TECHNICAL RESEARCH REPORT

Detection of Binary Sources Over Discrete Channels with Additive Markov Noise

by F. Alajaji, N. Phamdo,
N. Farvardin and T. Fuja

T.R. 94-31

INSTITUTE FOR SYSTEMS RESEARCH

Sponsored by
the National Science Foundation
Engineering Research Center Program,
the University of Maryland,
Harvard University,
and Industry

Detection of Binary Sources Over Discrete Channels with Additive Markov Noise*

Fady Alajaji†, *Nam Phamdo*‡, *Nariman Farvardin*† and *Tom Fuja*†

† Electrical Engineering Department
Institute for Systems Research
University of Maryland
College Park, MD 20742

‡ Electrical Engineering Department
State University of New York
at Stony Brook
Stony Brook, NY 11794-2350

Abstract

We consider the problem of directly transmitting a binary source with an inherent redundancy over a binary channel with additive stationary ergodic Markov noise. Our objective is to design an optimum receiver which fully utilizes the source redundancy in order to combat the channel noise.

We investigate the problem of detecting a binary iid non-uniform source transmitted across the Markov channel. Two *maximum a posteriori* (MAP) formulations are considered: a *sequence* MAP detection and an *instantaneous* MAP detection. The two MAP detection problems are implemented using a modified version of the Viterbi decoding algorithm and a recursive algorithm. Necessary and sufficient conditions under which the sequence MAP detector becomes useless as well as simulation results are presented. A comparison between the performance of the proposed system with that of a (substantially more complex) traditional tandem source-channel coding scheme exhibits a better performance for the proposed scheme at relatively high channel bit error rates.

The same detection problem is then analyzed for the case of a binary symmetric Markov source. Analytical and simulation results show the existence of a "mismatch" between the source and the channel. This mismatch is reduced by the use of a rate-one convolutional encoder. Finally, the detection problem is generalized for the case of a binary non-symmetric Markov source.

Keywords: Non-uniform iid source, Markov source, additive channels with memory, source redundancy, MAP detection.

* The work of the first, third and fourth authors is supported in part by National Science Foundation grant NSFD CD-8803012. Parts of this paper will be presented at the 1994 IEEE International Symposium on Information Theory, Trondheim, Norway, June 1994.

1 Introduction and Motivation

A source with memory as well as a memoryless source with a non-uniform distribution are sources with *redundancy.* For a finite alphabet of size J, a uniformly distributed *iid* random process contains a maximal amount of information and exhibits no redundancy. Its entropy rate is equal to $\log_2 J$ bits/sample. The total redundancy a stationary ergodic J-ary alphabet source $\{X_n\}_{n=1}^{\infty}$ possesses is equal to the difference between $\log_2 J$ and its entropy rate $H_\infty(X)$ [8]: $\rho_T = \log_2 J - H_\infty(X)$, where $H_\infty(X) \triangleq \lim_{n\to\infty} \frac{1}{n} H(X_1, X_2, \ldots, X_n)$. More specifically, we can write $\rho_T = \rho_D + \rho_M$ where $\rho_D \triangleq \log_2 J - H(X_1)$ denotes the redundancy in the form of non-uniform distribution and $\rho_M \triangleq H(X_1) - H_\infty(X)$ denotes the redundancy in the form of memory [8].

In many practical signal compression schemes, after some transformation, the transform coefficients are turned into bit streams (binary source). Due to the suboptimality of the compression algorithm, the bit stream might contain some redundancy (in the form of memory and/or non-uniformity). This paper addresses the advantages of using this redundancy in controlling channel noise as opposed to further removal of this redundancy to leave room for error control coding.

We consider a binary source with an inherent redundancy which we make no attempt to eliminate. The source is directly transmitted over a discrete noisy channel. Our objective is to design an optimum receiver which fully exploits the source redundancy in order to combat the impairments introduced by the channel noise.

The channel considered is a binary channel with additive noise modeled according to a finite version of the Polya contagion urn scheme [1]. The errors in this channel propagate in a fashion similar to the spread of a contagious disease through a population; the occurrence of each "unfavorable" event (i.e., an error) increases the probability of future unfavorable events. The resulting noise process is a stationary ergodic homogeneous Markov process with memory (order) M, where M is a positive constant. The motivation for the use of such a channel is founded in the fact that most real-world communication channels – in particular the digital cellular channel – have memory; our contagion-based model offers an interesting and less complex alternative to the Gilbert-Elliott model [4] and others [5].

We first investigate the problem of detecting a binary iid non-uniform source transmitted across the contagion Markov channel of order one ($M = 1$). The optimum receiver that minimizes the probability of error is a *maximum a posteriori* (MAP) detector. In a manner similar to the use of channel codes for error correction, the redundancy, due here to the non-uniform distribution of the source, is used by the MAP detector to provide some protection against channel errors. We present two MAP formulations: a *sequence* MAP detection which involves a large delay, and an *instantaneous* MAP detection which involves no delay. In sequence MAP detection, we determine the most probable transmitted *sequence* or *vector* given a received vector that we observe at the channel output. In instantaneous MAP detection, we estimate the most probable transmitted *bit* at a particular time given all the received bits up to that time. The solution of the first problem results in a "Viterbi-like" implementation while the latter problem yields a recursive implementation. Necessary and sufficient conditions under which the sequence MAP detector is not useful are derived. These results are in the same spirit as previous results on MAP detection of Markov sources over discrete memoryless channels [2, 9]. Simulation results for different values of the source and channel parameters, as well as for different orders of the Markov noise process ($M = 1, 2, 5$) indicate an improvement in the performance of the MAP detectors as the channel capacity increases. We also show that for channels with relatively high bit error rates (e.g., digital cellular channels) the performance of this system (with low complexity) is superior to that of a traditional tandem source-channel coding scheme where the source and channel codes are separately designed with the assumption that the Markov channel is rendered memoryless by means of an interleaver and de-interleaver.

We next analyze the same detection problem with the variation that the source is a binary symmetric stationary ergodic Markov process. In this case, the redundancy in the source is introduced by the Markov dependence between successive source symbols. The two MAP formulations above (sequence and instantaneous) are also studied for this system. As for the case of the binary iid source, the two MAP detection problems can be implemented using a modified version of the Viterbi decoding algorithm and a recursive algorithm. For the case of $M = 1$, we establish a necessary and sufficient condition under which the sequence MAP detector does not offer any improvement. The condition establishes the existence of a *mismatch* between the binary symmetric Markov source and the Markov channel; this causes a deterioration in the performance of the sequence MAP detector as the noise correlation parameter (and hence channel capacity) increases. This is illustrated by simulation results for the sequence and instantaneous MAP decoders. We reduce the mismatch (which is significant for high values of the noise correlation parameter) by the use of a simple rate-one convolutional encoder, where by rate one, we mean that the encoder outputs as many symbols as it accepts.

The purpose of the convolutional encoder is to convert the symmetric Markov source into a non-uniform iid random process, by transforming its redundancy from the form of memory into redundancy in the form of non-uniform distribution. Simulation results showing considerable improvement by the use of this simple code are obtained.

We finally generalize our detection problem by assuming that the source is a binary non-symmetric stationary ergodic Markov process, hence containing redundancy both in the form of memory and non-uniform distribution. General conditions for the uselessness of the sequence MAP detector are derived. These conditions narrow down to the conditions obtained for the special cases of non-uniform iid source and symmetric Markov source respectively. Through simulation, it is portrayed that the non-symmetric Markov source behaves like a non-uniform iid source when $\rho_D >> \rho_M$.

2 Channel Model

Consider a discrete channel with memory, with common input, noise and output binary alphabet and described by the following equation: $Y_n = X_n \oplus Z_n$, for $n = 1, 2, 3, \ldots$ where:

- $\oplus$ represents the addition operation modulo 2.
- The random variables X_n, Z_n and Y_n are respectively the input, noise and output of the channel.
- $\{X_n\} \perp \{Z_n\}$, i.e. the input and noise sequences are independent from each other.
- The noise process $\{Z_n\}_{n=1}^{\infty}$ is a homogeneous stationary mixing (hence ergodic) Markov process of order M. By this we mean that the noise sample at time n, Z_n, depends statistically on the previous noise samples $(Z_{n-M}, \ldots, Z_{n-2}, Z_{n-1})$, i.e.,

$$Pr\{Z_n = e_n | Z_1 = e_1, \ldots, Z_{n-1} = e_{n-1}\} = Pr\{Z_n = e_n | Z_{n-M} = e_{n-M}, \ldots, Z_{n-1} = e_{n-1}\}.$$

We assume that the process $\{Z_n\}$ is generated by the finite-memory contagion urn scheme derived in [1]. The marginal distribution of the noise process or the channel bit error rate (BER) is then given by

$$Pr\{Z_n = 1\} = \epsilon = 1 - Pr\{Z_n = 0\}.$$

We assume that $\epsilon < 1/2$. Furthermore, its transition probability is governed according to

$$Pr\{Z_n = 1 | Z_{n-M} = e_{n-M}, \ldots, Z_{n-1} = e_{n-1}\} = \frac{\epsilon + (e_{n-1} + e_{n-2} + \cdots + e_{n-M})\delta}{1 + M\delta},$$

where $e_i = 0$ or 1, for $i = 1, 2, \ldots, n-1$ and where $n \geq M+1$. Here, δ is a positive parameter which determines the amount of correlation in $\{Z_n\}$. The correlation coefficient of the noise process is $\frac{\delta}{1+\delta}$. Note that if $\delta = 0$, the noise process $\{Z_n\}$ becomes independent and identically distributed (iid) and the resulting additive noise channel becomes a memoryless binary symmetric channel (BSC).

For the case of $M = 1$, we denote the transition probabilities $Pr\{Z_n = e_n | Z_{n-1} = e_{n-1}\}$ by $Q(e_n | e_{n-1})$:

$$Q = \begin{pmatrix} Q(0|0) & Q(1|0) \\ Q(0|1) & Q(1|1) \end{pmatrix} = \begin{pmatrix} \frac{1-\epsilon+\delta}{1+\delta} & \frac{\epsilon}{1+\delta} \\ \frac{1-\epsilon}{1+\delta} & \frac{\epsilon+\delta}{1+\delta} \end{pmatrix}.$$

We note that the transition matrix of this first order Markov model is *general*; it can represent any first order binary Markov chain with positive[1] transition matrix.

A. The Distribution of the Noise

For an input block $X^n = (X_1, X_2, \ldots, X_n)$ and an output block $Y^n = (Y_1, Y_2, \ldots, Y_n)$, the block transition probability of the resulting binary channel is as follows [1]:

- For blocklength $n \leq M$, we have

$$Pr\{Y^n = y^n \mid X^n = x^n\} = L(n, d, \epsilon, \delta), \qquad (1)$$

where

$$L(n, d, \epsilon, \delta) = \frac{\left[\prod_{i=0}^{d-1}(\epsilon + i\delta)\right] \left[\prod_{j=0}^{n-d-1}(1 - \epsilon + j\delta)\right]}{\left[\prod_{l=0}^{n-1}(1 + l\delta)\right]}, \qquad (2)$$

and $d = d_H(y^n, x^n)$ is the Hamming distance between x^n and y^n.

- For $n \geq M+1$, we obtain

$$\begin{aligned} Pr\{Y^n = y^n \mid X^n = x^n\} &= Pr\{Z^n = e^n\} \\ &= \prod_{i=1}^{n} Pr\{Z_i = e_i \mid Z_{i-1} = e_{i-1}, \ldots, Z_{i-M} = e_{i-M}\} \\ &= L(M, k, \epsilon, \delta) \prod_{i=M+1}^{n} \left[\frac{\epsilon + s_{i-1}\delta}{1 + M\delta}\right]^{e_i} \left[\frac{1 - \epsilon + (M - s_{i-1})\delta}{1 + M\delta}\right]^{1-e_i} \qquad (3) \end{aligned}$$

[1]A positive transition matrix is a matrix whose entries are all strictly positive.

where

$$e_i = x_i \oplus y_i,$$

$$k = e_1 + \cdots + e_M,$$

and

$$s_{i-1} = e_{i-1} + \cdots + e_{i-M}.$$

B. Capacity of the Channel

The capacity C_M of this additive Markov channel is increasing with the memory M and is given by [1]:

$$C_M = 1 - \sum_{k=0}^{M} \binom{M}{k} L(M,k,\epsilon,\delta)\; h_b\left(\frac{\epsilon + k\delta}{1 + M\delta}\right), \tag{4}$$

where $h_b(x) = -x\log_2(x) - (1-x)\log_2(1-x)$ is the binary entropy function. Note that C_M is also increasing with the correlation parameter δ.

3 Detection of Binary I.I.D. Sources

Let $\{X_n\}_{n=1}^{\infty}$ be a binary iid non-uniform source with probability distribution $Pr\{X_n = 0\} = p = 1 - Pr\{X_n = 1\}$, where $p > 1/2$. We denote $Pr\{X_n = 0\}$ and $Pr\{X_n = 1\}$ by $P(0)$ and $P(1)$, respectively. We investigate the problem of detecting this source when it is transmitted across an additive Markov channel of order one ($M = 1$). The receiver we design is a *maximum a posteriori* (MAP) detector which is optimum in the sense of minimizing the probability of error [10]. We present two MAP formulations:

- A *sequence* MAP detection which involves a large delay and minimizes the sequence probability of error.
- An *instantaneous* MAP detection which involves no delay and minimizes the bit probability of error.

