CAR-TR-814 January 1996
CS-TR-3604

Transmission of Two-Tone Images over Noisy Communication Channels with Memory

Philippe Burlina[1]
Fady Alajaji[2]
Rama Chellappa[1]

[1]Center for Automation Research
Electrical Engineering Department
University of Maryland
College Park, MD 20742

[2]Department of Mathematics and Statistics
Queen's University
Kingston, ON K7L 3N6, Canada

Abstract

We propose a joint source-channel coding scheme for the reliable transmission of two-tone images over a binary channel with additive Markov noise. We first quantify the natural redundancy inherent in two-tone images by modeling the pixel inter-spatial dependency using first and second order Markov models. We next investigate techniques by which this redundancy can be exploited by an appropriately designed channel decoder. A maximum a posteriori (MAP) detector which is optimal in terms of minimizing the sequence probability of error is proposed. The MAP detector fully exploits the redundancy of binary images in combating channel noise. It also exploits the larger capacity of the channel with memory as opposed to the interleaved (memoryless) channel. Experimental results indicate a superior performance of the proposed scheme as compared to more complex schemes that do not exploit the image redundancy, such as traditional tandem coding schemes. Finally, the scheme is applied to bit-plane encoding of grey-level images over bursty Markov channels.

Keywords: Joint source/channel coding, MAP decoding, natural image redundancy, Markov sources, bursty channels with memory.

1 INTRODUCTION

We present an alternate approach to coding information-bearing data for the reliable transmission of two-tone images over noisy communication channels with memory. It consists of jointly designing the source and channel codes, a technique referred to as joint source-channel coding.

Source and channel coding are two problems that have traditionally been implemented separately, forming what is known as a tandem source-channel coding system. The separation of source and channel coding is optimal only in an asymptotic sense, i.e., when no constraints exist on the coding block lengths (delay) and on the complexity of the encoder/decoder [1, 2]. Joint source-channel coding has recently received increased attention [3–15]. It has been shown that if delay and complexity are constrained, performance can be increased if the source and channel codes are jointly designed, as opposed to being treated independently.

With the exception of [17], most of the work on joint source-channel coding of images [5, 6, 8, 11–13] has dealt with the memoryless channel, disregarding the fact that real-world communication channels—in particular, land mobile radio (LMR) or satellite channels—often have memory. In this work, we propose a joint source-channel coding scheme for the reliable transmission of two-tone images over a binary channel with additive Markov noise. Applications of this work are in the transmission of facsimile documents over bursty LMR or satellite channels.

We begin by quantifying the amount of natural redundancy inherent in binary images. This is achieved by modeling the image as a first- or second-order causal Markov process. It is shown that a substantial amount of redundancy exists in many binary images. We next investigate the problem of optimal detection of binary images directly transmitted over the Markov channel. This is an extension and application of results obtained in [14] in which the MAP detection problem for *ideal* binary Markov sources over Markov channels is analyzed. A sequence *maximum a posteriori* (MAP) detector that minimizes the sequence probability of error is proposed. The MAP detector fully exploits the image characteristics in order to efficiently combat channel noise. It also exploits the larger capacity of the channel with memory as opposed to the interleaved (memoryless) channel. The MAP decoder is implemented using a modified version of the Viterbi algorithm.

It is shown via simulation that for some binary images with strong correlation redundancy, a *mismatch* exists between the Markov-modeled image and the Markov channel. This is because such images behave like *symmetric* Markov sources; it illustrates an analytical result demonstrated in [14], which states that if a binary symmetric Markov source is connected to a binary Markov

channel, a mismatch occurs in the MAP decoder and degrades its performance as the channel noise becomes more correlated. In this case, we substantially reduce the mismatch by employing an *adaptive* rate-one convolutional encoder which transforms Markov images with high correlation redundancy into non-uniform independent and identically distributed (iid) processes.

Experimental results of the proposed scheme show very good performance, in spite of its low complexity (which primarily resides in the MAP decoder). Comparisons with more complex traditional tandem coding schemes are also presented. Since the proposed MAP scheme employs a model-based decoding algorithm, we assume that the image statistics are provided to the decoder. This can be achieved by transmitting the source characteristics over the channel using a forward error-control code. The amount of overhead information needed for this purpose is computed.

As in all joint source-channel coding techniques, it is assumed here that the channel parameters are perfectly estimated at the receiver. We investigate the robustness of the MAP scheme when the decoder does not have perfect knowledge of the channel parameters. Finally, the proposed scheme is extended to bit-plane encoding of grey-level images over bursty Markov channels.

The rest of this paper is organized as follows. We introduce the Markov channel model in Section 2. In Section 3, three image models are proposed and the resulting estimates of the natural image redundancy are presented. The MAP joint source-channel coding scheme as well as the related analytical results are described in Section 4. In Section 5, experimental results and discussion are provided. Finally, conclusions are stated in Section 6.

2 CHANNEL MODEL

Consider a discrete channel with memory, with common input, noise and output binary alphabets and described by the equation $Y_i = X_i \oplus Z_i$, for $i = 1, 2, 3, \ldots$, where $\oplus$ represents the addition operation modulo 2, and the random variables X_i, Z_i and Y_i represent, respectively, the input, noise and output of the channel. The input and noise sequences are independent of each other, i.e. $\{X_i\} \perp \{Z_i\}$. The noise process $\{Z_i\}_{i=1}^{\infty}$ is assumed to be a stationary mixing (hence ergodic) Markov process of order M. By this we mean that the noise sample, Z_i, depends only on the previous M noise samples, i.e., for $i \geq M + 1$,

$$\Pr\{Z_i = e_i | Z_1 = e_1, \ldots, Z_{i-1} = e_{i-1}\} = \Pr\{Z_i = e_i | Z_{i-M} = e_{i-M}, \ldots, Z_{i-1} = e_{i-1}\}.$$

We assume that the marginal distribution of the noise process is given by $\Pr\{Z_i = 1\} = \epsilon = 1 - \Pr\{Z_i = 0\}$, where $\epsilon \in [0, 1/2)$ is the channel bit error rate (BER). Furthermore, we assume

that the process $\{Z_i\}$ is generated by the finite-memory contagion urn model described in [18]. According to this model, the noise sample Z_i depends only on the *sum* of the previous M noise samples.[1] Thus, for $i \geq M+1$,

$$\Pr\{Z_i = 1 | Z_{i-M} = e_{i-M}, \ldots, Z_{i-1} = e_{i-1}\} = \frac{\epsilon + (\sum_{j=i-M}^{i-1} e_j)\delta}{1 + M\delta},$$

where $e_j = 0$ or 1, for $j = i-M, \ldots, i-1$. The non-negative parameter δ determines the amount of correlation in $\{Z_i\}$. The correlation coefficient of the noise process is $\delta/(1+\delta)$. Note that if $\delta = 0$, the noise process $\{Z_i\}$ becomes iid and the resulting additive noise channel reduces to a binary symmetric channel (BSC). Finally, we note that the channel is entirely characterized by three parameters: ϵ, δ and M.

Distribution of the Noise: For an input block $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ and an output block $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$, we denote the block channel transition probability matrix $\Pr\{\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x}\}$ by $Q(\mathbf{y}|\mathbf{x})$.

- For block length $n \leq M$, we have [18] $Q(\mathbf{y}|\mathbf{x}) = L(n, d, \epsilon, \delta)$, where $d = d_H(\mathbf{x}, \mathbf{y})$ is the Hamming distance between $\mathbf{x}$ and $\mathbf{y}$ and

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

- For $n \geq M+1$, we have [18]

$$Q(\mathbf{y}|\mathbf{x}) = \Pr\{\mathbf{Z} = \mathbf{e}\} = L(M, s, \epsilon, \delta) \prod_{i=M+1}^{n} \left[\frac{\epsilon + s_i\delta}{1 + M\delta}\right]^{e_i} \left[1 - \frac{\epsilon + s_i\delta}{1 + M\delta}\right]^{1-e_i},$$

  where $\mathbf{e} = (e_1, e_2, \ldots, e_n)$, $e_i = x_i \oplus y_i$, $s = e_1 + \cdots + e_M$ and $s_i = e_{i-1} + \cdots + e_{i-M}$.

Channel Capacity: The capacity C of this channel is given by [18]

$$C = 1 - \sum_{s=0}^{M} \binom{M}{s} L(M, s, \epsilon, \delta)\, h_b\left(\frac{\epsilon + s\delta}{1 + M\delta}\right),$$

where $h_b(x) = -x\log_2(x) - (1-x)\log_2(1-x)$ is the binary entropy function. The capacity is monotonically increasing with δ (for fixed ϵ, M) and M (for fixed ϵ, δ), and monotonically decreasing with ϵ (for fixed δ, M). Note that for fixed ϵ and M,

$$\text{as} \quad \delta \longrightarrow \infty, \quad C \longrightarrow 1.$$

[1] For $M = 1$, the model is general, i.e., it can represent any binary first-order Markov chain with positive transition probabilities.

Furthermore, it can be shown [18] that for fixed δ and ϵ,

$$\text{as} \quad M \longrightarrow \infty, \quad C \longrightarrow 1 - \int_0^1 h_b(z)\, f(z)\, dz,$$

where $h_b(z) = -z\log_2(z) - (1-z)\log_2(1-z)$ is the binary entropy function,

$$f(z) = \begin{cases} \frac{\Gamma(1/\delta)}{\Gamma(\epsilon/\delta)\Gamma((1-\epsilon)/\delta)}\, z^{\frac{\epsilon}{\delta}-1}\, (1-z)^{\frac{(1-\epsilon)}{\delta}-1}, & \text{if } 0 < z < 1\ ; \\ 0, & \text{otherwise,} \end{cases}$$

and $\Gamma(\cdot)$ is the gamma function, $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$ for $x > 0$.

Comments: The motivation for the use of this contagion-based Markov channel as our model for a channel with memory, as opposed to the Gilbert-Elliott model or others [19], is based on the fact that the contagion model is mathematically more tractable than the Gilbert-Elliott model. It is completely described by only three parameters, whereas the Gilbert-Elliott model is described by four parameters. Furthermore, unlike the Gilbert model, the transition probability and capacity of the contagion channel have closed-form expressions that can be easily computed.

