

On Decoding Binary Perfect and Quasi-Perfect Codes Over Markov Noise Channels

Haider Al-Lawati and Fady Alajaji, *Senior Member, IEEE*

Abstract—We study the decoding problem when a binary linear perfect or quasi-perfect code is transmitted over a binary channel with additive Markov noise. After examining the properties of the channel block transition distribution, we derive sufficient conditions under which strict maximum-likelihood decoding is equivalent to strict minimum Hamming distance decoding when the code is perfect. Additionally, we show a near equivalence relationship between strict maximum likelihood and strict minimum distance decoding for quasi-perfect codes for a range of channel parameters and the code's minimum distance. As a result, an improved (complete) minimum distance decoder is proposed and simulations illustrating its benefits are provided.

Index Terms—Binary channels with memory, Markov noise, maximum likelihood decoding, minimum Hamming distance decoding, linear block codes, perfect and quasi-perfect codes.

I. INTRODUCTION

CONVENTIONAL communication systems employ coding schemes that are designed for memoryless channels. However, since most real world channels have memory, interleaving is used in an attempt to spread the channel noise in a uniform fashion over the set of received words so that the channel appears memoryless to the decoder. This in fact adds more complexity and delay to the system, while failing to exploit the benefits of the channel memory.

Progress has been achieved on the statistical and information theoretic modeling of channels with memory (e.g., see [2], [7], [11], [12]), as well as on the design of effective iterative decoders for such channels (e.g., see [3], [4], [8], [9]). However, little is known about the structure of optimal maximum likelihood (ML) decoders for such channels. We herein focus on one of the simplest models for a channel with memory, the binary channel with additive Markov noise and analyze the performance of binary perfect and quasi-perfect codes, which can be useful for complexity and delay constrained applications such as wireless sensor networks. Since it is well known that ML decoding of binary codes over the memoryless binary symmetric channel (with bit error rate less than 1/2) is equivalent to minimum Hamming distance decoding, it is natural to investigate whether a relation exists between these two decoding methods for the Markov noise channel. For such

a channel (with memory), one would expect that ML decoding is not equivalent to minimum distance decoding for general codes; however, for certain codes with good (coset) properties (such as perfect and quasi-perfect codes), some equivalency may be established.

Indeed, we provide a partial answer to this problem by showing (after elucidating some properties of the Markov channel distribution) that the strict ML decoding of a binary linear perfect code can be equivalent to its strict minimum distance decoding while the strict ML decoding of a quasi-perfect code can be nearly equivalent to its strict minimum distance decoding. As a result, we propose a (complete) decoder which is an improved version of the minimum distance decoder, and we illustrate its performance via simulation results.

In a related work [5], the optimality of the binary perfect Hamming codes and the near-optimality of subcodes of Hamming codes are demonstrated for the same Markov noise channel.

II. SYSTEM DEFINITION AND PROPERTIES

We consider a binary additive noise channel whose output symbol Y_k at time k is described by $Y_k = X_k \oplus Z_k$, $k = 1, 2, \dots$, where \oplus denotes addition modulo-2, $X_k \in \{0, 1\}$ is the k th input symbol and $Z_k \in \{0, 1\}$ is the k th noise symbol. We assume that the input and noise processes are independent of each other, and that the noise process $\{Z_k\}_{k=1}^{\infty}$ is a stationary (first-order) Markov source with transition probability matrix given by

$$Q = [Q_{ij}] = \begin{bmatrix} \varepsilon + (1-\varepsilon)(1-p) & (1-\varepsilon)p \\ (1-\varepsilon)(1-p) & \varepsilon + (1-\varepsilon)p \end{bmatrix}, \quad (1)$$

where $Q_{ij} \triangleq \Pr(Z_k = j | Z_{k-1} = i)$, $i, j \in \{0, 1\}$. Here $p = \Pr(Z_k = 1)$ is the channel bit error rate (CBER), and $\varepsilon \triangleq \text{Cov}(Z_k, Z_{k-1}) / \text{Var}(Z_k) = [\Pr(Z_k = 1, Z_{k-1} = 1) - p^2] / [p(1-p)]$ is the correlation coefficient of the noise process, where $\text{Cov}(Z_k, Z_{k-1}) \triangleq E[Z_k Z_{k-1}] - E[Z_k]E[Z_{k-1}]$ is the covariance of Z_k and Z_{k-1} and $\text{Var}(Z_k) \triangleq E[Z_k^2] - E[Z_k]^2$ is the variance of Z_k . We assume that $0 < p < 1/2$ and that $0 \leq \varepsilon < 1$, ensuring that the noise process is irreducible. When $\varepsilon = 0$, the noise process becomes independent and identically distributed (i.i.d.) and the resulting channel reduces to the (memoryless) binary symmetric channel with crossover probability or CBER p (which we denote by $\text{BSC}(p)$). Note that this (memory-one) Markov noise channel is a special case of the Gilbert-Elliott channel [7] (realized when the probability for causing an error equals zero in the “good state” and one in the “bad state”).