3.1 Implementation of the MAP Detectors

A. Sequence MAP Decoding

Given that we observe $Y^n = y^n = (y_1, y_2, \ldots, y_n)$ at the output of the channel, we desire to determine the most probable transmitted *sequence* $\hat{x}^n$ where

$$\hat{x}^n = \arg \max_{x^n \in \{0,1\}^n} Pr\{X^n = x^n | Y^n = y^n\}. \tag{5}$$

But (5) is equivalent to

$$\begin{aligned}
\hat{x}^n &= \arg \max_{x^n \in \{0,1\}^n} Pr\{Y^n = y^n | X^n = x^n\} Pr\{X^n = x^n\} \\
&= \arg \max_{x^n \in \{0,1\}^n} Pr\{Z^n = y^n \oplus x^n\} Pr\{X^n = x^n\} \\
&= \arg \max_{x^n \in \{0,1\}^n} \left[Pr\{Z_1 = y_1 \oplus x_1\} \prod_{k=2}^{n} Pr\{Z_k = y_k \oplus x_k | Z_{k-1} = y_{k-1} \oplus x_{k-1}\} \right] \\
&\qquad \times \left[\prod_{k=1}^{n} Pr\{X_k = x_k\} \right]. \qquad (6)
\end{aligned}$$

Since the logarithm function is monotonic, the above equation is equivalent to

$$\hat{x}^n = \arg \max_{x^n \in \{0,1\}^n} \left[\log \left(Pr\{Z_1 = x_1 \oplus y_1\} P(x_1) \right) + \sum_{k=2}^{n} \log \left(Q(y_k \oplus x_k | y_{k-1} \oplus x_{k-1}) P(x_k) \right) \right]. \tag{7}$$

As expressed in equation (7), the sequence MAP detector can be implemented using the Viterbi algorithm [3]. We let x^n be the state sequence. The trellis has two states, with two branches leaving and entering each state. For a branch leaving state x_{k-1} at time $k-1$ and entering state x_k at time k, the path metric is $\log \left(Q(y_k \oplus x_k | y_{k-1} \oplus x_{k-1}) P(x_k) \right)$. The surviving path for each state is the path with the largest cumulative metric up to that state.

The sequence MAP detector involves a large delay since it needs to observe the entire sequence y^n at the output of the channel in order to estimate x_1. For a given sequence length n, this detector minimizes the sequence probability of error.

B. Instantaneous MAP Decoding

Unlike the sequence MAP detector, the instantaneous MAP detector minimizes the bit probability of error. Furthermore, it carries no delay; it decodes x_n as soon as it observes y_n. Here, the problem is to determine the most probable transmitted *bit* $\hat{x}_n$ where

$$\hat{x}_n = \arg \max_{x_n \in \{0,1\}} Pr\{X_n = x_n | Y^n = y^n\}. \tag{8}$$

Solving the equation above is equivalent to solving

$$\hat{x}_n = \arg \max_{x_n \in \{0,1\}} Pr\{X_n = x_n, Y^n = y^n\}. \tag{9}$$

Let $f^{(n)}(x_n) \triangleq Pr\{X_n = x_n, Y^n = y^n\}$ denote the objective function[2] that we wish to maximize at time instant n. We can rewrite $f^{(n)}(x_n)$ as

$$\begin{aligned}
f^{(n)}(x_n) &= \sum_{x^{n-1} \in \{0,1\}^{n-1}} Pr\{X^n = x^n, Y^n = y^n\} \\
&= \sum_{x^{n-1} \in \{0,1\}^{n-1}} Pr\{Z^n = y^n \oplus x^n\} Pr\{X^n = x^n\} \\
&= \sum_{x^{n-1} \in \{0,1\}^{n-1}} \left[\prod_{k=2}^{n} Q(y_k \oplus x_k | y_{k-1} \oplus x_{k-1}) P(x_k)\right] [Pr\{Z_1 = y_1 \oplus x_1\} P(x_1)] \\
&= P(x_n) \sum_{x_{n-1} \in \{0,1\}} Q(y_n \oplus x_n | y_{n-1} \oplus x_{n-1}) \\
&\quad \times \sum_{x^{n-2} \in \{0,1\}^{n-2}} \left[\prod_{k=2}^{n-1} Q(y_k \oplus x_k | y_{k-1} \oplus x_{k-1}) P(x_k)\right] [Pr\{Z_1 = y_1 \oplus x_1\} P(x_1)].
\end{aligned}$$

If we examine the last two equations above, we realize that the second summation in the last equation is nothing but $f^{(n-1)}(x_{n-1})$. Therefore, the instantaneous MAP detector can be implemented using the following recursion:

$$f^{(1)}(x_1) = Pr\{Z_1 = y_1 \oplus x_1\} P(x_1), \tag{10}$$

$$f^{(n)}(x_n) = P(x_n) \sum_{x_{n-1} \in \{0,1\}} Q(y_n \oplus x_n | y_{n-1} \oplus x_{n-1}) f^{(n-1)}(x_{n-1}), \quad n = 2, 3, \ldots, \tag{11}$$

where

[2]Of course, $f^{(n)}(x_n)$ depends also on the observed sequence y^n. Yet, we omit showing this dependency explicitly for notational simplicity.

$$\hat{x}_n = \arg \max_{x_n \in \{0,1\}} f^{(n)}(x_n), \qquad n = 1, 2, 3, \ldots. \tag{12}$$

We note that there exists a gap between the sequence and instantaneous MAP detectors: one has large delay while the other has no delay. In [9], a detector whose delay is in between that of the sequence and instantaneous MAP detectors, is proposed. This detector mitigates the large delay problem of the sequence MAP detector at the cost of some performance loss.

3.2 Analytical Results for Sequence MAP Detection

When we use a MAP detector, a natural question arises: Under what conditions on the system parameters, will the MAP detector not be useful; that is, it will perform no better than if we use no decoder at all at the channel output ("accept what you see" decoding rule) ?

Since we are dealing with a channel with memory, it is difficult (if not impossible) to completely answer this question in an analytical manner, particularly for the case where the decoder is an instantaneous MAP detector. We nevertheless shed some light on this question for the case of the sequence MAP detector, with the following results.

In the following theorem, we will need to assume that the first transmitted bit is not affected by the channel noise. This is known *a priori* by the MAP detector. Thus, the MAP detector will assume that the first bit is received without error. Any such restriction on the first bit will only have a diminishingly small effect on the system performance as the sequence length becomes large.

Theorem 1 Given $p \in (\frac{1}{2}, 1)$, $\epsilon \in (0, \frac{1}{2})$, and $\delta \geq 0$, assume that the channel noise does not affect the first bit (i.e., $X_1 = Y_1$ almost surely). If

$$\left[\frac{1-\epsilon+\delta}{\epsilon+\delta}\right] \left[\frac{1-p}{p}\right] \geq 1. \tag{13}$$

then $\hat{X}^n = Y^n$, $n \geq 2$, is an optimum sequence (MAP) detection rule.

Remark: Note that (13) is equivalent to

$$\delta \leq \delta_1 \triangleq \frac{1-\epsilon-p}{2p-1}, \tag{14}$$

which holds only if $1 - \epsilon \geq p$.

Proof 1 We need to show that if (13) holds, then $\forall x^n, y^n \in \{0,1\}^n$ with $x_1 = y_1$,

$$Pr\{X^n = y^n | Y^n = y^n\} \geq Pr\{X^n = x^n | Y^n = y^n\},$$

or

$$\alpha \triangleq \frac{Pr\{X^n = y^n | Y^n = y^n\}}{Pr\{X^n = x^n | Y^n = y^n\}} \geq 1.$$

Using the noise distribution given in (3), we can write α as

$$\begin{aligned} \alpha &= \frac{Pr\{Y^n = y^n | X^n = y^n\} Pr\{X^n = y^n\}}{Pr\{Y^n = y^n | X^n = x^n\} Pr\{X^n = x^n\}} \\ &= \left[\frac{Pr\{Z_1 = 0\}}{Pr\{Z_1 = x_1 \oplus y_1\}} \, \frac{P(y_1)}{P(x_1)}\right] \left[\prod_{k=2}^{n} \frac{Q(0|0)}{Q(e_k|e_{k-1})} \, \frac{P(y_k)}{P(x_k)}\right], \end{aligned} \tag{15}$$

where $e_k = x_k \oplus y_k$, $k = 1, 2, \ldots, n$. Note that the first factor above is unity since $x_1 = y_1$. Defining $\tilde{Q}(e_k|e_{k-1}) \triangleq (1+\delta) Q(e_k|e_{k-1})$, [i.e., $\tilde{Q}(0|0) = 1 - \epsilon + \delta$, $\tilde{Q}(0|1) = 1 - \epsilon$, $\tilde{Q}(1|0) = \epsilon$ and $\tilde{Q}(1|1) = \epsilon + \delta$], we get

$$\alpha = \prod_{k=2}^{n} \left[\frac{1 - \epsilon + \delta}{\tilde{Q}(e_k|e_{k-1})} \, \frac{P(y_k)}{P(x_k)}\right]. \tag{16}$$

We define

$$\mathcal{K} = \{2, 3, \ldots, n\}, \tag{17}$$

$$\mathcal{A} = \{k \in \mathcal{K} : x_k = y_k, x_{k-1} = y_{k-1}\}, \tag{18}$$

$$\mathcal{B} = \mathcal{A}^c = \{k \in \mathcal{K} : k \notin \mathcal{A}\} = \{k \in \mathcal{K} : e_k = 1 \text{ or } e_{k-1} = 1\}. \tag{19}$$

Therefore, equation (16) becomes

$$\alpha = \prod_{k \in \mathcal{B}} \left[\frac{1 - \epsilon + \delta}{\tilde{Q}(e_k|e_{k-1})} \, \frac{P(y_k)}{P(x_k)}\right]. \tag{20}$$

Case 1: $e_n = 0$

If $e_n = 0$, we partition $\mathcal{B}$ as follows:

$$\mathcal{B} = \bigcup_{i=1}^{N} \mathcal{B}_i; \qquad \mathcal{B}_i \bigcap \mathcal{B}_j = \emptyset, \quad i \neq j, \ i,j = 1,2,\ldots,N,$$

where

$$\mathcal{B}_i = \{m_i + 1, m_i + 2, \ldots, m_i + L_i\},$$

$$e_{m_i} = e_{m_i+L_i} = 0,$$

$$e_{m_i+1} = e_{m_i+2} = \cdots = e_{m_i+L_i-1} = 1,$$

with N denoting the number of partition sets and L_i denoting the cardinality of $\mathcal{B}_i$ ($L_i = |\mathcal{B}_i|$). To illustrate the partition above, consider for $n = 24$ the noise sequence

$$e^n = (000111001100001111010000).$$

Then $\mathcal{B} = \{4, 5, 6, 7, 9, 10, 11, 15, 16, 17, 18, 19, 20, 21\}$ and its partitioning sets are $\mathcal{B}_1 = \{4, 5, 6, 7\}$, $\mathcal{B}_2 = \{9, 10, 11\}$, $\mathcal{B}_3 = \{15, 16, 17, 18, 19\}$ and $\mathcal{B}_4 = \{20, 21\}$. Here $N = 4$, $L_1 = 4$, $L_2 = 3$, $L_3 = 5$ and $L_4 = 2$.

Note that $L_i \geq 2 \quad \forall i \in \{1, 2, \ldots, N\}$ and thus

$$|\mathcal{B}| = \sum_{i=1}^{N} |\mathcal{B}_i| = \sum_{i=1}^{N} L_i \geq 2N, \tag{21}$$

with equality if and only if the errors occur in isolation (i.e., if and only if $L_1 = L_2 = \cdots = L_N = 2$).

We now can rewrite (20) as

$$\alpha = \prod_{i=1}^{N} \alpha_i, \tag{22}$$

where

$$\begin{aligned} \alpha_i &= \prod_{k \in \mathcal{B}_i} \left[\frac{1 - \epsilon + \delta}{\tilde{Q}(e_k | e_{k-1})} \frac{P(y_k)}{P(x_k)} \right] = \prod_{k=m_i+1}^{m_i+L_i} \left[\frac{1 - \epsilon + \delta}{\tilde{Q}(e_k | e_{k-1})} \frac{P(y_k)}{P(x_k)} \right] \\ &= \frac{1 - \epsilon + \delta}{\tilde{Q}(e_{m_i+1} | e_{m_i})} \frac{P(y_{m_i+1})}{P(x_{m_i+1})} \left[\prod_{k=m_i+2}^{m_i+L_i-1} \frac{1 - \epsilon + \delta}{\tilde{Q}(e_k | e_{k-1})} \frac{P(y_k)}{P(x_k)} \right] \frac{1 - \epsilon + \delta}{\tilde{Q}(e_{m_i+L_i} | e_{m_i+L_i-1})} \frac{P(y_{m_i+L_i})}{P(x_{m_i+L_i})}. \end{aligned}$$

Noting that $p > \frac{1}{2}$ and that $(e_{m_i}, e_{m_i+1}, e_{m_i+2}, \ldots, e_{m_i+L_i-1}, e_{m_i+L_i}) = (0,1,1,\ldots,1,0)$, we obtain

$$\alpha_i \geq \left[\frac{1-\epsilon+\delta}{\epsilon}\,\frac{1-p}{p}\right]\left[\frac{1-\epsilon+\delta}{\epsilon+\delta}\,\frac{1-p}{p}\right]^{L_i-2}\left[\frac{1-\epsilon+\delta}{1-\epsilon}\right] \tag{23}$$

with equality if and only if $(x_{m_i+1}, x_{m_i+2}, \ldots, x_{m_i+L_i-1}) = (0,0,\ldots,0)$. Substituting (23) in (22) yields

$$\begin{aligned}\alpha &= \prod_{i=1}^{N}\alpha_i \\ &\geq \left[\frac{1-\epsilon+\delta}{\epsilon}\,\frac{1-p}{p}\right]^{N}\left[\frac{1-\epsilon+\delta}{\epsilon+\delta}\,\frac{1-p}{p}\right]^{|\mathcal{B}|-2N}\left[\frac{1-\epsilon+\delta}{1-\epsilon}\right]^{N}.\end{aligned} \tag{24}$$

The second factor above is ≥ 1 by hypothesis and by (21). The last factor is ≥ 1 since $\delta \geq 0$. Finally, the first factor is ≥ 1 since $\frac{1}{\epsilon} \geq \frac{1}{\epsilon+\delta}$. Therefore $\alpha \geq 1$.

