MDS codes with large symbols developed in [3], [4] are ideally suited to holographic recording. All this, possibly combined with two-dimensional interleaving [2], provides an extremely powerful coding scheme for holographic memory systems.

References

[1] S. Al-Bassam and B. Bose, "Conservative codes," in *IEEE Int. Symp. on Information Theory* (San Antonio, TX, 1993).

[2] M. Blaum and J. Bruck, "Correcting two-dimensional clusters by interleaving of symbols," in *IEEE Int. Symp. on Information Theory* (Trondheim, Norway, 1994).

[3] M. Blaum, J. Bruck, and A. Vardy, "MDS array codes with independent parity symbols," *IEEE Trans. Inform. Theory*, vol. 42, no. 2, Mar. 1996, to be published.

[4] M. Blaum and R. M. Roth, "New array codes for multiple phased burst correction," *IEEE Trans. Inform. Theory*, vol. 39, pp. 66–77, 1993.

[5] M. Blaum, P. H. Siegel, G. T. Sincerbox, and A. Vardy, "Method and apparatus for modulation of multi-dimensional data in holographic storage," filed with the US Patent Office, Aug. 1994.

[6] D. Brady and D. Psaltis, "Control of volume holograms," *J. Opt. Soc. Amer. A*, vol. 9, pp. 1167–1182, 1992.

[7] R. A. Brualdi, *Introductory Combinatorics*. New York: North-Holland, 1977.

[8] P. S. Henry, "Zero disparity coding system," US Patent 4 309 694, 1982.

[9] D. E. Knuth, "Efficient balanced codes," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 51–53, 1986.

[10] S. A. Lis and P. D. Henshaw, "Ultra-dense optical mass storage," AD-A232 767, report prepared for USAF, AFSC, Feb. 1991.

[11] Y. Ofek, "A family of conservative codes with block delimiters for decoding without a phase-locked loop," in *ACM Comput. Sci. Conf.* (Baltimore, MD, 1987).

[12] ——, "The conservative code for bit synchronization," *IEEE Trans. Commun.*, vol. 38, pp. 1107–1113, 1990.

Detection of Binary Markov Sources Over Channels with Additive Markov Noise

Fady Alajaji, Nam Phamdo, *Member, IEEE*, Nariman Farvardin, *Senior Member, IEEE*, and Thomas E. Fuja, *Member, IEEE*

***Abstract*— We consider maximum *a posteriori* (MAP) detection of a binary asymmetric Markov source transmitted over a binary Markov channel. Here, the MAP detector observes a long (but finite) sequence of channel outputs and determines the most probable source sequence. In some cases, the MAP detector can be implemented by simple rules such as the "believe what you see" rule or the "guess zero (or one) regardless of what you see" rule. We provide necessary and sufficient conditions under which this is true. When these conditions are satisfied, the exact bit error probability of the sequence MAP detector can be determined. We examine in detail two special cases of the above source: i) binary independent and identically distributed (i.i.d.) source and ii) binary symmetric Markov source. In case i), our simulations show that the performance of the MAP detector improves as the channel noise becomes more correlated. Furthermore, a comparison of the proposed system with a (substantially more complex) traditional tandem source-channel coding scheme portrays superior performance for the proposed scheme at relatively high channel bit error rates. In case ii), analytical as well as simulation results show the existence of a "mismatch" between the source and the channel (the performance *degrades* as the channel noise becomes more correlated). This mismatch is reduced by the use of a simple rate-one convolutional encoder.**

***Index Terms*— Markov source, Markov channel, source redundancy, MAP detection.**

I. Introduction and Motivation

A source with memory as well as a memoryless source with a nonuniform distribution are sources with *redundancy*. For a finite alphabet of size J, a uniformly distributed independent and identically distributed (i.i.d.) 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 $H_\infty(X)$ [9]: $\rho_T = \log_2 J - H_\infty(X)$, where

$$H_\infty(X) \triangleq \lim_{n\to\infty} \frac{1}{n} H(X_1, X_2, \cdots, X_n).$$

The redundancy may be attributed to the nonuniform source distribution or to the source memory (or both). 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 a nonuniform distribution and

Manuscript received March 27, 1995; revised June 26, 1995. The work of F. Alajaji, N. Farvardin, and T. E. Fuja was supported in part by the National Science Foundation under Grant NSFD CD-8803012. The work of N. Phamdo was supported in part by NTT Corporation. The material in this correspondence was presented in part at the IEEE International Symposium on Information Theory, Trondheim, Norway, June 1994.

F. Alajaji is with the Department of Mathematics and Statistics, Queen's University, Kingston, ON K7L 3N6, Canada.

N. Phamdo is with the Electrical Engineering Department, State University of New York, Stony Brook, NY 11794-2350 USA.

N. Farvardin and T. E. Fuja are with the Electrical Engineering Department, University of Maryland, College Park, MD 20742 USA.

Publisher Item Identifier S 0018-9448(96)00016-8.

$\rho_M \triangleq H(X_1) - H_\infty(X)$ denotes the redundancy in the form of memory [9].

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 nonuniformity). This correspondence addresses the advantages of using this redundancy in controlling 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 Markov process with memory order M. The motivation for the use of such a channel is founded in the fact that most real-world communication channels have memory; this contagion-based model offers an interesting and less complex alternative to the Gilbert model [5] and others [6].

We first investigate the problem of detecting a binary asymmetric first-order Markov source ($\rho_D > 0$ and $\rho_M > 0$) transmitted across the contagion Markov channel of order one ($M = 1$). Two *maximum a posteriori* (MAP) formulations are considered: 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. In instantaneous MAP detection, we estimate the most probable transmitted *bit* at a particular time given all the received bits up to that time [10].

In general, the sequence MAP detector is implemented by the Viterbi algorithm and the instantaneous MAP detector is implemented recursively. However, in some special cases, these implementations are not necessary. For example, the optimal choice may be to "believe what you see" (singlet decoding) or to always estimate "zero" regardless of "what you see." We say, in these cases, that MAP detection is *useless*. In this correspondence, we provide necessary and sufficient conditions under which the sequence MAP detector is useless. These results are in the same spirit as previous results on MAP detection of Markov sources over discrete memoryless channels [3], [10].

Two special cases of the above asymmetric Markov source are considered. In the first case, we consider a binary nonuniform i.i.d. source (redundancy strictly in the form of nonuniform distribution, $\rho_D > 0$ and $\rho_M = 0$). The above necessary and sufficient conditions are simplified for this case. Simulation results which confirm these theoretical conditions are given. We observe that the performance of the MAP detector improves as the channel noise becomes more correlated. We also show, via simulations, that for channels with relatively high bit error rates, the performance of this scheme (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.

