis decoder. The input to the pseudo channel is a k-tuple. Thus we can design k LDPC codes for this pseudo channel. At the decoder we will employ k corresponding LDPC decoders. The first decoder decodes using the first bit in each k tuple without assuming anything about the other bits. The second decoder decodes using the second bit in each k tuple assuming the first bit is successfully decoded (The trellis decoder gives an updated estimate on the second bit assuming the first bit is known), etc.

The rate of the outer code is $r = \frac{1}{k}\sum_i r_i$, where r_i's are defined by

$$r_1 = \lim_{N\to\infty} \frac{1}{N} I(X_1^N(1); \mathbf{Y}_1^N)$$

$$
\begin{aligned}
r_2 &= \lim_{N\to\infty} \frac{1}{N} I(X_1^N(2); \mathbf{Y}_1^N | X_1^N(1)) \\
&\vdots \\
r_k &= \lim_{N\to\infty} \frac{1}{N} I(X_1^N(k); \mathbf{Y}_1^N | X_1^N(1), \cdots, X_1^N(k-1)), \qquad (4.1)
\end{aligned}
$$

where $X_1^N(l)$ denotes the bits in the positions $l, l+k, l+2k, \cdots$.

The degree sequences of the LDPC codes are optimized so that we get the lowest thresholds for all the k codes, while their information rates as given by equation 4.1 are close to the target rate. The tools of density evolution as discussed in [14, **?**] are used for this optimization. The overall threshold of the codes is $E_s/N_0^* = \max.(E_s/N_{0_1}^*, E_s/N_{0_2}^* \cdots, E_s/N_{0_k}^*)$. Here $E_s/N_{0_l}^*$ is the threshold of the individual LDPC code.

In [16], the authors design the optimal outer irregular LDPC codes for the example we are considering. The codes were designed for rate 3/4 and had thresholds 0.17 dB away from the inner trellis code capacity at target rate of 0.5. This scheme had a gain of around 0.25 dB over the i.i.d capacity of $(1-D)/\sqrt{2}$ channel. The authors showed similar gains over the i.i.d capacity of the $(1-D+0.8D^2)/\sqrt{2.64}$ at a target rate of 0.5.

C. Summary

The matched information rate coding scheme discussed in this chapter has been shown to achieve high target rates over partial response channels with a considerable gain over the i.i.d capacity of the channel. However both the design process and implementation of the codes is quite complex (practically un-implementable).

CHAPTER V

MATCHED SPECTRUM CODES

A. Motivation

If the PSD of the encoded data matches the waterfilling spectrum, the information rate is maximized.[1] Thus, we expect the codes to perform better if their PSD resembles the waterfilling spectrum.

In Section D of Chapter II, we found the optimum transition probabilities of a 2 state Markov source over the $(1-D)/\sqrt{2}$ channel (equation 2.26). In Figure 9, we plot the power spectral density of that source. We notice that the PSD of the source

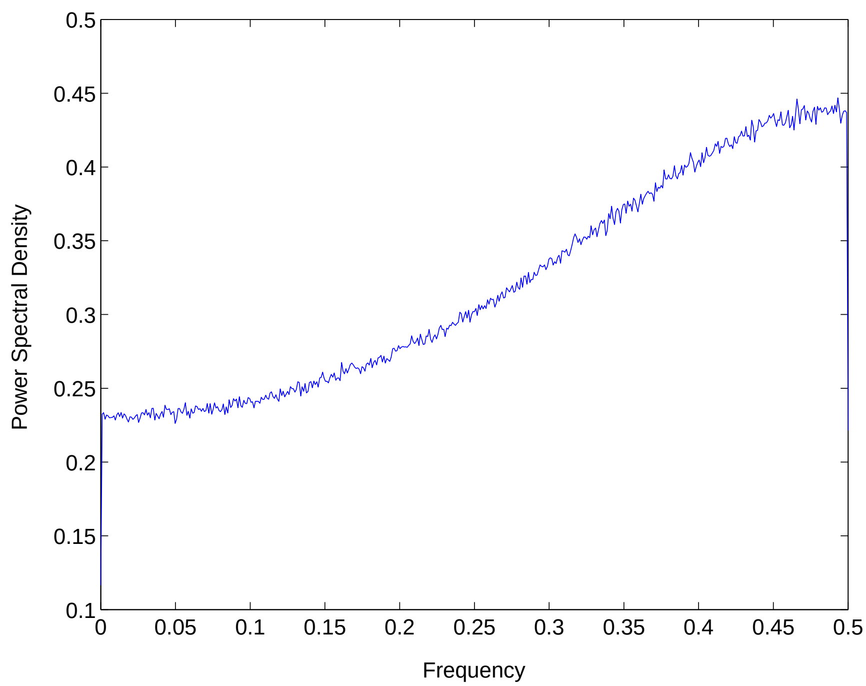

Fig. 9. Power Spectral Density of the Markov source defined by equation 2.26.

[1]This is usually not doable for finite alphabet sources.

peaks at high frequency and thus in some sense tries to match the frequency response of the channel. This gives us an intuitive motivation that for optimal binary sources over partial response channels the PSD of the source should in some sense match the channel frequency response. In [17], the authors show substantial increase than the i.i.d. capacity of the channel for block codes with spectrum matched to the channel response.

1. Performance Index of Codes

Now let us consider the problem of encoding for partial response channels for very low rates $1/n$ with $n \to \infty$. For each binary bit $\{\pm 1\}$, we should assign a binary n-tuple. Obviously these n tuples should be such that the output of these n-tuple's from the channel filter should have maximum energy. Thus again we are trying to match the PSD with the channel frequency response. The best we can hope to achieve is when the spectrum of the data bits is an impulse at the frequency where the channel frequency response is maximum. Thus for any coding scheme we can define a performance index (p.i.)

$$\text{p.i.} = \frac{\text{Power of the sequence at the oputput of noiseless partial response channel}}{\text{Input Power} \times \max|g(f)|^2}. \tag{5.1}$$

We expect that codes with higher p.i. should perform better than codes with lower p.i. (at least asymptotically for low rates).

In the previous chapter we discussed a coding scheme in which an encoder tries to “mimic” the performance of an optimal Markov source. The technique led to very complex encoders. However, the arguments presented in this section motivate that we can design our codes to match the channel spectrum and achieve good performance. It would be beneficial if we have some techniques by which we can design simple encoders which match the channel spectrum. In this chapter we discuss such a scheme for spectrum matching. The scheme is expected to perform better for low rates.