3 IMAGE MODELING AND REDUNDANCY

Consider a two-tone image $U = [U_{i,j}]$ of row size J and column size K, where $U_{i,j} = 0$ or 1, $i = 1, \ldots, J$, $j = 1, \ldots, K$. We assume that the image pixels follow a causal second-order Markov dependency such that the pixel at location (i,j) depends on the pixels at locations $(i-1,j)$ and $(i,j-1)$:

$$\begin{aligned} \Pr\{U_{i,j} = u_{i,j} | U_{i,j-1} = u_{i,j-1}, \ldots, U_{i,1} = u_{i,1}, U_{i-1,j} = u_{i-1,j}, \ldots, U_{1,1}\} \\ = \Pr\{U_{i,j} = u_{i,j} | U_{i,j-1} = u_{i,j-1}, U_{i-1,j} = u_{i-1,j}\}. \end{aligned}$$

Since two-tone images are here transmitted over communication channels as one continuous bit stream, we will represent the image pixels by a second-order Markov process $\{X_n\}$ described by

$$\Pr\{X_n = x_n | X_{n-1} = x_{n-1}, \ldots, X_1 = x_1\} = \Pr\{X_n = x_n | X_{n-1} = x_{n-1}, X_{n-K} = x_{n-K}\},$$

for $n > K$. For the boundary pixels, the Markov dependency is of first order only: if $n \leq K$, X_n depends only on X_{n-1}, and if $n = m * K + 1$, where $m = 1, 2, \ldots$, then X_n depends only on X_{n-K}. The above second-order Markov process constitutes a causal Markov random field. Assuming $\{X_n\}$

is stationary,[2] its transition probabilities are given by

$$\Pr\{X_n = 0 | X_{n-1} = 0, X_{n-K} = 0\} \stackrel{\triangle}{=} p_0, \quad \Pr\{X_n = 0 | X_{n-1} = 0, X_{n-K} = 1\} \stackrel{\triangle}{=} p_1,$$

$$\Pr\{X_n = 0 | X_{n-1} = 1, X_{n-K} = 0\} \stackrel{\triangle}{=} p_2, \quad \Pr\{X_n = 0 | X_{n-1} = 1, X_{n-K} = 1\} \stackrel{\triangle}{=} p_3,$$

where p_0, p_1, p_2 and $p_3 \in (0,1)$. We will refer to the above second-order Markov image model as *Model 0*. The following special cases of *Model 0* will be of particular interest to us:

- *Model 1*: $p_0 = p_1 \stackrel{\triangle}{=} q_0 \neq 1/2$ and $p_2 = p_3 \stackrel{\triangle}{=} q_1 \neq 1/2$. In this case, the source $\{X_n\}$ reduces to a first-order Markov process with one-step transition probabilities:

  $$\Pr\{X_n = 0 | X_{n-1} = 0\} = q_0, \quad \Pr\{X_n = 1 | X_{n-1} = 1\} = 1 - q_1.$$

  Note that if $q_0 = 1 - q_1$, then the first-order Markov process becomes *symmetric*; i.e., it has a uniform marginal probability $\Pr\{X_n = 0\} = 1/2$.

- *Model 2*: $p_0 = p_1 = p_2 = p_3 \stackrel{\triangle}{=} p \neq 1/2$. In this case, $\{X_n\}$ reduces to a non-uniform iid process with marginal $\Pr\{X_n = 0\} = p$.

Note that in this work we do not consider non-causal Markov random field (MRF) models, which are widely used for image processing and analysis [16]. However, the search for the MAP estimate of MRF-modeled images involves the use of computationally intensive algorithms such as Simulated Annealing [7] or Graduated Non-Convexity. Since these algorithms are inappropriate for real-time applications, we restrict ourselves to causal models that are easily implemented via sequential decoding algorithms.

We next turn to quantifying the natural *redundancy* that is inherent in a two-tone image if it is represented by a random process $\{X_n\}$ described according to each of the three proposed models. Let $H_\infty(X)$ denote the entropy rate of process $\{X_n\}$: $H_\infty(X) \stackrel{\triangle}{=} \lim_{n\to\infty} \frac{1}{n} H(X^n)$, where $X^n \stackrel{\triangle}{=} (X_1, X_2, \ldots, X_n)$ and $H(X^n)$ is the entropy of the block X^n:

$$H(X^n) = - \sum_{x^n \in \{0,1\}^n} \Pr\{X^n = x^n\} \log_2 \left(\Pr\{X^n = x^n\}\right).$$

The total redundancy the binary process $\{X_n\}$ contains is $\rho_T = 1 - H_\infty(X)$. This total redundancy can be written as $\rho_T = \rho_D + \rho_M$ [14], where $\rho_D \stackrel{\triangle}{=} 1 - H(X_1)$, and $\rho_M \stackrel{\triangle}{=} H(X_1) - H_\infty(X)$,

[2] Of course, image pixels do not form a stationary process. We will indeed "approximate" the pixel bit stream of each image line by a stationary process; the extent to which our model "fits" will be judged on the basis of how well decoding schemes matched to the model perform.

where $H(X_1)$ is the entropy of random variable X_1. Here ρ_D denotes the redundancy due to the non-uniformity of the marginal distribution ($\Pr\{X_1 = x\}$) of the process, and ρ_M denotes the redundancy due to the memory of the process. Note that if the redundancy due to the non-uniformity of a process is high relative to its redundancy in the form of memory ($\rho_D \gg \rho_M$), then the process behaves like a non-uniform iid source. Similarly, if $\rho_M \gg \rho_D$, the process tends to behave like a symmetric Markov source [20]. Note that if the process is uniformly distributed ($\Pr\{X_1 = 0\} = 1/2$), then $\rho_D = 0$. Furthermore, if the process is independent (e.g., Model 2), then $\rho_M = 0$.

We compute the redundancy two-tone images possess when they are modeled according to Model 0, Model 1, and Model 2. The two-tone test images we use are shown in Figure 1. Figure 1.a displays a (flat) binary version of the image Lena, while Figure 1.b represents a halftone version of Lena. Halftone images are binary (or two-tone) images that give a grey scale rendition [21]. The Headscan in Figure 1.c is a binary plane of a medical magnetic resonance (MRI) image, and Figure 1.d shows the image of one of the eight standard facsimile test documents recommended by CCITT[3] [21]. Although images are highly non-stationary, image pixels may be considered to be locally stationary. We assume that $\{X_n\}$ is stationary only within each image line; this requires an updating of the transition probabilities for each image line.

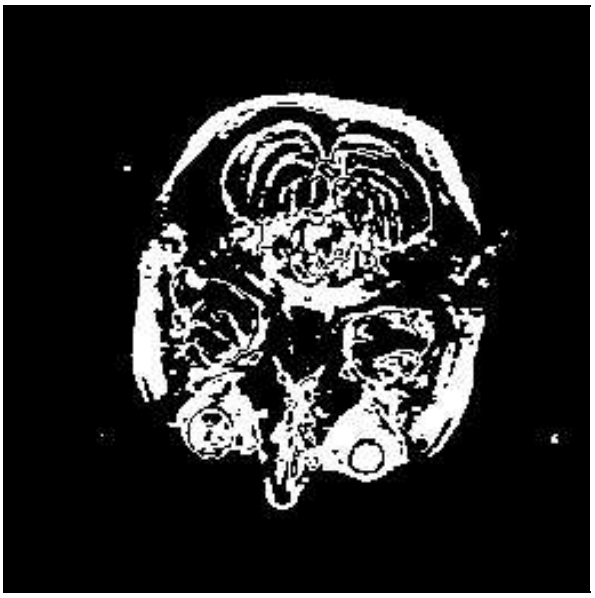
For each model, the relative frequency of transitions between the binary pixel values is compiled to extract the transition and marginal probabilities in each line. The entropy rate of the resulting line process is computed to arrive at an estimate of the total redundancy ρ_T in each line. The results for the four test images are compiled in Figures 2.a to 2.d in which we provide the histogram (or distribution) of the line redundancy ρ_T under each of the proposed models: Model 0 (2D Markov), Model 1 (1D Markov), and Model 2 (IID). Examination of the redundancy results leads to the following observations:

- Figures 2.a, 2.b, and 2.c clearly indicate that the images Lena 1, Headscan and CCITT are strongly redundant when modeled by a first-order (Model 1) or second-order (Model 0) Markov chain. For example, in Figure 2.a, Model 0 yields around 95% of the image lines with $\rho_T \geq 0.6$ bits, and Model 1 yields nearly 90% of the lines with $\rho_T \geq 0.5$ bits. Furthermore, images Lena 1 and Headscan has little redundancy due to non-uniformity since ρ_T of the IID model (Model 2) is less than 0.1 bits 95% of the time for Lena 1 and less than 0.4 bits

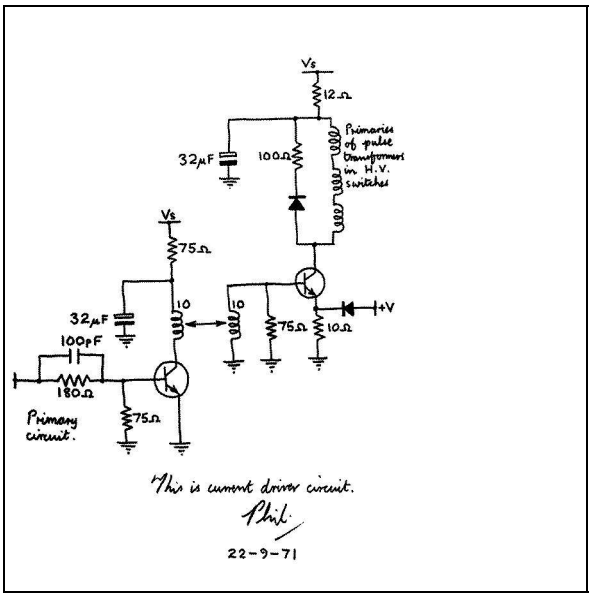
[3] "Comité Consultatif International de Téléphonie et Télégraphie"—recently renamed ITU.

(a) Lena 1 ($512 * 512$)

(b) Lena 2 ($512 * 512$)



(c) Headscan ($256 * 256$)

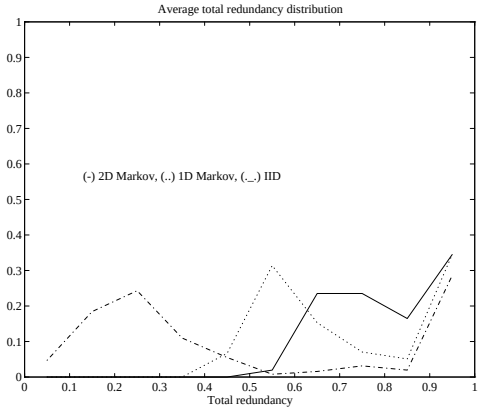


(d) CCITT ($594 * 432$)

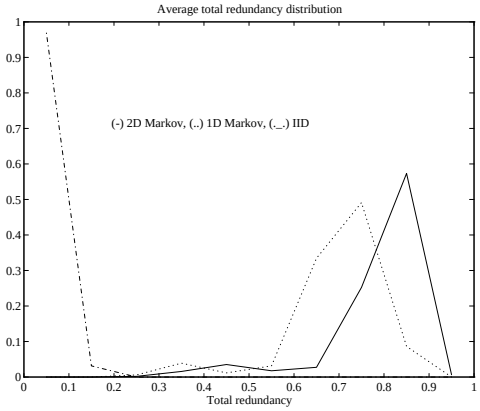
Figure 1: Two-Tone Test Images: Lena, Headscan and CCITT.

60% of the time for Headscan. This implies that ρ_M is much higher than ρ_D in Lena 1 and moderately higher in Headscan, hence these images behave like *symmetric* Markov chains. This is not the case, however, in Figure 2.c, where the distribution due to non-uniformity is not negligible.

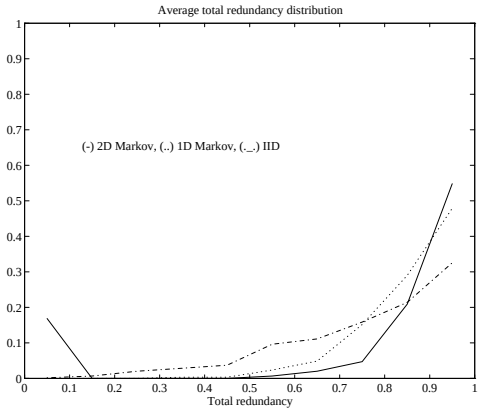
- The redundancy results shown in Figure 2.d demonstrate that the halftone image Lena 2 contains little redundancy under all three models ($\rho_T \leq 0.3$). This may be due to the fact that the halftone technique introduces more randomness in the black/white pixel distribution in order to increase the picture resolution.