In the second case, we consider a binary symmetric Markov source (redundancy strictly in the form of memory, $\rho_D = 0$ and $\rho_M > 0$). Again, we simplify the necessary and sufficient conditions for the uselessness of the sequence MAP detector. The simplified condition predicts the existence of a *mismatch* between the binary symmetric Markov source and the Markov channel. That is, unlike the first case, the performance of the MAP detector *degrades* as the channel noise becomes more correlated. This is illustrated by simulation results for the sequence and instantaneous MAP detectors. 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 bits as it accepts. The purpose of the convolutional encoder is to convert the symmetric Markov source into a nonuniform i.i.d. random process, by transforming its redundancy from the form of memory into redundancy in the form of nonuniform distribution. Simulation results showing considerable improvement by the use of this simple code are obtained.

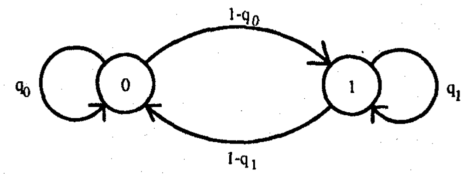

Fig. 1. Binary asymmetric Markov source model.

II. Source and Channel Models

A. Source Model

Consider a stationary ergodic binary first-order Markov source $\{X_n\}_{n=1}^{\infty}$. Fig. 1 illustrates the Markov chain, where $q_0 \in (0,1)$ and $q_1 \in (0,1)$ are the probabilities of remaining in states "zero" and "one," respectively. Denote

$$P(x_n \mid x_{n-1}) \triangleq \Pr\{X_n = x_n \mid X_{n-1} = x_{n-1}\}$$

and

$$P(x_n) \triangleq \Pr\{X_n = x_n\}$$

where $x_n, x_{n-1} \in \{0,1\}$. It can be easily shown that

$$P(0) = 1 - P(1) = (1-q_1)/(2-q_0-q_1).$$

Note that, in general, the Markov chain is asymmetric and the source redundancy is in the form of memory as well as in the form of a nonuniform distribution. We will be particularly interested in two special cases of the above source.

- *Special Case 1*: $q_0 = 1 - q_1 \neq 1/2$. Here, the source becomes a *nonuniform i.i.d.* source with distribution $P(0) = q_0$. The source redundancy is strictly in the form of a nonuniform distribution ($\rho_M = 0$ and $\rho_D > 0$).
- *Special Case 2*: $q_0 = q_1 \neq 1/2$. Here, the source becomes a *symmetric* Markov chain and the source redundancy is strictly in the form of memory ($\rho_M > 0$ and $\rho_D = 0$).

We will investigate these two cases in Sections IV and V, respectively.

B. Channel Model

The source $\{X_n\}_{n=1}^{\infty}$ is transmitted across a binary additive noise channel described by

$$Y_n = X_n \oplus Z_n$$

where $\oplus$ represents modulo-2 addition, Z_n is the channel noise, and Y_n is the channel output. The noise $\{Z_n\}_{n=1}^{\infty}$ is assumed to be independent of the source $\{X_n\}_{n=1}^{\infty}$. Furthermore, we assume that $\{Z_n\}_{n=1}^{\infty}$ is generated by the finite-memory contagion urn model derived in [1]. According to this model, the noise process is an Mth-order Markov chain in which the noise sample Z_n depends on the past only through the *sum* of the previous M noise samples. More specifically, for $n > M$

$$\Pr\{Z_n = 1 \mid Z_{n-M} = z_{n-M}, \cdots, Z_{n-1} = z_{n-1}\} = \frac{\epsilon + \delta \sum_{i=1}^{M} z_{n-i}}{1 + M\delta}$$

where $z_{n-i} \in \{0,1\}$ for $i = 1, 2, \cdots, M$. Here $\epsilon = \Pr\{Z_n = 1\}$ is the channel bit error rate (BER) (we assume that $0 < \epsilon \leq 1/2$) and δ is an nonnegative parameter which determines the amount of correlation in $\{Z_n\}_{n=1}^{\infty}$. The correlation coefficient of two adjacent noise samples is $\delta/(1+\delta)$. Note that if $\delta = 0$, the noise process becomes i.i.d. and the resulting additive noise channel becomes a memoryless binary symmetric channel (BSC) with BER ϵ.

We are particularly interested in the special case where $M = 1$ (the noise is a first-order Markov process). In this case, we denote

$$Q(z_n \mid z_{n-1}) \triangleq \Pr\{Z_n = z_n \mid Z_{n-1} = z_{n-1}\}$$

and

$$Q(z_n) \triangleq \Pr\{Z_n = z_n\}.$$

Note that

$$\begin{bmatrix} Q(0 \mid 0) & Q(1 \mid 0) \\ Q(0 \mid 1) & Q(1 \mid 1) \end{bmatrix} = (1+\delta)^{-1} \begin{bmatrix} 1-\epsilon+\delta & \epsilon \\ 1-\epsilon & \epsilon+\delta \end{bmatrix}$$

and $Q(1) = \epsilon = 1 - Q(0)$. In this correspondence, all theoretical results are given for this special case. However, some simulation results will be given for $M > 1$.

III. MAP Detection of Asymmetric Markov Sources

We investigate the problem of optimal detection of the binary asymmetric Markov source when it is transmitted across an additive Markov channel of memory order one ($M = 1$). Two *maximum a posteriori* (MAP) detection formulations are considered:

- 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.

A. Sequence MAP Detection

Given that we observe $Y^n = y^n = (y_1, y_2, \cdots, y_n)$ at the output of the channel,[1] we desire to determine the most probable transmitted *sequence* $\hat{x}^n$ where

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

It can be easily shown (see [2], [10]) that the above is equivalent to

$$\hat{x}^n = \arg \min_{x^n \in \{0,1\}^n} \left[-\log(Q(z_1)P(x_1)) - \sum_{k=2}^{n} \log(Q(z_k \mid z_{k-1})P(x_k \mid x_{k-1})) \right] \quad (1)$$

where $z_k = x_k \oplus y_k$ for $k = 1, 2, \cdots, n$. As expressed in (1), the sequence MAP detector can be implemented using the Viterbi algorithm [4]. We let $\{x_k\}_{k=1}^{n}$ 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

$$-\log(Q(y_k \oplus x_k \mid y_{k-1} \oplus x_{k-1})P(x_k \mid x_{k-1})).$$

The surviving path for each state is the path with the smallest 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.

[1] In this correspondence, the superscript n denotes a vector of dimension n.

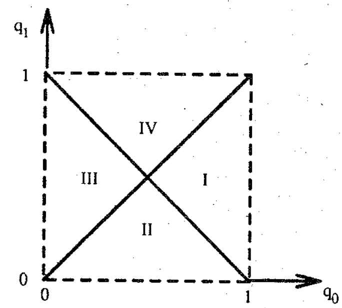

Fig. 2. Four regions for the parameters (q_0, q_1) of the binary asymmetric Markov source.