Case 2: $e_n = 1$

If $e_n = 1$, we again partition $\mathcal{B}$ as follows:

$$\mathcal{B} = \bigcup_{i=1}^{N}\mathcal{B}_i; \qquad \mathcal{B}_i \bigcap \mathcal{B}_j = \emptyset, \quad i \neq j, \; i,j = 1,2,\ldots,N,$$

where $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_{N-1}$ are defined as in Case 1, and $\mathcal{B}_N$ is defined as:

$$\mathcal{B}_N = \{n - L_N + 1, n - L_N + 2, \ldots, n\},$$

such that $e_{n-L_N} = 0$ and $e_{n-L_N+1} = e_{n-L_N+2} = \cdots = e_n = 1$. For $i = 1,2,\ldots,N-1$, we can lower bound $\alpha_i \triangleq \prod_{k\in\mathcal{B}_i}\left[\frac{1-\epsilon+\delta}{\tilde{Q}(e_k|e_{k-1})}\,\frac{P(y_k)}{P(x_k)}\right]$ by expression (23). For $i = N$, we get the following lower bound on α_N

$$\alpha_N = \prod_{k\in\mathcal{B}_N}\left[\frac{1-\epsilon+\delta}{\tilde{Q}(e_k|e_{k-1})}\,\frac{P(y_k)}{P(x_k)}\right] \geq \left[\frac{1-\epsilon+\delta}{\epsilon}\,\frac{1-p}{p}\right]\left[\frac{1-\epsilon+\delta}{\epsilon+\delta}\,\frac{1-p}{p}\right]^{L_N-1},$$

with equality if and only if $(x_{n-L_N+1}, x_{n-L_N+2}, \ldots, x_n) = (0,0,\ldots,0)$. We then get

$$\begin{aligned}\alpha &= \prod_{i=1}^{N} \alpha_i \\ &\geq \left[\frac{1-\epsilon+\delta}{\epsilon}\,\frac{1-p}{p}\right]^{N} \left[\frac{1-\epsilon+\delta}{\epsilon+\delta}\,\frac{1-p}{p}\right]^{|\mathcal{B}|-(2N-1)} \left[\frac{1-\epsilon+\delta}{1-\epsilon}\right]^{N-1}. \end{aligned} \tag{25}$$

By the same arguments as before and the fact that $|\mathcal{B}| \geq 2N-1$ (with equality if and only if $L_1 = L_2 = \cdots = L_{N-1} = 2$ and $L_N = 1$), each of the three factors above is ≥ 1. Therefore, $\alpha \geq 1$.

■

Note that if $p = \frac{1}{2}$, i.e., if the iid source is uniform (and hence exhibits no redundancy), then we can clearly observe from equation (16) that α is always ≥ 1 (similarly $\delta_1 = +\infty$) and thus the sequence MAP detector is *always* useless. This can also be explained by the fact that if the source contains no redundancy at all, then the MAP detector fails to provide any protection against channel noise. Also, note that if the inequality in (13) is strict then the sequence MAP detector is *unique* and is given by $\hat{X}^n = Y^n$. In the remaining of this paper, we will not emphasize this point since the uniqueness of the MAP detector will be clear from the context.

Theorem 1 is *only* a sufficient condition. A necessary condition could not be derived analytically, except for the following theorem, which is an asymptotic "weak" converse.

Theorem 2 Given $p \in (\frac{1}{2}, 1)$ and $\epsilon \in (0, \frac{1}{2})$, if

$$\left[\frac{1-\epsilon+\delta}{\epsilon+\delta}\right] \left[\frac{1-p}{p}\right] < 1, \tag{26}$$

then $\exists\ n_0 > 0$ sufficiently large, such that for $n \geq n_0$, $\hat{X}^n = Y^n$ is not an optimal detection rule.

Proof 2 The result is proved by a counterexample: we show that $\exists\ n_0, x^n$ and y^n such that $\alpha < 1$ for $n \geq n_0$.
Let $x^n = (0, 0, 0, \ldots, 0)$ and $y^n = (0, 1, 1, \ldots, 1)$. Using these two sequences in (16), we obtain

$$\alpha = \left[\frac{1-p}{p}\right]^{n-1} \left[\frac{1-\epsilon+\delta}{\epsilon}\right] \left[\frac{1-\epsilon+\delta}{\epsilon+\delta}\right]^{n-2}.$$

Now by hypothesis, $\exists\ n_0$, such that $\forall\ n \geq n_0$,

$$\left[\frac{1-p}{p}\right]^{n-2} \left[\frac{1-\epsilon+\delta}{\epsilon+\delta}\right]^{n-2} < \left[\frac{p}{1-p}\right] \left[\frac{\epsilon}{1-\epsilon+\delta}\right],$$

and thus $\forall\ n \geq n_0$, $\alpha < 1$. ■

We refered to the above theorem as an asymptotic "weak" converse, since the counterexample we provided in proving the result, utilizes input and output sequences which yield a non-typical noise sequence (i.e., the sequence can occur with low probability). We now derive a necessary and sufficient condition for which the all-zero sequence is the optimal sequence. We will assume that the first transmitted bit is zero. First we need the following lemma.

Lemma 1 Let $p \in (\frac{1}{2}, 1)$, $\epsilon \in (0, \frac{1}{2})$, $\delta \geq 0$ and $L \geq 2$. Assume that $x_0^L = (x_0, x_1, \ldots, x_L) = (0, 1, 1, \ldots, 1, 0)$. If

$$l \triangleq \frac{\epsilon(1-\epsilon)}{(1-\epsilon+\delta)^2} \, \frac{p}{1-p} \geq 1, \tag{27}$$

then $\forall\ y_0^L \in \{0,1\}^{L+1}$,

$$\alpha \triangleq \prod_{k=1}^{L} \frac{\tilde{Q}(y_k|y_{k-1})}{\tilde{Q}(e_k|e_{k-1})} \, \frac{P(0)}{P(x_k)} \geq 1, \tag{28}$$

where $e_0^L = x_0^L \oplus y_0^L$, $P(0) = p = 1 - P(1)$, $\tilde{Q}(0|0) = 1 - \epsilon + \delta$, $\tilde{Q}(1|0) = \epsilon$, $\tilde{Q}(0|1) = 1 - \epsilon$ and $\tilde{Q}(1|1) = \epsilon + \delta$.

In addition, the above also holds when $x_0^L = (0, 1, 1, \ldots, 1)$.

Proof See Appendix.

Theorem 3 Given $p \in (\frac{1}{2}, 1)$, $\epsilon \in (0, \frac{1}{2})$, $\delta \geq 0$ and $n \geq 3$, assume that $X_1 = 0$ almost surely. Then $\hat{X}^n = 0^n$, is an optimal sequence (MAP) detection rule if and only if

$$\frac{\epsilon(1-\epsilon)}{(1-\epsilon+\delta)^2} \, \frac{p}{1-p} \geq 1. \tag{29}$$

Proof 3 We first show that if (29) holds, then $\forall x^n, y^n \in \{0,1\}^n$ with $x_1 = 0$,

$$\alpha \triangleq \frac{Pr\{Y^n = y^n | X^n = 0^n\} Pr\{X^n = 0^n\}}{Pr\{Y^n = y^n | X^n = x^n\} Pr\{X^n = x^n\}} \geq 1.$$

As in Theorem 1, we rewrite α as

$$\alpha = \left[\frac{Pr\{Z_1 = y_1\} \; P(0)}{Pr\{Z_1 = x_1 \oplus y_1\} \; P(x_1)} \right] \left[\prod_{k=2}^{n} \frac{\tilde{Q}(y_k | y_{k-1})}{\tilde{Q}(e_k | e_{k-1})} \frac{P(0)}{P(x_k)} \right], \tag{30}$$

where $e_k = x_k \oplus y_k$, $k = 2, 3, \ldots, n$. The first factor in the above expression is unity since $x_1 = 0$. Analogous to Theorem 1, we write $\mathcal{K} = \{2, 3, \ldots, n\}$, $\mathcal{B} = \{k \in \mathcal{K} : x_k = 1 \text{ or } x_{k-1} = 1\}$ and partition $\mathcal{B}$ as

$$\mathcal{B} = \bigcup_{i=1}^{N} \mathcal{B}_i; \qquad \mathcal{B}_i \bigcap \mathcal{B}_j = \emptyset, \quad i \neq j, \; i, j = 1, 2, \ldots, N,$$

where

1. $\mathcal{B}_i = \{m_i + 1, m_i + 2, \ldots, m_i + L_i\}$,

2. $x_{m_i} = x_{m_i + L_i} = 0$,

3. $x_{m_i+1} = x_{m_i+2} = \cdots = x_{m_i+L_i-1} = 1$,

for $i = 1, 2, \ldots, N-1$. For $i = N$, conditions $(1) - (3)$ above hold with the exception that $x_{m_N + L_N}$ may be 0 or 1. In any case, we write $\alpha = \prod_{i=1}^{N} \alpha_i$ where

$$\alpha_i = \prod_{k \in \mathcal{B}_i} \frac{\tilde{Q}(y_k | y_{k-1})}{\tilde{Q}(e_k | e_{k-1})} \frac{P(0)}{P(x_k)}.$$

For each i, we shift the index (in $\mathcal{B}_i$) by m_i, let $L = L_i$ and apply the above Lemma to show that $\alpha_i \geq 1$. Therefore, $\alpha \geq 1$.

To prove the converse, assume that (29) does not hold and let $x^n = y^n = (0, 0, \ldots, 0, 1, 0, \ldots, 0)$. Then

$$\alpha = \frac{\epsilon(1-\epsilon)}{(1-\epsilon+\delta)^2} \frac{p}{1-p} < 1.$$

■

Observation: In the case of sequence MAP detection, we obtained the estimate of the transmitted source sequence x^n given that we received y^n at the channel output. It is pertinent to remark that this is the same as adding the received sequence y^n to the estimate of the noise sequence (given y^n); i.e.,

$$\arg \max_{x^n \in \{0,1\}^n} Pr\{X^n = x^n | Y^n = y^n\} = y^n \oplus \arg \max_{z^n \in \{0,1\}^n} Pr\{Z^n = z^n | Y^n = y^n\}.$$

3.3 Simulation Results

In Figures 2-7, simulation results for the sequence and instantaneous MAP detectors are plotted. Each simulation was performed on 1,000 samples of the iid source and the experiment was repeated 500 times. In Figures 2-6 , $\hat{\epsilon} = \Pr\{\text{bit error}\} = Pr\{\hat{X}_n \neq X_n\}$ and the average values of $\hat{\epsilon}$ (over the 500 experiments) are plotted versus the channel bit error rate ϵ. The straight line labeled "w/o MAP" indicates the probability of bit error when no MAP detection is performed (i.e., $\hat{\epsilon} = \epsilon$). Figure 7 shows the plot of the average value of $\hat{\epsilon}$ versus the channel correlation parameter δ.

In Figures 2, 3 and 7, the performances of sequence MAP detection for the iid source (with $p = 0.95$ and 0.99 respectively) over the contagion Markov channel (with $M = 1$), are presented. We can remark in each of the figures, that as δ increases, the performance of the MAP detector improves. This is due to the fact that as δ increases, the noise correlation in the channel increases (hence decreasing the noise entropy rate and increasing the channel capacity) which enhances the detector's capability in estimating the transmitted sequence. We furthermore observe that for fixed δ and ϵ, the gain achieved in Figure 3 is higher than the one attained in Figure 2. This can be explained by the fact that the iid source in Figure 3 (with $p = 0.99$) is more non-uniform than the iid source in Figure 2 (with $p = 0.95$); it thus contains a larger amount of redundancy which is used by the MAP detector to combat the channel noise.

In Figures 4 and 5, the performances of instantaneous MAP detection for the iid source (with $p = 0.95$ and 0.99 respectively) over the contagion Markov channel are presented. As before, the performance for the source with $p = 0.99$ is better than the one for $p = 0.95$. Figure 6 shows the effect of the order M of the Markov channel on the performance of the sequence MAP detection for a source with $p = 0.95$ and a channel with $\delta = 1.0$. Here again, the performance improves as M increases, since the channel capacity increases with the memory M. In the implementation of the sequence MAP detector, the state at time k consists of the vector $(x_k, x_{k-1}, \ldots, x_{k-M+1})$. Thus, there are 2^M states with two branches entering and leaving each state.

The analytical results of Theorems 1 and 3 are illustrated in Figures 2, 3 and 7. Theorem 1 is illustrated in Figure 7 where the performance is given versus the values of δ. With $p = 0.95$ and $\epsilon = 0.01$, the sufficient range on δ for which the MAP detector is useless (i.e., $\hat{\epsilon} = \epsilon = 0.01$), is $\delta \leq \delta_1 = 0.0444 = 10^{-1.35}$. As we can note from Figure 7, the curve of $\hat{\epsilon}$ diverges from the constant value of $\epsilon = 10^{-2}$ for a δ larger than δ_1; this is because Theorems 1 offers only a sufficient condition. The asymptotic converse given in Theorem 2 relies on a non-typical noise sequence; it thus has a low chance of occurring in a simulation.

Furthermore, the simulations shown in Figures 2 and 3 agree with Theorem 3. Theorem 3 offers a necessary and sufficient condition for which the all-zero sequence is the optimal sequence. Note that (29) is equivalent to

$$\epsilon \geq \epsilon^{(\delta)} = \frac{1}{2}\left[1 + (1-p)(1+2\delta) - \sqrt{p^2 - 4p\delta(1-p)(1+\delta)}\right].$$

We plot in Figures 2 and 3 the values of $\epsilon^{(0.5)}$, $\epsilon^{(1)}$ and $\epsilon^{(2)}$. Note that the performance curves flatten out exactly at $\epsilon = \epsilon^{(\delta)}$ and $\hat{\epsilon} = 1 - p$. Finally, note that for $\delta = 10$ and $p \geq 0.95$, the all-zero sequence is never the optimal sequence since (29) does not hold.

3.4 Comparison with Tandem Source-Channel Coding Schemes

The system we design in this paper, which consists of directly transmitting the source over the Markov channel and utilizing a MAP detector at the receiver, is in a "loose" sense, a joint source-channel scheme. We now compare this proposed system against the traditional tandem source-channel coding scheme where the source and channel codes are designed separately.

