

On the Equivalence Between Maximum Likelihood and Minimum Distance Decoding for Binary Contagion and Queue-Based Channels with Memory

Ghady Azar and Fady Alajaji, *Senior Member, IEEE*

Abstract

We study the optimal maximum likelihood (ML) block decoding of general binary codes sent over two classes of binary additive noise channels with memory. Specifically, we consider the infinite and finite memory Polya contagion and queue-based channel models which were recently shown to approximate well binary modulated correlated fading channels used with hard-decision demodulation. We establish conditions on the codes and channels parameters under which ML and minimum Hamming distance decoding are equivalent. We also present results on the optimality of classical perfect and quasi-perfect codes when used over the channels under ML decoding. Finally, we briefly apply these results to the dual problem of syndrome source coding with and without side information.

Index Terms—Binary channels with finite and infinite memory, Markov noise, ML and minimum distance decoding, block codes, source-channel coding duality, syndrome source coding.

This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada. Parts of this work were presented at the Thirteenth Canadian Workshop on Information Theory, Toronto, June 2013.

The authors are with the Department of Mathematics and Statistics, Queen's University, Kingston, ON, Canada (email: {ghady, fady}@masc.queensu.ca).

I. INTRODUCTION

The fundamental results in coding theory are primarily derived under the assumption that communication channels are memoryless in the sense that their noise is an independent and identically distributed process (e.g., see [1]). In fact, for such channels and in particular the memoryless binary symmetric channel (BSC), the development of high-performing codes with rich algebraic structures has reached an advanced stage. However, these codes are not necessarily good for other channel models. Moreover, most real life channels have statistical memory [2], [3] and cannot be adequately represented via memoryless channel models. As a result, current systems commonly employ interleaving to spread channel error bursts over the set of received codewords so that block decoding can recover most of the corrupted codewords (if the number of channel errors within a codeword is within the code's error correction capability) [1], [4], [5]. 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 (since channels with memory have larger capacity than their memoryless counterparts in a wide range of cases [6], [7]) while adding delay to the system (which is substantial when long error bursts occur in the channel) [5].

It is therefore of interest to understand the structures of ML decoders and the properties of optimal block codes for channels with memory. Since the ML decoding of binary codes over the BSC with crossover probability less than $1/2$ is equivalent to minimum Hamming distance decoding, it is natural to investigate whether a similar relation exists when the channel has memory. Indeed when such equivalence holds, any optimal or good code for the BSC will work well on the channel with memory. In this work, we derive conditions on arbitrary binary codes and on the channel characteristics, under which the equivalence holds. The channel models considered are the binary infinite and finite memory Polya-contagion channels [8] and the queue-based channel (QBC) [7]. These additive noise models, which subsume the BSC as a special cases, feature desirable properties for tractable system analysis.

The infinite memory contagion channel (IMCC) is non-ergodic, has a closed-form expression

for its epsilon-capacity¹ and admits a simple ML decoding rule. The finite memory contagion channel (FMCC) and the QBC (which generalizes the FMCC) both feature ergodic M^{th} -order Markov noise processes and have a single-letter capacity expression. They were shown to accurately model (in terms of replicating channel capacity and noise autocorrelation function) ergodic discrete fading channels composed of a binary modulator, a time-correlated flat Rayleigh or Rician fading channel and a hard-decision demodulator [7], [11]. Furthermore, it is demonstrated in [12] that the QBC with small values of M approximates well in terms of codeword error probability the packet error process of the above discrete fading channels when block-coded via Reed-Solomon codes.

The use of these binary channel models is relevant to the following scenarios. In a wireless sensor network where hard-decision demodulation is used at the sensors due to their limited computational and storage resources, the end-to-end binary system between any pair of sensors can be better represented by the above models than the BSC by virtue of the time-correlated nature of the fading links. Also, digital storage devices where errors occur in clusters can be well approximated by these models; indeed the IMCC and FMCC were originally studied in light of the observation that failures in semi-conductor chips are better characterized by a Polya “contagion” distribution [8]. Furthermore, the noise processes of the FMCC and QBC are viable alternatives (in light of their explicit Markovian structure and low number of parameters) to the typically used Gilbert-Elliott channel (GEC) [13] error process and other hidden Markov processes (e.g., see [14]) for capturing the packet loss behavior in data networks, since in practical networks, packet losses are often time-dependent due to a variety of causes including congestion and buffer overflow.

The contributions of this work include the following. For general binary codes of block length n sent over the IMCC or the QBC with memory $M \geq n$, we establish both necessary and sufficient conditions for which minimum distance (MD) and ML decoding are equivalent. We

¹For a fixed epsilon between 0 and 1, the epsilon-capacity of a channel is the largest asymptotic rate at which the channel can be encoded via a sequence of block codes with error probability less than epsilon [9], [10]. Channel capacity is then the limit of epsilon-capacity as epsilon tends to 0.

also give sufficient conditions under which classical perfect and quasi-perfect codes are optimal (under ML decoding). For the QBC with $M < n$, we restrict our study to two cases: $M = 1$ (or equivalently the binary first-order Markov noise channel or the FMCC with $M = 1$) and $M = 2$. In both cases, we determine sufficient conditions for the equivalence of strict MD (SMD) and ML decoding. We also present sufficient conditions for the optimality of classical perfect codes.