B. Cavers-Marchetto Coding Technique

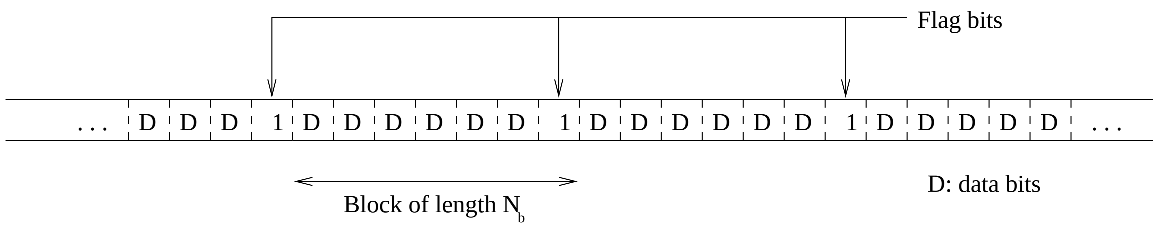

Fig. 10. Coding technique of Cavers-Marchetto

In [2], the authors describe a coding technique of rate $(N_b - 1)/N_b$, where N_b is an integer, to shape the spectrum of the code. Let us consider a binary data source, in which the data bits are divided into blocks of length $N_b - 1$ (Figure 10). We add a constant flag bit, say "1", at the end of each block. We now have blocks of length N_b. With the i^{th} such block we associate a flag index $F(i)$, where $F(i) \in \{-1, +1\}$. Before transmitting over the channel, the i^{th} block is multiplied by the flag index $F(i)$. We have the freedom to select the flag indices to shape the spectrum of the code.

Suppose that we want the spectrum of the encoded stream to match a filter $g_1(n)$. We want to choose the flag indices so that the energy of the encoded stream when passed through $g_1(n)$ is maximum. A symbolic representation of this idea is shown in Figure 11. In [2], the authors solve this problem by formulating a Viterbi type algorithm at the encoder.

We will equivalently solve this problem by representing the source as a trellis encoder. In the encoder, there are $2^{\text{Length}(g_1(n))-1}$ states corresponding to the last

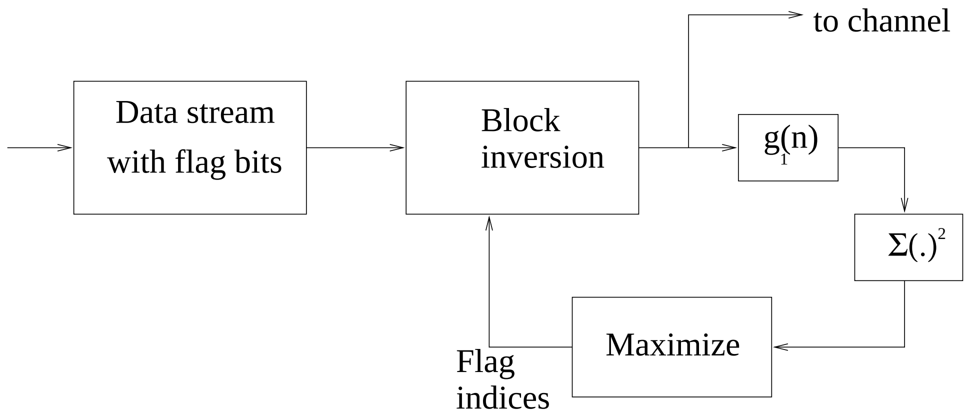

Fig. 11. Structure of Cavers-Marchetto encoder.

Length($g_1(n)$) $-$ 1 bits transmitted over the channel. Each branch in the trellis represents a $N_b - 1$ binary-tuple input and an output of length N_b which is transmitted over the channel. Thus each branch will have a transition probability of $2^{-(N_b-1)}$. An example of the rate 1/2 code matched to $g_1(D) = 1 - D$ is shown in Figure 12. However it may turn out that the encoder is not unique. In which case we select the encoder for which the average distance between the outputs of branches originating from a state is maximum (at the output of the channel).

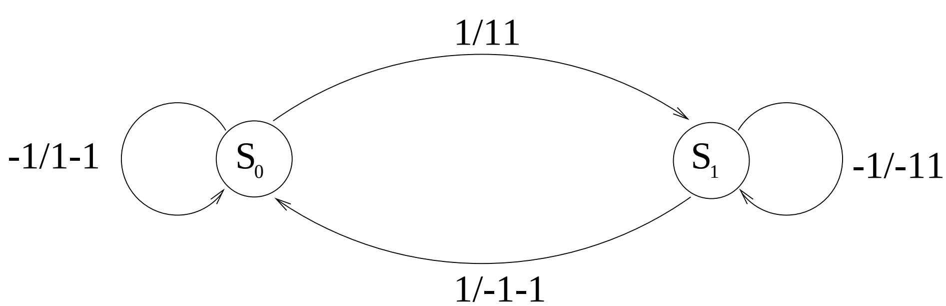

Fig. 12. Rate 1/2 code matched to $1 - D$.

The advantage of representing the source as a trellis encoder is that we can estimate the information rate at the output of the channel. Figure 13 shows the performance of codes matched to $1-D$ when transmitted over a $(1-D)/\sqrt{2}$ channel. We also show the i.i.d capacity of the channel for comparison. We notice that for low rate regions these encoders surpass the i.i.d. capacity.

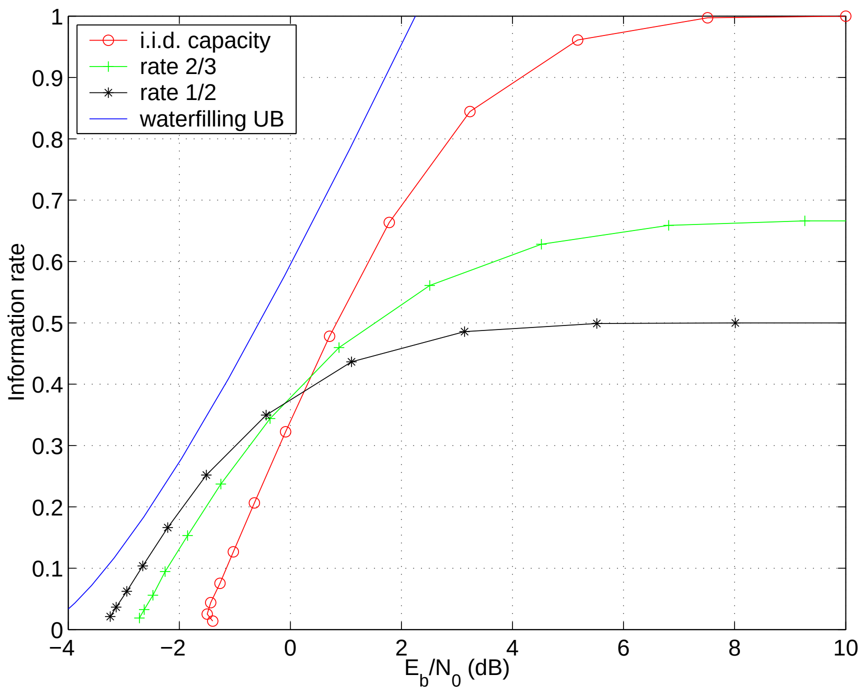

Fig. 13. Performance of codes matched to $1-D$, when transmitted over $(1-D)/\sqrt{2}$ channel (Cavers-Marchetto technique).