B. Instantaneous MAP Detection

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

$$\begin{aligned} \hat{x}_n &= \arg \max_{x_n \in \{0,1\}} \Pr\{X_n = x_n \mid Y^n = y^n\} \\ &= \arg \max_{x_n \in \{0,1\}} \Pr\{X_n = x_n, Y^n = y^n\}. \end{aligned}$$

Let

$$f^{(n)}(x_n) \triangleq \Pr\{X_n = x_n, Y^n = y^n\}$$

denote the objective function that we wish to maximize at time instant n. It is straightforward to show that $f^{(n)}(x_n)$ can be determined recursively according to (see [2], [10])

$$\begin{aligned} f^{(1)}(x_1) &= Q(y_1 \oplus x_1)P(x_1) \\ f^{(n)}(x_n) &= \sum_{x_{n-1} \in \{0,1\}} Q(y_n \oplus x_n \mid y_{n-1} \oplus x_{n-1}) \\ &\quad \cdot P(x_n \mid x_{n-1}) f^{(n-1)}(x_{n-1}), \quad n = 2, 3, \cdots. \end{aligned}$$

Instantaneous MAP detection is also referred to in the literature as MAP filtering.

C. Analytical Results for Sequence MAP Detection

In [10], it was observed that in some instances the output of the MAP detector is identical to its input, i.e., $\hat{X}_n = Y_n$ for all n. In such cases, we say that the MAP detector is "useless." That is, the MAP detector need not be implemented at all. In the following, we provide necessary and sufficient conditions for the uselessness of the sequence MAP detector.

Recall that the pair (q_0, q_1) can take values anywhere on the square $(0,1) \times (0,1)$. At this point, it is convenient to partition this square into four regions, as shown in Fig. 2. We first consider Region I. The other three regions will be considered subsequently.

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

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

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

and

$$\frac{1-\epsilon+\delta}{\epsilon+\delta} \frac{q_1}{q_0} \geq 1. \tag{3}$$

ii) If (2) does not hold, then $\hat{X}^n = Y^n$ is not an optimal sequence detection rule.

iii) If (3) 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: See Appendix I. □

From the proof of this theorem, it can be easily seen that if one of the inequalities in (2) and (3) is strict then the sequence MAP detector is *unique* and is given by $\hat{X}^n = Y^n$. In the remaining part of this correspondence, we will not emphasize this point since the uniqueness of the MAP detector will be clear from the context.

Condition iii) can be refered to as an asymptotic "weak" converse to i), since the counterexample we provided in proving it, utilizes input and output sequences which yield a nontypical noise sequence (i.e., the sequence can occur with low probability). If $q_0 = q_1 = 1/2$, (2) and (3) always hold and the MAP detector is useless. This can be clearly seen from the fact that in this case, the source becomes an i.i.d. uniformly distributed process, hence containing no redundancy.

In Region II ($q_1 \in (0, \frac{1}{2}]$, $q_0 \in [q_1, 1-q_1]$), an analogous theorem can be proven. The necessary and sufficient conditions for Region II are

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

and

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

Note that (5) is identical to (3). For Region III, the conditions are the same as (4) and (5) except that q_0 and q_1 are interchanged. Finally, for Region IV, the conditions are the same as (2) and (3) except that q_0 and q_1 are interchanged.

When $P(0)$ is close to unity, a reasonable guess is to choose $\hat{X}^n = 0^n$ regardless of the observations. The following theorem provides necessary and sufficient conditions under which the rule $\hat{X}^n = 0^n$ is optimal.[2] Here, we assume that the first and last transmitted bits are zero.

Theorem 2: Given $q_0 \in (0,1)$, $q_1 \in (0,1)$, $\epsilon \in (0, \frac{1}{2}]$, $\delta \geq 0$, and $n \geq 3$, assume that $X_1 = X_n = 0$. Then

i) $\hat{X}^n = 0^n$ is an optimal sequence (MAP) detection rule if

$$\frac{\epsilon(1-\epsilon)}{(1-\epsilon+\delta)^2} \frac{q_0^2}{(1-q_0)(1-q_1)} \geq 1 \tag{6}$$

and

$$\frac{\epsilon+\delta}{1-\epsilon+\delta} \frac{q_0}{q_1} \geq 1. \tag{7}$$

[2] In this correspondence, 0^n denotes the all-zero n-tuple. Similarly, 1^n denotes the all-one n-tuple.

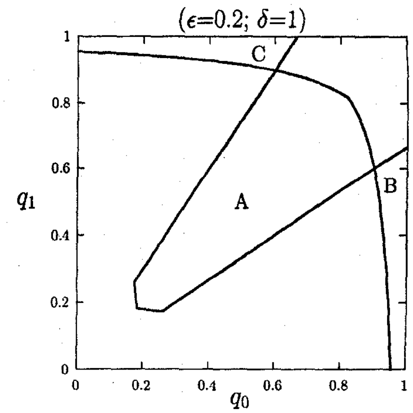

Fig. 3. An example of the regions where sequence MAP detection is useless with $\epsilon = 0.2$ and $\delta = 1$. In Region A, $\hat{X}^n = Y^n$ is optimal. In Region B, $\hat{X}^n = 0^n$ is optimal. In Region C, $\hat{X}^n = 1^n$ is optimal.

ii) If (6) does not hold, then $\hat{X}^n = 0^n$ is not an optimal sequence detection rule.

iii) If (7) does not hold, then $\exists\, n_0 > 0$ such that $\forall\, n \geq n_0$ $\hat{X}^n = 0^n$ is not an optimal sequence detection rule.

Proof: See Appendix II. □

It is interesting to note that conditions (6) and (7) are the converses (with the exception of equality) of (2) and (3), respectively. Note that Theorem 2 holds for all four regions in Fig. 2. However, in Regions III and IV, (7) never holds if $\epsilon < 1/2$, and hence $\hat{X}^n = 0^n$ is never optimal for large n. This is because, in these regions, $P(0) \leq \frac{1}{2}$ and a better guess would be $\hat{X}^n = 1^n$. The necessary and sufficient conditions for the optimality of $\hat{X}^n = 1^n$ (assuming that $X_1 = X_n = 1$) are the same as (6) and (7) except that q_0 and q_1 are interchanged.

It is easier to understand the above theorems by examining (2)–(7) graphically. In Fig. 3 we plot, for the case $\epsilon = 0.2$ and $\delta = 1$, the regions of the unit square where sequence MAP detection is useless. In these regions, we know the bit error probability, $\hat{\epsilon} = \Pr\{\hat{X}_n \neq X_n\}$, exactly. In Region A ($\hat{X}^n = Y^n$), $\hat{\epsilon} = \epsilon$; in Region B ($\hat{X}^n = 0^n$), $\hat{\epsilon} = P(1) = (1-q_0)/(2-q_0-q_1)$; and in Region C ($\hat{X}^n = 1^n$), $\hat{\epsilon} = P(0) = (1-q_1)/(2-q_0-q_1)$.