In a previous related work [15], it is proven that SMD and strict ML (SML) decoding are equivalent for perfect codes of minimum distance 3 over the first-order Markov noise channel (i.e., the FMCC or QBC with $M = 1$). In [16], sufficient conditions, under which SMD decoding of binary linear perfect codes becomes equivalent to SML decoding, are derived for the same channel. A near equivalence relationship between SMD and SML decoding is also obtained for binary linear quasi-perfect codes for a range of channel parameters and the codes' minimum distance. In this work, we generalize the equivalence conditions of [15], [16] to produce even tighter sufficient conditions that apply for any binary code (linear or non-linear). We also provide similar results for the QBC with $M = 2$. In [17], a sufficient condition on the IMCC is given on the MD-ML decoding equivalence for linear codes containing the all-one codeword; this result is herein improved via both necessary and sufficient conditions for general codes.

The rest of this paper is organized as follows. In Section II, we describe the channel models. In Section III, we introduce a generalized likelihood distance for additive noise channels with memory. We then generalize the basic notions in coding theory to this new distance and present two lemmas on the optimality of classical perfect and quasi-perfect codes over channels with memory. In Section IV, the MD/SMD-ML decoding equivalence results are established for the considered channel models. In Section V, we briefly discuss the numerical verification of the derived results and translate them for practical codes with short block lengths and covering radii. We also highlight how our results apply in the contexts of syndrome source coding with and without side information at the receiver. Finally, we conclude the paper in Section VI.

II. CHANNEL MODELS

We consider binary additive noise channels with memory. We use the notation $\mathbf{F}_2 = GF(2)$ throughout. The output Y_n at time instant n of the channel is given by $Y_n = X_n \oplus Z_n$, $n = 1, 2, \dots$, 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. We study several channel models for representing the noise process with memory.

A. Infinite Memory Contagion Channel

The IMCC is a communication channel with stationary non-ergodic additive noise characterized by two parameters. It is based on the contagion urn scheme of George Polya, a two-color ball sampling scheme empirically shown to describe well defects in semiconductor chips [8]. The channel noise propagates and corrupts the transmitted signal by mimicking the spread of an infectious disease through a population, in the sense that the occurrence of an error (or "infection", if we use the contagion interpretation) increases the probability of future errors, and hence may lead to a clustering or burst of errors (i.e., an "epidemic" in the population). For the IMCC, the probability of an n -bit error pattern $\mathbf{z}_1^n = (z_1, z_2, \dots, z_n)$ can be written as follows [8]:

$$P(\mathbf{Z}_1^n = \mathbf{z}_1^n) = \frac{\Gamma(\frac{1}{\delta})\Gamma(\frac{p}{\delta} + d)\Gamma(\frac{1-p}{\delta} + n - d)}{\Gamma(\frac{p}{\delta})\Gamma(\frac{1-p}{\delta})\Gamma(\frac{1}{\delta} + n)} \quad (1)$$

where $p = P(Z_i = 1)$ is the channel bit error rate (BER), $\delta > 0$ is a noise correlation parameter, d is the Hamming weight of the error pattern, and $\Gamma(\cdot)$ is the Gamma function. We can see that the probability distribution of an error pattern depends only on its Hamming weight, and does not depend on how the errors are clustered. The correlation $\text{Cor}(\cdot, \cdot)$ between any two distinct noise bits is given by

$$\epsilon \triangleq \text{Cor}(Z_i, Z_j) = \frac{\mathbb{E}[Z_i Z_j] - \mathbb{E}[Z_i]\mathbb{E}[Z_j]}{\text{Var}(Z_i)} = \frac{\delta}{1 + \delta} \quad \forall i \neq j$$

where $\text{Var}(Z_i) = \mathbb{E}[Z_i^2] - \mathbb{E}[Z_i]^2$ is the variance of Z_i . It is shown in [8] that, for this channel, ML decoding reduces to either minimum Hamming distance decoding or maximum Hamming distance decoding. It is also proven in [8] that the all-zero error word is the most likely among

all error words of length n generated by the IMCC. When we set $\delta = 0$, the IMCC reduces to a BSC with crossover probability p (BSC(p)). It is also shown that the IMCC (for $\delta > 0$) belongs to the class of (non-ergodic) averaged channels with memory and admits a closed-form expression for its epsilon-capacity.

B. Finite-Memory Contagion Channel

The FMCC with memory M is a channel model derived from the IMCC (it has one more parameter than the IMCC, which is its memory order M). It differs from the IMCC by making the current noise sample independent of “older” samples given the last M error bits. It is shown in [8, Section VI] that the generated noise process is stationary, ergodic and M^{th} -order Markovian yielding a positive channel capacity that increases with the memory M . The FMCC and its queue-based generalization, which will be described next, were also shown in [11] to model ergodic Rician fading channels more accurately than the GEC [13].

Note that for a block length $n \leq M$, the FMCC becomes analytically equivalent to the IMCC and the probability of an n -bit error word \mathbf{z}_1^n is given by (1). If $n > M$, the word error probability is given by

$$P(\mathbf{Z}_1^n = \mathbf{z}_1^n) = L^{(M)} \prod_{i=M+1}^n \left[\frac{p + s_{i-1}\delta}{1 + M\delta} \right]^{z_i} \left[\frac{1 - p + (M - s_{i-1})\delta}{1 + M\delta} \right]^{1-z_i} \quad (2)$$

where

$$L^{(M)} = \frac{\prod_{i=0}^{s_M-1} (p + i\delta) \prod_{j=0}^{M-s_M-1} (1 - p + j\delta)}{\prod_{l=1}^{M-1} (1 + l\delta)}$$

and for $k \geq M$,

$$s_k = \sum_{i=k-M+1}^k z_i.$$

The correlation coefficient of the noise process is given by

$$\epsilon = \frac{\mathbb{E}[Z_i Z_j] - \mathbb{E}[Z_i]\mathbb{E}[Z_j]}{\text{Var}(Z_i)} = \frac{\delta}{\delta + 1} \geq 0 \quad (3)$$

for any $i, j = 1, 2, \dots, M$ and for any $j = i + 1$ with $i \geq M$. As the IMCC, when $\delta = 0$ (or equivalently $\epsilon = 0$), the FMCC reduces to the BSC(p).