C. Generalization of Cavers-Marchetto Technique

We ask ourselves the following questions. Is the Cavers-Marchetto technique the ideal way to form a trellis encoder to maximize the energy of an encoded stream through a filter $g_1(n)$? Can the Cavers-Marchetto technique be generalized to form encoders for any arbitrary rate?

The questions above can be answered by the following "greedy" algorithm to design a rate k/n trellis encoder.

1. Consider a trellis with $2^{\text{Length}(g_1(n))-1}$ states. Each state corresponds to the last $\text{Length}(g_1(n)) - 1$ bits sent through the channel.

2. There are 2^k branches originating from each state. Assign an input k-tuple to each branch.

3. There are 2^n possible output n-tuples for each state. Out of these, select 2^k possible outputs which when passed through $g_1(n)$, have the maximum energy. Assign 2^k such n-tuples to all the branches originating from the particular state.

4. Terminate the branches into the states corresponding to the last $\text{Length}(g_1(n))-1$ bits transmitted.

5. Discard any redundant states i.e., states with no incoming branches.

6. If more than one optimal trellis encoder is possible, select the one for which the average distance between the outputs of branches originating from a state is maximum (at the output of the channel).

It is expected that the generalized algorithm will perform better if $g_1(D)$ resembles the channel transfer function. The simulations confirmed the same. In Figure 14, we show the normalized transfer functions of various filters that we will be using.

The performance of rate 1/3 codes over a $(1-D)/\sqrt{2}$ channel, designed by using the generalized scheme and matched to various different filters is shown in Figure 15. We also evaluated the performance index of each code and tabulated them in Table II. We notice from Table II and Figure 15, that asymptotically the codes with higher performance index perform better. In the figure we have also shown the performance

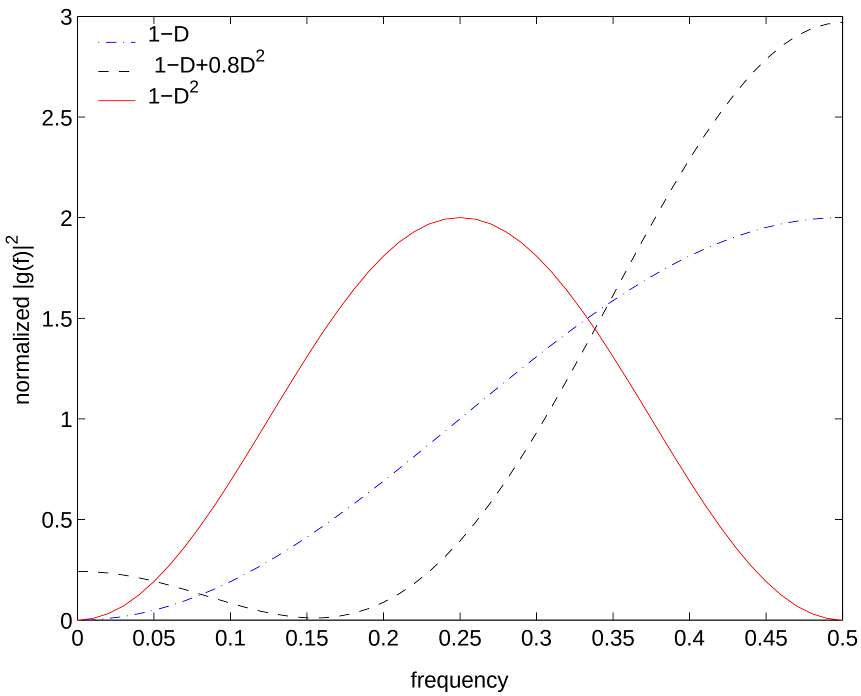

Fig. 14. Normalized energy spectrum of various filters.

of the encoder designed to match the 3 tap approximation to the waterfilling frequency response.

Similar simulation results were found for the $(1-D+0.8D^2)/\sqrt{2.64}$ channel and are shown in Figure 16 and Table III.

The results discussed above motivate us that matching the channel leads to better performance (at least asymptotically for low rates). In Figure 17, we show the performance of codes (with different rates), matched to the channel $(1-D)/\sqrt{2}$. Table IV shows the corresponding peroformance indices. For the $(1-D)/\sqrt{2}$ channel the Shannon's low rate limit is -4.59 dB. The figures suggest that the Shannon's limit to be around -4 dB. Figure 17 further justifies that the coding scheme is asymptotically optimal, as the codes tend to the Shannon's limit as the rate decreases.

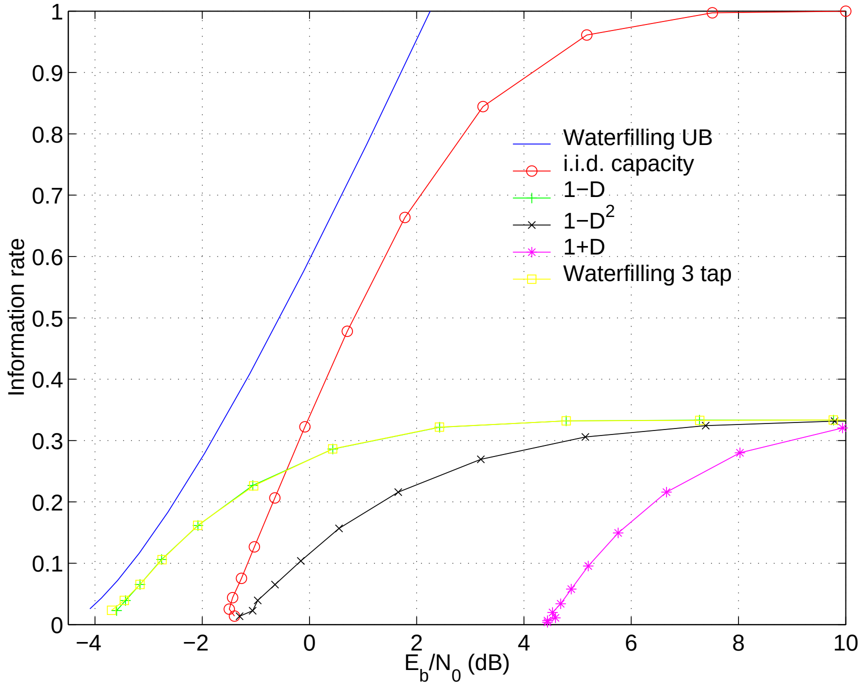

Fig. 15. Performance of rate 1/3 codes matched to various filters, when transmitted over $(1-D)/\sqrt{2}$ channel.