The traditional approach to handling a channel with memory is to use an interleaver and a de-interleaver. The purpose of the interleaver and de-interleaver is to convert the channel with memory into a memoryless channel. This is because most well-known channel codes are designed for the memoryless channel. The tandem scheme considered includes an interleaver/de-interleaver pair as it is depicted in Figure 1. It consists of the following

- Huffman encoder: We assume that the binary iid source has distribution $p = 0.9$; thus its entropy rate is $h_b(0.9) = 0.469$ bits/sample. Grouping the source binary stream in blocks of 4 bits, we encode the source stream using a 4'th order Huffman code with average code length of 0.49255 bits/sample.

- Convolutional Encoder: We match the output of the Huffman encoder to a convolutional encoder of rate $R = \frac{k}{n} = \frac{1}{2}$. The convolutional code has an input memory $m = 2$ (4 states) and the following tap coefficients $(1,0,1)$ and $(1,1,1)$ [6]. Its minimum free distance is $d_{free} = 5$, its minimum distance is $d_{min} = 3$ and its constraint length is $n(m+1) = 6$ bits.

- Interleaver, Markov channel, de-interleaver: This renders the channel memoryless; i.e., it transforms the bursts of errors in the Markov channel into isolated errors and thus enhances the error correction capability of the convolutional code.

- Decoders: ML decoder implemented using the Viterbi algorithm [3, 6] and a Huffman decoder.

It is pertinent to point out that the complexity of the proposed system is substantially lower than that of the tandem scheme: the tandem scheme contains 2 decoders (Viterbi and Huffman), 2 encoders, an interleaver and a de-interleaver, while the proposed system contains only a MAP decoder. Furthermore, the use of the interleaver/de-interleaver in the tandem scheme introduces a larger delay than in the proposed system.

In Figure 8, we compare the performance of the proposed scheme using sequence MAP detection for $p = 0.9$ and $\delta = 10$, with that of two tandem schemes with interleaving lengths $L = 100$ and $L = 1$ (no interleaving) respectively. We use the same interleaving procedure as in [6]. The simulations were run 50 times on 10,000 samples of the iid source. We observe that the non-interleaved tandem scheme ($L = 1$) behaves very badly; this is expected because the convolutional code is designed for a *memoryless* channel, and hence our need for interleaving. Indeed, the tandem scheme using interleaving (with $L = 100$) performs much better than the non-interleaved scheme.

More importantly, we remark that the proposed scheme outperforms the tandem scheme with $L = 100$ when the channel bit error rate is high ($\epsilon \geq 10^{-2}$). The performance of the tandem scheme is excellent for very low values of ϵ (all simulation errors are corrected for $\epsilon < 10^{-2}$). However, as the channel becomes more noisy, the tandem scheme breaks down; this is due to the effect of error propagation in the Huffman decoder. This suggests to us that for noisy channels with relatively high bit error rates ($\epsilon \geq 10^{-2}$) (e.g., cellular channels), the proposed system beats the tandem scheme while being substantially less complex.

4 Detection of Binary Markov Sources

4.1 Symmetric Markov Sources

In Section 3, the source is assumed to be a non-uniform iid random process. Therefore, the redundancy it contains is strictly in the form of a non-uniform distribution ($\rho_D > 0$ and $\rho_M = 0$). We now consider the same problem studied in Section 3, with the exception that the binary source is a symmetric stationary ergodic Markov process. In this case, the source redundancy is strictly in the form of memory ($\rho_M > 0$ and $\rho_D = 0$). We assume that the Markov source has a uniform marginal distribution ($Pr\{X_n = 0\} = Pr\{X_n = 1\} = 1/2$) and the following transition probabilities:

$$Pr\{X_n = 1|X_{n-1} = 1\} = Pr\{X_n = 0|X_{n-1} = 0\} = q,$$

and

$$Pr\{X_n = 1|X_{n-1} = 0\} = Pr\{X_n = 0|X_{n-1} = 1\} = 1 - q,$$

where $q > 1/2$. For notational simplicity we denote $Pr\{X_n = u|X_{n-1} = v\}$ by $P(u|v)$.

As for the iid case, we present two MAP formulations: a sequence MAP detection and an instantaneous MAP detection. The implementation of both MAP detectors is exactly the same as the ones described by equations (7), (10) and (11) in Section 3.1 upon replacing $P(x_1)$ by 1/2 in (7) and (10), $P(x_k)$ by $P(x_k|x_{k-1})$ in (7) and $P(x_n)$ by $P(x_n|x_{n-1})$ in (11) (with $P(x_n|x_{n-1})$ placed inside the summation in (11)).

A. Analytical Results for Sequence MAP Detection

In the following theorem, we derive a necessary and sufficient condition under which the sequence MAP detector is useless.

Theorem 4 Given $q \in (\frac{1}{2}, 1)$, $\epsilon \in (0, \frac{1}{2})$, $\delta \geq 0$ and $n \geq 3$, assume that $X_1 = Y_1$ almost surely. Then $\hat{X}^n = Y^n$, is an optimal sequence (MAP) detection rule if and only if

$$\frac{(1-\epsilon+\delta)^2}{\epsilon(1-\epsilon)} \left(\frac{1-q}{q}\right)^2 \geq 1. \tag{31}$$

Remark: (31) is equivalent to

$$\delta \geq \delta_2 \triangleq \left(\frac{q}{1-q}\right) \sqrt{\epsilon(1-\epsilon)} + \epsilon - 1. \tag{32}$$

Proof 4 We start by proving the direct part; i.e., we need to show that if condition (31) holds, then $\forall\, x^n, y^n \in \{0,1\}^n$ with $x_1 = y_1$,

$$\alpha \triangleq \frac{Pr\{X^n = y^n | Y^n = y^n\}}{Pr\{X^n = x^n | Y^n = y^n\}} \geq 1.$$

Following the same method as in Proof 1 with the same definitions of the sets $\mathcal{K}, \mathcal{A}$ and $\mathcal{B}$ (given by equations (17), (18) and (19) respectively), we have

$$\alpha = \prod_{k \in \mathcal{B}} \left[\frac{1-\epsilon+\delta}{\tilde{Q}(e_k|e_{k-1})} \frac{P(y_k|y_{k-1})}{P(x_k|x_{k-1})} \right],$$

where $\tilde{Q}(0|0) = 1-\epsilon+\delta$, $\tilde{Q}(0|1) = 1-\epsilon$, $\tilde{Q}(1|0) = \epsilon$ and $\tilde{Q}(1|1) = \epsilon+\delta$.

Case 1: $e_n = 0$

If $e_n = 0$, we partition $\mathcal{B}$ in the same way as in Case 1 of Proof 1. We can then write α as

$$\alpha = \prod_{i=1}^{N} \alpha_i, \tag{33}$$

where

$$\begin{aligned} \alpha_i &= \prod_{k \in \mathcal{B}_i} \left[\frac{1-\epsilon+\delta}{\tilde{Q}(e_k|e_{k-1})} \frac{P(y_k|y_{k-1})}{P(x_k|x_{k-1})} \right] = \prod_{k=m_i+1}^{m_i+L_i} \left[\frac{1-\epsilon+\delta}{\tilde{Q}(e_k|e_{k-1})} \frac{P(y_k|y_{k-1})}{P(x_k|x_{k-1})} \right] \\ &= \left[\frac{1-\epsilon+\delta}{\tilde{Q}(e_{m_i+1}|e_{m_i})} \frac{P(y_{m_i+1}|y_{m_i})}{P(x_{m_i+1}|x_{m_i})} \right] \left[\prod_{k=m_i+2}^{m_i+L_i-1} \frac{1-\epsilon+\delta}{\tilde{Q}(e_k|e_{k-1})} \frac{P(y_k|y_{k-1})}{P(x_k|x_{k-1})} \right] \\ &\quad \times \left[\frac{1-\epsilon+\delta}{\tilde{Q}(e_{m_i+L_i}|e_{m_i+L_i-1})} \frac{P(y_{m_i+L_i}|y_{m_i+L_i-1})}{P(x_{m_i+L_i}|x_{m_i+L_i-1})} \right]. \end{aligned} \tag{34}$$

Recalling that $q > \frac{1}{2}$ and that $(e_{m_i}, e_{m_i+1}, e_{m_i+2}, \ldots, e_{m_i+L_i-1}, e_{m_i+L_i}) = (0,1,1,\ldots,1,0)$, we get

$$\alpha_i \geq \left[\frac{1-\epsilon+\delta}{\epsilon} \frac{1-q}{q} \right] \left[\prod_{k=m_i+2}^{m_i+L_i-1} \frac{1-\epsilon+\delta}{\epsilon+\delta} \frac{P(y_k|y_{k-1})}{P(x_k|x_{k-1})} \right] \left[\frac{1-\epsilon+\delta}{1-\epsilon} \frac{1-q}{q} \right].$$

We now note that for $k \in \{m_i+2, m_i+3, \ldots, m_i+L_i-1\}$, $e_k = e_{k-1} = 1$; this implies that $y_k \oplus y_{k-1} = (1 \oplus x_k) \oplus (1 \oplus x_{k-1}) = x_k \oplus x_{k-1}$. Therefore,

$$\frac{P(y_k|y_{k-1})}{P(x_k|x_{k-1})} = 1.$$

Note that the above equation holds because the Markov source is *symmetric*; the result is not true for non-symmetric Markov sources. We then obtain

$$\alpha_i \;\geq\; \left[\frac{(1-\epsilon+\delta)^2}{\epsilon(1-\epsilon)} \left(\frac{1-q}{q}\right)^2\right] \left[\frac{1-\epsilon+\delta}{\epsilon+\delta}\right]^{L_i-2}. \tag{35}$$

Substituting (35) in (33) yields

$$\alpha \;\geq\; \left[\frac{(1-\epsilon+\delta)^2}{\epsilon(1-\epsilon)} \left(\frac{1-q}{q}\right)^2\right]^N \left[\frac{1-\epsilon+\delta}{\epsilon+\delta}\right]^{|\mathcal{B}|-2N}. \tag{36}$$

The first term above is ≥ 1 by hypothesis and the second term is ≥ 1 because $|\mathcal{B}| \geq 2N$ and $1-\epsilon \geq \epsilon$. Thus $\alpha \geq 1$.

Case 2: $e_n = 1$

If $e_n = 1$, we again partition $\mathcal{B}$ as in Case 2 of Proof 1. For $i = 1, 2, \ldots, N-1$, we can similarly lower bound $\alpha_i \triangleq \prod_{k \in \mathcal{B}_i} \left[\frac{1-\epsilon+\delta}{\tilde{Q}(e_k|e_{k-1})} \frac{P(y_k|y_{k-1})}{P(x_k|x_{k-1})}\right]$ by expression (35). For $i = N$, we get the following lower bound on α_N

$$\alpha_N = \prod_{k \in \mathcal{B}_N} \left[\frac{1-\epsilon+\delta}{\tilde{Q}(e_k|e_{k-1})} \frac{P(y_k|y_{k-1})}{P(x_k|x_{k-1})}\right] \geq \left[\frac{1-\epsilon+\delta}{\epsilon} \frac{1-q}{q}\right] \left[\frac{1-\epsilon+\delta}{\epsilon+\delta}\right]^{L_N-1}.$$

Therefore,

$$\begin{aligned}
\alpha \;&\geq\; \left[\frac{1-\epsilon+\delta}{\epsilon} \frac{1-q}{q}\right]^N \left[\frac{1-\epsilon+\delta}{\epsilon+\delta}\right]^{|\mathcal{B}|-(2N-1)} \left[\frac{1-\epsilon+\delta}{1-\epsilon} \frac{1-q}{q}\right]^{N-1} \\
&\geq\; \left[\frac{1-\epsilon+\delta}{\epsilon} \frac{1-q}{q}\right] \left[\frac{(1-\epsilon+\delta)^2}{\epsilon(1-\epsilon)} \left(\frac{1-q}{q}\right)^2\right]^{N-1},
\end{aligned} \tag{37}$$

where the last inequality holds since $|\mathcal{B}| \geq 2N-1$ and $\frac{1-\epsilon+\delta}{\epsilon+\delta} > 1$. The second term in (37) is ≥ 1 by hypothesis. The first term is > 1 because $\frac{1}{\epsilon} > \frac{1}{\sqrt{\epsilon(1-\epsilon)}}$ and because of the hypothesis. Therefore, $\alpha \geq 1$.

To prove the converse, we employ the following counterexample: let $x^n = (0,0,\ldots,0)$ and $y^n = (0,0,\ldots,0,1,0,\ldots,0)$. If (31) does not hold, then we get

$$\alpha = \left[\frac{1-\epsilon+\delta}{\epsilon}\,\frac{1-q}{q}\right]\left[\frac{1-\epsilon+\delta}{1-\epsilon}\,\frac{1-q}{q}\right] < 1.$$

■

Observations:

- Note that if $q = \frac{1}{2}$, then the Markov source becomes a uniformly distributed iid random process with zero redundancy. Thus, as mentioned in Section 3, the sequence MAP detector is *always* useless. This can also be seen from equations (36) and (37) which indicate that $\alpha > 1$ for $q = \frac{1}{2}$.
- For $\delta = 0$, (31) reduces to the necessary and sufficient condition given in [9] for the BSC.
- The necessary and sufficient condition derived in Theorem 4, indicates that for fixed values of ϵ and q (hence for a fixed value of δ_2), as the channel correlation parameter δ increases (hence as the channel capacity increases), the likeliness of the uselessness of the sequence MAP detector increases (cf. (32)). Therefore, the performance of the sequence MAP detector deteriorates with increasing δ; this shows the existence of a *mismatch* between the symmetric Markov source and the contagion Markov channel which prevents the MAP detector to fully exploit the capacity of the channel.

B. Simulation Results

In Figures 9-11, we present simulation results for the sequence and instantaneous MAP detectors. We performed each simulation on 1000 samples of the symmetric Markov source and repeated the experiment 500 times.