For $x^n = (x_1, \dots, x_n) \in \{0, 1\}^n$ and $y^n = (y_1, \dots, y_n) \in \{0, 1\}^n$, the channel block transition probability $\Pr(Y^n =$

Paper approved by M. Skoglund, the Editor for Source/Channel Coding of the IEEE Communications Society. Manuscript received August 24, 2007; revised March 31, 2008.

This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada and the Premier's Research Excellence Award (PREA) of Ontario. Parts of this paper were presented at the Tenth Canadian Workshop on Information Theory, Edmonton, Alberta, June 2007.

The authors are with the Department of Mathematics and Statistics, Queen's University, Kingston, Ontario K7L 3N6, Canada (E-mail: {haider, fady}@mst.queensu.ca).

$y^n | X^n = x^n$) can be expressed in terms of the channel noise block distribution as follows

$$\begin{aligned} \Pr(Y^n = y^n | X^n = x^n) &= \Pr(Z^n = z^n) \\ &= L \prod_{k=2}^n [z_{k-1}\varepsilon + (1-\varepsilon)p]^{z_k} \\ &\quad \times [(1-z_{k-1})\varepsilon + (1-\varepsilon)(1-p)]^{1-z_k} \end{aligned}$$

where $z_k = x_k \oplus y_k$, $k = 1, \dots, n$ and $L = \Pr(Z_1 = z_1) = p^{z_1}(1-p)^{1-z_1}$.

Given $z^n = (z_1, \dots, z_n) \in \{0, 1\}^n$, let $t_{ij}(z^n)$ denote the number of times two consecutive bits in z^n are equal to (i, j) , where $i, j \in \{0, 1\}$; i.e.,

$$\begin{aligned} t_{00}(z^n) &= \sum_{k=1}^{n-1} (1-z_k)(1-z_{k+1}), \quad t_{11}(z^n) = \sum_{k=1}^{n-1} z_k z_{k+1}, \\ t_{10}(z^n) &= \sum_{k=1}^{n-1} z_k(1-z_{k+1}), \quad t_{01}(z^n) = \sum_{k=1}^{n-1} (1-z_k)z_{k+1}. \end{aligned}$$

In terms of the $t_{ij}(z^n)$'s $\Pr(Z^n = z^n)$ can be written as

$$\begin{aligned} \Pr(Z^n = z^n) &= L [\varepsilon + (1-\varepsilon)(1-p)]^{t_{00}} [(1-\varepsilon)p]^{t_{01}} \\ &\quad \times [(1-\varepsilon)(1-p)]^{t_{10}} [\varepsilon + (1-\varepsilon)p]^{t_{11}}. \quad (2) \end{aligned}$$

But from the definition of the $t_{ij}(z^n)$'s, we have the following.

$$t_{10}(z^n) = n - 1 - w(z^n) - t_{00}(z^n) + z_1 \quad (3)$$

$$t_{01}(z^n) = w(z^n) - z_1 - t_{11}(z^n), \quad (4)$$

where $w(z^n) = \sum_{k=1}^n z_k$ is the Hamming weight of z^n . Substituting (3) and (4) into (2) yields the following expression for the noise block distribution, which will be instrumental in our analysis:

$$\begin{aligned} \Pr(Z^n = z^n) &= (1-\varepsilon)^{(n-1)} (1-p)^n \left[\frac{p}{1-p} \right]^{w(z^n)} \\ &\quad \times \left[\frac{\varepsilon + (1-\varepsilon)(1-p)}{(1-\varepsilon)(1-p)} \right]^{t_{00}(z^n)} \left[\frac{\varepsilon + (1-\varepsilon)p}{(1-\varepsilon)p} \right]^{t_{11}(z^n)}. \quad (5) \end{aligned}$$

The properties of $t_{00}(z^n)$ and $t_{11}(z^n)$ in terms of only n and $w(z^n)$ are as follows.

- 1) If $w(z^n) = 0$, then $t_{00}(z^n) = n - 1$ and $t_{11}(z^n) = 0$.
- 2) If $0 < w(z^n) = l \leq n - 1$, then $t_{00}(z^n) \leq n - l - 1$ with equality if and only if all the 0's in z^n occur consecutively. Also $t_{11}(z^n) \leq l - 1$ with equality if and only if all the 1's in z^n occur consecutively.
- 3) If $0 < w(z^n) = l \leq \frac{n}{2}$, then $t_{00}(z^n) \geq \max\{n - 2l - 1, 0\}$ and $t_{11}(z^n) \geq 0$.
- 4) If $\frac{n}{2} < w(z^n) = l \leq n - 1$, then $t_{00}(z^n) \geq 0$ and $t_{11}(z^n) \geq 2l - n - 1$.
- 5) If $w(z^n) = n$, then $t_{11}(z^n) = n - 1$ and $t_{00}(z^n) = 0$.

When there is no possibility for confusion, we will write $t_{00}(z^n)$ and $t_{11}(z^n)$ as t_{00} and t_{11} , respectively. We also assume throughout that the blocklength $n \geq 2$.

