When are Maximum Likelihood and Minimum Distance Decoding Equivalent for Binary Contagion Channels?

Ghady Azar and Fady Alajaji
Department of Mathematics and Statistics
Queen's University, Kingston, ON, Canada
Email: {ghady, fady}@mast.queensu.ca

Abstract—We study optimal maximum likelihood block decoding of binary codes sent over binary contagion channels with infinite and finite memory. We derive conditions on the codes and channels parameters under which maximum likelihood and minimum Hamming distance are equivalent. We also note that under these conditions, minimum distance decoding can perform better without the use of channel interleaving.

I. INTRODUCTION

Most of the results in coding theory are derived under the assumption that the communication channel is memoryless. The fact that most real life channels have memory has not been efficiently exploited in current communication systems. Instead, interleaving is most commonly used to spread channel error bursts over the set of received codewords so that block decoding can overcome most of the corrupted codewords (if the number of channel errors within a codeword is within the code's error correcting capability). In other words, the use of interleaving makes the channel appear memoryless to the block decoder. This method has immediate shortcomings as it fails to exploit the channel memory while adding delay to the system.

It is well known that the maximum likelihood (ML) decoding of binary codes over the memoryless binary symmetric channel (BSC) with crossover probability $p < \frac{1}{2}$ is equivalent to minimum Hamming distance decoding. When the communication channel has memory, the above equivalence does not necessarily hold. In this paper, we derive conditions on error words and on the channel characteristics, under which the equivalence holds. We deal with finite and infinite memory Polya-contagion channels introduced in [1]. The infinite memory contagion channel is a non-ergodic binary additive channel which can be used in modeling non-ergodic fading channels [2], [3]. It has a closed-form expression for its epsilon-capacity and admits a simple ML decoding rule. Alternatively, the finite memory contagion channel (of order M), which is stationary and ergodic, was shown along with its generalizations based on a finite queue [4]–[6] to accurately model ergodic correlated Rayleigh and Rician fading channels. Furthermore, it has been recently observed that, in the context of LDPC

coding, iterative decoders designed for this channel and its queue-based extension can outperfom the theoretical limit that is achievable on the equivalent BSC (realized via ideal interleaving) [7] (see also [8] and [9] for decoders designed for Gilbert-Elliott and finite-state Markov channels). For the finite-memory channel, we determine sufficient conditions on any binary code under which strict minimum Hamming distance decoding is equivalent to strict ML decoding; while for the infinite-memory channel, we show both necessary and sufficient conditions for which minimum distance and ML decoding are equivalent.

In related works [10], [11], it was proven that strict minimum Hamming distance decoding is equivalent to strict ML decoding for perfect codes of minimum distance 3 over the first-order Markov channel (finite memory contagion channel with $M = 1$) with a positive correlation coefficient. In [12], sufficient conditions, under which strict minimum Hamming distance decoding of binary linear perfect codes becomes equivalent to strict ML decoding, are derived for the same channel. We extend the provided conditions to obtain even tighter sufficient conditions that apply for any binary code (linear or non-linear). We also provide similar results for the finite memory contagion channel with $M = 2$. In an another work [13], a sufficient condition on the infinite memory contagion channel is provided, under which ML block decoding is equivalent to minimum Hamming distance block decoding for linear codes containing the all-one codeword. We also improve these results by obtaining necessary and sufficient conditions on any binary codes over the same channel.

The remainder of the paper is organized as follows. Section II states the general problem and presents the considered channel models with memory. Section III presents the main results. Section IV provides some numerical illustrations. Finally, Section V gives concluding remarks.