C. Queue-Based Channel

Introduced in [7] via a ball sampling mechanism involving a finite queue, the QBC also features binary additive stationary ergodic M^{th} -order Markov noise. It is a generalization of the FMCC² (having one additional parameter) while remaining mathematically tractable. Specifically, it has four parameters: memory order M , bias parameter α , noise correlation coefficient ϵ and BER p (see [18, Section 2.2.3]). For this channel, the probability of an n -bit error pattern \mathbf{z}_1^n is given by:

- If $n \leq M$, then

$$P(\mathbf{Z}_1^n = \mathbf{z}_1^n) = \frac{\prod_{j=0}^{n-d_1^n-1} \left[(1-p) + j \frac{\epsilon}{1-\epsilon} \right] \prod_{j=0}^{d_1^n-1} \left(p + j \frac{\epsilon}{1-\epsilon} \right)}{\prod_{j=0}^{n-1} \left(1 + j \frac{\epsilon}{1-\epsilon} \right)}, \quad (4)$$

where $d_a^b = \sum_{i=a}^b z_i$ ($d_a^b = 0$ if $a > b$).

- If $n > M$, then

$$P(\mathbf{Z}_1^n = \mathbf{z}_1^n) = \frac{L^{(M)}}{\left[1 + (M-1 + \alpha) \frac{\epsilon}{1-\epsilon} \right]^{n-M}} \prod_{i=M+1}^n \left[\frac{(d_{i-M+1}^{i-1} + \alpha z_{i-M}) \epsilon}{1-\epsilon} + p \right]^{z_i} \left\{ \frac{[M-1 - d_{i-M+1}^{i-1} + \alpha(1 - z_{i-M})] \epsilon}{1-\epsilon} + 1 - p \right\}^{1-z_i} \quad (5)$$

where

$$L^{(M)} = \frac{\prod_{j=0}^{M-d_1^M-1} \left[(1-p) + j \frac{\epsilon}{1-\epsilon} \right] \prod_{j=0}^{d_1^M-1} \left(p + j \frac{\epsilon}{1-\epsilon} \right)}{\prod_{j=0}^{M-1} \left(1 + j \frac{\epsilon}{1-\epsilon} \right)}.$$

Lemma 1: The all-zero error word $\mathbf{0}^n$ is the most likely among all error words of length n generated by the QBC.

Proof: See [18, Lemma 2.2]. ■

²Note that both the FMCC and the QBC are finite-state Markov channels.

Remark 1: (Summary of the QBC features)

- If $M = 1$, the QBC reduces to the FMCC with $M = 1$ (or equivalently, the binary first-order Markov additive noise channel³) with identical BER p and noise correlation coefficient ϵ .
- If $\epsilon = 0$, the QBC reduces to the BSC(p).
- If $\alpha = 1$, the QBC reduces to the FMCC with the same memory parameter M , the same BER p , and with the same correlation coefficient ϵ , or with $\delta = \frac{\epsilon}{1-\epsilon}$.
- If $\alpha = 0$, the QBC with parameters (memory M , BER p and correlation coefficient ϵ) reduces to the FMCC with parameters $(M - 1, p, \epsilon)$.

III. GENERALIZED CONCEPTS OF CODING THEORY

Consider a general binary additive noise communication channel with a block probability transition matrix $P_{\mathbf{Y}|\mathbf{X}}(\cdot|\cdot)$ where $\mathbf{X} = (X_1, X_2, \dots, X_n)$ and $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$ denote the channel's input and output vectors of lengths n , respectively. Let D_n be the following generalized distance:

$$\begin{aligned} D_n : \mathbf{F}_2^n \times \mathbf{F}_2^n &\mapsto \mathbb{R} \\ D_n(\mathbf{x}, \mathbf{y}) &= -\log_k \frac{P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})}{P(\mathbf{Z} = \mathbf{0}^n)} \\ &= -\log_k \frac{P(\mathbf{Z} = \mathbf{y} \oplus \mathbf{x})}{P(\mathbf{Z} = \mathbf{0}^n)} \end{aligned}$$

where $k > 1$ is a constant, the addition \oplus is component-wise and $\mathbf{Z} = (Z_1, Z_2, \dots, Z_n)$ is the channel's noise vector of length n . We denote by $\mathcal{K}(D_n)$ the domain of this distance, i.e.

$$\mathcal{K}(D_n) = \{t \in \mathbb{R} : \exists \mathbf{x}, \mathbf{y} \in \mathbf{F}_2^n, D_n(\mathbf{x}, \mathbf{y}) = t\}.$$

It is natural to associate with the distance D_n a weight function W_n defined as follows:

$$\begin{aligned} W_n : \mathbf{F}_2^n &\mapsto \mathbb{R} \\ W_n(\mathbf{e}) &= -\log_k \frac{P(\mathbf{Z} = \mathbf{e})}{P(\mathbf{Z} = \mathbf{0}^n)}. \end{aligned}$$

³The first-order Markov additive noise channel is simply the Gilbert channel, a special case of the GEC [13] realized by setting the probability for causing an error to zero in the "good state" and to one in the "bad state."