III. ANALYSIS OF THE NOISE BLOCK DISTRIBUTION

Lemma 1: Let 0^n be the all-zero word (of length n) and let $z^n \neq 0^n$ be any non-zero binary word. Then

$$\Pr(Z^n = z^n) < \Pr(Z^n = 0^n).$$

Proof: Using (2), we have

$$\begin{aligned} \Pr(Z^n = z^n) &= L [\varepsilon + (1-\varepsilon)(1-p)]^{t_{00}} [(1-\varepsilon)p]^{t_{01}} \\ &\quad \times [(1-\varepsilon)(1-p)]^{t_{10}} [\varepsilon + (1-\varepsilon)p]^{t_{11}} \\ &< (1-p) [\varepsilon + (1-\varepsilon)(1-p)]^{t_{00}} \\ &\quad \times [\varepsilon + (1-\varepsilon)(1-p)]^{t_{01}} \\ &\quad \times [\varepsilon + (1-\varepsilon)(1-p)]^{t_{10}} \\ &\quad \times [\varepsilon + (1-\varepsilon)(1-p)]^{t_{11}} \\ &= (1-p) [\varepsilon + (1-\varepsilon)(1-p)]^{t_{00}+t_{01}+t_{10}+t_{11}} \\ &= (1-p) [\varepsilon + (1-\varepsilon)(1-p)]^{n-1} \\ &= \Pr(Z^n = 0^n), \end{aligned}$$

where the strict inequality holds since $L = p < 1-p$ if $z_1 = 1$, and since $p < 1-p$ with $t_{01} > 0$ (since $z^n \neq 0^n$) if $z_1 = 0$. ■

Lemma 2: Let $z_1^n \neq 0^n$ be a non-zero noise word with Hamming weight $w(z_1^n) < n$, $t_{00} = n - w(z_1^n) - 1$ and $t_{11} = w(z_1^n) - 1$ (i.e., z_1^n is of the form $(11 \dots 100 \dots 0)$ or $(00 \dots 011 \dots 1)$). Let z_2^n be another non-zero noise word with $w(z_2^n) = w(z_1^n)$ but with different t_{00} and/or t_{11} . Then, if $\varepsilon > 0$,

$$\Pr(Z^n = z_1^n) > \Pr(Z^n = z_2^n).$$

Proof: From (5), we note that $\Pr(Z^n = z^n)$ strictly increases with both t_{00} and t_{11} when the weight is kept constant and $\varepsilon > 0$. Since z_1^n has maximum values for both t_{00} and t_{11} amongst all noise words of weight $w(z_1^n)$ (but with different t_{00} and/or t_{11}), the strict inequality above follows. ■

Note that when $\varepsilon = 0$, obviously all noise words with the same weight have identical distributions (since the channel reduces to the BSC(p)).

Lemma 3: Suppose that

$$u < u^* \triangleq \frac{\ln \left[\frac{\varepsilon + (1-\varepsilon)(1-p)}{(1-\varepsilon)(1-p)} \right] + \ln \left[\frac{1-p}{p} \right]}{\ln \left[\frac{\varepsilon + (1-\varepsilon)(1-p)}{(1-\varepsilon)(1-p)} \right] + \ln \left[\frac{\varepsilon + (1-\varepsilon)p}{(1-\varepsilon)p} \right]} - 1$$

and

$$0 < \varepsilon < \frac{1-2p}{2(1-p)}.$$

Let z^n be a noise word of weight $w(z^n) = m$ such that $0 \leq m \leq u + 1 \leq \frac{n}{2}$. Then $\Pr(Z^n = z^n) > \Pr(Z^n = \bar{z}^n)$ where \bar{z}^n is any noise word with weight $w(\bar{z}^n) = l > m$.

Proof: First, note that the result directly holds if $m = 0$ by Lemma 1. Now let z^n be a noise word of nonzero weight $m \leq u + 1$, and let \bar{z}^n be another noise word with $w(\bar{z}^n) > m$.

Case 1: Assume that $w(\bar{z}^n) = m + i$ where $i \in \{1, 2, \dots, n - m - 1\}$. Then by (5), we have

$$\begin{aligned} \frac{\Pr(Z^n = \bar{z}^n)}{\Pr(Z^n = z^n)} &\leq \left[\frac{\varepsilon + (1 - \varepsilon)(1 - p)}{(1 - \varepsilon)(1 - p)} \right]^{m-i} \\ &\quad \times \left[\frac{\varepsilon + (1 - \varepsilon)p}{(1 - \varepsilon)p} \right]^{m+i-1} \left(\frac{p}{1 - p} \right)^i \\ &\leq \left[\frac{\varepsilon + (1 - \varepsilon)(1 - p)}{(1 - \varepsilon)(1 - p)} \right]^{m-1} \\ &\quad \times \left[\frac{\varepsilon + (1 - \varepsilon)p}{(1 - \varepsilon)p} \right]^m \left(\frac{p}{1 - p} \right) \triangleq f(m). \end{aligned}$$