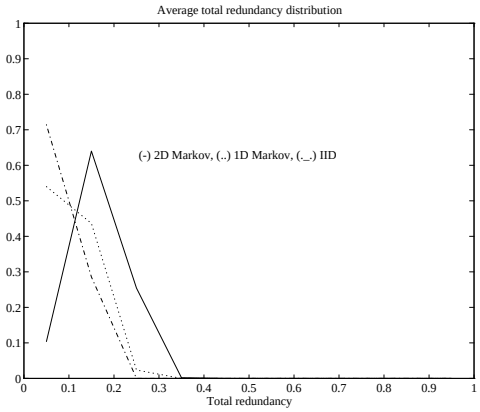
(a) Headscan – Histogram of Redundancy ρ_T per Line.



(b) Lena 1 – Histogram of Total Redundancy ρ_T per Line.



(c) CCITT – Histogram of Redundancy ρ_T per Line.



(d) Lena 2 – Histogram of Total Redundancy ρ_T per Line.

Figure 2: Histogram of Total Redundancy for Lena 1, Lena 2, Headscan and CCITT 2 images.

4 MAP CHANNEL DECODING OF IMAGES

4.1 Implementation

Consider the problem of transmitting the binary second-order Markov source $\{X_n\}$ described in the previous section, over a channel with memory. We will restrict our analysis to the Markov channel of memory order one ($M = 1$). Extending the analysis to Markov channels of order $M > 1$ is straightforward.

Given that a sequence of bits is observed at the channel output, the goal of the receiver is to determine the most probable transmitted *sequence*. The optimal detection method that minimizes the sequence probability of decoding error is the sequence *maximum a posteriori* (MAP) method [14]. More specifically, if $Y^n = y^n = (y_1, y_2, \ldots, y_n)$ denotes the received binary sequence at the output of the channel, the MAP detector "guesses" the transmitted sequence $\hat{x}^n$ according to

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

But (1) is equivalent to

$$\begin{aligned}
\hat{x}^n &= \arg \max_{x^n \in \{0,1\}^n} \Pr\{Y^n = y^n | X^n = x^n\} \Pr\{X^n = x^n\} \\
&= \arg \max_{x^n \in \{0,1\}^n} \Pr\{Z^n = y^n \oplus x^n\} \Pr\{X^n = x^n\} \\
&= \arg \max_{x^n \in \{0,1\}^n} \left[\Pr\{Z_1 = y_1 \oplus x_1\} \prod_{k=2}^{n} \Pr\{Z_k = y_k \oplus x_k | Z_{k-1} = y_{k-1} \oplus x_{k-1}\} \right] \\
&\qquad \times \Pr\{X_1 = x_1\} \left[\prod_{k=2}^{K} \Pr\{X_k = x_k | X_{k-1} = x_{k-1}\} \right] \\
&\qquad \times \left[\prod_{k=K+1}^{n} \Pr\{X_k = x_k | X_{k-1} = x_{k-1}, X_{k-K} = x_{k-K}\} \right].
\end{aligned} \tag{2}$$

For $M = 1$, the channel transition probability given in (1) can be easily expressed in terms of $Q(z_n|z_{n-1}) \stackrel{\triangle}{=} \Pr\{Z_n = z_n | Z_{n-1} = z_{n-1}\}$ and $Q(z_n) \stackrel{\triangle}{=} \Pr\{Z_n = z_n\}$, where

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

and $Q(1) = \epsilon = 1 - Q(0)$. We also denote the source distribution by $P(\cdot)$:

$$\begin{aligned}
\Pr\{X_k = x_k\} &\stackrel{\triangle}{=} P(x_k), \\
\Pr\{X_k = x_k | X_{k-1} = x_{k-1}\} &\stackrel{\triangle}{=} P(x_k | x_{k-1}), \\
\Pr\{X_k = x_k | X_{k-1} = x_{k-1}, X_{k-K} = x_{k-K}\} &\stackrel{\triangle}{=} P(x_k | x_{k-1}, x_{k-K}).
\end{aligned}$$

Since the logarithm function is monotone increasing, equation (2) is equivalent to

$$\hat{x}^n = \arg \max_{x^n \in \{0,1\}^n} \left[\log \left(Q(x_1 \oplus y_1) P(x_1) \right) + \sum_{k=2}^{K} \log \left(Q(y_k \oplus x_k | y_{k-1} \oplus x_{k-1}) P(x_k | x_{k-1}) \right) + \sum_{k=K+1}^{n} \log \left(Q(y_k \oplus x_k | y_{k-1} \oplus x_{k-1}) P(x_k | x_{x-1}, x_{k-K}) \right) \right]. \quad (3)$$

The sequence MAP detector described in (3) can be implemented in a straightforward way using the Viterbi algorithm [22]. Here, x_k denotes the state at time k; the trellis will hence have 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

$$- \log \left(Q(y_k \oplus x_k | y_{k-1} \oplus x_{k-1}) P(x_k | x_{x-1}) \right), \quad \text{for } k \leq K,$$

and

$$- \log \left(Q(y_k \oplus x_k | y_{k-1} \oplus x_{k-1}) P(x_k | x_{x-1}, x_{k-K}) \right), \quad \text{for } k > K.$$

The surviving path for each state is the path with the smallest cumulative metric up to that state. The sequence MAP decoder observes the entire received sequence y^n in order to estimate $x_1, x_2, \ldots, x_n$.

4.2 Analytical Results for First-Order Markov Sources

In [14], we provide necessary and sufficient conditions under which the sequence MAP detection for ideal stationary binary first-order Markov sources transmitted over the Markov channel (with $M = 1$) is useless. We quote these theorems here as they will prove useful to us in interpreting the image transmission experiments.

Theorem 1 (Asymmetric Markov Source) Consider an asymmetric binary Markov source with parameters q_0 and q_1 as described by Model 1. The source is sent over the Markov channel with parameters ϵ, δ, and $M = 1$.

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 $X_1 = Y_1$ and $X_n = Y_n$. Then

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

$$\frac{(1 - \epsilon + \delta)^2}{\epsilon(1 - \epsilon)} \, \frac{(1 - q_0)(1 - q_1)}{q_0^2} \geq 1, \quad (4)$$

and

$$\frac{1-\epsilon+\delta}{\epsilon+\delta}\ \frac{q_1}{q_0} \geq 1. \tag{5}$$

(ii) If (4) does not hold, then $\hat{X}^n = Y^n$ is not an optimal sequence detection rule.

(iii) If (5) 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.

Note that if $q_0 = q_1 = 1/2$, then equations (4) and (5) always hold; hence the sequence MAP detector is *always* useless. This is expected since if $q_0 = q_1 = 1/2$, the Markov source becomes a uniformly distributed iid random process with *zero* redundancy. When the source contains no redundancy, the MAP detector reduces to a maximum likelihood (ML) decoder and fails to provide any protection against channel noise.

The above theorem yields the conditions under which the sequence MAP decoder is useless. It gives necessary and sufficient conditions for which the "believe what you see rule" is the optimal MAP detection rule.

Corollary 1 (IID Source) Consider the case where the source transition distributions q_0 and q_1 are such that $q_0 = q_1 \stackrel{\triangle}{=} p \neq 1/2$; this results in a binary iid non-uniform source with probability distribution $P(0) = q_0 \stackrel{\triangle}{=} p$ (Model 2). Assume without loss of generality that $p > 1/2$.

Given $p \in (\frac{1}{2}, 1)$, $\epsilon \in (0, \frac{1}{2}]$, $\delta \geq 0$, and $n \geq 3$, assume that $X_1 = Y_1$ and $X_n = Y_n$. Then

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

$$\frac{1-\epsilon+\delta}{\epsilon+\delta}\ \frac{1-p}{p} \geq 1. \tag{6}$$

(ii) If (6) 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.

Remark: Expression (6) is equivalent to

$$\delta \leq \delta_1 \stackrel{\triangle}{=} \frac{1-\epsilon-p}{2p-1}, \tag{7}$$

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

Corollary 2 (Symmetric Markov Source) Consider the case where $q_0 = 1 - q_1 = q > 1/2$. This results in a binary symmetric Markov source.

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

$$\frac{(1 - \epsilon + \delta)^2}{\epsilon(1 - \epsilon)} \left(\frac{1 - q}{q}\right)^2 \geq 1. \tag{8}$$

Remark: Condition (8) is equivalent to

$$\delta \geq \delta_2 \stackrel{\triangle}{=} \left(\frac{q}{1 - q}\right) \sqrt{\epsilon(1 - \epsilon)} + \epsilon - 1. \tag{9}$$

Comments: The above necessary and sufficient condition indicates that for fixed ϵ and q (hence fixed δ_2), as the channel correlation parameter δ increases (hence as the channel capacity increases), the likelihood of the uselessness of the sequence MAP detector increases (cf. (9)). Hence, sequence MAP detection becomes useless for sufficiently large δ. 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 from fully exploiting the correlation (and thus capacity) of the channel.

5 EXPERIMENTAL RESULTS

5.1 Images Modeled as Second-Order Markov Sources

We present simulation results to determine the performance of the proposed MAP decoding scheme. We start by modeling the two-tone images according to the second-order causal Markov chain described by Model 0. Image lines are each represented by a Markov chain with parameters p_0, p_1, p_2 and p_3 (which are computed empirically), and transmitted uncompressed over the Markov channel in a lexicographic order. At the receiver, the sequence MAP decoder is implemented according to (3) via a modified version of the Viterbi algorithm to determine the most probable transmitted sequence.

In Figures 3 and 4, simulation results for the images Lena 1 and Headscan are displayed. Each experiment was repeated seven times and the average values of the bit probability of decoding error $P_e = P\{\hat{X}_n \neq X_n\}$ are plotted versus the channel bit error rate (BER) ϵ for different values of channel correlation parameter δ. We observe from the plots that as δ increases (and hence

channel capacity increases), the performance of the MAP detector deteriorates. This indicates that there exists a mismatch between the source and the channel that prevents the decoder from fully exploiting the noise correlation. This result is expected since—as seen in Section 3—the images Lena 1 and Headscan behave like binary symmetric Markov processes; the image redundancy due to the pixel memory dominates the redundancy due to the non-uniformity of the source distribution. Corollary 2 shows the existence of this mismatch when the source is first-order Markov.

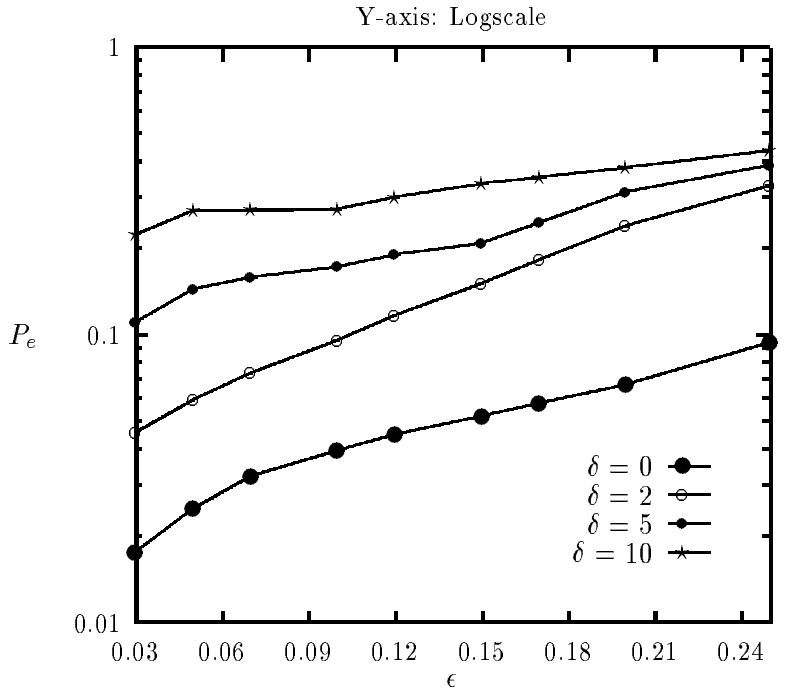


Figure 3: Probability of error vs. BER (ϵ); Lena 1 modeled according to Model 0.

