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

Haider Al-Lawati and Fady Alajaji
Department of Mathematics and Statistics
Queen's University, Kingston, ON K7L 3N6, Canada
Email: {haider,fady}@mast.queensu.ca

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

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 [1], [6], [10], [11]), as well as on the design of effective iterative decoders for such channels (e.g., see [2], [3], [7], [8]). 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. 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. 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 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 [4], 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.

*This work was supported in part by NSERC of Canada.

II. System Definition and Properties

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

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

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 $\varepsilon \triangleq [\Pr(Z_k = 1, Z_{k-1} = 1) - p^2]/[p(1-p)]$ is the correlation coefficient of the noise process. We assume that $0 < p < 1/2$ and that $0 \le \varepsilon < 1$, 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 BSC(p)). Note that this (memory-one) Markov noise channel is a special case of the Gilbert-Elliott channel [6] (realized when the probability for causing an error equals zero in the "good state" and one in the "bad state").

For $x^n = (x_1, \cdots, x_n) \in \{0,1\}^n$ and $y^n = (y_1, \cdots, y_n) \in \{0,1\}^n$, the channel block transition probability $\Pr(Y^n = y^n | X^n = x^n)$ can be expressed in terms of the channel noise block distribution as follows

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

where $z_k = x_k \oplus y_k$, $k = 1, \cdots, n$ and $L = \Pr(Z_1 = z_1) = p^{z_1}(1-p)^{1-z_1}$. Given $\mathrm{z}^n = (z_1, \cdots, z_n) \in \{0,1\}^n$, let $t_{ij}(\mathrm{z}^n)$ denote the number of times two consecutive bits in z^n are equal to (i,j), where $i, j \in \{0,1\}$; more specifically

$$t_{00}(\mathrm{z}^n) = \sum_{k=1}^{n-1} (1-z_k)(1-z_{k+1}), \qquad t_{11}(\mathrm{z}^n) = \sum_{k=1}^{n-1} z_k z_{k+1},$$

$$t_{10}(\mathrm{z}^n) = \sum_{k=1}^{n-1} z_k(1-z_{k+1}), \qquad t_{01}(\mathrm{z}^n) = \sum_{k=1}^{n-1} (1-z_k) z_{k+1}.$$

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

$$\Pr(Z^n = z^n) = L\left[\varepsilon + (1-\varepsilon)(1-p)\right]^{t_{00}} \left[(1-\varepsilon)p\right]^{t_{01}} \cdot \left[(1-\varepsilon)(1-p)\right]^{t_{10}} \left[\varepsilon + (1-\varepsilon)p\right]^{t_{11}}. \quad (2)$$

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.

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

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

- If $w(z^n) = 0$, then $t_{00}(z^n) = n-1$ and $t_{11}(z^n) = 0$.
- If $0 < w(z^n) = l \le n-1$, then
$$t_{00}(z^n) \le n - l - 1$$
with equality iff all the 0's in z^n occur consecutively, and
$$t_{11}(z^n) \le l - 1$$
with equality iff all the 1's in z^n occur consecutively.
- If $0 < w(z^n) = l \le \frac{n}{2}$, then
$$t_{00}(z^n) \ge \max\{n - 2l - 1, 0\}$$
and $t_{11}(z^n) \ge 0$.
- if $\frac{n}{2} < w(z^n) = l \le n-1$, then $t_{00}(z^n) \ge 0$ and
$$t_{11}(z^n) \ge 2l - n - 1.$$
- 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 $n \ge 2$.