Let $\mathcal{C} \subset \mathbf{F}_2^n$ be an (n, M) binary block code with M codewords, each of length n . If the code has a minimum Hamming distance d_{min} , we use the notation (n, M, d_{min}) to describe the code. If a codeword from \mathcal{C} is sent over the channel and is received as \mathbf{y} at the channel output, one can use one of the following familiar decoding rules to recover the transmitted codeword.

- *ML Decoding*: \mathbf{y} is decoded into codeword $\mathbf{c}_0 \in \mathcal{C}$ if $P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{c}_0) \geq P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{c})$ for all $\mathbf{c} \in \mathcal{C}$.
- *MD Decoding*: \mathbf{y} is decoded into codeword $\mathbf{c}_0 \in \mathcal{C}$ if $w_H(\mathbf{c}_0 \oplus \mathbf{y}) \leq w_H(\mathbf{c} \oplus \mathbf{y})$ for all $\mathbf{c} \in \mathcal{C}$, where $w_H(\cdot)$ denotes the Hamming weight.
- *SMD Decoding*: It is identical to the MD rule with the exception of requiring that the inequality holds strictly for all $\mathbf{c} \neq \mathbf{c}_0$; if no codeword \mathbf{c}_0 satisfies the strict inequality, the decoder declares a decoding failure.

Recall that the ML and MD decoders are complete decoders (i.e., they always select a codeword to decode the received word) while the SMD decoder (like bounded distance decoders) is an incomplete decoder as it declares a decoding failure when there is no unique codeword of minimal decoding metric [1]. The later decoder is useful for situations where the decoder can seek retransmission via a feedback link and where ties in the decoding metric are best left unresolved (unlike complete decoders which typically break ties at random) so that undetected errors are reduced and the decoder's reliability in correcting errors is increased.

Definition 1: The generalized minimum distance of an (n, M) code \mathcal{C} is:

$$\rho_{min} \triangleq \min\{D_n(\mathbf{c}, \mathbf{c}') : \mathbf{c}, \mathbf{c}' \in \mathcal{C} \text{ and } \mathbf{c} \neq \mathbf{c}'\}.$$

Definition 2: The generalized packing radius ρ_{pac} of an (n, M) code \mathcal{C} with generalized minimum distance ρ_{min} is:

$$\rho_{pac} \triangleq \max\{t \in \mathcal{K}(D_n) : \forall \mathbf{y} \in \mathbf{F}_2^n, \exists \text{ at most one error word } \mathbf{z} \in \mathcal{S}_{\mathcal{C}}(\mathbf{y}) \text{ such that } W_n(\mathbf{z}) \leq t\},$$

where

$$\mathcal{S}_{\mathcal{C}}(\mathbf{y}) \triangleq \{\mathbf{e} \in \mathbf{F}_2^n : \mathbf{e} \oplus \mathbf{y} \in \mathcal{C}\}.$$

In other words, ρ_{pac} is the maximum generalized weight of an error word that the code \mathcal{C} can correct under minimum generalized distance decoding, which is equivalent to ML decoding [19].

Definition 3: The generalized covering radius ρ_{cov} of an (n, M) code \mathcal{C} is:

$$\rho_{cov} = \max_{\mathbf{y} \in \mathbf{F}_2^n} \min_{\mathbf{c} \in \mathcal{C}} D_n(\mathbf{c}, \mathbf{y}).$$

In other words, any received word \mathbf{y} is within at most ρ_{cov} from at least one codeword in \mathcal{C} .

Definition 4 (Generalized perfect and quasi-perfect codes): A code \mathcal{C} is called a generalized perfect code iff $\rho_{cov} = \rho_{pac}$. For such a code, every received word \mathbf{y} is within at most ρ_{cov} from *exactly one* codeword in \mathcal{C} .

A code \mathcal{C} is called a generalized quasi-perfect code iff $\rho_{pac} < \rho_{cov}$ and $\exists t^* \in \mathcal{K}(D_n)$ such that $\rho_{pac} < t^* < \rho_{cov}$.

Remark 2: The generalized definition of perfect (respectively, quasi-perfect) codes reduces to the conventional definition when the distance D_n is given by the Hamming distance.

Theorem 1 ([19, Theorem 1]): Generalized perfect and quasi-perfect codes are optimal (i.e., have minimal codeword error probability) under ML decoding among all codes with the same length and dimension (i.e., number of codewords or size).

We close this section by proving the following two results.

Lemma 2 (Optimality of classical perfect codes over channels with memory): Let \mathcal{C} be an (n, M, d_{min}) perfect code (in the classical sense) to be used over the general binary additive noise channel. Define the set

$$\mathcal{D}_{\mathcal{C}} = \left\{ \mathbf{e} \in \mathbf{F}_2^n : w_H(\mathbf{e}) \leq r_{cov} = \left\lfloor \frac{d_{min} - 1}{2} \right\rfloor \right\}$$

where r_{cov} denotes the code's (classical) covering radius. Now consider the following condition.

Condition (*): For any $\mathbf{e} \in \mathcal{D}_{\mathcal{C}}$ and any $\mathbf{e}' \in \mathbf{F}_2^n$, $w_H(\mathbf{e}) < w_H(\mathbf{e}') \implies P(\mathbf{Z} = \mathbf{e}) > P(\mathbf{Z} = \mathbf{e}')$.