For a fixed δ, as ϵ decreases Region A expands while Regions B and C shrink. This is consistent with our intuition. When the channel is clean (small ϵ), we would expect to use the singlet decoding rule ($\hat{X}^n = Y^n$). When the channel is very noisy (large ϵ), we would expect to just guess $\hat{X}^n = 0^n$ or 1^n depending on which one is more probable. On the other hand, for a fixed ϵ, as δ increases, Region A shrinks along the $q_1 = 1 - q_0$ axis but *expands* along the $q_1 = q_0$ axis. This is a rather unexpected result and more will be said about it in Section V.

The above two theorems are for the sequence MAP detector. We do not give any theoretical result for the instantaneous MAP detector in this correspondence. For the BSC ($\delta = 0$), some results are given in [3]. For the general case, we would conjecture that the instantaneous MAP detector will be useless whenever the sequence MAP detector is useless.

IV. Special Case 1: Binary I.I.D. Sources

We next consider the case where the source transition distributions q_0 and q_1 are such that $q_0 = 1 - q_1 \neq 1/2$ (the diagonal line with slope -1 in Fig. 2). This results in a binary i.i.d. nonuniform

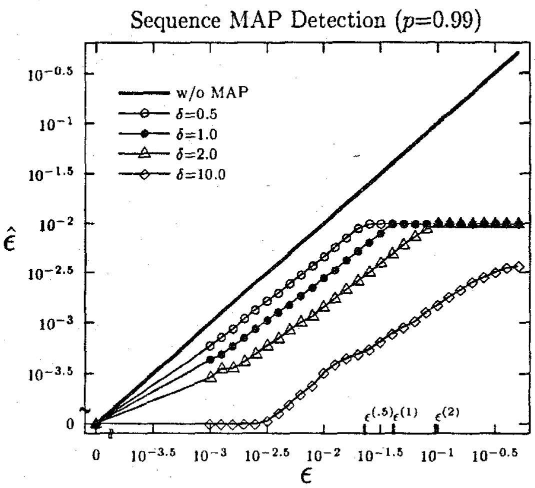

Fig. 4. Performance of sequence MAP detector for i.i.d. binary source with $p = 0.99$ over the contagion Markov channel ($M = 1$). $\epsilon =$ channel bit error rate, $\hat{\epsilon} = \Pr\{\text{bit error}\}$, and $\delta =$ correlation parameter of the channel.

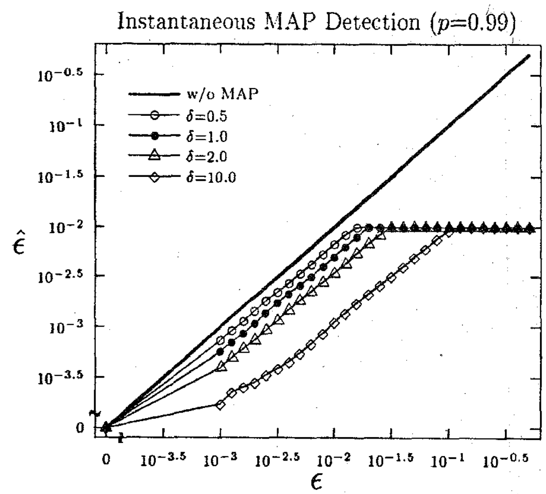

Fig. 5. Performance of instantaneous MAP detector for i.i.d. binary source with $p = 0.99$ over the contagion Markov channel ($M = 1$). $\epsilon =$ channel bit error rate, $\hat{\epsilon} = \Pr\{\text{bit error}\}$, and $\delta =$ correlation parameter of the channel.

source with probability distribution $P(0) = q_0 \triangleq p$. We will assume without loss of generality that $p > 1/2$.

Using $q_0 = 1 - q_1 = p > 1/2$ in (2) and (3), we obtain that (3) implies (2); this yields the following corollary of Theorem 1.

Corollary 1: 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

$$\frac{1-\epsilon+\delta}{\epsilon+\delta}\frac{1-p}{p} \geq 1. \tag{8}$$

ii) If (8) 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 (8) is equivalent to

$$\delta \leq \delta_1 \triangleq \frac{1-\epsilon-p}{2p-1} \tag{9}$$

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

Similarly, we realize that (6) implies (7) if we assume that $q_0 = 1 - q_1 = p > 1/2$. A corollary of Theorem 2 is therefore obtained.

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

$$\frac{\epsilon(1-\epsilon)}{(1-\epsilon+\delta)^2}\frac{p}{1-p} \geq 1. \tag{10}$$

A. Simulation Results

In Figs. 4–7, simulation results for the sequence and instantaneous MAP detectors are plotted. Each simulation was performed on 1000 samples of the i.i.d. source and the experiment was repeated 500 times. In Figs. 4–6, $\hat{\epsilon} = \Pr\{\hat{X}_n \neq X_n\}$ 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., $\hat{\epsilon} = \epsilon$). Fig. 7 shows the plot of the average value of $\hat{\epsilon}$ versus the channel correlation parameter δ.

In Figs. 4 and 7, the performances of sequence MAP detection for the iid source (with $p = 0.99$ and 0.95, respectively) over the Markov channel (with $M = 1$), are presented. We can remark that, in general, as δ increases, the performance of the MAP detector improves (or at least it does not degrade). This is due to the fact that as δ increases, the noise correlation in the channel increases (hence increasing the channel capacity) which enhances the detector's capability in estimating the transmitted sequence. It should be noted that the more redundant the source is (i.e., the closer to one is p), the better is the performance of the MAP detector. Additional simulation results can be found in [2].

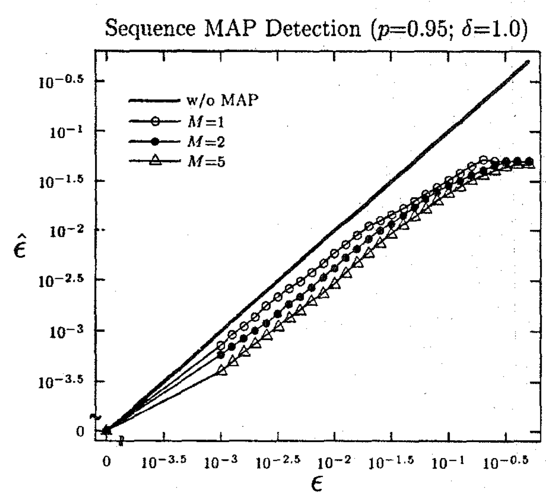

Fig. 6. Performance of sequence MAP detector for i.i.d. binary source with $p = 0.95$ over the contagion Markov channel of order M with $\delta = 1.0$. $\epsilon =$ channel bit error rate and $\hat{\epsilon} = \Pr\{\text{bit error}\}$.