The first inequality follows from (5) and by applying the bounds on t_{00} and t_{11} described at the end of the previous section, while the second inequality follows by noting that the right hand side of the first inequality decreases in i for a fixed m . Since $f(m)$ is strictly increasing in m (when $\varepsilon > 0$), and $m \leq u + 1 < u^* + 1$, we obtain that

$$f(m) < f(u^* + 1) = 1 \Rightarrow \frac{\Pr(Z^n = \bar{z}^n)}{\Pr(Z^n = z^n)} < 1.$$

Case 2: Assume that $w(\bar{z}^n) = n$. Let \hat{z}^n be another noise word with $w(\hat{z}^n) = n - 1$, $t_{11}(\hat{z}^n) = n - 2$ and $t_{00}(\hat{z}^n) = 0$. Then

$$\begin{aligned} \frac{\Pr(Z^n = \bar{z}^n)}{\Pr(Z^n = z^n)} &= \frac{\Pr(Z^n = \hat{z}^n) \Pr(Z^n = \bar{z}^n)}{\Pr(Z^n = z^n) \Pr(Z^n = \hat{z}^n)} \\ &< \frac{\Pr(Z^n = \bar{z}^n)}{\Pr(Z^n = \hat{z}^n)} \\ &= \left[\frac{\varepsilon + (1 - \varepsilon)p}{(1 - \varepsilon)p} \right] \left(\frac{p}{1 - p} \right) \\ &= \left[\frac{\varepsilon + (1 - \varepsilon)p}{(1 - \varepsilon)(1 - p)} \right] < 1 \end{aligned}$$

where the first strict inequality holds since $\Pr(Z^n = \hat{z}^n) < \Pr(Z^n = z^n)$ by Case 1, and the last strict inequality holds since $\varepsilon < \frac{1-2p}{2(1-p)}$. ■

Remark: Note that the weight limit u^* in the above lemma is independent of the codeword length.

IV. DECODING PERFECT AND QUASI-PERFECT CODES

We next study the relationship between strict maximum likelihood (SML) decoding and strict minimum (Hamming) distance decoding for binary linear perfect and quasi-perfect codes sent over the additive Markov noise channel. SML decoding is an (incomplete) optimal decoder where optimality is in the sense of minimizing the probability of codeword error (PCE) when the codewords are equally likely (which we herein assume).

Let $\mathcal{F}_2^n = \{0, 1\}^n$ denote the set of all binary words of length n . A non-empty subset \mathcal{C} of \mathcal{F}_2^n is called a binary linear code if it is a subgroup of \mathcal{F}_2^n . The elements of \mathcal{C} are called codewords. We usually describe \mathcal{C} with the triplet (n, M, d) to indicate that n is the blocklength of its codewords, M is its size and d is its minimum Hamming distance; in other words, $d \triangleq \min_{c_1, c_2 \in \mathcal{C}: c_1 \neq c_2} d(c_1, c_2)$ where $d(c_1, c_2) = w(c_1 \oplus c_2)$ is the Hamming distance between c_1 and c_2 and the modulo-2 operation is applied component-wise on c_1 and c_2 .

Definition 1: [10], [6] An (n, M, d) linear code \mathcal{C} is said to be a *perfect code* if, for some non-negative integer t , it has all patterns (i.e., elements of $\{0, 1\}^n$) of Hamming weight t or less and no others as coset leaders.

Definition 2: [10], [6] An (n, M, d) binary linear code \mathcal{C} is said to be *quasi-perfect* if, for some non-negative integer t , it has all patterns of weight t or less, some of weight $t + 1$, and none of greater weight as coset leaders.

An equivalent definition for quasi-perfectness is that, for some non-negative integer t , \mathcal{C} has a packing radius equal to t and a covering radius equal to $t + 1$; i.e., the spheres with (Hamming) radius t around the codewords of \mathcal{C} are disjoint, and the spheres with radius $t + 1$ around the codewords cover $\{0, 1\}^n$. On the other hand, perfectness means that both packing and covering radii are equal. For these two classes of codes, $t = \lfloor \frac{d-1}{2} \rfloor$ (with $d = 2t + 1$ for perfect codes and $d = 2t + 1$ or $d = 2t + 2$ for quasi-perfect codes).

The $(2^m - 1, 2^{2^m - 1 - m}, 3)$ Hamming codes ($m \geq 2$), the $(n, 2, n)$ repetition code with n odd and the $(23, 2^{12}, 7)$ Golay code are the only members of the family of binary perfect linear codes. Examples of quasi-perfect binary linear codes include the $(n, 2, n)$ repetition codes with n even, the $(2^m, 2^{2^m - 1 - m}, 4)$ extended Hamming codes as well as the $(2^m - 2, 2^{2^m - 2 - m}, 3)$ shortened Hamming codes ($m \geq 2$), the $(2^m - 1, 2^{2^m - 1 - 2m}, 5)$ double-error correcting BCH codes ($m \geq 3$), and the $(24, 2^{12}, 8)$ extended Golay code.

Perfect codes as well as quasi-perfect codes are not powerful error-correcting codes due to their small Hamming distances. However, they can be useful in complexity and delay constrained applications where codes with short blocklengths are needed.