If condition (*) holds, then \mathcal{C} is a generalized perfect code and hence is optimal among all codes of the same length and dimension under minimum generalized distance decoding (which is equivalent to ML decoding).

Proof: Let \mathcal{C} be an (n, M, d_{min}) perfect code satisfying condition (*). Its generalized covering radius satisfies

$$\begin{aligned}
\rho_{cov} &= \max_{\mathbf{y} \in \mathbf{F}_2^n} \min_{\mathbf{c} \in \mathcal{C}} D_n(\mathbf{c}, \mathbf{y}) \\
&= \max_{\mathbf{y} \in \mathbf{F}_2^n} \min_{\mathbf{c} \in \mathcal{C}} W_n(\mathbf{c} \oplus \mathbf{y}) \\
&= \max_{\mathbf{y} \in \mathbf{F}_2^n} \min_{\mathbf{z} \in \mathcal{S}_C(\mathbf{y})} W_n(\mathbf{z}) \\
&= \max_{\mathbf{y} \in \mathbf{F}_2^n} W_n(\mathbf{z}^*(\mathbf{y})), \tag{6}
\end{aligned}$$

where

$$\mathbf{z}^*(\mathbf{y}) = \arg \min_{\mathbf{z} \in \mathcal{S}_C(\mathbf{y})} W_n(\mathbf{z}).$$

Now since \mathcal{C} is a perfect code in the classical sense, then $\forall \mathbf{y} \in \mathbf{F}_2^n, \exists$ a unique error pattern $\hat{\mathbf{z}}(\mathbf{y}) \in \mathcal{S}_C(\mathbf{y})$ (of minimal Hamming weight) such that $w_H(\hat{\mathbf{z}}(\mathbf{y})) \leq r_{cov}$. From condition (*), $\forall \mathbf{z} \in \mathcal{S}_C(\mathbf{y})$ such that $\mathbf{z} \neq \hat{\mathbf{z}}(\mathbf{y}), P(\mathbf{Z} = \hat{\mathbf{z}}(\mathbf{y})) > P(\mathbf{Z} = \mathbf{z})$. Hence, $\mathbf{z}^*(\mathbf{y}) = \hat{\mathbf{z}}(\mathbf{y})$. Since \mathcal{C} is a perfect code, then

$$\{\hat{\mathbf{z}}(\mathbf{y}) : \mathbf{y} \in \mathbf{F}_2^n\} = \{\mathbf{z} \in \mathbf{F}_2^n : w_H(\mathbf{z}) \leq r_{cov}\} = \mathcal{D}_C,$$

which is the set of all coset leaders for \mathcal{C} . Therefore,

$$\begin{aligned}
\rho_{cov} &= \max_{\mathbf{y} \in \mathbf{F}_2^n} W_n(\hat{\mathbf{z}}(\mathbf{y})) \\
&= \max_{\mathbf{z} \in \mathcal{D}_C} W_n(\mathbf{z}) \\
&= \max_{\mathbf{z} \in \mathcal{D}_C : w_H(\mathbf{z}) = r_{cov}} W_n(\mathbf{z}).
\end{aligned}$$

The last equality is a result of condition (*).

We now prove that the generalized packing radius ρ_{pac} of \mathcal{C} is the same as its generalized covering radius ρ_{cov} . By definition, $\rho_{pac} \leq \rho_{cov}$. Assume $\rho_{pac} < \rho_{cov}$, then there exists at least one word $\mathbf{y} \in \mathbf{F}_2^n$ with two error patterns \mathbf{z}_1 and $\mathbf{z}_2 \in \mathcal{S}_C(\mathbf{y})$ such that $W_n(\mathbf{z}_1) \leq \rho_{cov}$ and $W_n(\mathbf{z}_2) \leq \rho_{cov}$. Now for any $\mathbf{z}' \in \mathbf{F}_2^n$,

$$\begin{aligned}
W_n(\mathbf{z}') \leq \rho_{cov} &\iff W_n(\mathbf{z}') \leq \max_{\mathbf{z} \in \mathcal{D}_C : w_H(\mathbf{z}) = r_{cov}} W_n(\mathbf{z}) \\
&\iff P(\mathbf{Z} = \mathbf{z}') \geq \max_{\mathbf{z} \in \mathcal{D}_C : w_H(\mathbf{z}) = r_{cov}} P(\mathbf{Z} = \mathbf{z}) \\
&\iff P(\mathbf{Z} = \mathbf{z}') \geq P(\mathbf{Z} = \mathbf{z}^*),
\end{aligned}$$

where

$$\mathbf{z}^* = \arg \max_{\mathbf{z} \in \mathcal{D}_C : w_H(\mathbf{z}) = r_{cov}} P(\mathbf{Z} = \mathbf{z}).$$

Since $\mathbf{z}^* \in \mathcal{D}_C$, from condition (*):

$$P(\mathbf{Z} = \mathbf{z}') \geq P(\mathbf{Z} = \mathbf{z}^*) \iff w_H(\mathbf{z}') \leq w_H(\mathbf{z}^*) = r_{cov}.$$

Therefore, both \mathbf{z}_1 and \mathbf{z}_2 have a Hamming weight of at most r_{cov} and they both belong to the error set of \mathbf{y} , $\mathcal{S}_C(\mathbf{y})$. This is a contradiction since \mathcal{C} is a perfect code and hence any error set can contain at most one error word with a Hamming weight less than or equal to the covering radius of the code. Therefore, $\rho_{pac} = \rho_{cov}$ and hence \mathcal{C} is a generalized perfect code. ■

Lemma 3 (Optimality of classical quasi-perfect codes over channels with memory): Let \mathcal{C} be an (n, M, d_{min}) quasi-perfect code (in the classical sense) with covering radius r_{cov} to be used over the general binary additive noise channel. Consider the set

$$\Gamma_C = \left\{ \mathbf{e} \in \mathbf{F}_2^n : w_H(\mathbf{e}) \leq r_{cov} = \left\lfloor \frac{d_{min} - 1}{2} \right\rfloor + 1 \right\}$$

and the following condition.