In Fig. 5, the performance of instantaneous MAP detection for the i.i.d. source (with $p = 0.99$) over the Markov channel is presented. The instantaneous MAP decoder behaves very much like the sequence MAP detector, except that its performance is slightly inferior to the later. This is because the instantaneous decoder admits no delay and decodes bit by bit, while the sequence decoder observes an entire vector before decoding it. Fig. 6 shows the effect of the order M of the Markov channel on the performance of the sequence MAP detection[3] for a source with $p = 0.95$ and a channel with $\delta = 1.0$.

[3] In the implementation of the sequence MAP detector, the state at time k

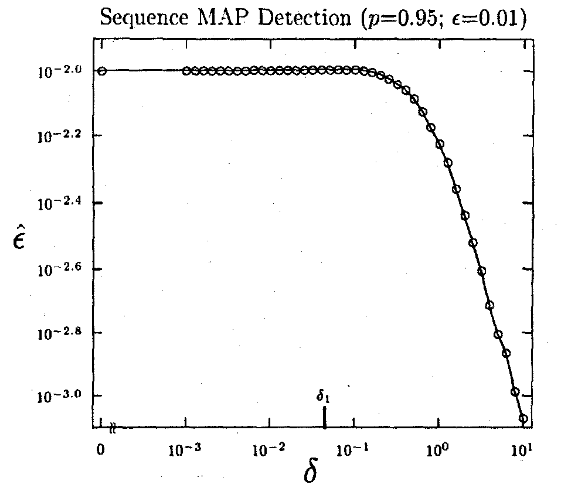

Fig. 7. Performance of sequence MAP detector for i.i.d. binary source with $p = 0.95$ over the contagion Markov channel ($M = 1$) with $\epsilon = 0.01$. $\hat{\epsilon} = \Pr\{\text{bit error}\}$, and $\delta =$ correlation parameter of the channel.

Here again, the performance improves as M increases, since the channel capacity increases with the memory M.

The analytical results of Corollaries 1 and 2 are illustrated in Figs. 4 and 7. Corollary 1 is illustrated in Fig. 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 Fig. 7, the curve of $\hat{\epsilon}$ diverges from the constant value of $\hat{\epsilon} = 10^{-2}$ for a δ slightly larger than δ_1; this is because Corollary 1 offers a sufficient condition and an asymptotic converse that relies on a nontypical noise sequence; it thus has a low chance of occurring in a simulation.

Furthermore, the simulations shown in Fig. 4 agree with Corollary 2. Corollary 2 offers a necessary and sufficient condition for which the all-zero sequence is the optimal sequence. Note that (10) is equivalent to

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

We plot in Fig. 4 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 $\delta = 10$, the all-zero sequence is never the optimal sequence since (10) does not hold.

B. Comparison with Tandem Source–Channel Coding Schemes

The system we consider in this correspondence can be thought of as a joint source–channel coding scheme. The MAP detector exploits the inherent source redundancy to correct channel errors. We now compare this system against a 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

consists of the vector $(x_k, x_{k-1}, \cdots, x_{k-M+1})$. Thus there are 2^M states with two branches entering and leaving each state.

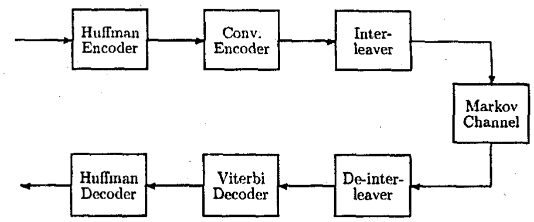

Fig. 8. Block diagram of the tandem scheme.

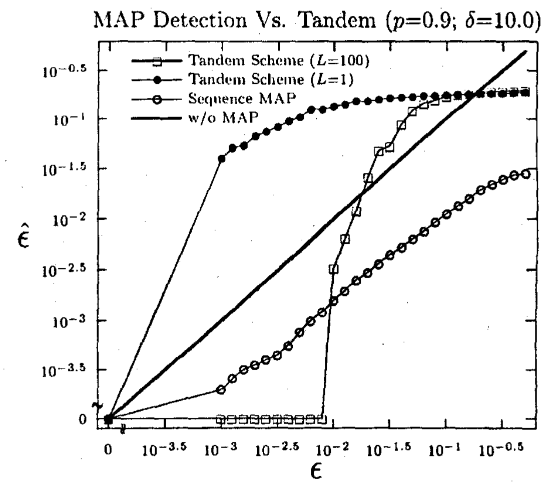

Fig. 9. Comparisons of proposed sequence MAP detection system versus Tandem source–channel coding system. Binary i.i.d. source with $p = 0.9$. $\hat{\epsilon} = \Pr\{\text{bit error}\}$, $\epsilon =$ channel bit error rate, $\delta =$ correlation parameter of the channel, and $L =$ interleaving length.

considered includes an interleaver/de-interleaver pair as it is depicted in Fig. 8. It consists of the following:

- Huffman encoder: We assume that the i.i.d. source has distribution $p = 0.9$; thus its entropy rate is 0.469 bit/sample. Grouping the source stream in blocks of 4 bits, we encode it using a fourth-order Huffman code with average code length of 0.492 bits/sample.
- Convolutional encoder: We match the output of the Huffman encoder to a convolutional encoder of rate $\frac{1}{2}$. It has an input memory of two and the following tap coefficients $(1,0,1)$ and $(1,1,1)$ [7]. Its minimum free distance is equal to five.
- Interleaver, Markov channel, de-interleaver, decoders: The interleaver 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. The interleaver size is $L \times L$. The decoders used are, respectively, an ML Viterbi decoder [4], [7] 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 two decoders (Viterbi and Huffman), two 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 may lead to a larger delay.

In Fig. 9, 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 [7]. The simulations were run 50 times on 10000 samples of the i.i.d. source. We observe that the noninterleaved 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 noninterleaved 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 memory at relatively high bit error rates ($\epsilon \geq 10^{-2}$), the proposed system beats the tandem scheme while being substantially less complex.

V. Special Case 2: Binary Symmetric Markov Sources

In this section we consider the case where $q_0 = q_1 = q > 1/2$ (the diagonal line with slope 1 in Fig. 2). This results in a binary symmetric Markov source. With this new condition, Theorem 1 yields the following corollary.