Table II. Performance index of rate 1/3 codes matched to $g_1(D)$ when transmitted over $(1-D)/\sqrt{2}$ channel.

match filter $g_1(D)$	p.i.
$1-D$	0.83
$1-D^2$	0.53
$1+D$	0.17
IID	0.5

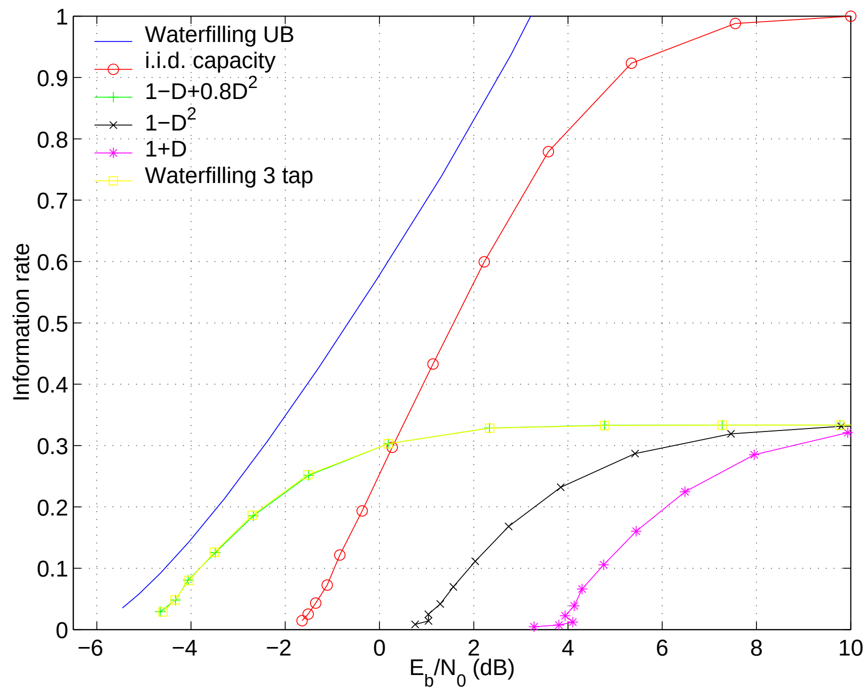

Fig. 16. Performance of rate 1/3 codes matched to various filters, when transmitted over $(1-D+0.8D^2)/\sqrt{2.64}$ channel.

Table III. Performance index of rate 1/3 codes matched to $g_1(D)$ when transmitted over $(1-D+0.8D^2)/\sqrt{2.64}$ channel.

match filter $g_1(D)$	p.i.
$1-D+0.8D^2$	0.71
$1-D^2$	0.23
$1+D$	0.1
IID	0.34

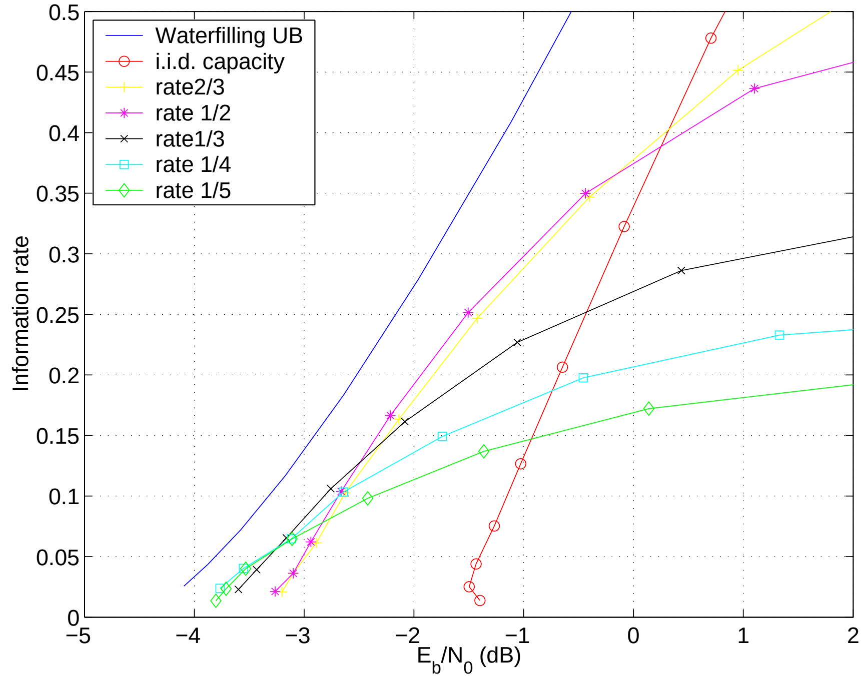

Fig. 17. Performance of codes matched to $(1-D)/\sqrt{2}$, when transmitted over $(1-D)/\sqrt{2}$ channel.

Table IV. Performance index of codes matched to $(1-D)/\sqrt{2}$ when transmitted over $(1-D)/\sqrt{2}$ channel.

rate	p.i.
2/3	0.75^{-}
1/2	0.75^{+}
1/3	0.83
1/4	0.87
1/5	0.90

Now let us reconsider the code design problem and exploit the results shown in Figure 17. Suppose that we target a rate of 0.1 over the channel. We notice that if we use an inner trellis code of rate 1/3 matched to $(1-D)/\sqrt{2}$, then we expect to gain around 1.6 dB with respect to the i.i.d capacity. Thus in the next section we will be discussing how to design the outer LDPC code for the inner rate 1/3 code over the $(1-D)/\sqrt{2}$ channel.

Similar results for the $(1-D+0.8D^2)/\sqrt{2.64}$ channel are shown in Figure 18 and Table V. From the figure, we estimate that the Shannon's low rate limit should be around -5.5 dB (The actual limit for the channel is -6.3 dB). Further, to target a rate of 0.1, we will be using an inner trellis code of rate 1/3, so that we expect to achieve a gain of around 2.7 dB over the i.i.d. capacity.