As predicted by Theorem 4, we remark from the plots that the performance of the MAP detectors deteriorates as the value of δ increases. This is clearly illustrated in Figure 11, where $\hat{\epsilon}$ increases as a function of δ and then reaches a constant value of $\hat{\epsilon} = \epsilon = 0.01$ (MAP is useless) at the point corresponding to $\delta = \delta_2 = 0.9 = 10^{-0.046}$, as given by Theorem 4. We note that, contrary to our expectation, the curve in Figure 11 is not exactly a zero slope straight line for $\delta \geq \delta_2$; this may be due to some slight inaccuracies in the simulation.

C. Rate-One Convolutional Encoding

If we directly connect a binary symmetric Markov source to the contagion Markov channel, a mismatch occurs between the source and channel as the correlation parameter δ increases. In this section, we attempt to reduce this mismatch by the use of a rate-one convolutional code. More specifically, we attempt to improve the performance of the sequence MAP detector for high values of δ. This is achieved by the use of a simple rate-one convolutional code, where by rate-one, we mean that the convolutional encoder produces as many bits as it receives. The purpose of this code is not to introduce additional redundancy but to transform the redundancy in the symmetric Markov source from the form of memory into redundancy in the form of non-uniform distribution. This is because, if the source redundancy is in the form of non-uniform distribution, no such mismatch occurs between the source and the channel, as we have seen in Section 3.

We employ a rate-one convolutional code described by $V_n = X_n \oplus X_{n-1}$, $n = 1, 2, \ldots$, where $\{X_n\}_{n=1}^{\infty}$ is the symmetric Markov source studied in this section and $\{V_n\}_{n=1}^{\infty}$ represents the output of the convolutional encoder. We assume that $X_0 = 0$ almost surely; that is $V_1 = X_1$. Due to the symmetry in the source, we can easily verify that $Pr\{V_k = v_k\} = q^{v_k}\,(1-q)^{1-v_k}$, where v_k is 0 or 1, $k = 1, 2, \ldots$. Furthermore, we can write

$$\begin{aligned} Pr\{V_k = v_k | V_1 = v_1, \ldots, V_{k-1} = v_{k-1}\} &= Pr\{X_k = e^{(k)} | X_1 = e^{(1)}, \ldots, X_{k-1} = e^{(k-1)}\} \\ &= Pr\{X_k = e^{(k)} | X_{k-1} = e^{(k-1)}\} \\ &= q^{v_k}\,(1-q)^{1-v_k} \\ &= Pr\{V_k = v_k\} \end{aligned}$$

where $e^{(l)} \triangleq v_1 \oplus \ldots \oplus v_l$, $l = 1, 2, \ldots, k$. Therefore, $\{V_n\}$ is a non-uniform iid process with distribution given by $Pr\{V_k = 0\} = q$, where $q > 1/2$.

The new system functions as follows. A sequence of N samples of the symmetric Markov source X^N is fed into the rate-one convolutional encoder. The output of the encoder is then sent over the Markov channel. At the receiver, we use the sequence MAP detector which estimates the most likely transmitted sequence $\hat{V}^N$. The convolutional decoder is described by the following relation $\hat{X}_k = \hat{V}_k \oplus \hat{X}_{k-1}$, $k = 1, 2, \ldots, N$ with $\hat{X}_1 = \hat{V}_1$. We therefore obtain $\hat{X}^N$.

Note however, that decoding errors in the MAP sequence detector cause error propagations in the convolutional decoder, which may be significant if an odd number of decoding errors occur. We limit the effect of the propagation by grouping the N source samples into small blocks of length n.

The performance of this system for $q = 0.99$ and $\delta = 10$ is shown in Figure 12. We performed the simulations on $N = 500,000$ source samples with $N = n \cdot T$ where T is the number of trials and n is the number of source samples transmitted per trial. The results clearly indicate that the coded system outperforms the uncoded system. Furthermore, for large ϵ, the performance of the coded system improves as n decreases, as expected, since for small n the effect of the error propagation in the convolutional decoder is limited.

4.2 Non-Symmetric Markov Sources

We generalize our detection problem by considering a binary non-symmetric stationary ergodic Markov source. In this case, the source redundancy is in the form of memory ($\rho_M > 0$) as well as in the form of a non-uniform distribution ($\rho_D > 0$). The transition and marginal distributions of the source are given as follows

$$Pr\{X_n = 0|X_{n-1} = 0\} \triangleq P(0|0) = q_0,$$

$$Pr\{X_n = 1|X_{n-1} = 1\} \triangleq P(1|1) = q_1,$$

and

$$Pr\{X_n = 0\} = 1 - Pr\{X_n = 1\} = \frac{1 - q_1}{2 - q_0 - q_1},$$

where we assume that $q_0 > 1/2$ and $1 - q_0 \leq q_1 \leq q_0$. Note that

- If $q_0 = q_1$, then the source becomes a *symmetric* Markov source.
- If $q_0 = 1 - q_1$, then the source becomes a *non-uniform iid* source with distribution $Pr\{X_n = 0\} = q_0$.

As in the case of the symmetric source, the implementation of the sequence MAP detector for the non-symmetric Markov source is according to equation (7) with replacing $P(x_1)$ by $Pr\{X_1 = x_1\}$ and $P(x_k)$ by $P(x_k|x_{k-1})$.

We derive a theorem offering general conditions for the uselessness of the sequence MAP detector. But first we need the following lemma.

Lemma 2 Let $q_0 \in (\frac{1}{2}, 1)$, $q_1 \in [1-q_0, q_0]$, $\epsilon \in (0, \frac{1}{2})$, $\delta \geq 0$ and $L \geq 2$. Assume that $e_0^L = (e_0, e_1, \ldots, e_L) = (0, 1, 1, \ldots, 1, 0)$. Then $\forall\, y_0^L \in \{0,1\}^{L+1}$,

$$s \triangleq \prod_{k=1}^{L} \frac{P(y_k|y_{k-1})}{P(x_k|x_{k-1})} \geq \frac{(1-q_0)(1-q_1)}{q_0^2} \left(\frac{q_1}{q_0}\right)^{L-2}, \tag{38}$$

where $x_0^L = e_0^L \oplus y_0^L$, $P(0|0) = q_0$, $P(1|0) = 1-q_0$, $P(0|1) = 1-q_1$ and $P(1|1) = q_1$.

In addition, the above also holds when $e_0^L = (0, 1, 1, \ldots, 1)$.

Proof See Appendix.

Theorem 5 Given $q_0 \in (\frac{1}{2}, 1)$, $q_1 \in [1-q_0, q_0]$, $\epsilon \in (0, \frac{1}{2})$, $\delta \geq 0$, and $n \geq 3$, assume that $X_1 = Y_1$ almost surely. Then

(i) $\hat{X}^n = Y^n$ is an optimal sequence (MAP) detection rule if

$$\frac{(1-\epsilon+\delta)^2}{\epsilon(1-\epsilon)} \frac{(1-q_0)(1-q_1)}{q_0^2} \geq 1, \tag{39}$$

and

$$\frac{1-\epsilon+\delta}{\epsilon+\delta} \frac{q_1}{q_0} \geq 1. \tag{40}$$

(ii) If (39) does not hold, then $\hat{X}^n = Y^n$ is not an optimal sequence detection rule.

(iii) If (40) does not hold, then $\exists\, n_0 > 0$ such that $\forall\, n \geq n_0$ $\hat{X}^n = Y^n$ is not an optimal sequence detection rule.

Proof 5 (i) Assuming (39) and (40) hold and following the same method as in Theorem 4, we obtain for $e_n = 0$ (Case 1) that $\alpha = \prod_{i=1}^{N} \alpha_i$, where

$$\alpha_i = \left[\frac{(1-\epsilon+\delta)^2}{\epsilon(1-\epsilon)}\right] \left[\frac{1-\epsilon+\delta}{\epsilon+\delta}\right]^{L_i-2} \left[\prod_{k=m_i+1}^{m_i+L_i} \frac{P(y_k|y_{k-1})}{P(x_k|x_{k-1})}\right].$$

Applying Lemma 2 yields

$$\alpha_i \geq \left[\frac{(1-\epsilon+\delta)^2}{\epsilon(1-\epsilon)} \frac{(1-q_0)(1-q_1)}{q_0^2}\right] \left[\frac{1-\epsilon+\delta}{\epsilon+\delta} \frac{q_1}{q_0}\right]^{L_i-2} \geq 1,$$

where the last inequality follows by hypothesis. Hence, $\alpha = \prod_{i=1}^{N} \alpha_i \geq 1$. Note that for $e_n = 1$ (Case 2), we get

$$\begin{aligned}\alpha_N &= \left[\frac{1-\epsilon+\delta}{\epsilon}\right]\left[\frac{1-\epsilon+\delta}{\epsilon+\delta}\right]^{L_N-1}\left[\prod_{k=m_N+1}^{m_N+L_N}\frac{P(y_k|y_{k-1})}{P(x_k|x_{k-1})}\right] \\ &\geq \left[\frac{1-\epsilon+\delta}{\epsilon}\,\frac{(1-q_0)(1-q_1)}{q_0^2}\right]\left[\frac{1-\epsilon+\delta}{\epsilon+\delta}\right]\left[\frac{1-\epsilon+\delta}{\epsilon+\delta}\,\frac{q_1}{q_0}\right]^{L_N-2} \geq 1.\end{aligned}$$

Thus $\alpha = \prod_{i=1}^{N} \alpha_i \geq 1$.

(ii) Suppose that (39) does not hold. Take $y^n = (0,0,\ldots,0,1,0,\ldots,0)$ and $x^n = (0,0,\ldots,0)$. Then

$$\alpha = \left[\frac{1-\epsilon+\delta}{\epsilon}\,\frac{1-q_0}{q_0}\right]\left[\frac{1-\epsilon+\delta}{1-\epsilon}\,\frac{1-q_1}{q_0}\right] < 1.$$

(iii) Suppose that (40) does not hold. Let $y^n = (0,1,1,\ldots,1)$ and $x^n = (0,0,\ldots,0)$. Then

$$\alpha = \left[\frac{1-\epsilon+\delta}{\epsilon}\,\frac{1-q_0}{q_0}\right]\left[\frac{1-\epsilon+\delta}{\epsilon+\delta}\,\frac{q_1}{q_0}\right]^{n-2}.$$

Then for n sufficiently large, $\alpha < 1$.

■

Observations:

- If $q_0 = q_1 = q$ (symmetric Markov source), (40) always hold and (39) reduces to

$$\frac{(1-\epsilon+\delta)^2}{\epsilon(1-\epsilon)}\left(\frac{1-q}{q}\right)^2 \geq 1,$$

  which along with (ii) is the same as Theorem 4.

- If $q_0 = 1 - q_1 = p$ (non-uniform iid source with distribution p), (39) and (40) reduce to

$$\frac{(1-\epsilon+\delta)^2}{\epsilon(1-\epsilon)}\,\frac{1-p}{p} \geq 1, \tag{41}$$

and

$$\frac{1-\epsilon+\delta}{\epsilon+\delta}\,\frac{1-p}{p} \geq 1, \tag{42}$$

respectively. Note that (42) implies (41), and hence (42) and (iii) are equivalent to Theorems 1 and 2, respectively.

Simulation results of the performance of the sequence MAP detector are presented in Figures 13 and 14. The simulations were performed on 1,000 samples of the non-symmetric Markov source and the experiment was repeated 500 times. In both figures, we notice that the performance of the MAP detector improves with δ, we have no mismatch in these cases. This can be explained by the fact that for these cases, the non-symmetric Markov source behaves like a non-uniform iid source because its redundancy in the form of non-uniform distribution is much more dominant than its redundancy in the form of memory ($\rho_D >> \rho_M$). Indeed, using the following redundancy formulas

$$\rho_D = 1 - H(X_1) = 1 - h_b\left(\frac{1-q_1}{2-q_0-q_1}\right),$$

$$\rho_M = H(X_1) - H(X_2|X_1) = h_b\left(\frac{1-q_1}{2-q_0-q_1}\right) - \frac{1-q_1}{2-q_0-q_1}h_b(q_0) - \left(1 - \frac{1-q_1}{2-q_0-q_1}\right)h_b(q_1),$$

we obtain

- For $q_0 = 0.99$ and $q_1 = 0.45$, $\rho_D = 0.8585 >> \rho_M = 0.044$.
- For $q_0 = 0.99$ and $q_1 = 0.1$, $\rho_D = 0.9193 >> \rho_M = 0.0048$.

Finally, we can similarly assert that in cases where $\rho_M >> \rho_D$, the source tends to behave like a symmetric Markov source; this will result in a mismatch between the source and the channel as demonstrated in the previous subsection.

5 Conclusions

In this paper, we analyzed the MAP detection problems (sequence and instantaneous) of a source with an inherent redundancy transmitted over a discrete channel with additive Markov noise. The proposed MAP detectors exploit the source redundancy in order to combat channel errors. The problem was investigated for three cases: (i) non-uniform iid source, (ii) symmetric Markov source, and (iii) non-symmetric Markov source. In all cases, analytical results giving conditions for the uselessness of the sequence MAP detector as well as simulation results were presented. For the case of the non-uniform iid source, we showed that our proposed simple system beats a traditional tandem source-channel coding scheme for high channel bit error rates. A mismatch was established for case (ii) between the source and the channel. This mismatch was reduced for high values of the channel correlation parameter by the use of a rate-one convolutional encoder.

Applications of the MAP detection problem in a combined source-channel coding system are currently under investigation [7]. Future work may consist of comparing the results above to those obtained by detecting binary sources over the Gilbert-Elliott channel with potential applications to digital cellular channels.

6 Appendix

In this Appendix, we prove Lemmas 1 and 2 stated in Sections 3.2 and 4.2 respectively.