Corollary 3: 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

$$\frac{(1-\epsilon+\delta)^2}{\epsilon(1-\epsilon)}\left(\frac{1-q}{q}\right)^2 \geq 1. \tag{11}$$

Remark: Condition (11) is equivalent to

$$\delta \geq \delta_2 \triangleq \left(\frac{q}{1-q}\right)\sqrt{\epsilon(1-\epsilon)} + \epsilon - 1. \tag{12}$$

Observations:

- The above necessary and sufficient condition indicates that for fixed ϵ and q (hence fixed δ_2), sequence MAP detection becomes useless for sufficiently large δ (cf. (12)). Thus the sequence MAP detector performance will be no better than $\hat{\epsilon} = \epsilon$ even though the noise is highly correlated. This shows that a *mismatch* exists between the symmetric Markov source and the Markov channel which prevents the MAP detector from exploiting the noise correlation (see Section IV-A).
- For $q_0 = q_1$, Theorem 2 has no corollary since (7) never holds (if $\epsilon < 1/2$). This can be indeed verified in Fig. 3; the line $q_0 = q_1$ never intersects with Regions B and C which are the respective regions for the all-zero and all-one sequences being the optimal MAP solutions.

A. Simulation Results

In Figs. 10 and 11, we present simulation results for the sequence MAP detector. We performed each simulation on 1000 samples of the symmetric Markov source and repeated the experiment 500 times. We remark from the plots that the performance of the MAP detectors deteriorates as the value of δ increases and read a constant value of $\hat{\epsilon} = \epsilon$ for large δ. This is clearly illustrated in Fig. 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 Corollary 3.[4]

[4]We note that, contrary to our expectation, the curve in Fig. 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.

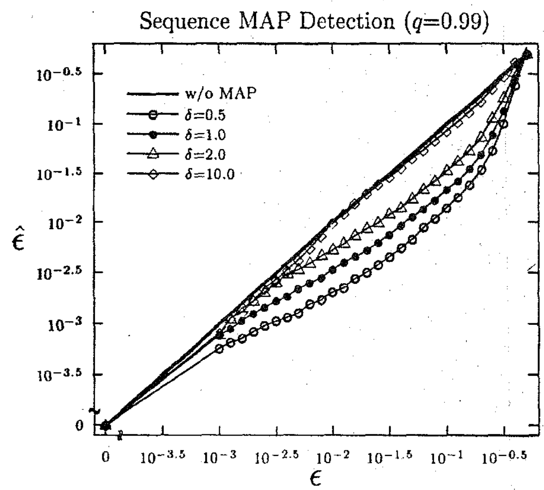

Fig. 10. Performance of sequence MAP detector for binary symmetric Markov source with $q = 0.99$ over the contagion Markov channel ($M = 1$). $\epsilon =$ channel bit error rate, $\hat{\epsilon} = \Pr\{\text{bit error}\}$, and $\delta =$ correlation parameter of the channel.

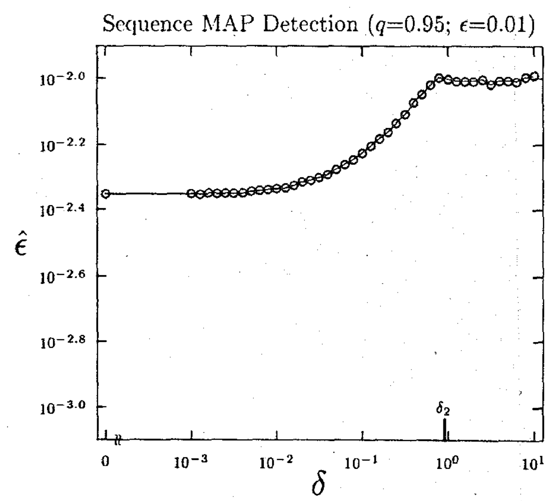

Fig. 11. Performance of sequence MAP detector for binary symmetric Markov source with $q = 0.95$over the contagion Markov channel ($M = 1$) with $\epsilon = 0.01$. $\hat{\epsilon} = \Pr\{\text{bit error}\}$, and $\delta =$ correlation parameter of the channel.

B. Rate-One Convolutional Encoding

If we directly connect a binary symmetric Markov source to the 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 nonuniform distribution. This is because, if the source redundancy is in the form of nonuniform distribution, no such mismatch occurs between the source and the channel, as seen in Section IV.

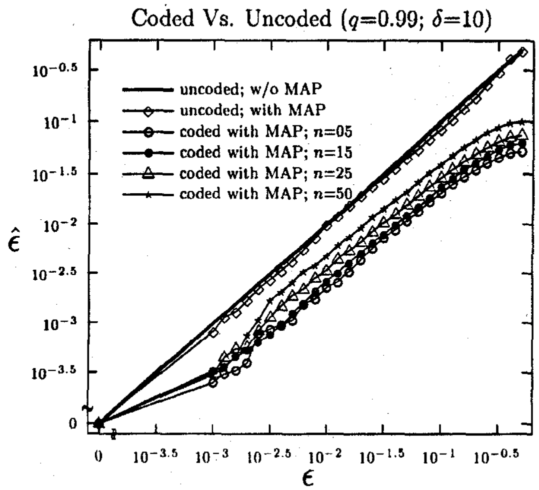

Fig. 12. Performance of the coded system with sequence MAP detection for binary symmetric Markov source with $q = 0.99$ over the contagion Markov channel ($M = 1$) with $\delta = 10$. $\hat{\epsilon} = \Pr\{\text{bit error}\}$, $\epsilon =$ channel bit error rate, and $\delta =$ correlation parameter of the channel.

We employ a rate-one convolutional code described by $V_n = X_n \oplus X_{n-1}$, $n = 1, 2, \cdots$, 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 $V_1 = X_1$. It can be easily checked that $\{V_n\}$ is a nonuniform binary i.i.d. process with distribution given by $\Pr\{V_k = 0\} = q > 1/2$.

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 $\hat{X}_k = \hat{V}_k \oplus \hat{X}_{k-1}$, $k = 1, 2, \cdots, N$ 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. 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 Fig. 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.

VI. Conclusion

In this correspondence, we considered 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) asymmetric Markov source, ii) nonuniform i.i.d. source, and iii) symmetric Markov source. Analytical results giving conditions for the uselessness of the sequence MAP detector as well as simulation results were presented. For the case of the nonuniform i.i.d. source, we showed that, for certain source and channel parameters, the proposed simple system beats a traditional tandem source–channel coding scheme for high channel bit error rates. A mismatch was established for case iii) 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 investigated in [8]. Future work may consist of comparing the results above to those obtained by detecting sources over the Gilbert channel with applications to digital cellular channels.

Appendix I

In this appendix, we prove Theorem 1. First, we need to prove the following lemma which provides a lower bound on the ratio of probabilities of two binary sequences which agree on the first and last bits but disagree everywhere else.