II. PROBLEM FORMULATION AND CHANNEL MODELS

A. ML decoding problem

We first present the general problem of ML decoding of binary block codes over modulo-2 additive noise channels with memory. The output Y_n at time instant n of the binary additive

channel is given by $Y_n = X_n \oplus Z_n, n = 0, 1, 2, \cdots$, where $\oplus$ denotes modulo-2 addition and X_n and Z_n are binary-valued input and noise (error) symbols, respectively. The input and noise processes are assumed to be independent from each other. Let $\mathbf{F} := GF(2)$ denote the binary Galois field and define the function $D_n : \mathbf{F}^n \times \mathbf{F}^n \to \mathbb{R}$ as follows:

$$
\begin{aligned}
D_n(\mathbf{x}^n, \mathbf{y}^n) &= -\log_k \frac{\mathrm{P}\left(\mathbf{Y}^n = \mathbf{y}^n | \mathbf{X}^n = \mathbf{x}^n\right)}{\mathrm{P}(\mathbf{Z}^n = \mathbf{0}^n)} \\
&= -\log_k \frac{\mathrm{P}\left(\mathbf{Z}^n = \mathbf{y}^n \oplus \mathbf{x}^n\right)}{\mathrm{P}(\mathbf{Z}^n = \mathbf{0}^n)}
\end{aligned}
$$

where $\mathbf{x}^n = (x_0, x_1, \cdots, x_{n-1})$, $\mathbf{y}^n = (y_0, y_1, \cdots, y_{n-1})$ and modulo-2 addition is performed component-wise. In [14], Hamada proved that for any additive noise channel and for any blocklength n, $(\mathbf{F}^n, D_n)$ forms a metric space. It is natural to associate with this metric a weight function $W_n(\mathbf{z}^n) = D_n(\mathbf{z}^n, \mathbf{0}^n)$, $\mathbf{z}^n \in \mathbf{F}^n$. Let $\mathcal{C}$ be an (n, K, d) code where n is the code's length (i.e., blocklength of its codewords), K is its dimension (number of codewords) and d is its minimum distance. When a codeword in $\mathcal{C}$ is sent over the channel and received as $\mathbf{y}^n$ at the channel output, then it can be directly noted that the ML block decoder selects the following codeword as an estimate of the transmitted one: $\hat{\mathbf{c}}^n = \mathbf{y}^n \oplus \hat{\mathbf{z}}^n$, where $\hat{\mathbf{z}}^n = \arg\min_{\mathbf{z}^n \in \mathbf{F}^n : \mathbf{y}^n \oplus \mathbf{z}^n \in \mathcal{C}} W_n(\mathbf{z}^n)$. It can be shown that, for memoryless channels, the above weight reduces to the Hamming weight. Hence, we can use the new $D_n(\cdot, \cdot)$ metric to generalize basic definitions from coding theory such as a code's minimum distance, its packing and covering radii, and so on. Similarly, we can generalize the concept of perfect and quasi-perfect codes.[1] It was proven in [14] that, for any binary additive noise channel, generalized perfect and quasi-perfect codes are optimal under ML decoding among all codes with the same lengths and dimensions.

In Section III, we establish sufficient conditions (and necessary, where applicable) on binary codes over two classes of Polya contagion channels with memory, under which ML decoding reduces to minimum Hamming distance decoding (for the case of the finite-memory contagion channel, the equivalence holds under strict decoding).[2] Thus, from the optimality of generalized perfect codes [14], we conclude that, under the derived conditions, perfect codes in the classical (Hamming distance) sense are still optimal for the considered channel models, which we briefly describe next.

B. Polya Contagion Channels

In the binary contagion channel model [1], the noise, generated via Polya's contagion urn sampling scheme, propagates in the channel by mimicking the spread of a contagious disease through a population in the sense that the event of a channel bit

---

[1]Recall that a classical perfect code has identical covering and packing radii; while a quasi-perfect code has a covering radius equal to its packing radius plus one.

[2]By strict decoding, we mean that conventional incomplete decoding is used where decoding failure is declared if there are more than one codewords with minimal decoding metric (e.g., see [12]).

error increases the probability of future errors. This channel, which operates under two memory modes (infinite and finite), has some attractive statistical and information-theoretic properties and presents an alternative to the Gilbert-Elliott channel [15] and other finite-state channels with memory (e.g., cf. [16], [17]).