III. Analysis of the Noise Block Distribution

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

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

Proof: Using (2), we have

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

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

Lemma 2: Let $z_1^n \ne 0^n$ 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 $(11\cdots100\cdots0)$ or $(00\cdots011\cdots1)$). 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

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

and

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

Let z^n be a noise word of weight $w(z^n) = m$ such that $0 \le m \le t+1 \le \frac{n}{2}$. 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 $m \le t+1$, 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 $i \in \{1, 2, ..., n-m-1\}$. Then by (5), we have

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

Since $f(m)$ is strictly increasing in m (when $\varepsilon > 0$), and $m \le t+1 < t^*+1$, we obtain that

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

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

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

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

■

IV. Decoding of Quasi-Perfect Codes

We next study the relationship between strict maximum likelihood (SML) decoding and strict minimum (Hamming) distance decoding for binary linear quasi-perfect codes sent over the additive Markov noise channel. Strict maximum likelihood (SML) decoding is an optimal (incomplete) decoder in the sense of minimizing the code's frame error rate (FER) – i.e., the codeword error probability – when the codewords are operated on with equal probability (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.[1]

Definition 1: [9], [5] 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 $\mathcal{F}_2^n$. Clearly, for such a code $t = \left\lfloor \frac{d-1}{2} \right\rfloor$ (i.e., $d = 2t+1$ or $d = 2t+2$). 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.

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. [2]

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

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

and

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

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

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

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

(b) Let $\hat{c} \in \mathcal{C}$ such that $\Pr(Y^n = y^n | X^n = \hat{c}) > \Pr(Y^n = y^n | X^n = c)\ \forall\ c \neq \hat{c} \in \mathcal{C}$. Assume that $\exists \bar{c} \neq \hat{c} \in \mathcal{C}$ 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 $w(\bar{c} \oplus y^n) \leq \frac{n}{2}$ since the coset leader has weight less than or equal to $\frac{n}{2}$ (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.[3] In light of the above result and Lemma 2, we next propose the following complete decoder that improves over MD decoding. It includes SMD decoding and exploits

[1] 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.

[2] Recall 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.

[3] In 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 $\left\lfloor \frac{d-1}{2} \right\rfloor < \infty$, and $p < \frac{1}{2}$ (which is consistent with what was just mentioned).

p	ε_1	ε_2	ε_3
1×10^{-3}	0.3172	0.02843	0.08801
5×10^{-3}	0.3152	0.05628	0.02277
1×10^{-2}	0.3126	0.07297	0.03308
5×10^{-2}	0.2918	0.11492	0.06644
1×10^{-1}	0.2645	0.12367	0.07995

TABLE I

VALUES OF ϵ_t FOR DIFFERENT p AND t. LEMMA 4 HOLDS FOR ALL $\varepsilon \leq \epsilon_t$.

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 c_0 satisfying the MD decoding condition. If there is more than one such codeword, then the decoder chooses c_0 that maximizes $t_{00}(c_0 \oplus y^n) + t_{11}(c_0 \oplus y^n)$. If there is still a tie, then the decoder chooses c_0 that maximizes $t_{11}(c_0 \oplus y^n)$. Finally, if there is still a tie, then the codeword c_0 is picked at random.[4]

V. SIMULATION RESULTS

Given an (n, M, d) quasi-perfect code and a fixed CBER p, we let ϵ_t be the largest ε for which both conditions of Lemma 4 hold, 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.

We herein present simulation results for decoding the binary $(8, 2^4, 4)$ extended Hamming code and the $(15, 2^7, 5)$ BCH code over the additive Markov noise channel. A large sequence of a uniformly distributed binary i.i.d. source was generated, encoded via one of these codes and sent over the channel. For the Hamming code, $t = 1$; thus the values for ϵ_1 in Table I provide the largest values of ϵ for which Lemma 4 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 $\epsilon = 0.05$ and all values of p in Table I except $p = 10^{-3}$. A typical Hamming code simulation result is presented in Fig. 1 for $\varepsilon = 0.25$, and the BCH code simulation is shown in Fig. 2 for $\epsilon = 0.05$. 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 $\epsilon = 0$).

Finally, note that one limitation of Lemma 4 is that its conditions are too stringent to accommodate quasi-perfect codes with large minimum distance, unless if the channel correlation ϵ is substantially decreased towards 0, thus rendering the Markov channel nearly memoryless (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.