This leads us to conclude that when such images are modeled by a second-order Markov chain (Model 0) and sent over the binary Markov channel, the best performance is obtained when $\delta = 0$; i.e., when the channel is fully interleaved and transformed into a memoryless channel (BSC). Examples for Lena 1 and Headscan transmitted over the fully interleaved Markov channel with BER $\epsilon = 0.1$, are shown in Figures 5.a to 5.d. The received images are displayed in Figures 5.a and 5.c, and the decoded images are shown in Figures 5.b and 5.d. The resulting average decoding bit error probabilities are 0.039 for Lena 1 and 0.033 for Headscan.

We can express the above results in terms of the peak-to-peak signal-to-noise ratio (PSNR)

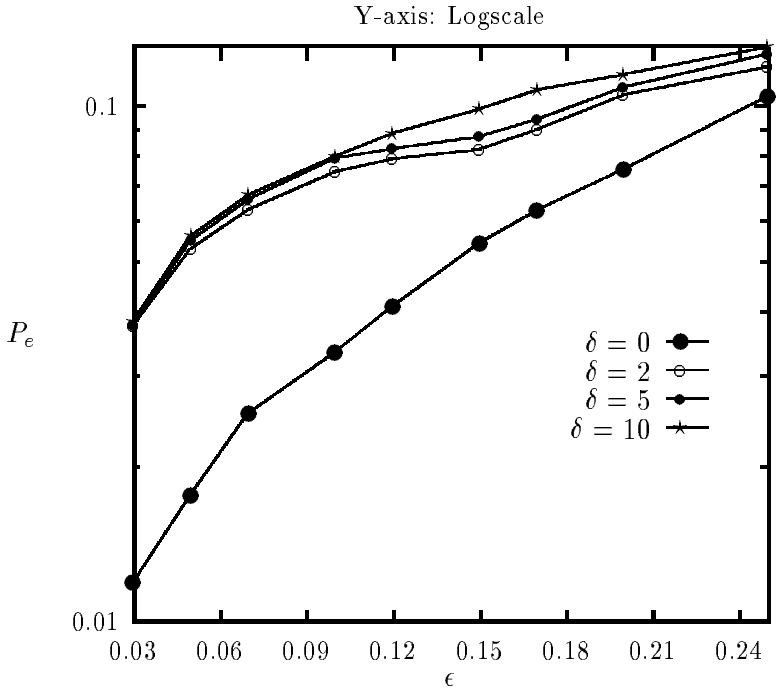


Figure 4: Probability of error vs. BER (ϵ); Headscan modeled according to Model 0.

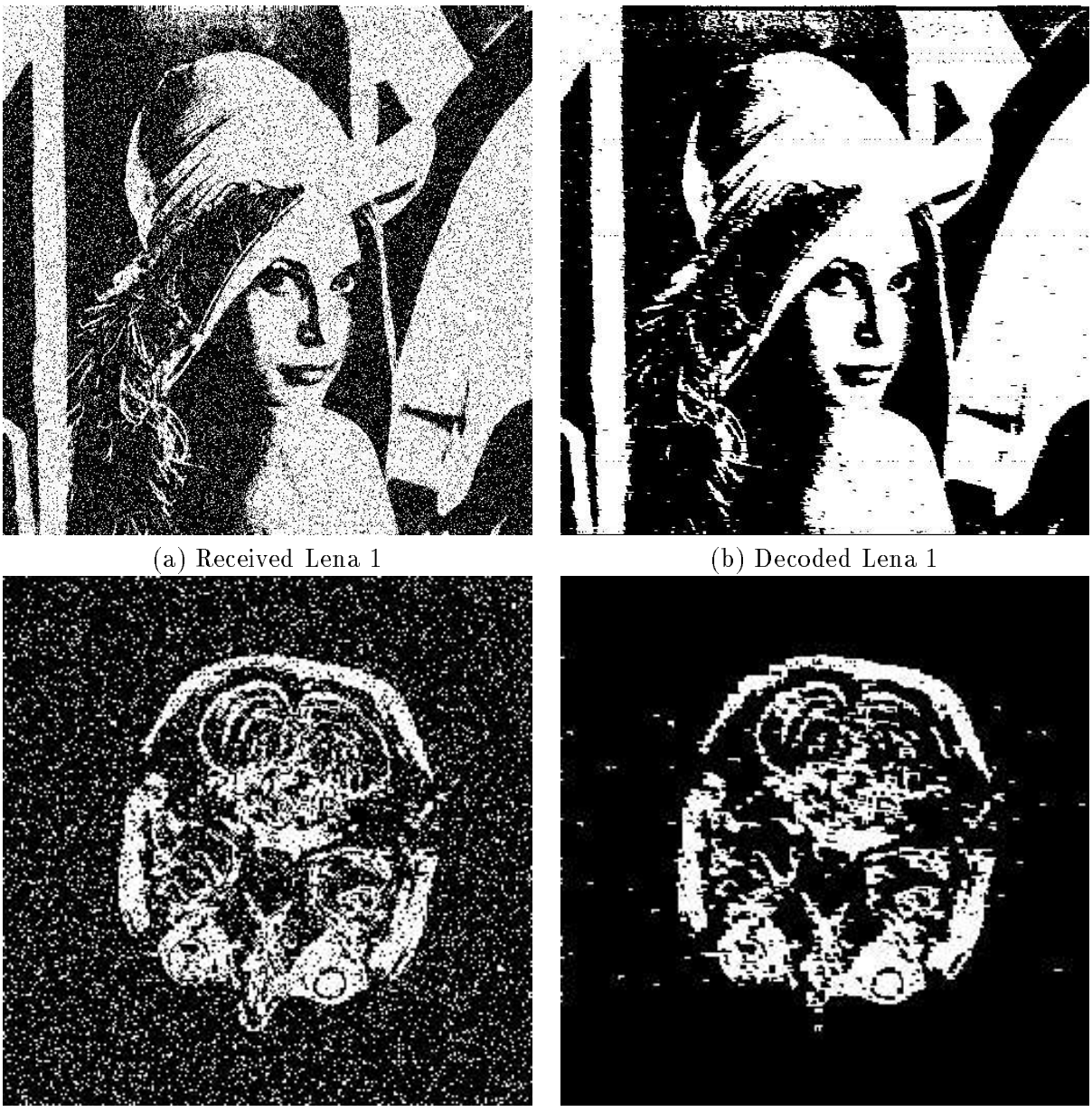
assuming a mean squared error distortion criterion:

$$\text{PSNR} \stackrel{\triangle}{=} 10 \log_{10} \frac{1}{\sigma_e^2} = 10 \log_{10} \frac{1}{P_e} \quad \text{dB},$$

where $\sigma_e^2 \stackrel{\triangle}{=} \frac{1}{JK} \sum_{i=1}^{JK} |x_i - \hat{x}_i|^2 = P_e$ since the image is binary. Hence, the results of Figures 5.b and 5.d can be expressed as follows. The channel BER is 0.1, i.e., if no MAP decoding is performed the image PSNR = 10 dB. With MAP decoding, PSNR = 14.08 dB for Lena 1 and 14.81 dB for Headscan. This results in a gain of more than 4 dB for both images.

5.2 Images Modeled as First-Order Markov Sources

We next evaluate the performance of the MAP decoding scheme when the image lines are modeled by first-order Markov chains as described by Model 1. Experiments were performed seven times for images Lena 1 and Headscan. The resulting bit error probability P_e is plotted versus ϵ for different values of δ in Figures 6 and 7. The dotted curve labeled "w/o MAP" indicates the probability of bit error when no MAP decoding is performed; i.e., $P_e = \epsilon$. The results shown in these figures lead to the following observations:



(a) Received Lena 1

(b) Decoded Lena 1

(c) Received Headscan

(d) Decoded Headscan, $P_e = 0.033$

Figure 5: Transmission of Lena 1 and Headscan, Model 0 with $\epsilon = 0.1$ and $\delta = 0$ (Interleaved Channel).

- For both images a mismatch exists between the source and the channel: the best performance is obtained when the channel is fully interleaved ($\delta = 0$).
- The mismatch illustrates Corollary 2; these images behave like symmetric Markov sources with nearly uniform marginal distributions. Indeed equation (12) is verified for the majority of the lines in these images.

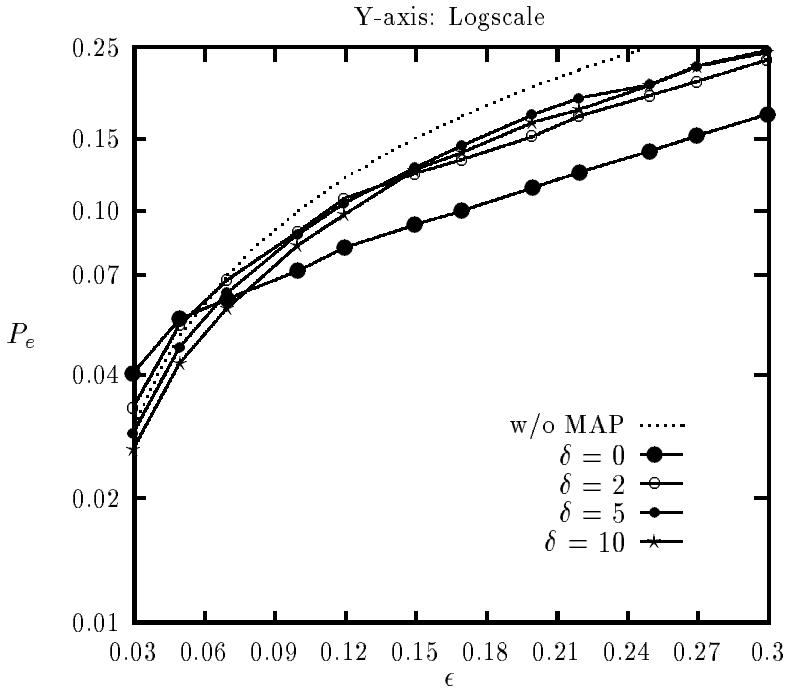


Figure 6: Probability of error vs. BER (ϵ); Lena 1 modeled according to Model 1.

Adaptive Encoding Scheme

When a binary symmetric Markov source is directly sent over a binary Markov channel, a mismatch occurs between the source and channel as the correlation parameter δ increases seen as, in Figures 6–7 and Corollary 2. This mismatch can be removed if the source is encoded using a 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 a non-uniform distribution. This is because, if the source redundancy is in the form of a non-uniform distribution, no such mismatch occurs between the source and the

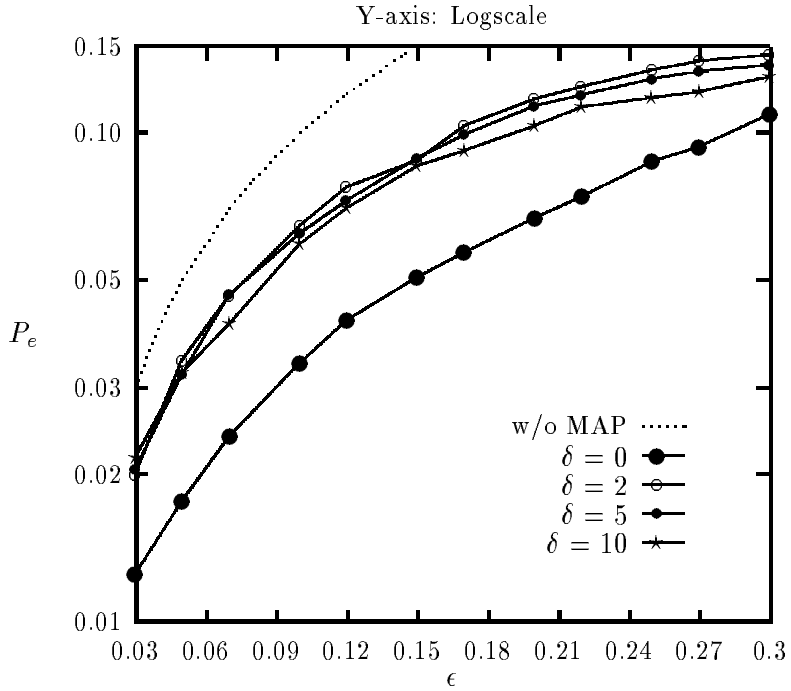


Figure 7: Probability of error vs. BER (ϵ); Headscan modeled according to Model 1.

channel, as predicted by Corollary 1 (cf. (10) and [14]).