Suppose that a codeword of a quasi-perfect code \mathcal{C} is transmitted over the Markov noise channel and that y^n is received at the decoder. The following are possible decoding rules one can use to recover the transmitted codeword.

- *ML Decoding:* y^n is decoded into codeword $c_0 \in \mathcal{C}$ if $\Pr(Y^n = y^n | X^n = c_0) \geq \Pr(Y^n = y^n | X^n = c)$ for all $c \in \mathcal{C}$. If there is more than one codeword for which the above condition holds, then the decoder picks one of such codewords at random.
- *Strict ML (SML) Decoding:* It is identical to the ML rule with the exception of replacing the inequality with a strict inequality; if no codeword c_0 satisfies the strict inequality, the decoder declares a decoding failure.
- *Minimum Distance (MD) Decoding:* y^n is decoded into codeword $c_0 \in \mathcal{C}$ if $w(c_0 \oplus y) \leq w(c \oplus y)$ for all $c \in \mathcal{C}$. If there is more than one codeword for which the above condition holds, then the decoder picks one of such codewords at random.
- *Strict Minimum Distance (SMD) Decoding:* It is identical to the MD rule with the exception of replacing the inequality with a strict inequality; if no codeword c_0 satisfies the strict inequality, the decoder declares a decoding failure.¹

¹Recall that the ML and MD decoders are complete decoders – i.e., they always select a codeword to decode the received word – while the SML and SMD decoders are incomplete decoders as they declare a decoding failure when there are more than one codeword with minimal decoding metric.

Lemma 4: Let \mathcal{C} be an (n, M, d) perfect code to be used over the Markov noise channel. Assume that

$$\left\lfloor \frac{d-1}{2} \right\rfloor < \frac{\ln \left[\frac{\varepsilon+(1-\varepsilon)(1-p)}{(1-\varepsilon)(1-p)} \right] + \ln \left[\frac{1-p}{p} \right]}{\ln \left[\frac{\varepsilon+(1-\varepsilon)(1-p)}{(1-\varepsilon)(1-p)} \right] + \ln \left[\frac{\varepsilon+(1-\varepsilon)p}{(1-\varepsilon)p} \right]}$$

and

$$0 < \varepsilon < \frac{1-2p}{2(1-p)}.$$

Then SMD and SML decoding are equivalent.

Proof: First note that for perfect codes, the element within each coset of minimum weight (i.e., the coset leader) is unique. Also notice that the coset leader is of weight less than or equal to $\lfloor (d-1)/2 \rfloor \leq n/2$. Assume that y^n is received; then $\exists \hat{c} \in \mathcal{C}$ which is unique such that $w(\hat{c} \oplus y^n) < w(c \oplus y^n) \forall c \in \mathcal{C} \setminus \{\hat{c}\}$. Using Lemma 3 with $u = \lfloor (d-1)/2 \rfloor - 1$, we conclude that $\forall c \in \mathcal{C} \setminus \{\hat{c}\}$

$$\Pr(Z^n = \hat{c} \oplus y^n) > \Pr(Z^n = c \oplus y^n)$$

\Leftrightarrow

$$\Pr(Y^n = y^n | X^n = \hat{c}) > \Pr(Y^n = y^n | X^n = c).$$

Hence, given a received word y^n , the codeword with the smallest Hamming distance to y^n will be the most likely codeword that was sent over the channel amongst all the codewords in \mathcal{C} . Therefore, SMD and SML decoding are equivalent. ■

Observations:

- The above lemma also proves that for perfect codes MD and ML decoding are equivalent under the same assumptions on d, ε and p . This is because for such codes SMD and MD are the same due to the uniqueness of their coset leaders which results in no ties in the MD decoder. Similarly, the uniqueness of coset leaders coupled with the proof of the above lemma also imply that SML and ML are equivalent for the perfect codes under the range of channel parameters given in the lemma.
- In a related work [5], Hamada showed that for the Markov channel with a non-negative noise correlation coefficient (i.e., $\varepsilon \geq 0$) and bit error rate $p < 1/2$, the binary perfect Hamming codes (of minimum distance 3) are optimal (under ML decoding) in the sense of minimizing the probability of decoding error amongst all codes having the same blocklength and rate provided that $\varepsilon < (1-2p)/2(1-p)$. Thus, in light of the above lemma, for a communication system employing codes with short blocklength due to delay constraints, Hamming codes used with MD decoding will be optimal over the Markov noise channel amongst all codes of the same blocklength and rate.

Lemma 5: Let \mathcal{C} be an (n, M, d) binary linear quasi-perfect code to be used over the Markov noise channel. Assume that

$$\left\lfloor \frac{d-1}{2} \right\rfloor < \frac{\ln \left[\frac{\varepsilon+(1-\varepsilon)(1-p)}{(1-\varepsilon)(1-p)} \right] + \ln \left[\frac{1-p}{p} \right]}{\ln \left[\frac{\varepsilon+(1-\varepsilon)(1-p)}{(1-\varepsilon)(1-p)} \right] + \ln \left[\frac{\varepsilon+(1-\varepsilon)p}{(1-\varepsilon)p} \right]} - 1$$

and

$$0 < \varepsilon < \frac{1-2p}{2(1-p)}.$$

Then, for a given word y^n received at the channel output, the following hold.