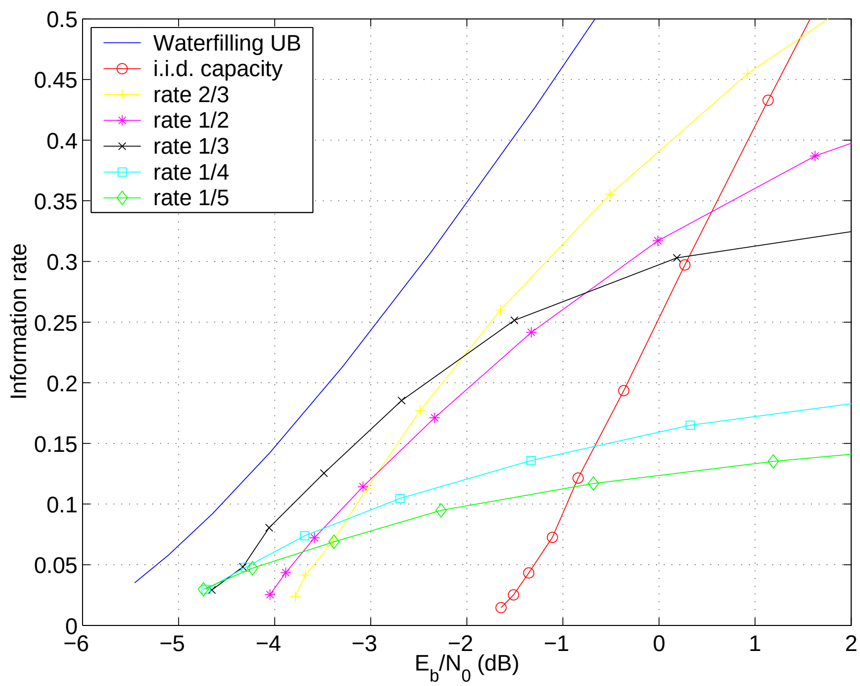

Fig. 18. Performance of codes matched to $(1-D+0.8D^2)/\sqrt{2}$, when transmitted over $(1-D+0.8D^2)/\sqrt{2.64}$ channel.

Table V. Performance index of codes matched to $(1 - D + 0.8D^2)/\sqrt{2.64}$ when transmitted over $(1 - D + 0.8D^2)/\sqrt{2.64}$ channel.

rate	p.i.
2/3	0.58
1/2	0.63
1/3	0.71
1/4	0.82
1/5	0.85

D. Design of Outer Irregular LDPC Codes

In this section we will concentrate on the design of an outer irregular LDPC code of rate 0.3. The LDPC code design problem can be viewed as a code design problem for the pseudo channel (Figure 8) where the pseudo channel consists of a trellis encoder of rate 1/3, a partial response channel, an AWGN source and a trellis decoder. The design technique follows from the Extrinsic Information Transfer (EXIT) chart technique as discussed in [18]. A good exposition of extrinsic information and EXIT charts can be found in [19, 20]. A basic introduction to LDPC codes can be found in [21]. Here, we will just mention how we used the concepts.

We use equation 2.7 to estimate the log likelihood ratios of the bits transmitted over the pseudo channel. We do so by using the BCJR algorithm. A typical distribution of the log likelihood ratios (LLR) is shown in Figure 19. We approximate the distribution to be Gaussian to ease calculations.

The LDPC decoder can be visualized as an iterative decoder for a serially concatenated code with an inner repeat code (Variable node - VND) and outer single

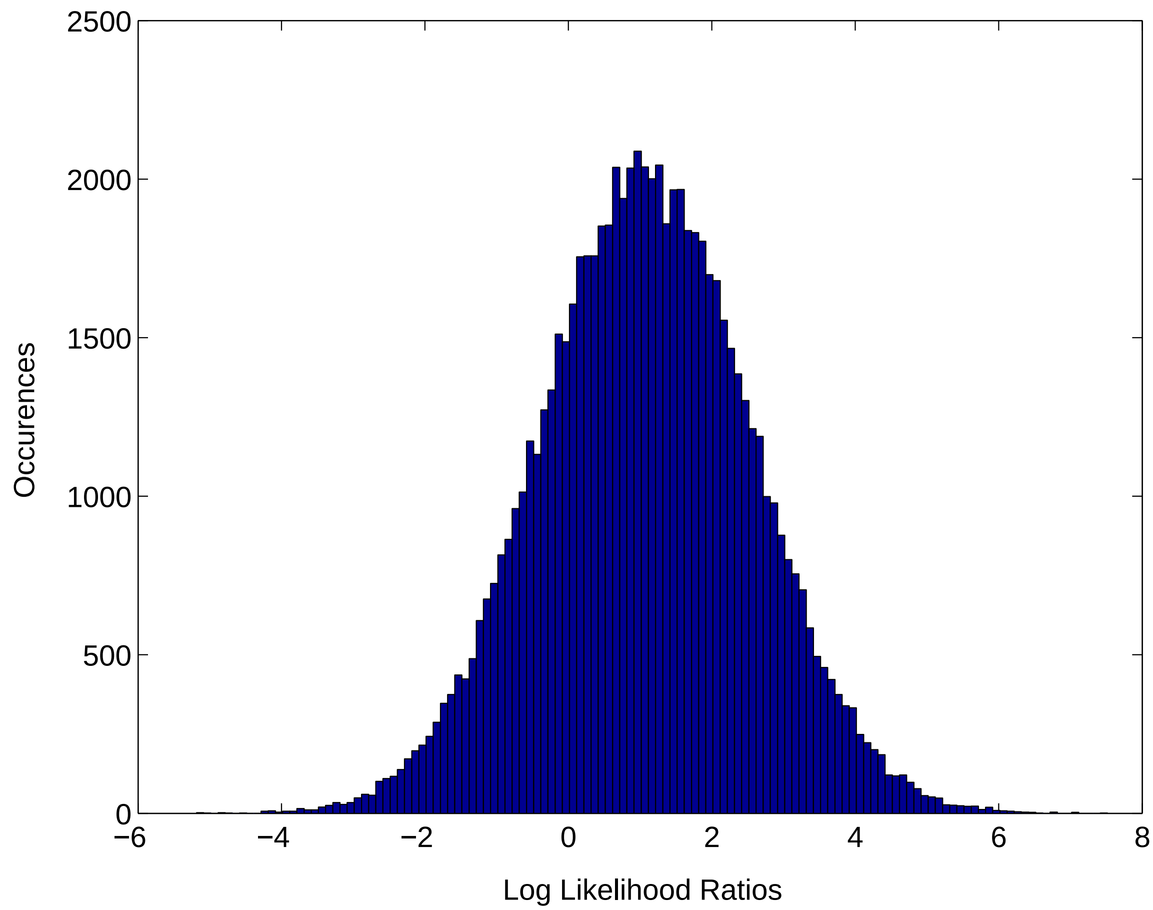

Fig. 19. Histogram of the log likelihood ratio values from the pseudo channel over the partial response channel $(1 - D + 0.8D^2)/\sqrt{2.64}$ for $E_b/N_0 = -3.55$ dB.

parity check codes (Check node - CND). For the decoder to successfully converge, the EXIT chart at the variable node decoder should lie above the EXIT chart of the check node decoder. We use the approximations mentioned in [18] to formulate the Exit chart curves as a function of the standard deviation of the LLR values, the degree sequence and the a priori information. We then optimized the degree distribution to maximize the rate for a particular standard deviation (and hence a particular E_b/N_0), under the constraint that the VND curve should lie above the CND curve. We repeat this procedure for various values of E_b/N_0 until we get a rate close to the target rate. Hence the threshold SNR is evaluated.