Lemma 1: Let $q_0 \in [\frac{1}{2}, 1)$, $q_1 \in [1 - q_0, q_0]$, and $L \geq 2$. Assume that $z_0^L = (z_0, z_1, \cdots, z_L) = (0, 1, 1, \cdots, 1, 0)$. Then $\forall\ y_0^L \in \{0,1\}^{L+1}$

$$\prod_{k=1}^{L} \frac{P(y_k \mid y_{k-1})}{P(x_k \mid x_{k-1})} \geq \frac{(1-q_0)(1-q_1)}{q_0^2} \left(\frac{q_1}{q_0}\right)^{L-2} \tag{13}$$

where $x_0^L = z_0^L \oplus y_0^L$, $P(0 \mid 0) = q_0$, $P(1 \mid 0) = 1 - q_0$, $P(0 \mid 1) = 1 - q_1$ and $P(1 \mid 1) = q_1$.

Proof: Write

$$\begin{aligned} s &\triangleq \prod_{k=1}^{L} \frac{P(y_k \mid y_{k-1})}{P(x_k \mid x_{k-1})} \\ &= \left[\frac{P(y_1 \mid y_0)}{P(\bar{y}_1 \mid y_0)}\right] \left[\prod_{k=2}^{L-1} \frac{P(y_k \mid y_{k-1})}{P(\bar{y}_k \mid \bar{y}_{k-1})}\right] \left[\frac{P(y_L \mid y_{L-1})}{P(y_L \mid \bar{y}_{L-1})}\right] \end{aligned}$$

where the overbar denotes the binary complement. Note that when $L = 2$

$$s = \frac{P(y_1 \mid y_0)\ P(y_2 \mid y_1)}{P(\bar{y}_1 \mid y_0)\ P(y_2 \mid \bar{y}_1)}$$

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 $\mathcal{K} = \{2, 3, \cdots, L-1\}$ as follows:

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

where

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

We then rewrite s as

$$\begin{aligned} s = &\left[\frac{P(y_1 \mid y_0)}{P(\bar{y}_1 \mid y_0)}\ \frac{P(y_L \mid y_{L-1})}{P(y_L \mid \bar{y}_{L-1})}\right] \\ &\times \left[\frac{q_1}{q_0}\right]^{|\mathcal{K}_{11}| - |\mathcal{K}_{00}|} \left[\frac{1-q_0}{1-q_1}\right]^{|\mathcal{K}_{01}| - |\mathcal{K}_{10}|}. \end{aligned} \tag{14}$$

The first factor in the right-hand side of (14) is defined as u.

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

Consider the sequence $(y_1, y_2, \cdots, y_{L-1})$. 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 $|\mathcal{K}_{11}| \leq |\mathcal{K}| = L - 2$. Furthermore, the minimum value of u in this case is

$$\frac{(1-q_0)(1-q_1)}{q_0^2}.$$

Hence

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

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

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

This completes the proof of the lemma. □

Proof of Theorem 1: i) We need to show that if (2) and (3) hold, then $\forall x^n, y^n \in \{0,1\}^n$ with $x_1 = y_1$ and $x_n = y_n$

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

We can rewrite α as

$$\begin{aligned} \alpha &= \frac{\Pr\{Y^n = y^n \mid X^n = y^n\}\Pr\{X^n = y^n\}}{\Pr\{Y^n = y^n \mid X^n = x^n\}\Pr\{X^n = x^n\}} \\ &= \left[\frac{Q(0)}{Q(z_1)}\frac{P(y_1)}{P(x_1)}\right]\left[\prod_{k=2}^{n}\frac{Q(0\mid 0)}{Q(z_k\mid z_{k-1})}\frac{P(y_k\mid y_{k-1})}{P(x_k\mid x_{k-1})}\right] \end{aligned}$$

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

$$\alpha = \prod_{k=2}^{n}\left[\frac{1-\epsilon+\delta}{\tilde{Q}(z_k\mid z_{k-1})}\frac{P(y_k\mid y_{k-1})}{P(x_k\mid x_{k-1})}\right].$$

We partition the index set as follows:

$$\begin{aligned} &\mathcal{K} \triangleq \{2, 3, \ldots, n\} \\ &\mathcal{A} \triangleq \{k \in \mathcal{K} : x_k = y_k, x_{k-1} = y_{k-1}\} \\ &\mathcal{B} \triangleq \mathcal{A}^c = \{k \in \mathcal{K} : k \notin \mathcal{A}\} = \{k \in \mathcal{K} : z_k = 1 \text{ or } z_{k-1} = 1\}. \end{aligned}$$

Thus

$$\begin{aligned} \alpha = &\prod_{k\in\mathcal{A}}\left[\frac{1-\epsilon+\delta}{\tilde{Q}(z_k\mid z_{k-1})}\frac{P(y_k\mid y_{k-1})}{P(x_k\mid x_{k-1})}\right] \\ &\cdot\prod_{k\in\mathcal{B}}\left[\frac{1-\epsilon+\delta}{\tilde{Q}(z_k\mid z_{k-1})}\frac{P(y_k\mid y_{k-1})}{P(x_k\mid x_{k-1})}\right]. \end{aligned}$$

By definition of $\mathcal{A}$, the product over $\mathcal{A}$ is unity. We next partition $\mathcal{B}$ as follows:

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

where

$$\begin{aligned} \mathcal{B}_i &= \{m_i+1, m_i+2, \cdots, m_i+L_i\} \\ z_{m_i} &= z_{m_i+L_i} = 0 \\ z_{m_i+1} &= z_{m_i+2} = \cdots = z_{m_i+L_i-1} = 1 \end{aligned}$$

with N denoting the number of partitions 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

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

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 $L_4 = 2$ We mention that this method of partitioning is possible because of the assumption that $z_1 = z_n = 0$.

Thus α can be written as

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

where

$$\begin{aligned} \alpha_i &= \prod_{k=m_i+1}^{m_i+L_i}\left[\frac{1-\epsilon+\delta}{\tilde{Q}(z_k\mid z_{k-1})}\frac{P(y_k\mid y_{k-1})}{P(x_k\mid x_{k-1})}\right] \\ &= \left[\frac{(1-\epsilon+\delta)^2}{\epsilon(1-\epsilon)}\right]\left[\frac{1-\epsilon+\delta}{\epsilon+\delta}\right]^{L_i-2} \\ &\quad\cdot\left[\prod_{k=m_i+1}^{m_i+L_i}\frac{P(y_k\mid y_{k-1})}{P(x_k\mid x_{k-1})}\right] \end{aligned}$$

since

$$(z_{m_i}, z_{m_i+1}, z_{m_i+2}, \cdots, z_{m_i+L_i-1}, z_{m_i+L_i}) = (0,1,1,\cdots,1,0).$$

Applying Lemma 1 yields

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

where the last inequality follows by hypothesis. Hence

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

ii) Suppose that (2) does not hold. Take $y^n = (0,0,\cdots,0,1,0,\cdots,0)$ 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 (3) does not hold. Let $y^n = (0,1,1,\cdots,1)$ and $x^n = (0,0,\cdots,0,1)$. Then

$$\alpha = \left[\frac{1-\epsilon+\delta}{\epsilon}\frac{1-q_0}{q_0}\right]\left[\frac{1-\epsilon+\delta}{\epsilon+\delta}\frac{q_1}{q_0}\right]^{n-3}\left[\frac{1-\epsilon+\delta}{1-\epsilon}\frac{q_1}{1-q_1}\right].$$

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

Appendix II

In this appendix, we prove Theorem 2. The proof is similar to the proof of Theorem 1 and thus we only give its outline. We will use the following lemma which is analogous to Lemma 1.