- If $\exists \hat{c} \in \mathcal{C}$ such that $w(\hat{c} \oplus y^n) < w(c \oplus y^n) \forall c \in \mathcal{C} \setminus \{\hat{c}\}$, then $\Pr(Y^n = y^n | X^n = \hat{c}) > \Pr(Y^n = y^n | X^n = c) \forall c \in \mathcal{C} \setminus \{\hat{c}\}$.
- If $\exists \hat{c} \in \mathcal{C}$ such that $\Pr(Y^n = y^n | X^n = \hat{c}) > \Pr(Y^n = y^n | X^n = c) \forall c \in \mathcal{C} \setminus \{\hat{c}\}$, then $w(\hat{c} \oplus y^n) \leq w(c \oplus y^n) \forall c \in \mathcal{C}$.

Proof: (a) Let $\hat{c} \in \mathcal{C}$ such that $w(\hat{c} \oplus y^n) < w(c \oplus y^n) \forall c \in \mathcal{C} \setminus \{\hat{c}\}$. Obviously, $\hat{c} \oplus y^n$ is a coset leader, thus $w(\hat{c} \oplus y^n) \leq \lfloor \frac{d-1}{2} \rfloor + 1 \leq \frac{n}{2}$ since \mathcal{C} is quasi-perfect. By Lemma 3, $\Pr(Z^n = \hat{c} \oplus y^n) > \Pr(Z^n = c \oplus y^n) \forall c \in \mathcal{C} \iff \Pr(Y^n = y^n | X^n = \hat{c}) > \Pr(Y^n = y^n | X^n = c) \forall c \in \mathcal{C} \setminus \{\hat{c}\}$.

(b) Let $\hat{c} \in \mathcal{C}$ such that $\Pr(Y^n = y^n | X^n = \hat{c}) > \Pr(Y^n = y^n | X^n = c) \forall c \in \mathcal{C} \setminus \{\hat{c}\}$. Assume that $\exists \bar{c} \in \mathcal{C} \setminus \{\hat{c}\}$ such that $w(\bar{c} \oplus y^n) < w(\hat{c} \oplus y^n)$; the existence of \bar{c} is always guaranteed by choosing it such that $\bar{c} \oplus y^n$ is the coset leader of $\mathcal{C} \oplus y^n$. Thus, we can assume that $w(\bar{c} \oplus y^n) \leq \frac{n}{2}$ (as \mathcal{C} is quasi-perfect). Then by Lemma 3, $\Pr(Z^n = \hat{c} \oplus y^n) < \Pr(Z^n = \bar{c} \oplus y^n) \iff \Pr(Y^n = y^n | X^n = \hat{c}) < \Pr(Y^n = y^n | X^n = \bar{c})$ which contradicts our assumption that \hat{c} maximizes $\Pr(y^n | c)$ over all codewords. Hence, $w(\hat{c} \oplus y^n) \leq w(c \oplus y^n) \forall c \in \mathcal{C}$. ■

Note the above lemma implies that if a quasi-perfect code has no decoding failures in its SMD decoder, then its SMD and SML decoders are equivalent under the stated conditions on the Markov channel parameters (p, ε) and the code's minimum distance.² Inspired by the above result, Lemma 2 and (5), we next propose the following complete decoder that improves over MD decoding. It includes SMD decoding and exploits the knowledge of t_{00} and t_{11} to resolve ties (which occur when there are more than one codeword that are closest to the received word).

MD+ Decoding: Assume that y^n is received at the channel output. Suppose the decoder outputs the codeword \tilde{c} satisfying the MD decoding condition. If there is more than one such codeword, then the decoder chooses \tilde{c} that maximizes $t_{00}(\tilde{c} \oplus y^n) + t_{11}(\tilde{c} \oplus y^n)$. If there is still a tie, then the decoder chooses \tilde{c} that maximizes $t_{11}(\tilde{c} \oplus y^n)$. Finally, if there is still a tie, then the codeword \tilde{c} is picked at random.³

V. SIMULATION RESULTS AND DISCUSSION

Given an (n, M, d) perfect (respectively, quasi-perfect) code and a fixed CBER p , we let ε_{t-1} (respectively, ε_t) be the largest ε for which Lemma 4 (respectively, Lemma 5) holds, where $t \triangleq \lfloor (d-1)/2 \rfloor$. In Table I, we provide the values of ε_t for $t = 1, 2, 3$ and different values of p .

²In contrast, recall that for the BSC(p) with $p < 1/2$, SML and SMD decoding are equivalent for all binary codes (the same equivalence also holds between ML and MD decoding). Note also that when $\varepsilon \downarrow 0$, the conditions in the above lemma reduce to $\lfloor \frac{d-1}{2} \rfloor < \infty$, and $p < \frac{1}{2}$ (which is consistent with what was just mentioned).