1. Results

The trellis code capacity of the rate 1/3 code for the $(1-D)/\sqrt{2}$ channel is $E_b/N_0 = -2.82$ dB at a rate of 0.1 (Figure 17). We designed the LDPC code with rate 0.3 and threshold $E_b/N_0 = -2.6$ dB. Thus the design is better than the i.i.d. capacity of the channel by 1.4 dB. The code had a check node degree $d_c = 5$. The generating function $\lambda(x)$ is shown in the Table VI and the EXIT chart plotted in Figure 20.

Similar results for the $(1-D+0.8D^2)/\sqrt{2.64}$ channel, are shown in Figure 21 and Table VII. The trellis code capacity was 0.1 at $E_b/N_0 = -3.81$ dB. The designed LDPC code has a threshold of -3.55 dB. With respect to the i.i.d. capacity, this implies a gain of 2.59 dB.

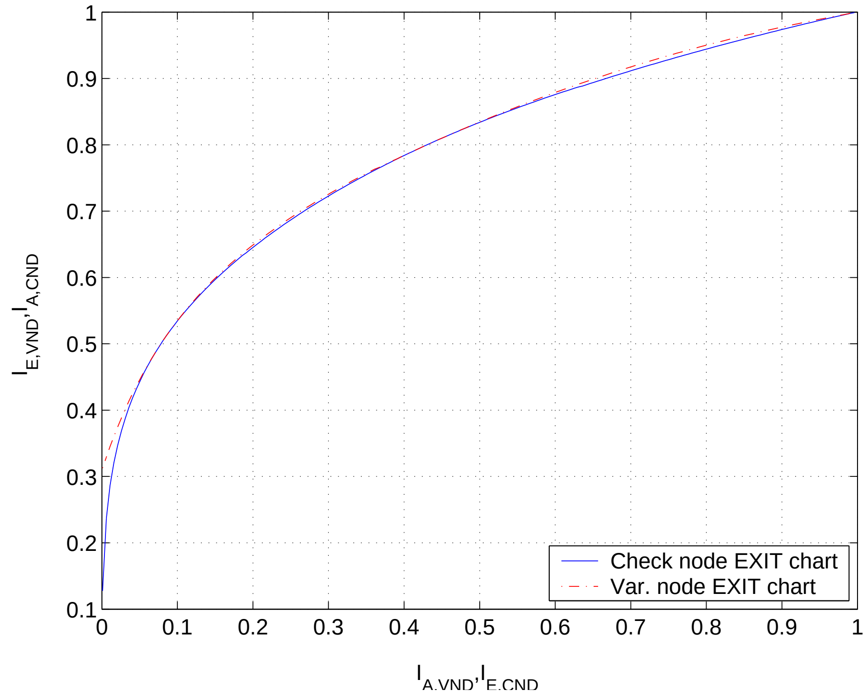

Fig. 20. EXIT chart for the check regular LDPC code with rate 0.3 and threshold $E_b/N_0 = -2.6$ dB. This code is deigned for the $(1-D)/\sqrt{2}$ channel.

Table VI. Degree sequence for the rate 0.3 LDPC code for the $(1-D)/\sqrt{2}$ channel. It has a threshold of $E_b/N_0 = -2.6$ dB.

i	λ_i	ρ_i
2	0.2617	
3	0.3566	
5		1
6	0.0236	
8	0.1655	
18	0.1926	

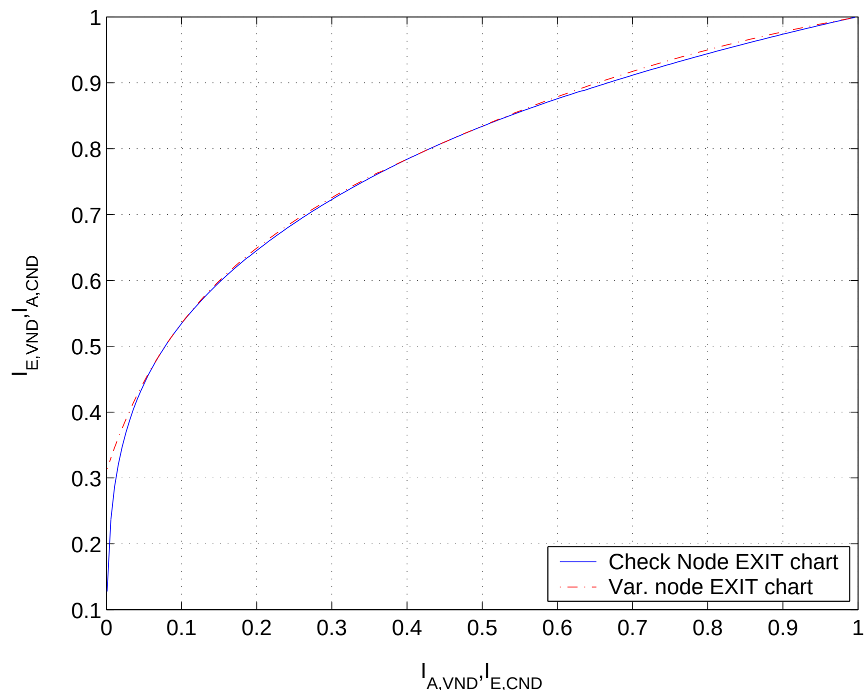

Fig. 21. EXIT chart for the check regular LDPC code with rate 0.3 and threshold $E_b/N_0 = -3.55$ dB. This code is designed for the $(1 - D + 0.8D^2)/\sqrt{2.64}$ channel.

Table VII. Degree sequence for the rate 0.3 LDPC code for the $(1-D+0.8D^2)/\sqrt{2.64}$ channel. It has a threshold of $E_b/N_0 = -3.55$ dB.

i	λ_i	ρ_i
2	0.2621	
3	0.3592	
5		1
6	0.0084	
8	0.1831	
18	0.1872	

CHAPTER VI

CONCLUSION

The main contributions of this thesis can be summed up as follows:

1. By simulations, we verified that matching the spectrum of the data stream with the channel frequency response leads to better information rates.