The infinite memory contagion channel (IMCC) is an information unstable channel with stationary non-ergodic binary noise process $\{Z_i\}_{i=0}^{\infty}$ admitting the following closed-form n-fold distribution:

$$
\mathrm{P}(\mathbf{Z}^n = \mathbf{z}^n) = \frac{\Gamma(\frac{1}{\delta})\Gamma(\frac{p}{\delta} + d)\Gamma(\frac{1-p}{\delta} + n - d)}{\Gamma(\frac{p}{\delta})\Gamma(\frac{1-p}{\delta})\Gamma(\frac{1}{\delta} + n)}
$$

where $\Gamma(\cdot)$ is the Gamma function, d is the Hamming weight of the error (or noise) pattern $\mathbf{z}^n$. Here, $\delta > 0$ and $0 \le p < \frac{1}{2}$ are the channel parameters. Specifically, $p = \mathrm{P}(Z_i = 1)$ is the channel bit error rate (BER) and δ is the channel correlation parameter resulting in the following correlation coefficient between any two noise bits at distinct time instants: $\epsilon = \frac{\mathrm{Cov}(Z_i, Z_j)}{\mathrm{Var}(Z_i)} = \frac{\delta}{1 + \delta}$ $\forall i \ne j$ where $\mathrm{Cov}(\cdot, \cdot)$ and $\mathrm{Var}(\cdot)$ denote covariance and variance, respectively. As $\delta \to 0$, the IMCC reduces to a BSC with BER p. It is also proven in [1] that the IMCC belongs to the class of averaged channels with memory, and a closed-form expression of its epsilon-capacity is obtained. However, the fact that the first noise samples have the same effect as "more recent" noise samples on future noise outcomes (due to its infinite memory) makes this model appropriate for only non-ergodic binary channels.

The finite memory contagion channel (FMCC) is a derivative model from the IMCC, introduced to limit the effects of the current noise sample to only a finite number M of future noise samples. Its noise process, generated by a simple modification of Polya's original urn scheme used to produce the IMCC noise process, is a stationary ergodic M'th order Markov source. If $M = 1$, the FMCC reduces to the first-order binary Markov additive noise channel (which is a special case of the Gilbert-Elliott channel). The probability of an error word $\mathbf{z}^n = (z_0, z_1, ..., z_{n-1}) \in \mathbf{F}^n$ generated by a FMCC is identical that of the IMMC noise for $M \ge n$, while for $M < n$, it is given by:

$$
\mathrm{P}(\mathbf{Z}^n = \mathbf{z}^n) = T \underbrace{\prod_{i=0}^{k-1}(p + i\delta) \prod_{j=0}^{M-1-k}(1 - p + j\delta)}_{\text{depends only on the first M samples}}
$$
$$
\times \prod_{i=M}^{n-1}\left[(1 - p + (M - s_{i-1})\delta)^{1-z_i}(p + s_{i-1}\delta)^{z_i}\right]
$$

where T is a constant depending on the channel parameters (BER p, correlation coefficient δ and memory order M), $k = z_0 + z_1 + ... + z_{M-1}$ and $s_{i-1} = z_{i-M} + ... + z_{i-1}$.

We now present, without proof, another useful property of contagion channels.

Proposition 1: For both IMCC and FMCC contagion channels with $p < \frac{1}{2}$, the all-0 error word is the most likely among all possible error words.

III. ML DECODING OVER CONTAGION CHANNELS

In this section, we study the problem of block ML decoding over the above contagion channels. For the FMCC (with $M = 1, 2$), we derive sufficient conditions under which strict ML (SML) and strict minimum distance (SMD) decoding are equivalent. Furthermore, for the IMCC, we provide necessary and sufficient conditions for which ML and MD decoding are equivalent. We first present the results for the FMCC ($M = 1, 2$) and then for the IMCC.

A. Finite Memory Contagion Channel

1) FMCC with $M = 1$: As mentioned earlier, the FMCC reduces to the first-order Markov channel when $M = 1$. We herein tighten the results already derived in [12]. The channel noise transition matrix is given by:

$$\mathbf{Q} = \begin{bmatrix} \epsilon + (1-\epsilon)(1-p) & (1-\epsilon)p \\ (1-\epsilon)(1-p) & \epsilon + (1-\epsilon)p \end{bmatrix}$$

where p is the channel noise BER and $\epsilon = \frac{\text{Cov}(Z_k, Z_{k+1})}{\text{Var}(Z_k)}$. We assume that $0 < \epsilon < 1$.