Proof of Lemma 1 First assume that $x_0^L = (0, 1, 1, \ldots, 1, 0)$. Rewrite α as

$$\alpha = \left[\frac{\tilde{Q}(y_1|y_0)}{\tilde{Q}(\overline{y}_1|y_0)} \frac{\tilde{Q}(y_L|y_{L-1})}{\tilde{Q}(y_L|\overline{y}_{L-1})}\right] \left[\prod_{k=2}^{L-1} \left(\frac{\tilde{Q}(y_k|y_{k-1})}{\tilde{Q}(\overline{y}_k|\overline{y}_{k-1})} \frac{p}{1-p}\right)\right] \left[\frac{p}{1-p}\right],$$

where the overbar denotes the binary complement. Define the terms inside the first set of brackets above as g. Note that when $L = 2$, $y_1 = y_{L-1}$ and the minimum value of g (under the constraint that $y_1 = y_{L-1}$) is $\frac{\epsilon(1-\epsilon)}{(1-\epsilon+\delta)^2}$, which is achieved when $(y_0, y_1, y_2) = (0, 1, 0)$. Thus, for $L = 2$,

$$\alpha = g \frac{p}{1-p} \geq \frac{\epsilon(1-\epsilon)}{(1-\epsilon+\delta)^2} \frac{p}{1-p} \geq 1.$$

We now assume without loss of generality (w.l.o.g.) that $L > 2$. Partition the index set $\mathcal{K} = \{2, 3, \ldots, L-1\}$ as follows

$$\mathcal{K} = \mathcal{K}_{00} \cup \mathcal{K}_{01} \cup \mathcal{K}_{10} \cup \mathcal{K}_{11},$$

where

$$\mathcal{K}_{ab} \triangleq \{k \in \mathcal{K} : y_{k-1} = a, y_k = b\}, \qquad a, b \in \{0, 1\}.$$

Therefore,

$$\begin{aligned}
m &\triangleq \prod_{k=2}^{L-1}\left[\frac{\tilde{Q}(y_k|y_{k-1})}{\tilde{Q}(\overline{y}_k|\overline{y}_{k-1})}\,\frac{p}{1-p}\right] \\
&= \left[\frac{p}{1-p}\right]^{L-2}\left[\prod_{k\in\mathcal{K}_{00}}\frac{1-\epsilon+\delta}{\epsilon+\delta}\right]\left[\prod_{k\in\mathcal{K}_{11}}\frac{\epsilon+\delta}{1-\epsilon+\delta}\right]\left[\prod_{k\in\mathcal{K}_{01}}\frac{\epsilon}{1-\epsilon}\right]\left[\prod_{k\in\mathcal{K}_{10}}\frac{1-\epsilon}{\epsilon}\right] \\
&= \left[\frac{p}{1-p}\right]^{L-2}\left[\frac{\epsilon+\delta}{1-\epsilon+\delta}\right]^{|\mathcal{K}_{11}|-|\mathcal{K}_{00}|}\left[\frac{\epsilon}{1-\epsilon}\right]^{|\mathcal{K}_{01}|-|\mathcal{K}_{10}|}.
\end{aligned}$$

Case 1: $y_1 = y_{L-1}$

Consider the sequence $(y_1, y_2, \ldots, y_{L-1})$. For every transition from 0 to 1, there is a corresponding transition from 1 to 0 (because $y_1 = y_{L-1}$). Thus, $|\mathcal{K}_{01}| = |\mathcal{K}_{10}|$. Also, note that $|\mathcal{K}_{00}| \geq 0$ and $|\mathcal{K}_{11}| \leq |\mathcal{K}| = L-2$. Hence

$$m \geq \left[\frac{p}{1-p}\right]^{L-2}\left[\frac{\epsilon+\delta}{1-\epsilon+\delta}\right]^{L-2} \geq l^{L-2} \geq 1,$$

where the second inequality follows from the fact that $\epsilon+\delta \geq \epsilon$ and $1 \geq \frac{1-\epsilon}{1-\epsilon+\delta}$. Finally, note that the minimum value of g under the constraint $y_1 = y_{L-1}$ is $\frac{\epsilon(1-\epsilon)}{(1-\epsilon+\delta)^2}$. Therefore,

$$\alpha = g\,\frac{p}{1-p}\,m \;\geq\; \left[\frac{\epsilon(1-\epsilon)}{(1-\epsilon+\delta)^2}\,\frac{p}{1-p}\right]m \;=\; l\,m \;\geq\; 1.$$

Case 2: $y_1 = 0,\ y_{L-1} = 1$

In this case: $|\mathcal{K}_{00}| \geq 0$, $|\mathcal{K}_{01}| \geq 1$, $|\mathcal{K}_{10}| = |\mathcal{K}_{01}| - 1$ and $|\mathcal{K}_{11}| \leq L-3$. Thus

$$\begin{aligned}
m &\geq \left[\frac{p}{1-p}\right]^{L-2}\left[\frac{\epsilon+\delta}{1-\epsilon+\delta}\right]^{L-3}\left[\frac{\epsilon}{1-\epsilon}\right] \\
&= \left[\frac{p}{1-p}\,\frac{\epsilon+\delta}{1-\epsilon+\delta}\right]^{L-3}\left[\frac{p}{1-p}\,\frac{\epsilon}{1-\epsilon}\right].
\end{aligned}$$

The first term in the right hand side of the above equation is $\geq l^{L-3}$ by the same argument as before. The second term is $\geq l$ because $\frac{\epsilon}{1-\epsilon} \geq \frac{\epsilon}{1-\epsilon+\delta} \geq \frac{\epsilon(1-\epsilon)}{(1-\epsilon+\delta)^2}$. Thus $m \geq 1$. The minimum value of g in this case is $\frac{(1-\epsilon)^2}{(1-\epsilon+\delta)(\epsilon+\delta)}$. Hence

$$\alpha = g \, \frac{p}{1-p} \, m \; \geq \; \frac{(1-\epsilon)^2}{(1-\epsilon+\delta)(\epsilon+\delta)} \, \frac{p}{1-p} \, m \; \geq \; l \, m \; \geq \; 1.$$

Case 3: $y_1 = 1,\ y_{L-1} = 0$

Here we have: $|\mathcal{K}_{00}| \geq 0$, $|\mathcal{K}_{10}| \geq 1$, $|\mathcal{K}_{01}| = |\mathcal{K}_{10}| - 1$ and $|\mathcal{K}_{11}| \leq L-3$. Thus

$$m \geq \left[\frac{p}{1-p} \, \frac{\epsilon+\delta}{1-\epsilon+\delta} \right]^{L-3} \left[\frac{p}{1-p} \, \frac{1-\epsilon}{\epsilon} \right].$$

The minimum value of g in this case is $\frac{\epsilon^2}{(1-\epsilon+\delta)(\epsilon+\delta)}$. Therefore,

$$\begin{aligned} \alpha \; &\geq \; \left[\frac{\epsilon^2}{(1-\epsilon+\delta)(\epsilon+\delta)} \, \frac{p}{1-p} \right] \left[\frac{1-\epsilon}{\epsilon} \, \frac{p}{1-p} \right] \left[\frac{p}{1-p} \, \frac{\epsilon+\delta}{1-\epsilon+\delta} \right]^{L-3} \\ &= \; \left[\frac{\epsilon(1-\epsilon)}{(1-\epsilon+\delta)(\epsilon+\delta)} \left(\frac{p}{1-p} \right)^2 \right] \left[\frac{p}{1-p} \, \frac{\epsilon+\delta}{1-\epsilon+\delta} \right]^{L-3}. \end{aligned}$$

The second term above is $\geq l^{L-3}$ by previous arguments. The first term is $\geq l^2$ because $\frac{1}{\epsilon+\delta} \geq \frac{1}{1-\epsilon+\delta}$ and $1 \geq \frac{\epsilon(1-\epsilon)}{(1-\epsilon+\delta)^2}$. Thus $\alpha \geq l^{L-1} \geq 1$. This proves the Lemma for $x_0^L = (0,1,1,\ldots,1,0)$. Note that in all three cases, $m \geq 1$.

Now, assume that $x_0^L = (0,1,1,\ldots,1)$. We write α as

$$\alpha = \left[\frac{\tilde{Q}(y_1|y_0)}{\tilde{Q}(\overline{y}_1|y_0)} \, \frac{p}{1-p} \right] \left[\prod_{k=2}^{L} \left(\frac{\tilde{Q}(y_k|y_{k-1})}{\tilde{Q}(\overline{y}_k|\overline{y}_{k-1})} \, \frac{p}{1-p} \right) \right].$$

The minimum value of $\tilde{Q}(y_1|y_0)/\tilde{Q}(\overline{y}_1|y_0)$ is $\frac{\epsilon}{1-\epsilon+\delta}$. Thus, the first term in the right hand side of the expression of α above is $\geq l \geq 1$. By replacing L by $L-1$, we can use the same arguments as before to show that the second term $(= m)$ is also ≥ 1. Thus $\alpha \geq 1$.

■

Proof of Lemma 2 First assume that $e_0^L = x_0^L \oplus y_0^L = (0,1,1,\ldots,1,0)$. Write

$$s = \prod_{k=1}^{L} \frac{P(y_k|y_{k-1})}{P(x_k|x_{k-1})} = \left[\frac{P(y_1|y_0)}{P(\overline{y}_1|y_0)} \right] \left[\prod_{k=2}^{L-1} \frac{P(y_k|y_{k-1})}{P(\overline{y}_k|\overline{y}_{k-1})} \right] \left[\frac{P(y_L|y_{L-1})}{P(y_L|\overline{y}_{L-1})} \right],$$

where the overbar denotes the binary complement. Note that when $L = 2$, $s = \frac{P(y_1|y_0)\ P(y_2|y_1)}{P(\overline{y}_1|y_0)\ P(y_2|\overline{y}_1)}$, the minimum value of which is $\frac{(1-q_0)(1-q_1)}{q_0^2}$. Thus, we may assume w.l.o.g. that $L > 2$. Now partition the set $\mathcal{K} = \{2, 3, \ldots, L-1\}$ as follows

$$\mathcal{K} = \mathcal{K}_{00} \cup \mathcal{K}_{01} \cup \mathcal{K}_{10} \cup \mathcal{K}_{11},$$

where

$$\mathcal{K}_{ab} \triangleq \{k \in \mathcal{K} : y_{k-1} = a, y_k = b\}, \qquad a, b \in \{0, 1\}.$$

We then rewrite s as

$$s = \left[\frac{P(y_1|y_0)}{P(\overline{y}_1|y_0)}\ \frac{P(y_L|y_{L-1})}{P(y_L|\overline{y}_{L-1})}\right] \left[\frac{q_1}{q_0}\right]^{|\mathcal{K}_{11}|-|\mathcal{K}_{00}|} \left[\frac{1-q_0}{1-q_1}\right]^{|\mathcal{K}_{01}|-|\mathcal{K}_{10}|}. \tag{43}$$

The first factor in the right hand side of (43) is defined as u. The product of the next two factors is defined as v. Note that $1-q_0 \leq q_1 \leq q_0$ implies that $1-q_0 \leq 1-q_1 \leq q_0$. Furthermore, since $q_0 > \frac{1}{2}$, $q_0(1-q_0) \leq q_1(1-q_1)$.

Case 1: $y_1 = y_{L-1}$

The minimum value of u in this case is $\frac{(1-q_0)(1-q_1)}{q_0^2}$. Furthermore, $|\mathcal{K}_{01}| = |\mathcal{K}_{10}|$, $|\mathcal{K}_{00}| \geq 0$ and $|\mathcal{K}_{11}| \leq L-2$. Therefore by (43)

$$s \geq \frac{(1-q_0)(1-q_1)}{q_0^2} \left(\frac{q_1}{q_0}\right)^{L-2}.$$

Case 2: $y_1 = 0$, $y_{L-1} = 1$

Here, the minimum value of u is $\frac{(1-q_1)^2}{q_0 q_1}$. Also, $|\mathcal{K}_{01}| \geq 1$, $|\mathcal{K}_{10}| = |\mathcal{K}_{01}| - 1$, $|\mathcal{K}_{00}| \geq 0$ and $|\mathcal{K}_{11}| \leq L-3$. Thus

$$\begin{aligned} s &\geq \left[\frac{(1-q_1)^2}{q_0 q_1}\right] \left[\frac{q_1}{q_0}\right]^{L-3} \left[\frac{1-q_0}{1-q_1}\right] \\ &= \left[\frac{(1-q_0)(1-q_1)}{q_0 q_1}\right] \left[\frac{q_1}{q_0}\right]^{L-3} \geq \left[\frac{(1-q_0)(1-q_1)}{q_0^2}\right] \left(\frac{q_1}{q_0}\right)^{L-2}. \end{aligned}$$

Case 3: $y_1 = 1,\ y_{L-1} = 0$

The minimum value of u is $\frac{(1-q_0)^2}{q_0 q_1}$. Also, $|\mathcal{K}_{10}| \geq 1$, $|\mathcal{K}_{01}| = |\mathcal{K}_{10}| - 1$, $|\mathcal{K}_{00}| \geq 0$ and $|\mathcal{K}_{11}| \leq L-3$. Thus

$$\begin{aligned} s &\geq \left[\frac{(1-q_0)^2}{q_0 q_1}\right]\left[\frac{q_1}{q_0}\right]^{L-3}\left[\frac{1-q_1}{1-q_0}\right] \\ &= \left[\frac{(1-q_0)(1-q_1)}{q_0 q_1}\right]\left[\frac{q_1}{q_0}\right]^{L-3} \geq \left[\frac{(1-q_0)(1-q_1)}{q_0^2}\right]\left(\frac{q_1}{q_0}\right)^{L-2}. \end{aligned}$$

Note that in each case, $v \geq \left(\frac{q_1}{q_0}\right)^{L-2}$. Now, assume that $e_0^L = (0,1,1,\ldots,1)$. Replacing $L-1$ by L in (43), we write

$$s = \frac{P(y_1|y_0)}{P(\overline{y}_1|y_0)}\, v \geq \frac{P(y_1|y_0)}{P(\overline{y}_1|y_0)}\left(\frac{q_1}{q_0}\right)^{L-2}.$$

Note that the minimum value of $\frac{P(y_1|y_0)}{P(\overline{y}_1|y_0)}$ is $\frac{1-q_0}{q_0}$. Thus

$$s \geq \left(\frac{1-q_0}{q_0}\right)\left(\frac{q_1}{q_0}\right)^{L-2} \geq \left[\frac{(1-q_0)(1-q_1)}{q_0^2}\right]\left(\frac{q_1}{q_0}\right)^{L-2}.$$

This completes the proof of the lemma.