[4]Clearly, 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).

REFERENCES

[1] F. Alajaji and T. Fuja, "A communication channel modeled on contagion," *IEEE Trans. Inform. Theory*, vol. 40, pp. 2035–2041, Nov. 1994.

[2] 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.

[3] 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.

[4] 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.

[5] F. J. MacWilliams and N. J.A. Sloane, *The Theory of Error-Correcting Codes*, North-Holland, Amsterdam, 1977.

[6] 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.

[7] V. Nagarajan and O. Milenkovic, "Performance analysis of structured LDPC over the Polya-urn channel with finite memory," in *Proc. CCECE*, May 2004.

[8] 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.

[9] W. W. Peterson and E. J. Weldon Jr., *Error-Correcting Codes*, 2nd Ed., MIT, 1972.

[10] C. Pimentel and I. F. Blake, "Modeling burst channels using partitioned Fritchman's Markov models," *IEEE Trans. Veh. Technol.*, pp. 885–899, Aug. 1998.

[11] L. Zhong, F. Alajaji and G. Takahara, "A queue-based model for wireless Rayleigh fading channels with memory," in *Proc. IEEE 62nd Semiannual Veh. Technol. Conf.*, pp. 1362-1366, Dallas, Sept. 2005.

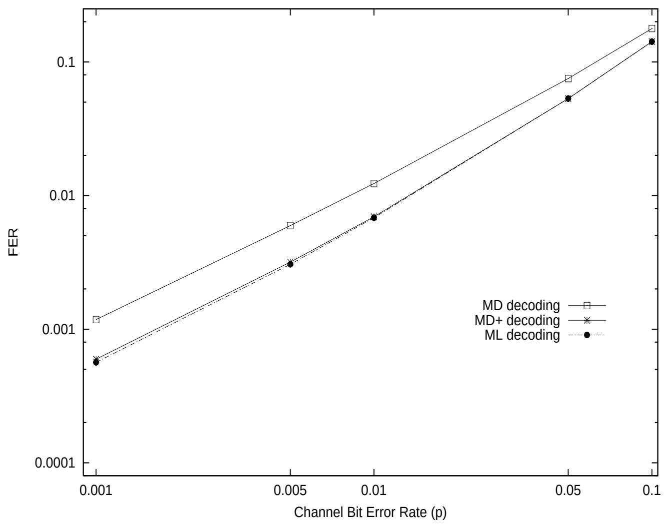

Fig. 1. FER vs CBER p under different decoding schemes for the Hamming $(8, 2^4, 4)$ code over the Markov channel with noise correlation $\epsilon = 0.25$.

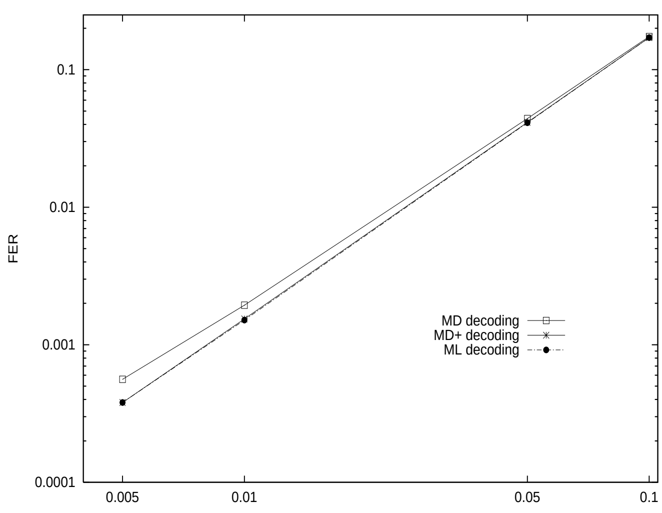

Fig. 2. FER vs CBER p under different decoding schemes for the BCH $(15, 2^7, 5)$ code over the Markov channel with noise correlation $\epsilon = 0.05$.