³Clearly, MD+ and MD decoding are equivalent for the BSC, since for this channel, it does not matter what codeword the decoder selects when there is a tie (as long as it is one of the codewords closest to the received word).

p	ε_0	ε_1	ε_2	ε_3
1×10^{-3}	499/999	0.3172	0.02843	0.08801
5×10^{-3}	99/199	0.3152	0.05628	0.02277
1×10^{-2}	49/99	0.3126	0.07297	0.03308
5×10^{-2}	9/19	0.2918	0.11492	0.06644
1×10^{-1}	4/9	0.2645	0.12367	0.07995

TABLE I
VALUES OF ε_t FOR DIFFERENT p AND t . LEMMA 4 HOLDS FOR ALL $\varepsilon \leq \varepsilon_{t-1}$ AND LEMMA 5 HOLDS FOR ALL $\varepsilon \leq \varepsilon_t$.

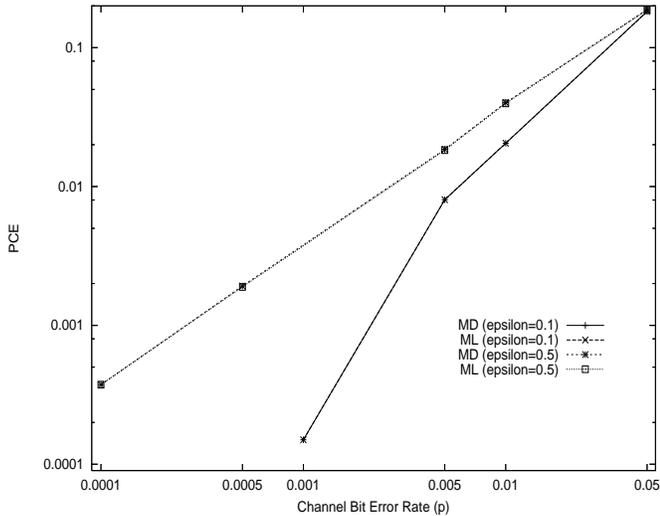


Fig. 1. PCE vs CBER p under different decoding schemes for the Hamming $(15, 2^{11}, 3)$ code over the Markov channel with noise correlation $\varepsilon = 0.1, 0.5$.

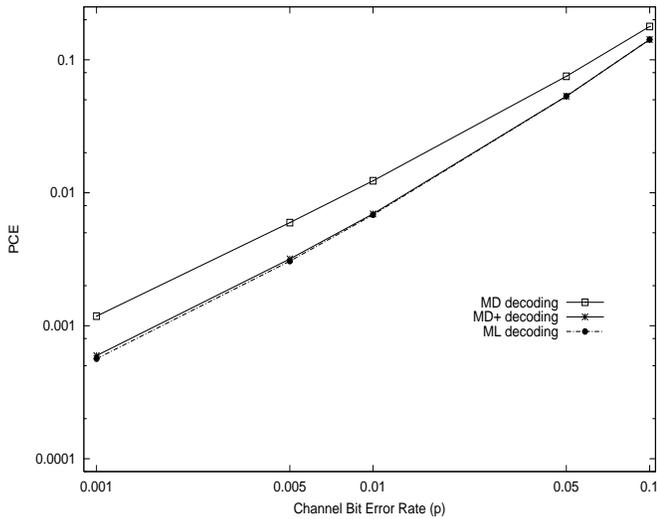


Fig. 2. PCE vs CBER p under different decoding schemes for the Hamming $(8, 2^4, 4)$ code over the Markov channel with noise correlation $\varepsilon = 0.25$.

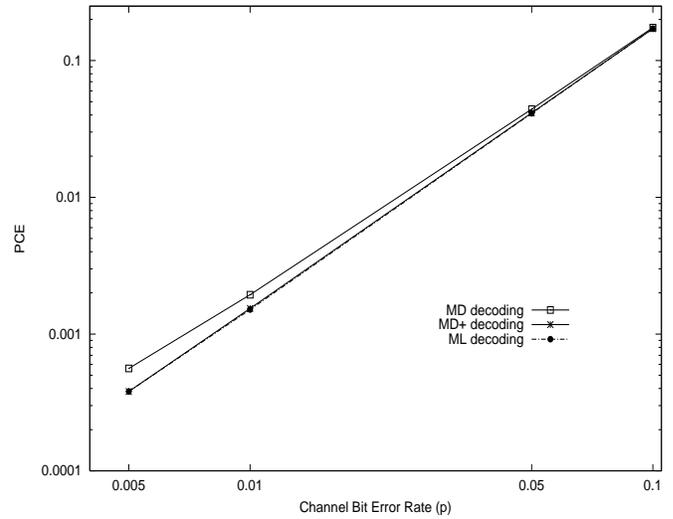


Fig. 3. PCE vs CBER p under different decoding schemes for the BCH $(15, 2^7, 5)$ code over the Markov channel with noise correlation $\varepsilon = 0.05$.