Proposition 2: Define:

$$m_1 \triangleq \frac{\ln\left(\dfrac{\epsilon + (1-\epsilon)(1-p)}{(1-\epsilon)p}\right)}{\ln\left(\dfrac{\epsilon + (1-\epsilon)p}{(1-\epsilon)p}\right) + \ln\left(\dfrac{\epsilon + (1-\epsilon)(1-p)}{(1-\epsilon)(1-p)}\right)}$$

and

$$m_2 \triangleq \frac{(n-1)\ln\left(\dfrac{\epsilon + (1-\epsilon)(1-p)}{\epsilon + (1-\epsilon)p}\right) + \ln\left(\dfrac{1-p}{p}\right)}{2\ln\left(\dfrac{\epsilon + (1-\epsilon)(1-p)}{(1-\epsilon)(1-p)}\right) + \ln\left(\dfrac{1-p}{p}\right)}.$$

For any two error words $\mathbf{z}^n$ and $\bar{\mathbf{z}}^n$ satisfying:

  i. $w_H(\mathbf{z}^n) = m$, where $0 < m < \frac{n}{2}$
  ii. $w_H(\bar{\mathbf{z}}^n) = m + i$, where $1 \leq i \leq n - m$

where $w_H(\cdot)$ denotes the Hamming weight, we have that $P[\mathbf{Z}^n = \mathbf{z}^n] > P[\mathbf{Z}^n = \bar{\mathbf{z}}^n]$ iff $m < m^* = \min(m_1, m_2)$.

The above proposition states that for given channel parameters p and ϵ and for a given blocklength n, the ML criterion used to compare two different error patterns is equivalent to the minimum Hamming weight criterion when the Hamming weight of at least one of these patterns is less than $m^*(\epsilon, p, n)$. Conversely, if we look only into error patterns with Hamming weights greater than or equal to $m^*(\epsilon, p, n)$, we can find patterns that do not satisfy this equivalence.

Proposition 2, which improves upon the sufficient condition in [12, Lemma 3], is fundamental for the following result.

Theorem 1: Let $\mathcal{C}$ be any (n, K, d) code. Denote by r_{cov} the classical covering radius of this code. If $r_{cov} < \min\left(m^*, \frac{n}{2}\right)$ then SMD and SML decoding are equivalent for this code.

Proof: Let $\mathbf{y}$ be the received word. Let $m \triangleq \min_{\mathbf{c} \in \mathcal{C}} d_H(\mathbf{y}, \mathbf{c})$, where d_H denotes the Hamming distance. Clearly, $m \leq r_{cov} < \frac{n}{2}$ (from the definition of the covering radius). If there exist a unique codeword $\hat{\mathbf{c}}$ such that $d_H(\mathbf{y}, \hat{\mathbf{c}}) = m$, then the SMD decoding gives a valid codeword. Since $m < m^*$, it follows from Proposition 2 that all other error words of larger Hamming weights have a smaller probability than the error word corresponding to the SMD decision. Therefore, SMD and SML decoding are equivalent. ∎

For a given p, ϵ and n, Theorem 1 gives a sufficient condition for binary codes under which SMD decoding is equivalent to SML decoding. Hence it can be shown that under these conditions, traditional perfect codes are also generalized perfect codes over this channel. Therefore, following the discussion in Section II, traditional perfect codes are optimal under the sufficient condition in Theorem 1.
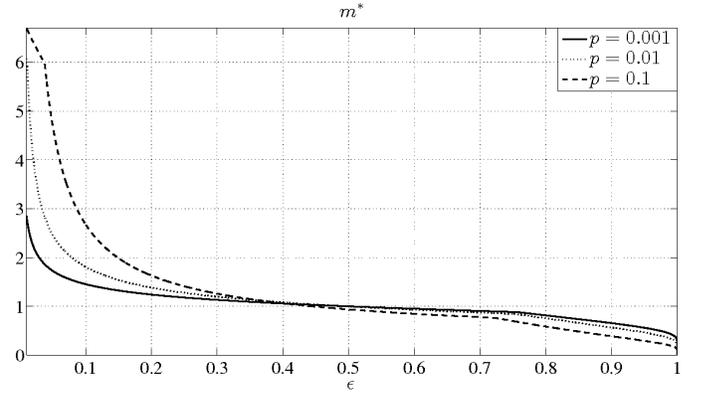

Fig. 1: Plot of m^* with respect to ϵ for different values of p.