Consider a rate-one convolutional code described by $V_n = X_n \oplus X_{n-1}$, $n = 1, 2, \ldots$, where $\{X_n\}$ is a symmetric Markov source with transition distribution $q_0 = 1 - q_1 \stackrel{\triangle}{=} q$, and $\{V_n\}$ 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 $\Pr\{V_k = v_k\} = q^{1-v_k}\ (1-q)^{v_k}$, where v_k is 0 or 1, $k = 1, 2, \ldots$. Furthermore, we can write

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

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

A new system employing a rate-one convolutional encoder will function as follows: A sequence of K samples of the symmetric Markov source X^K (one image line) 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 $\hat{V}^K$. The

convolutional decoder is described by the relation $\hat{X}_k = \hat{V}_k \oplus \hat{X}_{k-1}$, $k = 1, 2, \ldots, N$ with $\hat{X}_1 = \hat{V}_1$. The estimate $\hat{X}^K$ of the transmitted sequence is thus obtained. Note, however, that decoding errors in the sequence MAP detector cause error propagations in the convolutional decoder, which may be significant if an odd number of decoding errors occurs. The effect of error propagation can be effectively limited by grouping the K source samples into small blocks of length N and transmitting/decoding the N-tuples K/N times instead of the entire sequence of length K (assuming perfect synchronization between the receiver and the transmitter).

Since images are non-stationary, image lines modeled according to Model 1 can be classified in two ways:

1. Lines having $\rho_M \gg \rho_D$; they tend to behave like binary symmetric Markov chains. This occurs when q_0 is close to $1 - q_1$. In this case, equation (5) dominates (6) in Theorem 1.

2. Lines for which neither ρ_M nor ρ_D is dominant. In this case no mismatch occurs.

We hence employ a new adaptive encoding system on the image lines that takes into consideration the above observations. Each image line modeled according to Model 1 is processed as follows:

- After computing q_0 and q_1, evaluate ρ_M and ρ_D.

- If $\rho_M \geq T * \rho_D$, where T is some fixed threshold (e.g., $T = 10$), then use a rate-one convolutional encoder as described above. Transmit the output $\{V_n\}$ of the convolutional encoder in blocks of length N over the Markov channel and decode using the iid distribution of $\{V_n\}$.

- If $\rho_M < T * \rho_D$, proceed as usual: transmit the image line over the channel and decode via the MAP algorithm using the line statistics q_0 and q_1.

Simulation results on the performance of this new system are displayed in Figures 8–11 for all four test images. The experiments were repeated seven times and we used $T = 10$ and $N = 8$. The results clearly indicate very good performance and that the mismatch is eliminated; indeed, the performance of the MAP decoder improves as the channel correlation parameter δ increases. This is due to the fact that when the image pixels behave like a symmetric Markov chain, they are converted into an iid non-uniform process. Hence the MAP detector observes an iid non-uniform source corrupted by Markovian noise. As δ increases, the noise correlation in the channel increases (hence decreasing the noise entropy rate and increasing the channel capacity); this enhances the detector's capability to estimate the transmitted sequence.

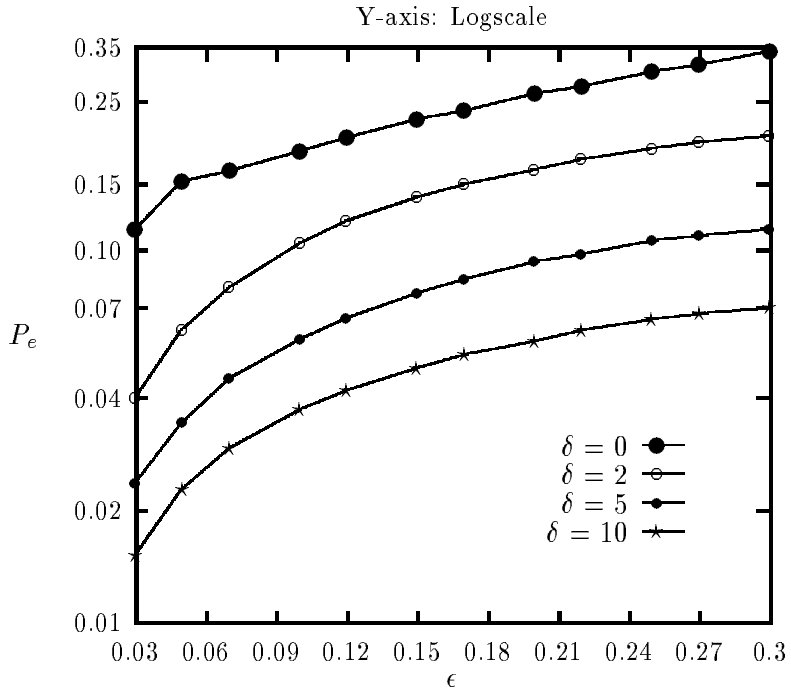


Figure 8: P_e vs. ϵ; Lena 1 modeled according to Model 1 using adaptive encoding scheme.

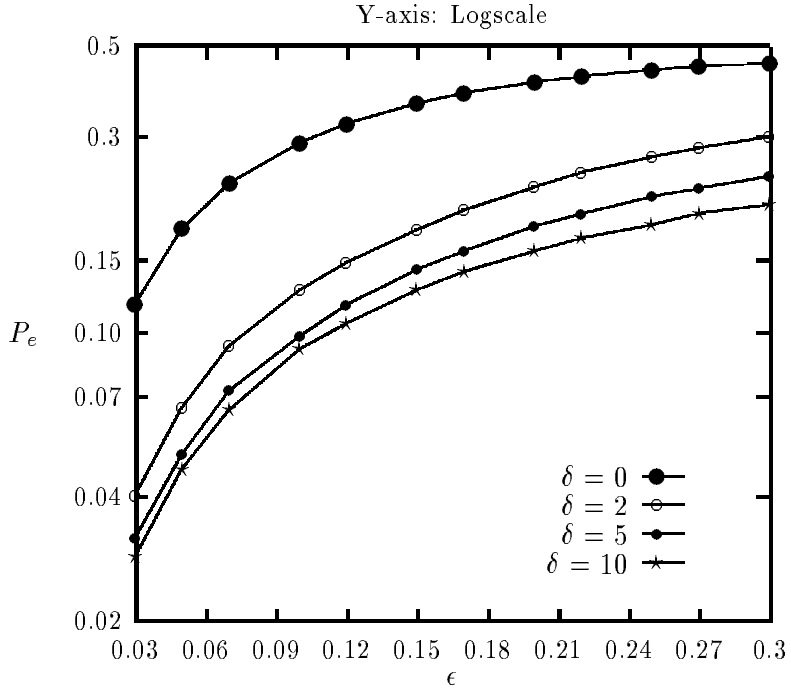


Figure 9: P_e vs. ϵ; Lena 2 modeled according to Model 1 using adaptive encoding scheme.

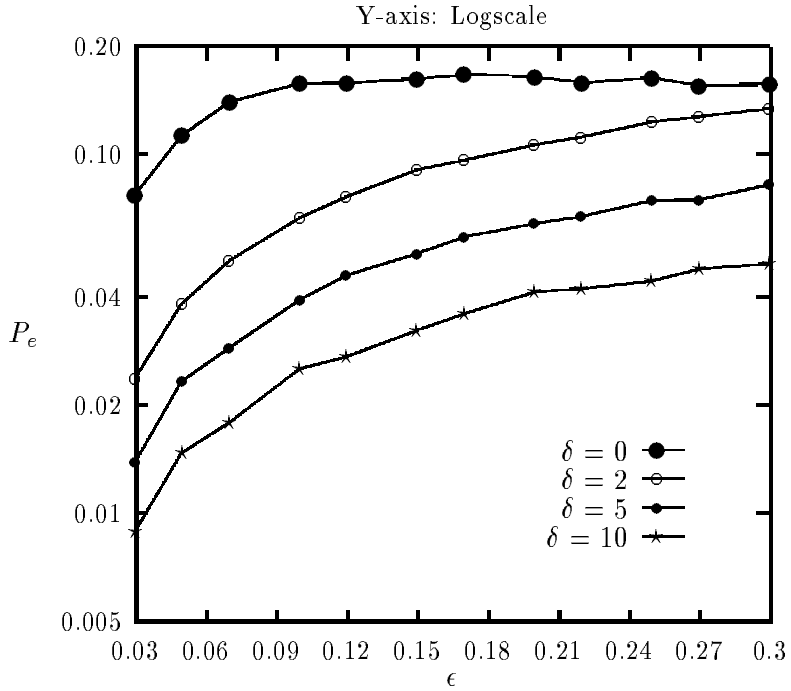


Figure 10: P_e vs. ϵ; Headscan modeled according to Model 1 using adaptive encoding scheme.

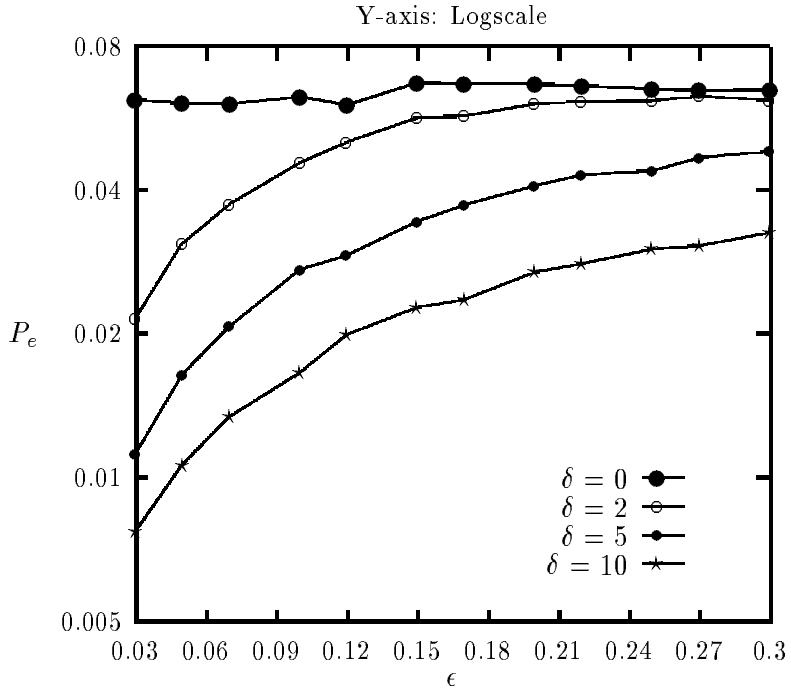


Figure 11: P_e vs. ϵ; CCITT modeled according to Model 1 using adaptive encoding scheme.