Lemma 2: Let $\epsilon \in (0, \frac{1}{2}]$, $\delta \geq 0$, and $L \geq 2$. Assume that

$$x_0^L = (x_0, x_1, \cdots, x_L) = (0,1,1,\cdots,1,0).$$

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

$$\prod_{k=1}^{L}\frac{\tilde{Q}(y_k\,|\,y_{k-1})}{\tilde{Q}(z_k\,|\,z_{k-1})} \geq \frac{\epsilon(1-\epsilon)}{(1-\epsilon+\delta)^2}\left(\frac{\epsilon+\delta}{1-\epsilon+\delta}\right)^{L-2} \tag{15}$$

where $z_0^L = x_0^L \oplus y_0^L$, $\tilde{Q}(0\,|\,0) = 1-\epsilon+\delta$, $\tilde{Q}(1\,|\,0) = \epsilon$, $\tilde{Q}(0\,|\,1) = 1-\epsilon$, and $\tilde{Q}(1\,|\,1) = \epsilon+\delta$.

Proof: The proof of this lemma follows exactly the same line of reasoning as the proof of Lemma 1. We will omit it for the sake of brevity.

Proof of Theorem 2: i) We need to show that if (6) and (7) hold, then $\forall x^n, y^n \in \{0,1\}^n$ with $x_1 = x_n = 0$

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

We can rewrite β as

$$\beta = \prod_{k\in\mathcal{B}}\left[\frac{\tilde{Q}(y_k\,|\,y_{k-1})}{\tilde{Q}(z_k\,|\,z_{k-1})}\,\frac{P(0\,|\,0)}{P(x_k\,|\,x_{k-1})}\right]$$

where $z_k = x_k \oplus y_k$ and $\mathcal{B} \triangleq \{k \in \mathcal{K} : x_k = 1 \text{ or } x_{k-1} = 1\}$. We now partition $\mathcal{B}$ as before and write

$$\beta = \prod_{i=1}^{N}\beta_i$$

where

$$\beta_i = \left[\frac{q_0^2}{(1-q_0)(1-q_1)}\right]\left[\frac{q_0}{q_1}\right]^{L_i-2}\left[\prod_{k=m_i+1}^{m_i+L_i}\frac{\tilde{Q}(y_k\,|\,y_{k-1})}{\tilde{Q}(z_k\,|\,z_{k-1})}\right]$$

since

$$(x_{m_i}, x_{m_i+1}, x_{m_i+2}, \cdots, x_{m_i+L_i-1}, x_{m_i+L_i}) = (0,1,1,\cdots,1,0).$$

Applying Lemma 2 yields

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

where the last inequality follows by hypothesis. Hence

$$\beta = \prod_{i=1}^{N}\beta_i \geq 1.$$

ii) Use $y^n = (0,0,\cdots,0,1,0,\cdots,0) = x^n$ as a counterexample.
iii) Use $y^n = (0,1,1,\cdots,1,0) = x^n =$ as a counterexample.

References

[1] F. Alajaji and T. E. Fuja, "A communication channel modeled on contagion," *IEEE Trans. Inform. Theory*, vol. 40, pp. 2035–2041, Nov. 1994.
[2] F. Alajaji, N. Phamdo, N. Farvardin, and T. E. Fuja, "Detection of binary sources over discrete channels with additive Markov noise," Institute for Systems Research Tech. Rep. T.R. 94-31, Univ. of Maryland, College Park, MD, 1994.
[3] J. L. Devore, "A note on the observation of a Markov source through a noisy channel," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 762–764, Nov. 1974.
[4] G. D. Forney, "The Viterbi algorithm," *Proc. IEEE*, vol. 61, pp. 268–278, Mar. 1973.
[5] E. N. Gilbert, "Capacity of burst-noise channels," *Bell Syst. Tech. J.*, vol. 39, pp. 1253–1265, 1960.
[6] 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.
[7] S. Lin and D. J. Costello Jr., *Error Control Coding*. Englewood Cliffs, NJ: Prentice-Hall, 1983.
[8] N. Phamdo, F. Alajaji, and N. Farvardin, "Quantization of memoryless and Gauss–Markov sources over binary Markov channels," submitted to *IEEE Trans. Commun.*, Nov. 1994.
[9] N. Phamdo and N. Farvardin, "Quantization over discrete noisy channels using rate-one convolutional codes," submitted to *IEEE Trans. Inform. Theory*, Aug. 1993.
[10] ——, "Optimal detection of discrete Markov sources over discrete memoryless channels—Applications to combined source-channel coding," *IEEE Trans. Inform. Theory*, vol. 40, pp. 186–193, Jan. 1994.

Simple Universal Lossy Data Compression Schemes Derived from the Lempel–Ziv Algorithm

En-hui Yang and John C. Kieffer, *Fellow, IEEE*

***Abstract*— Two universal lossy data compression schemes, one with fixed rate and the other with fixed distortion, are presented, based on the well-known Lempel–Ziv algorithm. In the case of fixed rate R, our universal lossy data compression scheme works as follows: first pick a codebook B_n consisting of all reproduction sequences of length n whose Lempel–Ziv codeword length is $\leq nR$, and then use B_n to encode the entire source sequence n-block by n-block. This fixed-rate data compression scheme is universal in the sense that for any stationary, ergodic source or for any individual sequence, the sample distortion performance as $n \to \infty$ is given almost surely by the distortion rate function. A similar result is shown in the context of fixed distortion lossy source coding.**

***Index Terms*— Universal lossy data compression, Lempel–Ziv algorithm, stationary sources, individual sequences.**

I. Introduction

Universal source coding theory [6], [13], [18], [28] aims at designing a sequence of codes, whose performance is asymptotically

Manuscript received February 16, 1994; revised April 19, 1995. This work was supported by the National Science Foundation under Grants NCR-9003106 and NCR-9304984.
E Yang is with the Department of Mathematics, Nankai University, Tianjin 300071, P. R. China.
J. C. Kieffer is with the Department of Electrical Engineering, University of Minnesota, Minneapolis, MN 55455 USA.
Publisher Item Identifier S 0018-9448(96)00434-8