In Fig. 1, we plot m^* versus the channel correlation coefficient ϵ for different values of the BER p. We notice from the figure that the condition of Theorem 1 is quite restrictive for channels with $\epsilon > 0.1$. In fact, for these channels, only codes with a covering radius $r_{cov} = 1$ satisfy the condition (e.g. the family of Hamming codes). For smaller ϵ, more codes satisfy the condition, and asymptotically, when $\epsilon \to 0$ (BSC), unsurprisingly all block codes satisfy it.

2) FMCC with $M = 2$: We next present results on the equivalence between SMD and SML decoding for codes sent over the FMCC with $M = 2$ using a similar (albeit more involved) method as with the case of $M = 1$.

Proposition 3: Define:

$$\tilde{m}_1 \triangleq \frac{\ln\left(\dfrac{1-p+2\delta}{p}\right) + \ln\left(\dfrac{p+2\delta}{p+\delta}\right)}{\frac{1}{2}\ln\left(\dfrac{(1-p+2\delta)^4(p+2\delta)^2}{p^2(1-p)(1-p+\delta)^3}\right)}$$

and

$$\tilde{m}_2 \triangleq \frac{2\ln\left(\dfrac{(1-p+2\delta)^{n+2}(1-p)^2}{p^2(p+\delta)(1-p+\delta)^3(p+2\delta)^{n-3}}\right)}{\ln\left(\dfrac{(1-p+2\delta)^6}{p^2(1-p)(1-p+\delta)^3}\right)}$$

where $\delta = \frac{\epsilon}{1-\epsilon}$. For any two error words $\mathbf{z}^n$ and $\bar{\mathbf{z}}^n$ satisfying

  i. $w_H(\mathbf{z}^n) = m$, where $0 < m < \frac{n}{2}$

  ii. $w_H(\bar{\mathbf{z}}^n) = m + i$, where $1 \le i \le n - m$

we have that $P[\mathbf{Z}^n = \mathbf{z}^n] > P[\mathbf{Z}^n = \bar{\mathbf{z}}^n]$ if $m < \tilde{m}^* = \min(\tilde{m}_1, \tilde{m}_2)$.

The above proposition is instrumental in proving the following result.

Theorem 2: Let $\mathcal{C}$ be any (n, K, d) code. Denote by r_{cov} the classical covering radius of this code. If $r_{cov} < \min\left(\tilde{m}^*, \frac{n}{3}\right)$ then SMD and SML decoding are equivalent for this code.
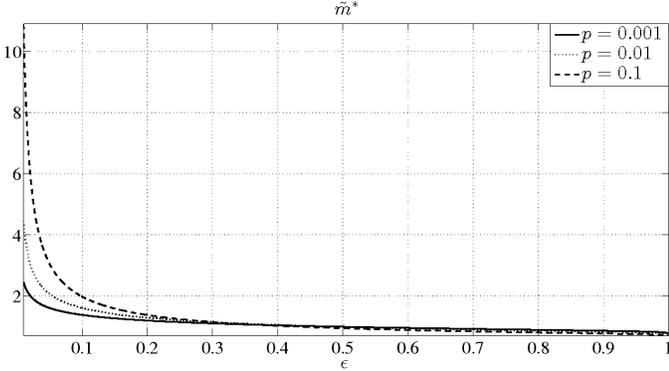
Fig. 2: Plot of $\tilde{m}^*$ with respect to ϵ for different values of p.

The possible values of $\tilde{m}^*$ are plotted in Fig. 2 for different values of p and ϵ.