In Figures 12 and 13, examples of the transmission of the four two-tone test images using the above adaptive system are displayed. In these examples the channel BER is $\epsilon = 0.1$ and the correlation parameter is $\delta = 10$. These parameters correspond to a very noisy channel with high noise correlation. The adaptive scheme encodes the image lines via the rate-one encoder if $\rho_M \geq 10\rho_D$ before sending them over the channel. For each of the images, the decoded image is displayed after the received image as if it were not encoded, for comparison purposes. The gains in dB achieved by the adaptive MAP decoder over the case when no MAP decoding is done are: 4.68 dB for Lena 1 (Figure 12.b), 1.94 dB for Lena 2 (Figure 12.d), 6.02 dB for Headscan (Figure 13.b), and 7.69 dB for CCITT (Figure 13.d).

5.3 Comparison with Tandem Schemes

In a traditional tandem coding scheme, the source and channel codes are designed separately. We show in this section that the performance of the above joint source-channel coding scheme outperforms that of a more complex traditional tandem source-channel coding method. The tandem scheme employed here includes the following elements: (i) Huffman encoder: Grouping the image pixel stream in blocks of 4 bits, we encode the source stream using a 4^{th} order Huffman code. (ii) Convolutional encoder: The output of the Huffman coder is sent through a convolutional encoder of rate $R = \frac{k}{n} = \frac{1}{2}$. The convolutional code has input memory $m = 2$ (4 states) and tap coefficients $(1, 0, 1)$ and $(1, 1, 1)$ [23]. Its minimum free distance is $d_{\text{free}} = 5$, its minimum distance is $d_{\min} = 3$, and its constraint length is $n(m + 1) = 6$ bits. (iii) Interleaver, Markov channel, de-interleaver: Since most channel codes are designed for the memoryless channel, the traditional approach to handling a channel with memory is to convert it into a memoryless channel by use of an interleaver/de-interleaver pair. We assume ideal interleaving which renders the channel memoryless ($\delta = 0$); i.e., it transforms bursts of channel errors into isolated errors and thus enhances the error correction capability of the convolutional code. (iv) Decoders: ML decoder implemented using the Viterbi algorithm, and a Huffman decoder.

The complexity of the tandem coding scheme is substantially higher than that of our proposed joint coding scheme. In addition to this, interleaving introduces delay. Note also that the above tandem scheme closely matches actual fax compression schemes which use run length coding followed by Huffman encoding [21].

In Figures 14 and 15, we compare the performance of the tandem scheme, with perfect interleaving, with two of our proposed MAP detection schemes on images Lena 1 and Headscan. The

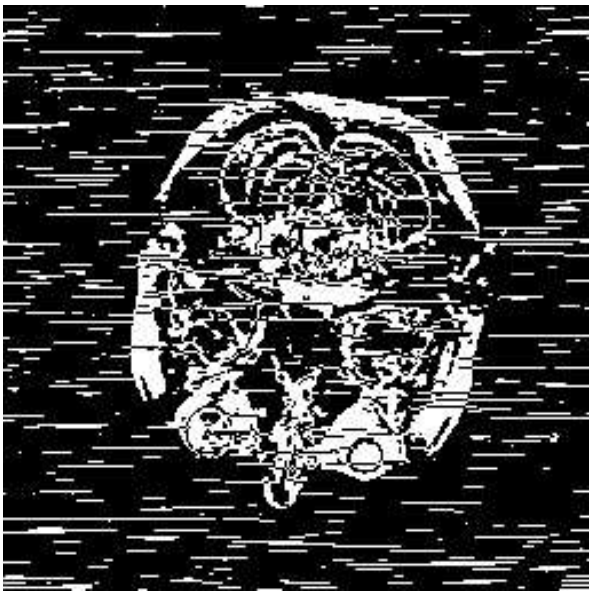
(a) Received uncoded Lena 1

(b) Decoded Lena 1, $P_e = 0.034$

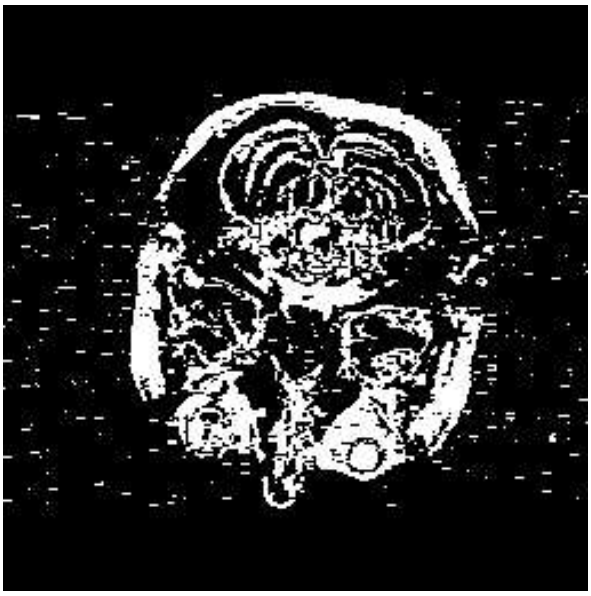
(c) Received uncoded Lena 2

(d) Decoded Lena 2, $P_e = 0.064$

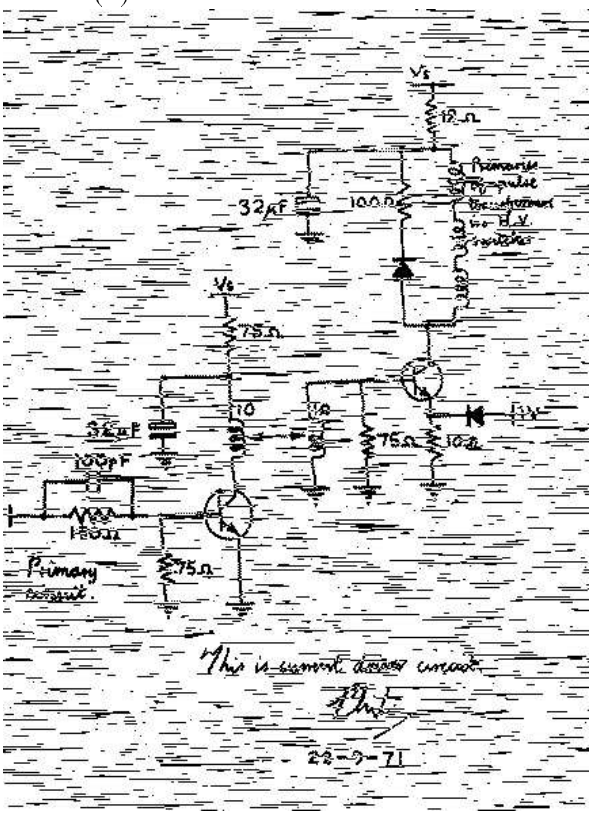
Figure 12: Transmission of Lena 1 and Lena 2 with adaptive scheme; $\epsilon = 0.1$, $\delta = 10$, $N = 4$, $T = 10$.



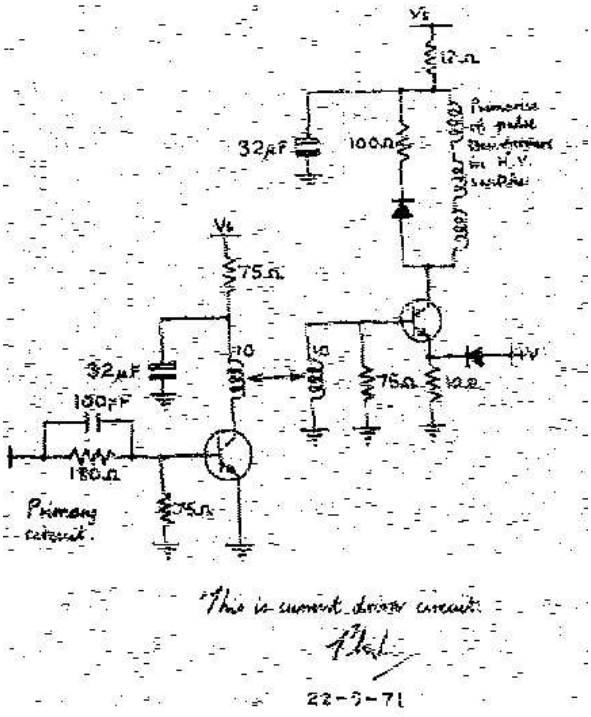
(a) Received uncoded Headscan



(b) Decoded Headscan, $P_e = 0.025$



(c) Received uncoded CCITT



(d) Decoded CCITT, $P_e = 0.017$

Figure 13: Transmission of Headscan and CCITT with adaptive scheme; $\epsilon = 0.1$, $\delta = 10$, $N = 4$, $T = 10$.

first MAP decoding scheme is the one that models the image lines as second-order Markov chains (according to Model 0) and uses perfect interleaving ($\delta = 0$). The second MAP scheme is the adaptive rate-one convolutional encoding scheme discussed in the previous section. Note that the first MAP scheme exploits only the image redundancy, while the second adaptive MAP scheme exploits both the image redundancy and the noise correlation. The simulations were run seven times and average values of the bit probability of error were compiled for different values of the channel BER. We remark that both MAP decoding schemes outperform the tandem scheme while being substantially less complex, with the adaptive MAP scheme yielding the highest gains. For example, when Lena 1 is sent over the channel with $\epsilon = 0.1$ and $\delta = 10.0$, the adaptive MAP decoding scheme outperforms the tandem scheme by as much as 8 dB.

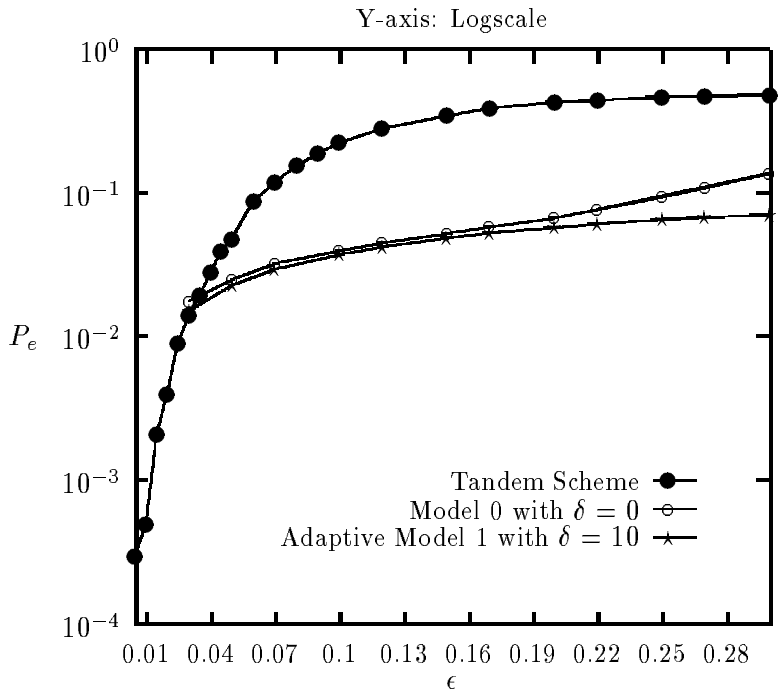


Figure 14: Lena 1: P_e vs. ϵ; Tandem scheme versus MAP decoding scheme using Model 0 and interleaving, and MAP decoding scheme using adaptive Model 1; $\delta = 10$.