2. We provided a qualitative argument for the optimality of the encoder design technique presented in Section C of Chapter V for low rates.

3. We formulated the concept of performance index of an encoder and verified its validity by simulation results.

4. Motivated by the above results, we designed a serially concatenated coding scheme with an outer irregular LDPC code and an inner matched spectrum trellis code. The performance was found to be 1.4 dB better than the i.i.d. capacity of the channel for a target rate of 0.1 for the $(1-D)/\sqrt{2}$ channel. For the $(1-D+0.8D^2)/\sqrt{2.64}$ channel, the performance was found to be 2.59 dB better. The coding scheme is considerably easier to implement than the existing schemes.

A. Future Work

We believe that the following problems may be very interesting to investigate.

1. We believe that it can be analytically shown that there is a threshold SNR for every channel, so that the information rate as a function of transition probabilities of a Markov source, will have just one global maximum, if the SNR is

greater than the threshold. This thought is motivated by the discussion at the end of Section C of Chapter II.

2. It would be beneficial to extend Kavčić's algorithm for the case when the transition probabilities of the Markov source are linearly dependent on each other. This would help us optimize any Markov source over any partial response channel. Perhaps the proof shown in Appendix A may be extended.

3. In [22], the authors find the closed form expression for the power spectral density of certain special Markov sources. However, in general there is no closed form expression for the power spectral density of a Markov source. The working in Section 4.4.3 of [23] is pretty confusing, and so we have included the general derivation of the power spectral density of a Markov source in Appendix B. Designing Markov sources so that the PSD matches a given function in the frequency domain is still an open problem.

4. We considered the problem of code designing for 2 channels. There was a wide difference in the gain over the i.i.d. capacity for the 2 channels. It would be interesting to investigate, for which channels should the proposed coding scheme perform better as compared to other channels.

REFERENCES

[1] R. G. Gallager, *Information Theory and Reliable Communication*, New York: John Wiley and Sons Inc., 1968.

[2] J. K. Cavers and R. F. Marchetto, "A New Coding Technique for Spectral Shaping of Data," *IEEE Transactions on Communications*, Vol. 40, No. 9, pp. 1418-1422, Septmeber 1992.

[3] R. Karabed and P. H. Siegel, "Matched Spectral-Null Codes for Partial-Response Channels," *IEEE Transactions of Information Theory*, Vol. 37, No. 3, pp. 818-855, May 1991.

[4] B. F. Uchôa Filho and M. A. Herro, "Good Convolutional Codes for Precoded $(1-D)(1+D)^n$ Partial-Response Channels," *IEEE Transactions on Information Theory*, Vol. 43, No. 2, pp. 441-453, March 1997.

[5] J. A. Thomas and T. M. Cover, *Elements of Information Theory*, New York: John Wiley and Sons Inc., 1991.

[6] L. R. Bahl, J. Cocke, F. Jelinek and J. Raviv, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate," *IEEE Transactions on Information Theory*, Vol. 20, pp. 284-287, March 1974.

[7] A. Kavčić, "On the Capacity of Markov Sources over Noisy Channels," in *Proceedings IEEE Global Communications Conference 2001*, San Antonio, Texas, November 2001, pp. 2997-3001.

[8] S. Arimoto, "An Algorithm for Computing the Capacity of Arbitrary Discrete Memoryless Channels," *IEEE Transactions on Information Theory*, Vol. IT-18, No. 1, pp. 14-20, January 1972.

[9] R. E. Blahut, *Digital Transmission of Information*, Reading, Massachusets: Addison-Wesley Publishing Company, 1990.

[10] B. H. Marcus, P. H. Siegel and J. K. Wolf, "Finite-State Modulation Codes for Data Storage," *IEEE Journal on Selected Areas in Communications*, Vol. 10, No. 1, pp. 5-37, January 1992.

[11] P. O. Vonotobel,"Maximization of the Information Rate under Constraints,", May 2001, (Personal Collection, D. Kumar).

[12] H. D. Pfister, J. B. Soriaga and P. H. Siegel, "On the Achievable Information Rates of Finite State ISI Channels," in *Proceedings IEEE Global Communications Conference 2001*, San Antonio, Texas, November 2001, pp. 2992-2996.

[13] D. M. Arnold and H. A. Loeliger, "On the Information Rate of Binary Input Channels with Memory," in *Proceedings IEEE International Conference on Communications 2001*, Helsinki, Finland, June 2001, pp. 2692-2695.

[14] N. Varnica and A. Kavčić, "Optimized Low-Density Parity Check Codes for Partial Response Channels," in *IEEE Communications Letters*, Vol. 7, No. 4, pp. 168-170, April 2003.

[15] P. O. Vontobel and D. M. Arnold, "An Upper Bound on the Capacity of Channels with Memory and Constraint Input," in *Information Theory Workshop 2001*, Cairns, Australia, September 2001, pp. 147-149.

[16] X. Ma, N. Varnica and A. Kavčić, "Matched Information Rate Codes for Binary ISI channels," in *Proceedings of IEEE International Symposium on Information Theory 2002*, Lausanne, Switzerland, July 2002, p. 269.

[17] D. N. Doan and K. R. Narayanan, "Design of Good Low Rate Codes for ISI Channels Based on Spectral Shaping,"to appear at the *3rd International Symposium on Turbo Codes and Related Topics*, Brest, France, September 2003.

[18] S. ten Brink, G. Kramer and A. Ashikhmin "Design of Low-Density Parity-Check Codes for Multi-Antenna Modulation and Detection," submitted to the *IEEE Transactions on Communications*, June 2002 (Personal collection, D. Kumar).

[19] C. Berrou, A. Glavieux and P. Thitimajshima, "Near Shannon Limit Error-Correcting Coding and Decoding: Turbo Codes(1),"in *Proceedings IEEE International Conference on Communications 1993*, Geneva, May 1993, pp. 1064-1070.

[20] S. ten Brink, "Convergence Behavior of Iteratively Decoded Parallel Concatenated Codes,"*IEEE Transactions on Communications*, Vol. 49, No. 10, pp. 1727-1737, October 2001.

[21] W. Ryan, *An Introduction to Low-Density Parity-Check Codes*, hand written lecture notes, April 2001 (Personcal collection, D. Kumar).

[22] A. Gallopoulos, C. Heegard and P. H. Siegel, "The Power Spectrum of Run-Length-Limited Codes,"*IEEE Transactions on Communications*, Vol. 37, No. 9, pp. 906-917, Septmeber 1989.

[23] J. G. Proakis, *Digital Communications*, 4th edition, New York: The McGraw-Hill Companies, Inc., 2001.