■

References

[1] F. Alajaji and T. Fuja, "A Communication Channel Modeled on Contagion," submitted to the *IEEE Trans. Inform. Theory*, August 1993.

[2] J. L. Devore, "A Note on the Observation of a Markov Source Through a Noisy Channel," *IEEE Trans. Inform. Theory*, vol. 20, pp. 762-764, Nov. 1974.

[3] G. D. Forney, "The Viterbi Algorithm," *Proc. IEEE*, vol. 61, pp. 268-278, March 1973.

[4] E. N. Gilbert, "Capacity of Burst-Noise Channels," *Bell Syst. Tech. Journal*, vol. 39, pp. 1253-1265, 1960.

[5] L. Kanal and A. Sastry, "Models for Channels with Memory and Their Applications to Error Control," *Proc. IEEE*, vol. 66, pp. 724-744, July 1978.

[6] S. Lin and D. J. Costello, Jr., *Error Control Coding*, Prentice-Hall Inc., Englewood Cliffs, NJ, 1983.

[7] N. Phamdo, F. Alajaji, N. Farvardin and T. Fuja, "Quantization of Memoryless Sources Over Discrete Markov Channels," *Proc. of Conference on Information Sciences and Systems,* Princeton University, Princeton, NJ, March 1994.

[8] N. Phamdo and N. Farvardin, "Quantization Over Discrete Noisy Channels Using Rate-One Convolutional Codes," submitted to the *IEEE Trans. Inform. Theory*, August 1993.

[9] N. Phamdo and N. Farvardin, "Optimal Detection of Discrete Markov Sources Over Discrete Memoryless Channels — Applications to Combined Source-Channel Coding," to appear in *IEEE Trans. Inform. Theory*, January 1994.

[10] H. V. Poor, *An Introduction to Signal Detection and Estimation*, Springer-Verlag, New York, 1988.

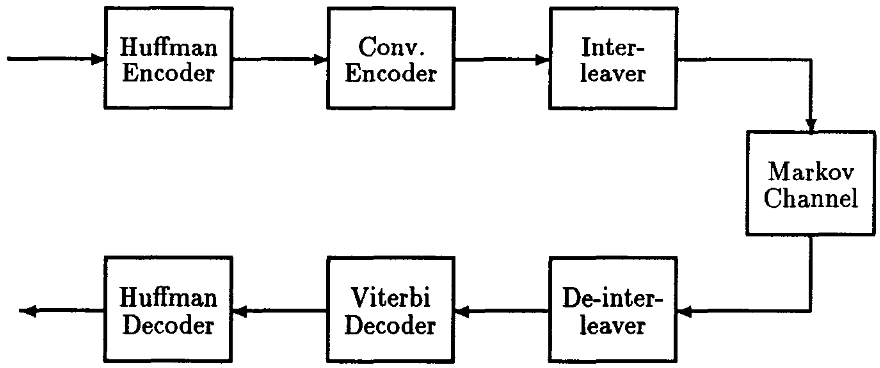

Figure 1: Block Diagram of the Tandem Scheme.

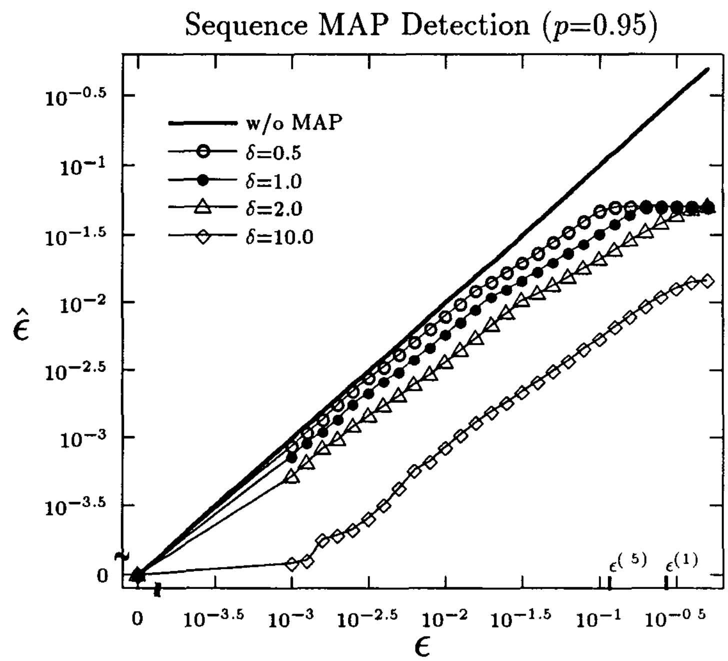

Figure 2: Performance of Sequence MAP Detector for IID Binary Source with $p = 0.95$ over the Contagion Markov Channel ($M = 1$); ϵ = Channel Bit Error Rate; $\hat{\epsilon}$ = Pr{bit error}; and δ = Correlation Parameter of the Channel.

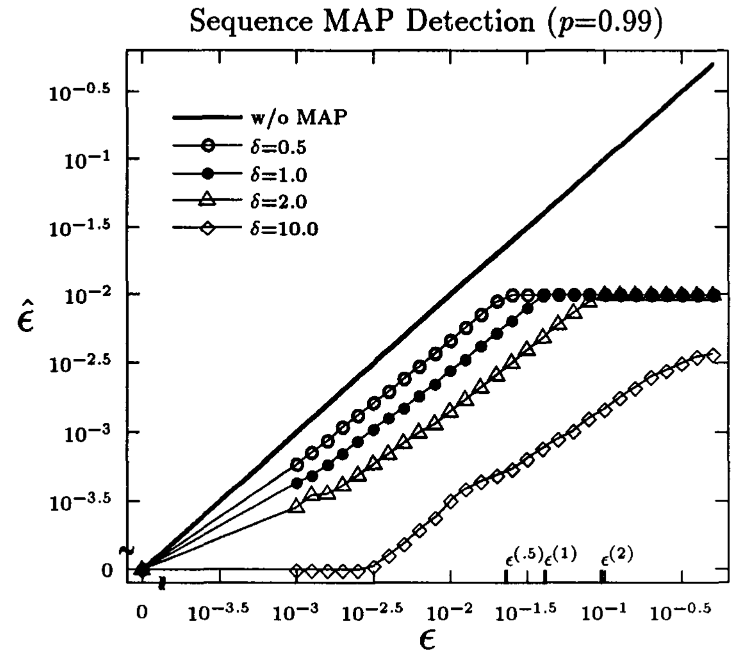

Figure 3: Performance of Sequence MAP Detector for IID Binary Source with $p = 0.99$ over the Contagion Markov Channel ($M = 1$); ϵ = Channel Bit Error Rate; $\hat{\epsilon}$ = Pr{bit error}; and δ = Correlation Parameter of the Channel.

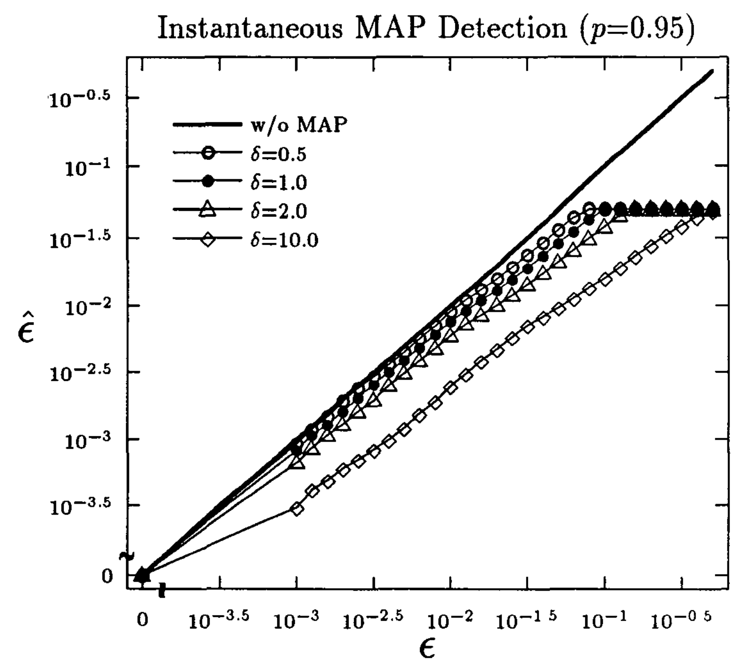

Figure 4: Performance of Instantaneous MAP Detector for IID Binary Source with $p = 0.95$ over the Contagion Markov Channel ($M = 1$); ϵ = Channel Bit Error Rate; $\hat{\epsilon}$ = Pr{bit error}; and δ = Correlation Parameter of the Channel.

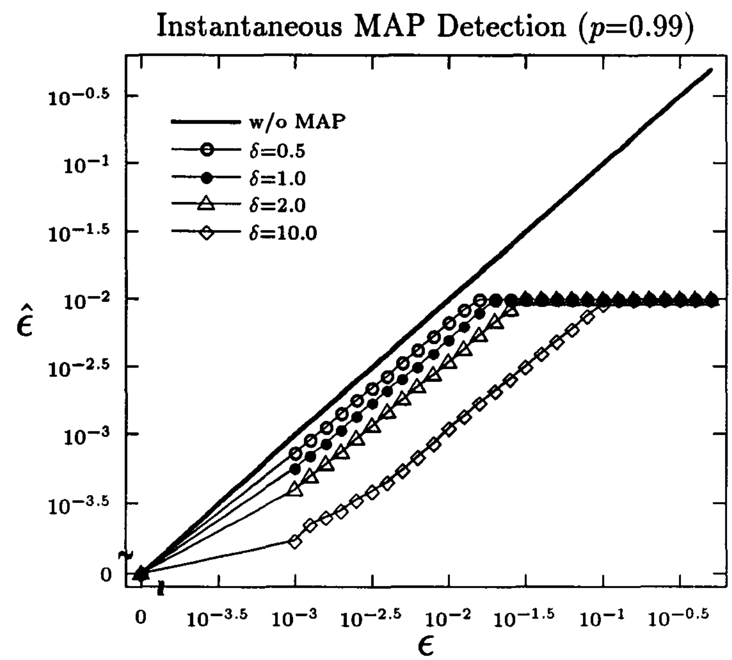

Figure 5: Performance of Instantaneous MAP Detector for IID Binary Source with $p = 0.99$ over the Contagion Markov Channel ($M = 1$); $\epsilon =$ Channel Bit Error Rate; $\hat{\epsilon} = \Pr\{\text{bit error}\}$; and $\delta =$ Correlation Parameter of the Channel.

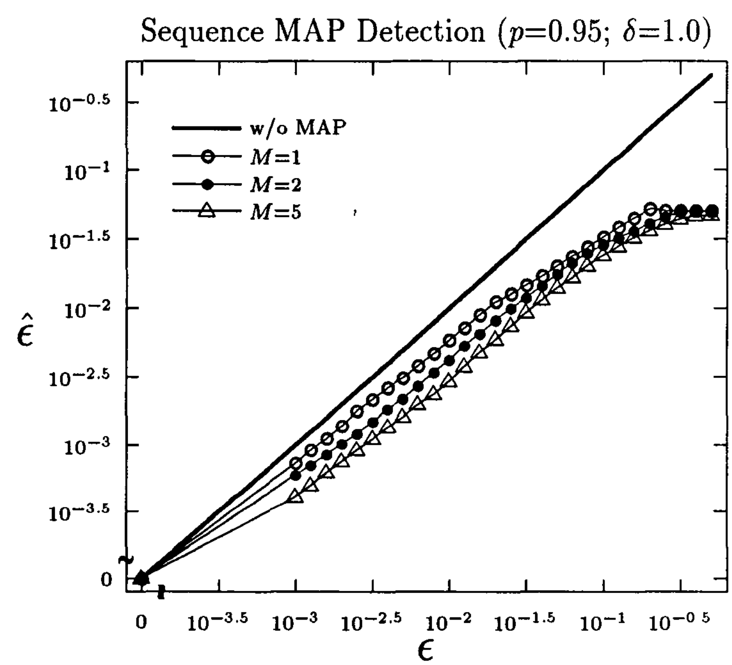

Figure 6: Performance of Sequence MAP Detector for IID Binary Source with $p = 0.95$ over the Contagion Markov Channel of Order M with $\delta = 1.0$; $\epsilon =$ Channel Bit Error Rate; and $\hat{\epsilon} = \Pr\{\text{bit error}\}$.

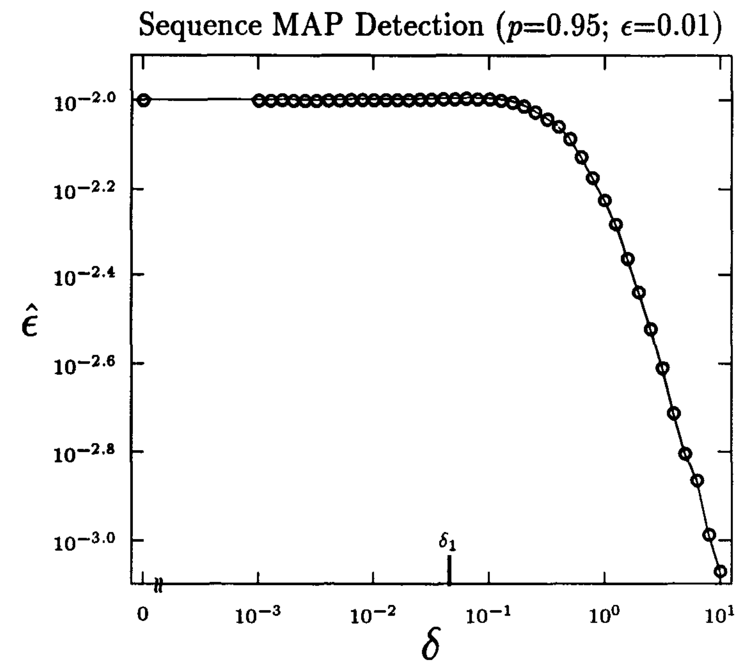

Figure 7: Performance of Sequence MAP Detector for IID Binary Source with $p = 0.95$ over the Contagion Markov Channel ($M = 1$) with $\epsilon = 0.01$; $\hat{\epsilon} = \Pr\{\text{bit error}\}$; and δ = Correlation Parameter of the Channel.