Condition (**): For any $\mathbf{e} \in \Gamma_C$ and any $\mathbf{e}' \in \mathbf{F}_2^n$,

$$w_H(\mathbf{e}) < w_H(\mathbf{e}') \iff P(\mathbf{Z} = \mathbf{e}) > P(\mathbf{Z} = \mathbf{e}').$$

If condition (**) holds, then \mathcal{C} is a generalized quasi-perfect code and hence is optimal among all codes of the same length and dimension under minimum generalized distance decoding (which is equivalent to ML decoding).

Proof: See [18, Lemma 3.4]. ■

IV. ML DECODING OVER CHANNELS WITH MEMORY

In this section, we study the problem of block ML decoding of length- n binary codes over the IMCC and the QBC. For the IMCC and the QBC with $M \geq n$, we provide necessary and sufficient conditions for which ML and MD decoding are equivalent. Furthermore, for the QBC with $M = 1, 2$, we derive sufficient conditions under which ML and SMD decoding are equivalent. We first present the results for the IMCC and for the QBC with $M \geq n$ together and then we present the results for the QBC with $M = 1, 2$ (note that for $M = 1$, the QBC and the FMCC are identical).

A. *Infinite Memory Contagion Channel and Queue-Based Channel with $M \geq n$*

Theorem 2: For any (n, M, d_{min}) code \mathcal{C} used over the IMCC or the QBC with $M \geq n$, if the code's (classical) covering radius satisfies

$$r_{cov} \leq \frac{1 - 2p}{\delta},$$

then the outputs of the MD and ML decoders are identical.

Proof: We only provide the proof for the IMCC (the proof is identical for the QBC with $M \geq n$; see [18, Section 6.1]).

Let $\mathbf{y}_1^n = (y_1, y_2, \dots, y_n)$ be the received word. Define

$$d_{min}(\mathbf{y}_1^n) = \min_{\mathbf{c} \in \mathcal{C}} d_H(\mathbf{c}, \mathbf{y}_1^n)$$

and

$$d_{max}(\mathbf{y}_1^n) = \max_{\mathbf{c} \in \mathcal{C}} d_H(\mathbf{y}_1^n, \mathbf{c}).$$

Then we have that $d_{min}(\mathbf{y}_1^n) \leq r_{cov}$ from the definition of the covering radius (with equality achieved for at least one word). On the other hand, $d_{max}(\mathbf{y}_1^n) \leq n$. (Note that there are codes for which both inequalities can be satisfied with equality for the same word \mathbf{y}_1^n .) Hence, $|d_{max}(\mathbf{y}_1^n) - \mu| \leq \frac{n}{2} - \frac{1-2p}{2\delta}$ and $|d_{min}(\mathbf{y}_1^n) - \mu| \geq \frac{n}{2} - \frac{1-2p}{2\delta}$, where

$$\mu = \frac{1 - 2p}{2\delta} + \frac{n}{2} \quad (7)$$

and we have used the assumption that $r_{cov} \leq \frac{1-2p}{\delta}$. Therefore, for any received word \mathbf{y}_1^n , $|d_{min}(\mathbf{y}_1^n) - \mu| \geq |d_{max}(\mathbf{y}_1^n) - \mu|$ which by [8, Section III] (see also [18, Section 4.1]) directly means that the MD decoding rule is always used. ■

We illustrate the condition of Theorem 2 in Fig. 1 by plotting, for different values of the IMCC BER p , the maximum allowable value for a code's covering radius for which ML and MD decoding are equivalent. It is noted that as the channel's noise correlation coefficient decreases, larger values of r_{cov} are permissible; this is expected, since when the noise correlation is set to zero, the channel reverts to the memoryless BSC for which the MD-ML decoding equivalence holds for all block codes. Furthermore, we can tighten the condition in Theorem 2 to obtain a necessary and sufficient condition on the code \mathcal{C} .