5.4 Overhead Information

As in all joint source-channel coding schemes, it is assumed that the image statistics are available at the decoder. This can be achieved by transmitting them along with the image using a forward error-correcting code, or estimating them at the decoder during good channel conditions. We

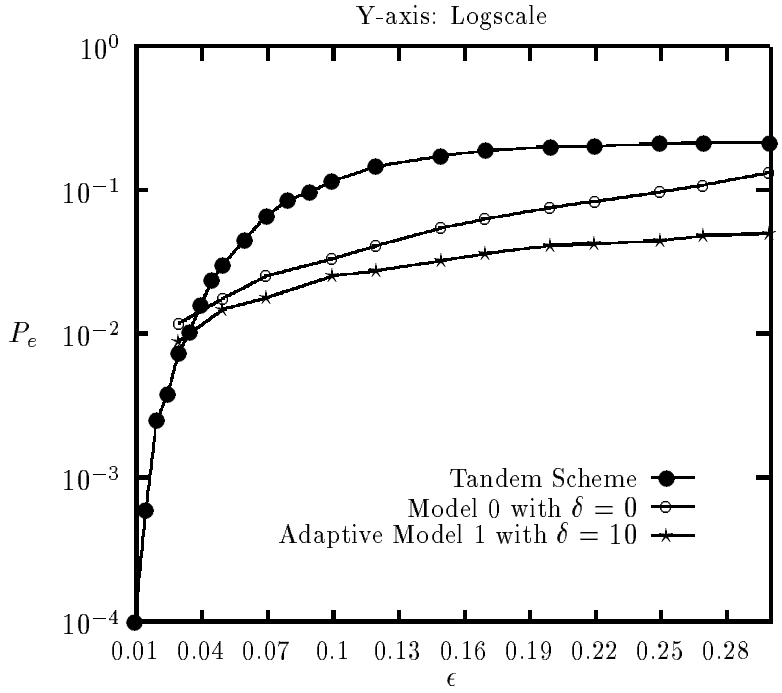


Figure 15: Headscan: P_e vs. ϵ; Tandem scheme versus MAP decoding scheme using Model 0 and interleaving, and MAP decoding scheme using adaptive Model 1; $\delta = 10$.

will concentrate on the first method and compute the amount of overhead information needed to transmit the image parameters.

We assume that a rate $1/R$ convolutional encoder is used to protect the source statistics. Note that if we represent the image using Model 0, we need to transmit four conditional probabilities per line, whereas we need to send two conditional probabilities per line for Model 1. Therefore the adaptive MAP scheme using Model 1—which yields the best performance—needs as much as half the amount of overhead required by the MAP scheme using Model 0. If we let l denote the number of accuracy digits for each source parameter, then we need $\lceil \log_2(10^l - 1) \rceil$ bits to describe the parameter.

The percentage of overhead information is equal to

$$\% \text{ Overhead } = \frac{mR\lceil \log_2(10^l - 1) \rceil}{K},$$

where K is the image width and m is the number of source statistics per line ($m = 4$ for Model 0, and $m = 2$ for Model 1). The amount of overhead needed for each of the two-tone test images is presented in Table 1 for $R = 2$ and $l = 1, 2$.

Table 1: Percentage of Overhead; $R = 2$.

l	Lena $K = 512$	CCITT $K = 432$	Headscan $K = 256$
1	6.25 %	7.41 %	12.5 %
2	10.94 %	12.96 %	21.87 %

(a) Model 0

l	Lena $K = 512$	CCITT $K = 432$	Headscan $K = 256$
1	3.13 %	3.70 %	6.25 %
2	5.47 %	6.48 %	10.94 %

(b) Model 1

Remark: Note that we can avoid transmitting overhead information about the source statistics by using training images to estimate the statistics and then evaluate the performance of the system on a distinct sequence of testing images. Of course, this approach is valid only for applications where the images belong to a particular class—e.g., in the transmission of medical magnetic resonance images (MRI) [17].

5.5 Robustness Under Imperfectly Known Channel Statistics

Up to now, we have assumed that the channel statistics (ϵ and δ) are known a priori at the receiver. In this section, we investigate the robustness of the MAP decoding system when these parameters are not known perfectly. This may occur due to inadequate estimation of the channel parameters, particularly when the channel is time-varying (e.g., LMR channels). Simulations using the adaptive MAP decoding scheme were performed for all four test images; however, we display the results only for Lena 1 since all images yielded similar performance behavior. In Table 2.a, we present PSNR results when the receiver misestimates the BER ϵ with the correlation parameter $\delta = 10$. Note that when $\epsilon_a = \epsilon_d = 0.0$, then $P_e = 0$; this results in a PSNR of infinite value (Table 2.a). In Table 2.b, we provide PSNR results when the receiver misestimates the correlation parameter δ with the channel BER $\epsilon = 0.1$. We observe from Table 2 that when the channel is noisy ($\epsilon_a \geq 0.05$) or when its correlation parameter $\delta_a \geq 2$, it is better to overestimate the true parameters than to underestimate them. Finally, we can conclude that the MAP scheme is *not* very sensitive to mistakes in estimating ϵ or δ, provided that we do not design ϵ or δ to be zero when the actual parameter is non-zero.

Table 2: Lena 1: Robustness results for adaptive MAP decoding scheme in PSNR (dB); $T = 10, N = 4$; ϵ_d = Design BER; ϵ_a = Actual BER; δ_d = Design Correlation Parameter; δ_a = Actual Correlation Parameter; (a) Robustness with BER ($\delta_d = \delta_a = 10$); (b) Robustness with Correlation Parameter ($\epsilon_d = \epsilon_a = 0.1$).

	$\epsilon_d = 0.00$	$\epsilon_d = 0.01$	$\epsilon_d = 0.05$	$\epsilon_d = 0.10$
$\epsilon_a = 0.00$	∞	27.71	21.89	19.33
$\epsilon_a = 0.01$	22.12	23.09	20.41	18.54
$\epsilon_a = 0.05$	14.94	17.35	17.45	16.27
$\epsilon_a = 0.10$	12.07	14.73	14.68	14.79

(a)

	$\delta_d = 0$	$\delta_d = 2$	$\delta_d = 5$	$\delta_d = 10$
$\delta_a = 0$	8.25	6.71	6.67	6.67
$\delta_a = 2$	8.49	10.69	10.65	10.61
$\delta_a = 5$	8.46	12.91	12.92	12.92
$\delta_a = 10$	8.55	14.57	14.58	14.79

(b)

5.6 Transmission of Bit-Plane Encoded Grey-Level Images

A grey-level image can be separated into a set of 8 one-bit planes, where each plane is subsequently compressed using binary image coding techniques such as run-length or Huffman coding [21]. This method is very sensitive to channel errors since it employs variable length encoding schemes. Furthermore, it typically yields a compression ratio on the order of 2, leaving little room for protection against channel noise.

We can apply the adaptive MAP decoding scheme discussed above to bit-plane encoded grey-level images transmitted over the Markov channel. Experimental results are shown in Figures 16–20 for three grey-level images: Lena, Headscan, and a bone x-ray. In Figures 16–18, the resulting PSNR plots for the MAP-decoded images show significant improvements over the received images. For $\delta = 10$, gains in excess of 6 dB are achieved when the channel BER is greater than or equal to 0.1.

6 CONCLUSIONS

In this paper we have presented a joint source-channel coding scheme for the reliable transmission of two-tone images over channels with memory. Three image models are proposed and image redundancy is characterized for each of these models. It is shown that a strong correlation exists between image pixels. We demonstrate empirically that for second-order Markov models, the best results

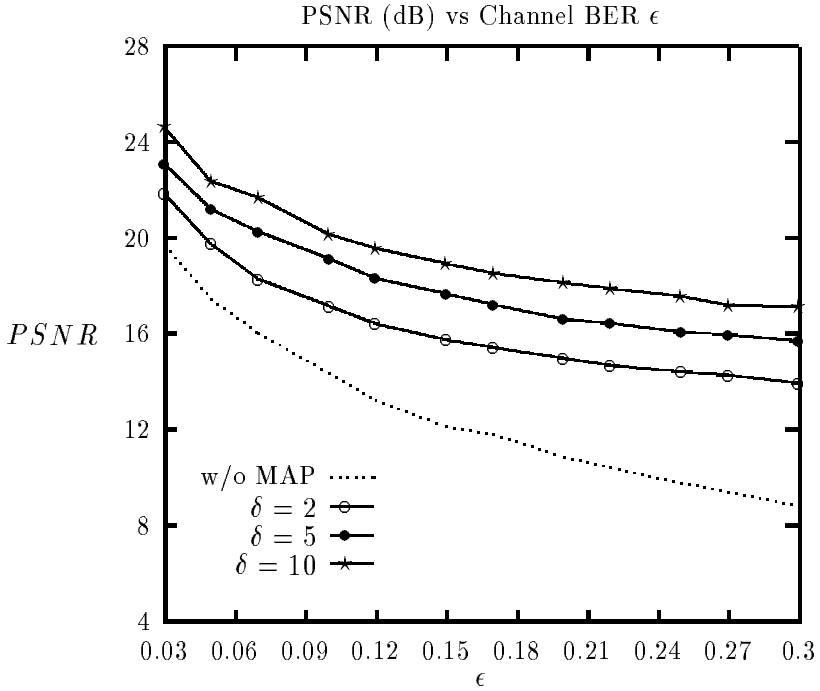


Figure 16: PSNR (dB) vs. ϵ; Grey Headscan modeled according to Model 1 using adaptive encoding scheme; $T = 10$, $N = 4$.

are obtained for the interleaved channel. An adaptive scheme is used for first-order Markov image models to avoid mismatch and exploit the channel memory. The sensitivity to misspecification of channel parameters and the amount of necessary overhead information are discussed. The MAP scheme is extended to the transmission of bit-plane-encoded grey-level images.

Current efforts focus on low-bit-rate joint source-channel coding of grey-level images using unequal error protection; this may be achieved by encoding the most significant bits in a way similar to that described above, and by employing a tandem scheme (transform coding followed by Reed-Solomon or convolutional coding) for the least significant bits. Future work will address the use of soft decision decoding in conjunction with trellis-coded modulation (TCM) for the MAP channel decoding of JPEG and MPEG signals over bursty fading channels.

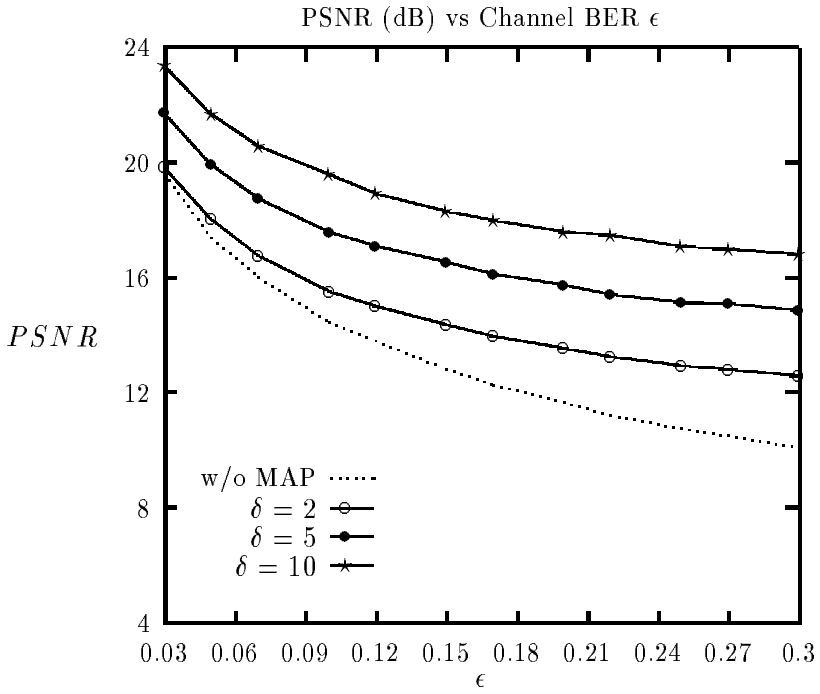


Figure 17: PSNR (dB) vs. ϵ; Grey Lena modeled according to Model 1 using adaptive encoding scheme; $T = 10$, $N = 4$.