B. Infinite Memory Contagion Channel

A curious property about the IMCC is that ML decoding always reduces to either minimum distance or maximum distance decoding, depending on the set of possible error words [1]. In fact, let $\mathbf{z_1}^n$ and $\mathbf{z_2}^n$ be two error words and let d_1 and d_2 be their respective Hamming weights. We have the following equivalence:

$$P[\mathbf{Z}^n = \mathbf{z_1}^n] \ge P[\mathbf{Z}^n = \mathbf{z_2}^n] \iff |d_1 - \mu| \ge |d_2 - \mu| \quad (1)$$

where $\mu = \frac{1-2p}{\delta} + \frac{n}{2}$.

From the above equivalence, we can obtain a sufficient condition on any binary code so that ML decoding becomes equivalent to minimum distance decoding. This is presented in the following theorem.

Theorem 3: For any (n, M, d) code $\mathcal{C}$, if the covering radius of this code is $r_{cov} \le \frac{1-2p}{\delta}$, then MD decoding is equivalent to ML decoding.

Proof: Let $\mathbf{y}^n$ be an arbitrary received word in $\mathbf{F}^n$. We denote by $\mathbf{z_{min}}^n(\mathbf{y}^n)$ and $\mathbf{z_{max}}^n(\mathbf{y}^n)$ the error word estimates of the minimum and maximum distance decoders, respectively. We denote by d_{min} and d_{max} their respective Hamming weights. From the definition of the covering radius, we know that $d_{min} \le r_{cov} \le \frac{1-2p}{\delta}$. Hence, $|d_{min} - \mu| \ge \frac{n}{2} - \frac{1-2p}{2\delta}$. On the other hand, $d_{max} \le n$. Hence, $|d_{max} - \mu| \le \frac{n}{2} - \frac{1-2p}{2\delta}$. Therefore, for any received word $\mathbf{y}$, $|d_{min} - \mu| \ge |d_{max} - \mu|$. From (1), we deduce that the minimum distance estimate
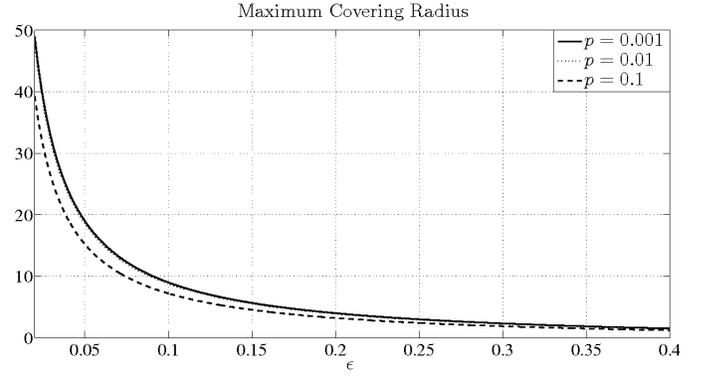
Fig. 3: Plot of the maximum allowable r_{cov} over the IMCC with respect to correlation coefficient ϵ for different values of BER p.

is always more likely than the maximum distance estimate. Hence, ML and MD decoding are equivalent. ∎

We illustrate the condition of Theorem 3 in Fig. 3 by plotting the maximum allowable value for a code's covering radius over the IMCC.

We can tighten the condition in Theorem 3 to obtain a necessary and sufficient condition on the code $\mathcal{C}$. For an arbitrary received word $\mathbf{y} \in \mathbf{F}^n$, let $d_{min}(\mathbf{y}^n) = \min_{\mathbf{c} \in \mathcal{C}} d_H(\mathbf{y}^n, \mathbf{c})$ and $d_{max}(\mathbf{y}^n) = \max_{\mathbf{c} \in \mathcal{C}} d_H(\mathbf{y}^n, \mathbf{c})$ as above. Define $d_{sum}(\mathbf{y}^n) \triangleq |d_{min}(\mathbf{y}^n) - d_{min}(\mathbf{1}^n \oplus \mathbf{y}^n)|$, where $\mathbf{1}^n = (1, \cdots, 1)$ is the all-one word of length n, and let $d_{sum}(\mathcal{C}) \triangleq \max_{\mathbf{y}^n \in \mathbf{F}^n} d_{sum}(\mathbf{y}^n)$. We now can state the necessary and sufficient condition.

Theorem 4: MD decoding is equivalent to ML decoding iff $d_{sum}(\mathcal{C}) \le \frac{1-2p}{\delta}$.