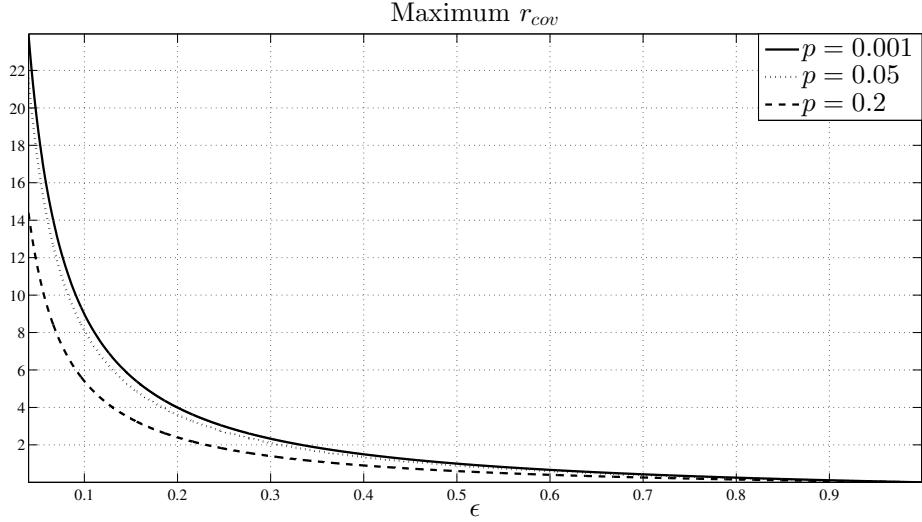


Fig. 1: Plot of the condition in Theorem 2: maximum allowable r_{cov} over the IMCC with respect to the correlation coefficient $\epsilon = \delta/(1 + \delta)$ and for different values of the BER p .

Definition 5: Let \mathbf{y}_1^n be the received word. We define:

$$d_{sum}(\mathbf{y}_1^n) \triangleq |d_{min}(\mathbf{y}_1^n) - d_{min}(\mathbf{1}^n \oplus \mathbf{y}_1^n)|,$$

and let

$$d_{sum}(\mathcal{C}) \triangleq \max_{\mathbf{y}_1^n \in \mathbf{F}_2^n} d_{sum}(\mathbf{y}_1^n),$$

where $\mathbf{1}^n$ is the all-one word of length n and $d_{min}(\mathbf{y}_1^n)$ is defined above.

Theorem 3: For any (n, M, d_{min}) code \mathcal{C} used over the IMCC or the QBC with $M \geq n$, the outputs of the MD and ML decoders are identical iff

$$d_{sum}(\mathcal{C}) \leq \frac{1 - 2p}{\delta}.$$

Proof: Again, we only provide the proof for the IMCC (as the proof is identical for the QBC with $M \geq n$; see [4, Section 6.1]).

We start by proving the first direction (\implies):

Assume $d_{sum}(\mathcal{C}) \leq \frac{1-2p}{\delta}$, and let \mathbf{y}_1^n be the received word. Then:

$$d_{min}(\mathbf{y}_1^n) + d_{max}(\mathbf{y}_1^n) = d_{min}(\mathbf{y}_1^n) + n - d_{min}(\mathbf{1}^n \oplus \mathbf{y}_1^n)$$

$$\begin{aligned}
&\leq n + d_{sum}(\mathbf{y}_1^n) \\
&\leq n + d_{sum}(\mathcal{C}) \\
&\leq n + \frac{1-2p}{\delta} \\
&= 2\mu
\end{aligned}$$

where μ is given in (7). Hence,

$$d_{min}(\mathbf{y}_1^n) + d_{max}(\mathbf{y}_1^n) \leq 2\mu \implies |d_{min}(\mathbf{y}_1^n) - \mu| \geq |d_{max}(\mathbf{y}_1^n) - \mu|,$$

which by [8, Section III] (see also [18, Section 4.1] directly means that for every $\mathbf{y}_1^n \in \mathbb{F}_2^n$, the ML decoder picks the minimum Hamming weight error word and hence reduces to the MD decoder.

We now prove the other direction (\Leftarrow):

Assume $d_{sum}(\mathcal{C}) > \frac{1-2p}{\delta}$. From the definition of $d_{sum}(\mathcal{C})$, we know that there exist at least one word $\bar{\mathbf{y}}_1^n$ such that:

$$d_{min}(\bar{\mathbf{y}}_1^n) - d_{min}(\mathbf{1}^n \oplus \bar{\mathbf{y}}_1^n) > \frac{1-2p}{\delta}.$$

For this received word, we have

$$\begin{aligned}
d_{min}(\bar{\mathbf{y}}_1^n) + d_{max}(\bar{\mathbf{y}}_1^n) &= d_{min}(\bar{\mathbf{y}}_1^n) + n - d_{min}(\mathbf{1}^n \oplus \bar{\mathbf{y}}_1^n) \\
&> n + \frac{1-2p}{\delta} \\
&= 2\mu.
\end{aligned}$$

Hence,

$$d_{min}(\bar{\mathbf{y}}_1^n) + d_{max}(\bar{\mathbf{y}}_1^n) > 2\mu \implies |d_{min}(\bar{\mathbf{y}}_1^n) - \mu| < |d_{max}(\bar{\mathbf{y}}_1^n) - \mu|.$$

The above means (by [8, Section III]) that for this received word $\bar{\mathbf{y}}_1^n$, the ML decoder picks the maximum Hamming weight error word and hence the ML decoder is not equivalent to the MD decoder. ■

Corollary 1: Let \mathcal{C} be an (n, M, d_{min}) perfect (respectively, quasi-perfect) code, in the classical sense, used over the IMCC or the QBC with $M \geq n$. If

$$r_{cov} \leq \frac{1-2p}{\delta},$$

then \mathcal{C} is a generalized perfect (resp., generalized quasi-perfect) code for this channel and hence is optimal under ML decoding among all codes of the same length and dimension sent over the same channel.

Proof: Immediate from Lemma 2 (resp., Lemma 3) and Theorem 2. ■

Lemma 4: Consider any (n, M, d) code \mathcal{C} that includes every codeword along with its complement, i.e., if $\mathbf{c} \in \mathcal{C}$ then $(\mathbf{1}^n \oplus \mathbf{c}) \in \mathcal{C}$. If $p < 0.5$ then ML decoding over the IMCC or the QBC with $M \geq n$ reduces to MD decoding.