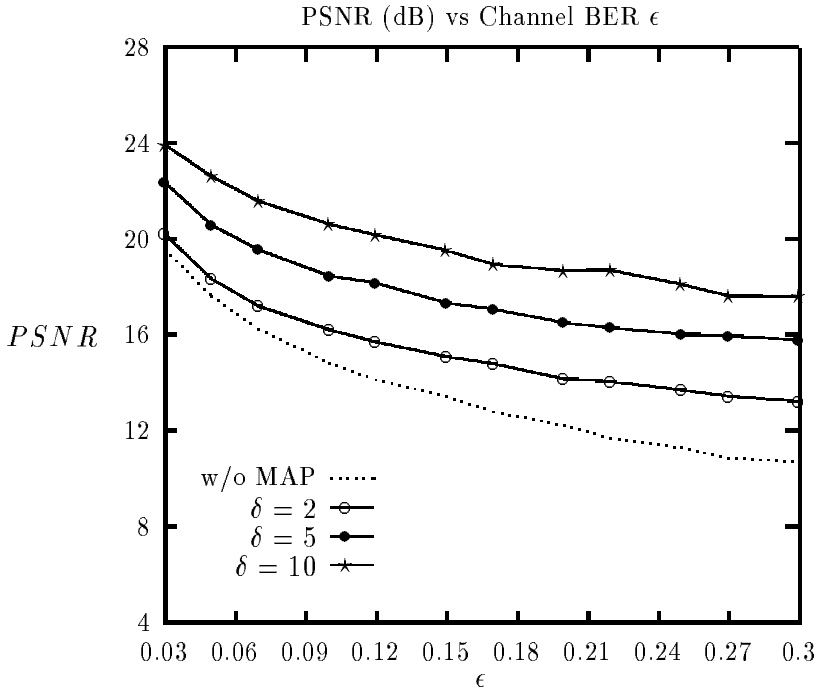
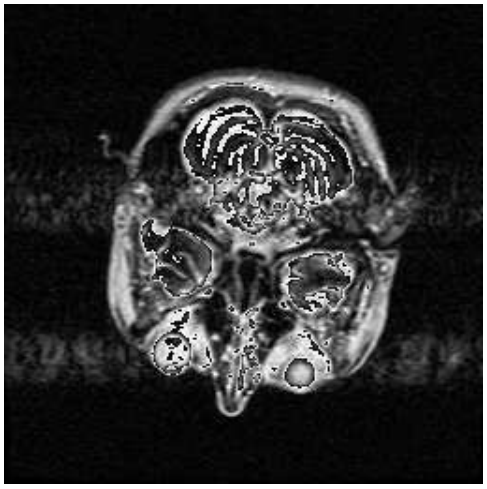


Figure 18: PSNR (dB) vs. ϵ; Grey MRI Bone modeled according to Model 1 using adaptive encoding scheme; $T = 10$, $N = 4$.

(a) Grey Lena: Original

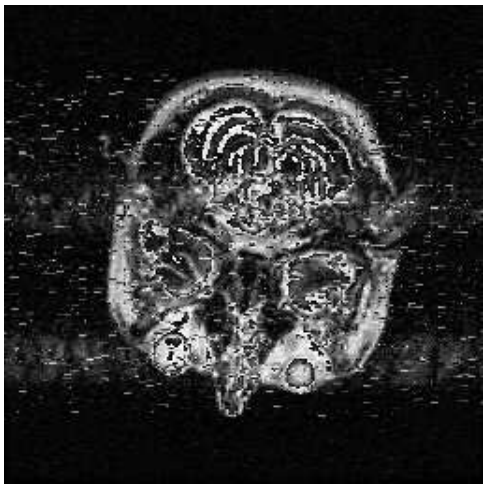


(b) Grey Headscan: Original

(c) Lena received uncoded; PSNR = 14.45 dB

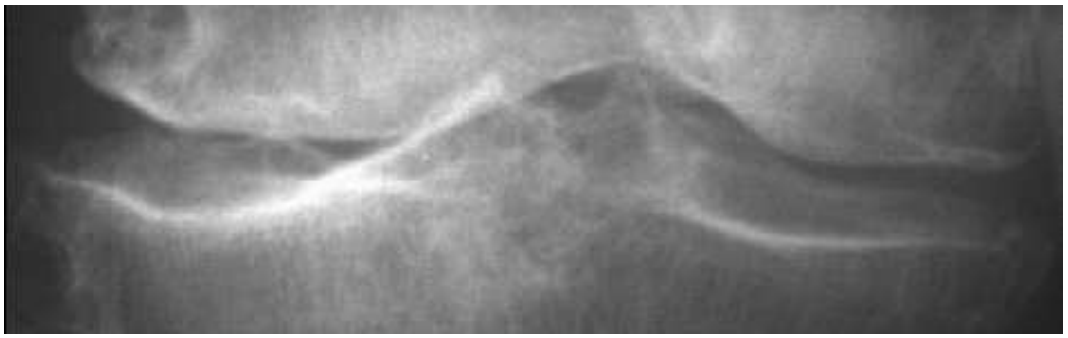
(d) Grey Headscan: Received uncoded; PSNR = 14.25 dB

(e) Decoded Grey Lena: PSNR = 19.53 dB

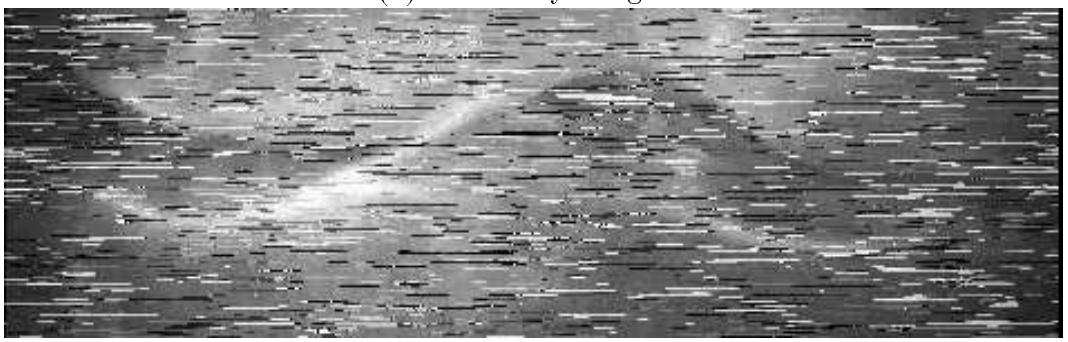


(f) Decoded Grey Headscan: PSNR = 20.25 dB

Figure 19: Transmission of Lena and Headscan with adaptive scheme; $\epsilon = 0.1$, $\delta = 10$, $N = 4$, $T = 10$.



(a) Bone x-ray: Original



(b) Bone x-ray received uncoded; PSNR = 14.82 dB

(c) Bone x-ray decoded: PSNR = 20.65 dB

Figure 20: Transmission of Bone x-ray with adaptive scheme; $\epsilon = 0.1$, $\delta = 10$, $N = 4$, $T = 10$.

References

[1] C.E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal,* Vol. 27, pp. 379–423, July 1948.

[2] C.E. Shannon, "Coding Theorems for a Discrete Source with a Fidelity Criterion," *IRE National Convention Record*, pp. 142–163, March 1959.

[3] E. Ayanoglu and R. Gray, "The Design of Joint Source and Channel Trellis Waveform Coders," *IEEE Transactions on Information Theory,* Vol. 33, pp. 885–865, November 1987.

[4] N. Phamdo, N. Farvardin, and T. Moriya, "A Unified Approach to Tree-Structured and Multi-Stage Vector Quantization for Noisy Channels," *IEEE Transactions on Information Theory*, Vol. 39, pp. 835–850, May 1993.

[5] J. Modestino, D. Daut, and A. Vickers, "Combined Source-Channel Coding of Images Using the Block Cosine Transform," *IEEE Transactions on Communications*, Vol. 29, pp. 1261–1274, September 1981.

[6] V. Vaishampayam and N. Farvardin, "Optimal Block Cosine Transform Image Coding for Noisy Channels," *IEEE Transactions on Communications*, Vol. 38, pp. 327–336, March 1990.

[7] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 6, pp. 721–741, 1984.

[8] W. Xu, J. Hagenauer and J. Hollmann, "Joint Source-Channel Decoding Using the Residual Redundancy in Compressed Images," submitted to *International Conference on Communications*, Dallas, TX, June 1996.

[9] J. Hagenauer, "Joint Source and Channel Coding for Broadcast Applications," *Audio and Video Digital Radio Broadcasting Systems and Techniques,* Elsevier, Amsterdam, 1994.

[10] J. Hagenauer, "Source-Controlled Channel Decoding," *IEEE Transactions on Communications*, Vol. 43, pp. 2449–2457, September 1995.

[11] K. Sayood and J.C. Borkenhagen, "Use of Residual Redundancy in the Design of Joint Source/Channel Coders," *IEEE Transactions on Communications*, Vol. 39, pp. 838–846, June 1991.

[12] K. Sayood, F. Liu and J.D. Gibson, "A Constrained Joint Source/Channel Coder Design," *IEEE Journal on Selected Areas in Communications,* Vol. 12, pp. 1584–1593, December 1994.

[13] S. Emmani and S.L. Miller, "DPCM Picture Transmission over Noisy Channels with the Aid of a Markov Model," *IEEE Transactions on Image Processing,* Vol. 4, pp. 1473–1481, November 1995.

[14] F. Alajaji, N. Phamdo, N. Farvardin and T. Fuja, "Detection of Binary Markov Sources Over Channels with Additive Markov Noise," *IEEE Transactions on Information Theory*, to appear, 1996.

[15] F. Alajaji, N. Phamdo and T. Fuja, "Channel Codes that Exploit the Residual Redundancy in CELP-Encoded Speech", *IEEE Transactions on Speech and Audio Processing*, accepted for publication, 1995.

[16] R. Chellappa and A.K. Jain (Eds.), *Theory and Applications of Markov Random Fields*, Academic Press, Boston, 1993.

[17] B.S. Srinivas, E.A. Riskin, R. Ladner and M. Azizoglu, "Progressive Image Transmission on a Channel with Memory," *Proceedings of the Allerton Conference on Communication, Control and Computing*, October 1995.

[18] F. Alajaji and T. Fuja, "A Communication Channel Modeled on Contagion," *IEEE Transactions on Information Theory*, Vol. 40, pp. 2035–2041, November 1994.

[19] L. Kanal and A. Sastry, "Models for Channels with Memory and their Applications to Error Control," *Proceedings of the IEEE*, Vol. 66, pp. 724–744, July 1978.

[20] F. Alajaji, N. Phamdo, N. Farvardin and T. Fuja, "Detection of Binary Sources Over Discrete Channels with Additive Markov Noise," University of Maryland Institute for Systems Research Technical Report 94-31, 1994.

[21] A.K. Jain, *Fundamentals of Digital Image Processing,* Prentice Hall, Englewood Cliffs, NJ, 1989.

[22] G.D. Forney, "The Viterbi Algorithm," *Proceedings of the IEEE*, Vol. 61, pp. 268–278, March 1973.

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