IV. NUMERICAL RESULTS

Let $\mathcal{C}$ be the $(15, 2^{11}, 3)$ perfect binary Hamming code. We simulate the performance of $\mathcal{C}$ over the binary first-order Markov channel and the FMCC with $M = 2$. We choose a correlation coefficient $\epsilon = 0.2$ for the Markov channel and we set $\delta = 0.3$ (i.e., with a noise correlation coefficient of 0.23) for the FMCC with $M = 2$. This choice of parameters ensures that both m^* and $\tilde{m}^*$ are greater than the covering radius of $\mathcal{C}$ for our range of p, as shown in Figs. 1 and 2, respectively. Hence, according to Theorems 1 and 2, SML decoding is equivalent to SMD decoding. The simulated communication system consists of a simple encoder, a channel and a minimum distance decoder. In Fig. 4, we compare its probability of codeword error (PCE) to that of a system that employs an infinite-depth interleaving scheme. In our simulation, we send 20000 codewords over each channel. The figure shows that, for relatively small channel BERs ($p < 0.1$), the performances of both systems are very comparable. Whereas for higher BERs, the system without interleaving is slightly better. We observed a similar behavior over the IMCC. Hence under the conditions of Theorems 1, 2 and 3, interleaving does not only add delay to the communication system, it also does not provide any

significant gain in performance and even in some cases it slightly reduces the performance of the system. Therefore, under these conditions, the use of interleaving is not justified.

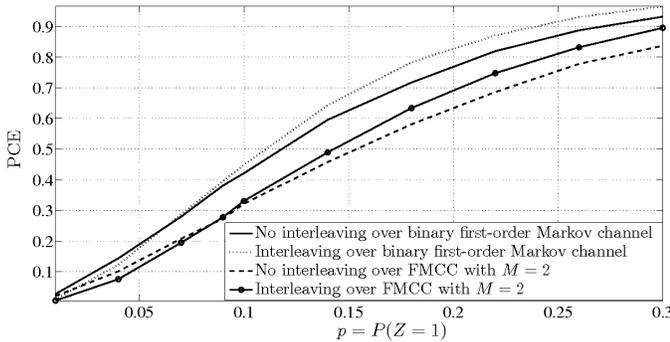We also simulate the performance of the $(7, 2^4, 3)$ perfect

Fig. 4: Performance of the $(15, 2^{11}, 3)$ Hamming code with and without interleaving over the first-order Markov channel and over the FMCC with $M = 2$.

Hamming code $\mathcal{C}'$ over the binary first-order Markov channel with $\epsilon = 0.45$, the FMCC with $M = 2$ and $\delta = 0.7$ and the IMCC with $\delta = 0.7$. The results are presented in Table I. The

p	0.001	0.08	0.1	0.2	0.3
m^*	1.0272	1.0079	0.9930	0.8966	0.7759
SMD over FMCC with $M = 1$	0.0018	0.14834	0.18688	0.37784	0.56074
SML over FMCC with $M = 1$	0.0018	0.14834	0.18666	0.37082	0.54466
$\tilde{m}^*$	1.0335	1.0227	1.0114	0.9410	0.8563
SMD over FMCC with $M = 2$	0.00066	0.0738	0.097	0.21654	0.36046
SML over FMCC with $M = 2$	0.00066	0.0738	0.097	0.21654	0.34882
MD over IMCC	0.00121	0.15091	0.16474	0.33869	0.48022
ML over IMCC	0.00121	0.15091	0.16474	0.33869	0.48022

TABLE I: Verifying the conditions of Theorems 1, 2 and 4 for the $(7, 2^4, 3)$ Hamming code over the FMCC and IMCC.

covering radius of the perfect Hamming code is 1. According to Theorem 1, if $m^* > r_{cov} = 1$ we have equivalence between SMD and SML decoding. Indeed, as the table shows, when $m^* > 1$ the probabilities of codeword error for SMD and SML decoding over the first-order binary Markov channel are identical. We start to see some discrepancy when $m^* < 1$. Similarly, according to Theorem 2, if $\tilde{m}^* > r_{cov} = 1$ then SMD and SML decoding are equivalent. This is again validated by simulation as shown in the table. We also notice that for $p = 0.2$, $\tilde{m}^* < 1$. However, the probabilities of codeword error are still identical. That is not surprising since the condition in Theorem 2 is only sufficient. Finally, since $\mathcal{C}'$ has the all-one codeword, we can show that $d_{sum}(\mathcal{C}') = 0 < \frac{1-2p}{\delta}$ as $p < \frac{1}{2}$. Hence, according to Theorem 4, MD and ML equivalent are always equivalent for $\mathcal{C}'$ as illustrated in the table.