We first examine the perfect $(15, 2^{11}, 3)$ Hamming code under different channel conditions, and show that indeed MD decoding and ML decoding are equivalent for the channel conditions specified by Lemma 4, as illustrated in Table I. A large sequence of a uniformly distributed binary i.i.d. source was generated, encoded via one of these codes and sent over the Markov channel. Typical values are shown for $\varepsilon \in \{0.1, 0.5\}$ in Fig. 1. Note that $\varepsilon = 0.1$ satisfies the conditions of Lemma 4 while $\varepsilon = 0.5$ does not. The simulation results show that MD and ML are identical for the case $\varepsilon = 0.1$ and almost identical at $\varepsilon = 0.5$.

We next present simulation results for decoding two quasi-perfect codes, the binary $(8, 2^4, 4)$ extended Hamming code and the $(15, 2^7, 5)$ BCH code. For the Hamming code, $t = 1$; thus the values for ε_1 in Table I provide the largest values of ε for which Lemma 5 holds for different CBERs p . As a result, we simulated the Hamming system for the 5 values of p listed in Table I and $\varepsilon \in \{0.05, 0.1, 0.2, 0.25\}$. Similarly, since $t = 2$ for the BCH code, the values for ε_2 apply, and the BCH system was simulated for $\varepsilon = 0.05$ and all values of p in Table I except $p = 10^{-3}$. A typical Hamming code simulation result is presented in Fig. 2 for $\varepsilon = 0.25$, and the BCH code simulation is shown in Fig. 3 for $\varepsilon = 0.05$. The results indicate that MD+ performs nearly identically to ML decoding and provides significant gain over MD decoding. By comparing the two figures, we also note that the performance gap between MD and ML decoding decreases with ε (which is consistent with the fact that MD and ML decoding are equivalent when $\varepsilon = 0$). Additional results are available in [1].

As this work is a basic first step towards understanding the structure of ML decoders for channels with memory, there are several directions for future work. For example, note that one limitation of Lemmas 4 and 5 is that their conditions are too stringent to accommodate codes with larger minimum distance, unless if the channel correlation ε is substantially decreased towards 0, thus rendering the Markov channel nearly

memoryless⁴ (e.g, see how ε_t decreases as t increases in Table I). The determination of less stringent conditions is an interesting topic for future work. Another possible future direction is to design a decoder that exploits the memory between blocks by using estimates of the previous noise samples. This can result in an improved performance over the block-by-block ML and MD+ methods (studied here) at a cost of increased complexity. Finally, extending this work to channels with M th order Markovian noise [12] or hidden Markovian noise [7], which are good models for correlated fading channels, may be a worthwhile endeavor.

REFERENCES

- [1] H. Al-Lawati, *Performance Analysis of Linear Block Codes Over the Queue-Based Channel*, M.Sc. Thesis, Mathematics and Engineering, Department of Mathematics and Statistics, Queen's University, Kingston, Ontario K7L 3N6, Canada, Aug. 2007. Available [Online] at <http://www.mast.queensu.ca/~web/publications.html/>
- [2] F. Alajaji and T. Fuja, "A communication channel modeled on contagion," *IEEE Trans. Inform. Theory*, pp. 2035–2041, Nov. 1994.
- [3] A. Ecford, F. Kschischang and S. Pasupathy, "Analysis of low-density parity-check codes for the Gilbert-Elliott channel," *IEEE Trans. Inform. Theory*, vol. 51, pp. 3872–3889, Nov. 2005.
- [4] J. Garcia-Frias, "Decoding of low-density parity-check codes over finite-state binary Markov channels," *IEEE Trans. Commun.*, vol. 52, pp. 1840–1843, Nov. 2004.
- [5] M. Hamada, "Near-optimality of subcodes of Hamming codes on the two-state Markovian additive channel," *IEICE Trans. Fundamentals Electronics, Commun. and Comp. Sciences*, vol. E84-A, pp. 2383–2388, Oct. 2001.
- [6] F. J. MacWilliams and N. J.A. Sloane, *The Theory of Error-Correcting Codes*, North-Holland, Amsterdam, 1977.
- [7] M. Mushkin and I. Bar-David, "Capacity and coding for the Gilbert-Elliott channel," *IEEE Trans. Inform. Theory*, vol. 35, pp. 1277–1290, Nov. 1989.
- [8] V. Nagarajan and O. Milenkovic, "Performance analysis of structured LDPC over the Poly-urn channel with finite memory," in *Proc. CCECE*, May 2004.
- [9] C. Nicola, F. Alajaji, and T. Linder, "Decoding LDPC codes over binary channels with additive Markov noise," in *Proc. CWIT'05*, pp. 187–190, Montreal, June 2005.
- [10] W. W. Peterson and E. J. Weldon Jr., *Error-Correcting Codes*, 2nd Ed., MIT, 1972.
- [11] C. Pimentel and I. F. Blake, "Modeling burst channels using partitioned Fritchman's Markov models," *IEEE Trans. Veh. Technol.*, pp. 885–899, Aug. 1998.
- [12] L. Zhong, F. Alajaji and G. Takahara, "A model for correlated Rician fading channels based on a finite queue," *IEEE Trans. Veh. Technol.*, vol. 57, no. 1, pp. 79–89, Jan. 2008.

⁴Note that as ε decreases towards 0, the channel noise becomes less and less bursty, behaving more and more like a memoryless process.