[24] X. Ma, A. Kavčić and M. Mitzenmacher, "The Power Spectra of Good Codes for Partial Response Channels,"accepted for *IEEE International Symposium on Information Theory 2003*, Yokohama, Japan, July 2003.

APPENDIX A

OUTLINE OF THE PROOF OF EQUATION 2.24

We present a brief outline of the proof to equation 2.24 as given in [11]. For the sake of completeness, let us restate the optimization problem.

Let S denote the set of states of a Markov source. Let $\mathbf{A}$ be the incidence matrix of the source, i.e., $A_{ij} = 1$ if the transition from state i to j is allowed, else $A_{ij} = 0$. The matrix $\mathbf{P}$ which maximizes the argument

$$V = \sum_{i \in S} \mu_i \sum_{j \in S} A_{ij} P_{ij} (\log_2 (\frac{1}{P_{ij}}) + T_{ij}) \tag{A.1}$$

under the conditions

$$\begin{aligned} \sum_{j \in S} A_{ij} P_{ij} &= 1 \qquad \text{for all } i \in S \\ \sum_{i \in S} \mu_i A_{ij} P_{ij} &= \mu_j \qquad \text{for all } j \in S \\ \sum_{i \in S} \mu_i &= 1, \end{aligned} \tag{A.2}$$

is given by

$$P_{ij} = \frac{b_j}{b_i} \frac{\tilde{A}_{ij}}{\rho}, \tag{A.3}$$

where the matrix $\tilde{\mathbf{A}}$ is defined by $\tilde{A}_{ij} = A_{ij} 2^{T_{ij}}$ and ρ is the maximal (real) eigenvalue of $\tilde{\mathbf{A}}$ with eigenvector $\mathbf{b}$.

Now that we have stated the problem, we set up the Lagrangian,

$$\begin{aligned} L &= \sum_{i \in S} \mu_i \sum_{j \in S} A_{ij} P_{ij} (\log_2 (\frac{1}{P_{ij}}) + T_{ij}) \\ &\quad + \sum_{i \in S} \lambda_i \sum_{j \in S} A_{ij} P_{ij} + \sum_{j \in S} \lambda'_j (\sum_{i \in S} \mu_i A_{ij} P_{ij} - \mu_j) + \lambda'' \sum_{i \in S} \mu_i, \end{aligned} \tag{A.4}$$

We need to solve the equations

$$\begin{aligned} \frac{\partial L}{\partial \mu_k} &= 0 \qquad \text{for all } k \in S \\ \frac{\partial L}{\partial P_{kl}} &= 0 \qquad \text{for all } k, l \in S \end{aligned} \tag{A.5}$$

under the constraints defined by equation A.2. By some manipulation, we get

$$P_{kl} = 2^{\lambda'_l - \lambda'_k + \lambda'' + T_{kl}} A_{kl} \tag{A.6}$$

and

$$\sum_{l \in S} A_{kl} 2^{T_{kl}} 2^{\lambda'_l} = 2^{-\lambda'' + \lambda'_k}. \tag{A.7}$$

In equation A.7, we identify $\tilde{A}_{ij} = A_{ij} 2^{T_{ij}}$, $\mathbf{b}$ as the eigenvector with entries $b_i = 2^{\lambda'_i}$ and $\rho = 2^{-\lambda''}$ as the eigenvalue. We thus get

$$\tilde{\mathbf{A}} \mathbf{b}^T = \rho \mathbf{b}^T. \tag{A.8}$$

Thus, using equation A.8 and equation A.6 we get the desired result of equation A.3.

APPENDIX B

POWER SPECTRAL DENSITY OF A MARKOV SOURCE

Suppose a Markov source has n states. Let the probability transition matrix be defined by $P_{ij} = P(s_{t+1} = j/s_t = i)$, where s_t is the state of the Markov source at time t. Let the matrix $\mathbf{I}$ be such that I_{ij} is the symbol transmitted by the Markov source for a transition from state i to j. Let Φ represent the autocorrelation function.

$$\begin{aligned}
PSD &= \sum_{k=-\infty}^{\infty} \Phi(k) \underbrace{e^{j2\pi k f T_s}}_{\theta_k} \\
&= \underbrace{\sum_{k=1}^{\infty} E(I_t I_{t+k}^*)\theta_k}_{T_1} + \underbrace{\sum_{k=-\infty}^{-1} E(I_t I_{t+k}^*)\theta_k}_{T_2} + \underbrace{E(|I|^2)}_{T_3} \\
T_3 &= \sum_{ij} \mu_i P_{ij} |I_{ij}|^2 \\
T_1 &= T_2^* \\
&= \sum_{k=1}^{\infty} \sum_{ij} E(I_t I_{t+k}^* / s_{t-1} = i; s_t = j) \mu_i P_{ij} \theta_k \\
&= \sum_{k=1}^{\infty} \sum_{ij} I_{ij} \mu_i P_{ij} \theta_k E(I_{t+k}^* / s_t = j) \\
&= \sum_{k=1}^{\infty} \sum_{ij} I_{ij} \mu_i P_{ij} \theta_k \sum_{lm} I_{lm}^* [P^{k-1}]_{jl} P_{lm}, \qquad \text{(B.1)}
\end{aligned}$$

where $[P^k]_{jl}$ is the jl^{th} entry of matrix P^k and μ_i is the steady state probability of state i.

Thus

$$PSD = \sum_{ij} \mu_i P_{ij} |I_{ij}|^2 + 2Re[\sum_{k=1}^{\infty} \sum_{ijlm} I_{ij} \mu_i P_{ij} \theta_k I_{lm}^* [P^{k-1}]_{jl} P_{lm}]. \qquad \text{(B.2)}$$

If $I_{ij} \in \{-1, +1\}$, then the above equation reduces to

$$PSD = 1 + 2 \sum_{k=1}^{\infty} \sum_{ijlm} I_{ij} \mu_i P_{ij} I_{lm} [P^{k-1}]_{jl} P_{lm} \cos 2\pi k f T_s. \qquad \text{(B.3)}$$

Still we do not have a closed form expression for the power spectral density. As discussed in [24], a good approximation is achieved if we restrict the summation of k index in the equation above to a large finite integer.

VITA

Deepak Kumar was born in India on October 20, 1979. He received his Bachelor of Technology in electrical engineering from the Indian Institute of Technology, Bombay in August 2001. In September 2001, he joined the Department of Electrical Engineering at Texas A&M University, College Station, to pursue a Master of Science degree. He can be reached at:

Deepak Kumar

106 Watson

MC 128-95

California Institute of Technology

Pasadena, CA 91125.

Phone: (626) 395-3878

Fax: (626) 229-7564