V. CONCLUSION

In this work, we presented sufficient conditions on general binary codes under which SMD and SML decoding are equivalent over the binary first-order Markov channel and the FMCC

with $M = 2$. For the IMCC, we provided both necessary and sufficient conditions for which MD and ML are equivalent. We have also noted by simulation that, under these conditions, the use of an infinite-depth interleaver does not necessarily offer performance gains.

Future work may include extending the results to other channel models with memory (e.g., the FMCC with $M > 2$ and the Gilbert-Elliott channel). Another interesting direction is to study optimal structures of binary block codes over channels with memory as we have established that the Hamming distance is not necessarily the most important parameter in the code design.

REFERENCES

[1] F. Alajaji and T. Fuja, "A communication channel modeled on contagion," *IEEE Trans. Inform. Theory*, vol. 40, no. 6, pp. 2035-2041, Nov. 1994.

[2] M. Effros, A. Goldsmith and Y. Liang, "Generalizing capacity: New definitions and capacity theorems for composite channels," *IEEE Trans. Inform. Theory*, vol. 56, no. 7, pp. 3069-3087, July 2010.

[3] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Euro. Trans. Telecomm. (ETT)*, vol. 10, no. 6, pp. 585-596, Nov. 1999.

[4] L. Zhong, F. Alajaji and G. Takahara, "A binary communication channel with memory based on a finite queue," *IEEE Trans. Inform. Theory,* vol. 53, pp. 2815-2840, Aug. 2007.

[5] L. Zhong, F. Alajaji, and G. Takahara, "A model for correlated Rician fading channels based on a finite queue," *IEEE Trans. Veh. Technology*, vol. 57, no. 1, pp. 79-89, Jan. 2008.

[6] C. Pimentel, F. Alajaji and P. Melo, "A discrete queue-based model for capturing memory and soft-decision information in correlated fading channels," *IEEE Trans. Commun.*, vol. 60, no. 6, pp. 1702-1711, June 2012..

[7] 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.

[8] A. Eckford, 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.

[9] J. Garcia-Frias, "Decoding of low-density parity check codes over finite-state binary Markov channels," in *IEEE trans. Commun.,*, vol. 52, pp. 1840–1843, Nov. 2004.

[10] M. Hamada, "Near-optimal codes on the two-state Markovian additive channel," in *Proc. IEEE Int. Symp. Inform. Theory*, 2001, p. 246.

[11] 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.

[12] H. Al-Lawati and F. Alajaji, "On decoding binary perfect and quasi-perfect codes over Markov noise channels," *IEEE Trans. Commun.*, vol. 57, no. 4, pp. 873-878, Apr. 2009.

[13] A. Cohen, F. Alajaji, N. Kashyap, and G. Takahara, "LP decoding for joint source-channel codes and for the non-ergodic Polya channel," *IEEE Commun. Letters, IEEE*, vol. 12, no. 9, pp. 678-680, Sep. 2008.

[14] M. Hamada, "A sufficient condition for a code to achieve the minimum decoding error probability–generalization of perfect and quasi-perfect codes," *IEICE Trans. Fundamentals Electronics, Commun. and Comp. Sciences*, vol. E83-A, no. 10, pp. 1870-1877, Oct. 2000.

[15] 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.

[16] R. G. Gallager, *Information Theory and Reliable Communication*, New York: Wiley, 1968.

[17] P. Sadeghi and P. Rapajic, "Capacity analysis for finite-state Markov mapping of Markov channels," *IEEE Trans. Commun.*, vol. 53, pp. 833-840, May 2005.