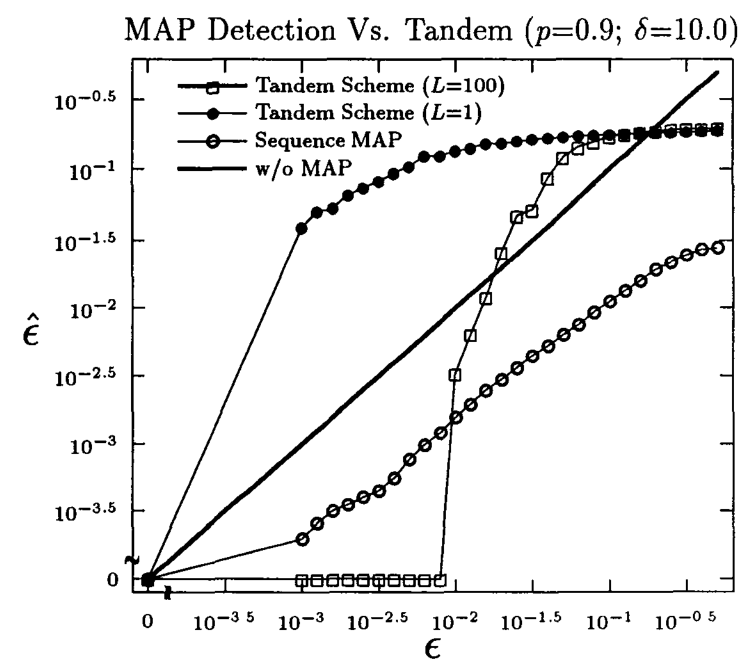

Figure 8: Comparisons of Proposed Sequence MAP Detection System Versus Tandem Source-Channel Coding System; Binary IID Source with $p = 0.9$; $\hat{\epsilon} = \Pr\{\text{bit error}\}$; ϵ = Channel Bit Error Rate; δ = Correlation Parameter of the Channel; and L = Interleaving Length.

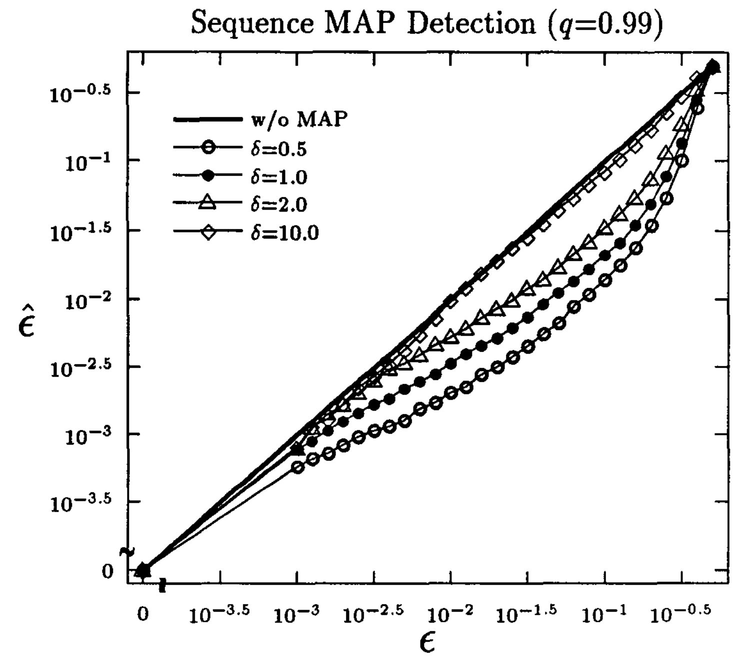

Figure 9: Performance of Sequence MAP Detector for Binary Symmetric Markov Source with $q = 0.99$ over the Contagion Markov Channel ($M = 1$); ϵ = Channel Bit Error Rate; $\hat{\epsilon}$ = Pr{bit error}; and δ = Correlation Parameter of the Channel.

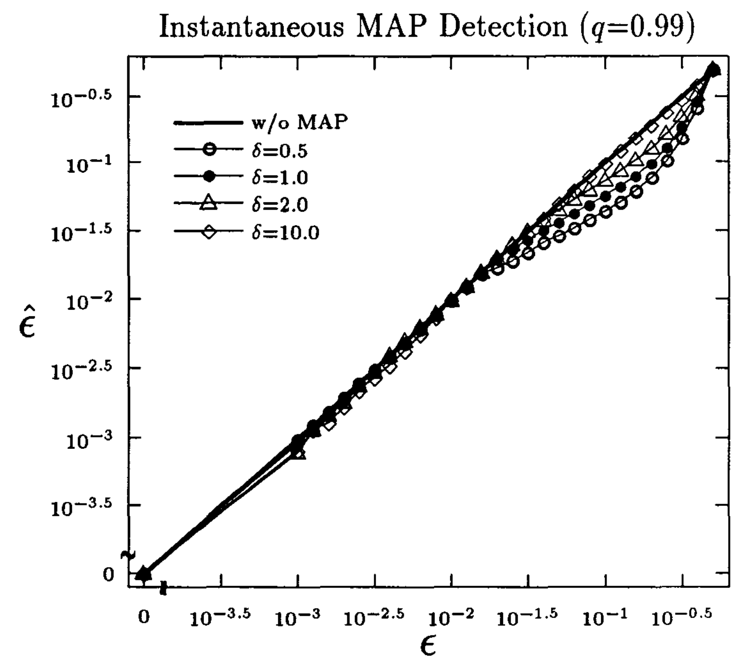

Figure 10: Performance of Instantaneous MAP Detector for Binary Symmetric Markov Source with $q = 0.99$ over the Contagion Markov Channel ($M = 1$); ϵ = Channel Bit Error Rate; $\hat{\epsilon}$ = Pr{bit error}; and δ = Correlation Parameter of the Channel.

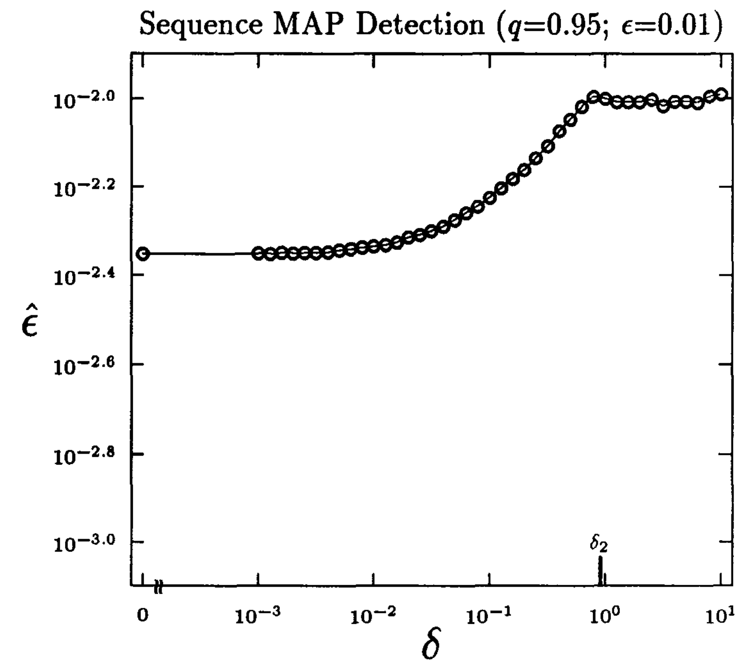

Figure 11: Performance of Sequence MAP Detector for Binary Symmetric Markov Source with $q = 0.95$ over the Contagion Markov Channel ($M = 1$) with $\epsilon = 0.01$; $\hat{\epsilon} = \Pr\{\text{bit error}\}$; and δ = Correlation Parameter of the Channel.

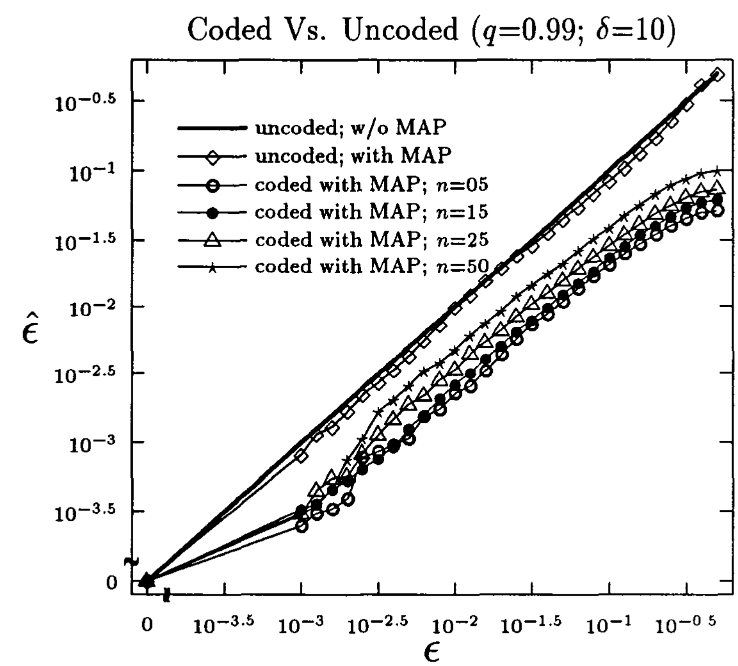

Figure 12: Performance of the Coded System with Sequence MAP Detection for Binary Symmetric Markov Source with $q = 0.99$ over the Contagion Markov Channel ($M = 1$) with $\delta = 10$; $\hat{\epsilon} = \Pr\{\text{bit error}\}$; ϵ = Channel Bit Error Rate; and δ = Correlation Parameter of the Channel.

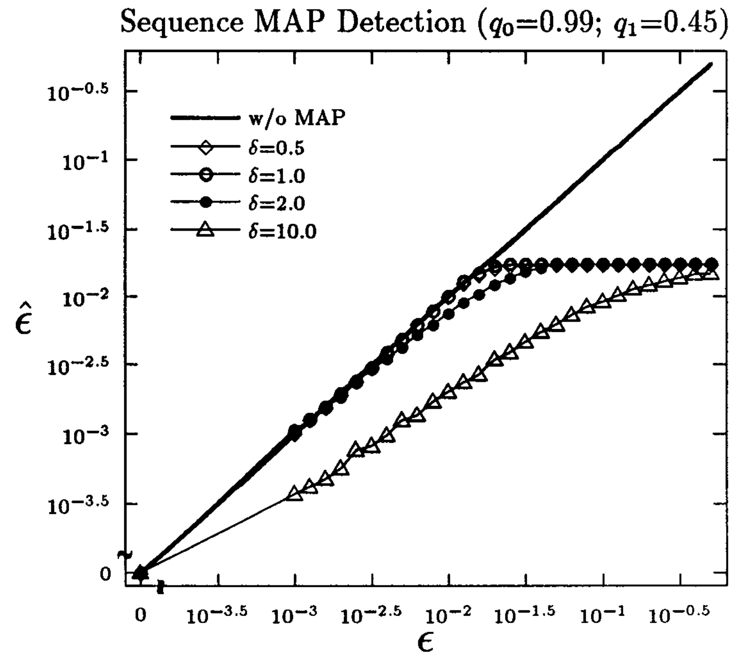

Figure 13: Performance of Sequence MAP Detector for Binary Non-Symmetric Markov Source with $q_0 = 0.99$ and $q_1 = 0.45$ over the Contagion Markov Channel ($M = 1$); $\hat{\epsilon} = \Pr\{\text{bit error}\}$; ϵ = Channel Bit Error Rate; and δ = Correlation Parameter of the Channel.

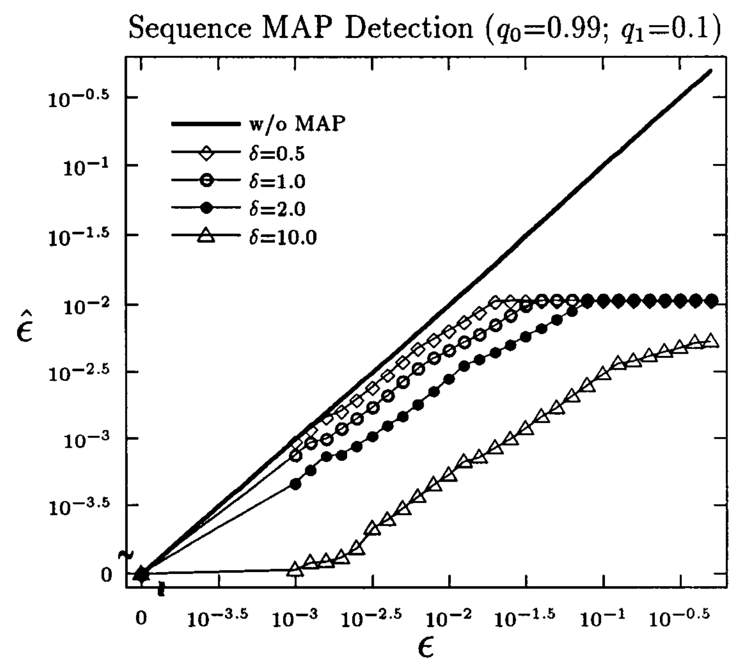

Figure 14: Performance of Sequence MAP Detector for Binary Non-Symmetric Markov Source with $q_0 = 0.99$ and $q_1 = 0.1$ over the Contagion Markov Channel ($M = 1$); $\hat{\epsilon} = \Pr\{\text{bit error}\}$; ϵ = Channel Bit Error Rate; and δ = Correlation Parameter of the Channel.