Proof: The proof follows directly from Theorem 3 by noting that $d_{sum}(\mathcal{C}) = 0$. ■

B. Queue Based Channel with $M = 1, 2$

The following quantities expressed in terms of the QBC parameters and block length will be needed for the main results (Theorems 4 and 5 and Corollary 2) of this subsection.

Definition 6: For the QBC with parameters $(M = 1, \epsilon, p)$ and a block length $n > 2$, define

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

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

and

$$m^*(n, \epsilon, p) \triangleq \min\{m_1(\epsilon, p), m_2(n, \epsilon, p)\}.$$

For the QBC with parameters ($M = 2$, ϵ , p and α) and a block length $n > 2$, define

$$\tilde{m}_1(\alpha, \epsilon, p) \triangleq \frac{\ln \left\{ \frac{\left[\frac{(1+\alpha)\epsilon}{1-\epsilon} + 1 - p \right] \left[\frac{(1+\alpha)\epsilon}{1-\epsilon} + p \right]}{p \left(\frac{\epsilon}{1-\epsilon} + p \right)} \right\}}{\ln \left\{ \frac{\left[\frac{(1+\alpha)\epsilon}{1-\epsilon} + p \right] \left[\frac{(1+\alpha)\epsilon}{1-\epsilon} + 1 - p \right]^2}{p \left(\frac{\alpha\epsilon}{1-\epsilon} + 1 - p \right) \left(\frac{\epsilon}{1-\epsilon} + 1 - p \right)} \right\}},$$

$$\tilde{m}_2(n, \alpha, \epsilon, p) \triangleq \begin{cases} \frac{D+A}{C}, & \text{if } \alpha > \frac{(1-2p)(1-\epsilon)}{\epsilon} - 1 \\ \frac{D+B}{C}, & \text{otherwise,} \end{cases}$$

and

$$\tilde{m}(n, \alpha, \epsilon, p) \triangleq \min\{\tilde{m}_1(\alpha, \epsilon, p), \tilde{m}_2(n, \alpha, \epsilon, p)\},$$

where

$$\begin{aligned} A &\triangleq \ln \left[\frac{\left[\frac{(1+\alpha)\epsilon}{1-\epsilon} + p \right]^2 (1-p) \left(\frac{\epsilon}{1-\epsilon} + 1 - p \right)}{p \left(\frac{\epsilon}{1-\epsilon} + p \right) \left[\frac{(1+\alpha)\epsilon}{1-\epsilon} + 1 - p \right]^2} \right], \\ B &\triangleq \ln \left[\frac{\left[\frac{(1+\alpha)\epsilon}{1-\epsilon} + p \right]^3 \left(\frac{\epsilon}{1-\epsilon} + 1 - p \right)}{p \left(\frac{\epsilon}{1-\epsilon} + p \right) \left[\frac{(1+\alpha)\epsilon}{1-\epsilon} + 1 - p \right]^2} \right], \\ C &\triangleq \ln \left[\frac{\left(\frac{(1+\alpha)\epsilon}{1-\epsilon} + 1 - p \right)^3}{p \left(\frac{\alpha\epsilon}{1-\epsilon} + 1 - p \right) \left(\frac{\epsilon}{1-\epsilon} + 1 - p \right)} \right], \\ D &\triangleq n \ln \left[\frac{\frac{(1+\alpha)\epsilon}{1-\epsilon} + 1 - p}{\frac{(1+\alpha)\epsilon}{1-\epsilon} + p} \right]. \end{aligned}$$

Theorem 4: For any two error words \mathbf{z}_1^n and $\bar{\mathbf{z}}_1^n$ generated by the QBC with $M = 1$ (resp., $M = 2$) satisfying

- i. $w_H(\mathbf{z}_1^n) = m$, where $0 \leq m < \frac{n}{2}$ (resp., $0 \leq m < \frac{n}{3}$),
- ii. $w_H(\bar{\mathbf{z}}_1^n) = m + i$, where $1 \leq i \leq n - m$

we have that

$$m < m^*(n, \epsilon, p) \quad (\text{resp.}, m < \tilde{m}(n, \alpha, \epsilon, p)) \iff \frac{P(\mathbf{Z}_1^n = \mathbf{z}_1^n)}{P(\mathbf{Z}_1^n = \bar{\mathbf{z}}_1^n)} > 1.$$

Proof: See the Appendix. ■

Theorem 5: Let \mathcal{C} be any (n, K, d) code used over the QBC with $M = 1$ (resp., $M = 2$) and parameters α, ϵ and p . If the code's classical covering radius r_{cov} satisfies

$$r_{cov} < \min \left\{ m^*(n, \epsilon, p), \frac{n}{2} \right\} \quad (\text{resp.}, r_{cov} < \min \left\{ \tilde{m}(n, \alpha, \epsilon, p), \frac{n}{3} \right\}),$$

then the output of the SMD decoder (when it does not declare a decoding failure) is identical to the output of the ML decoder for this code.

Proof: We only prove the result for $M = 2$ (the proof for $M = 1$ is identical). Let \mathbf{y} be the received word. Let $m \triangleq \min_{\mathbf{c} \in \mathcal{C}} d_H(\mathbf{y}, \mathbf{c})$. Clearly, $m \leq r_{cov} < \frac{n}{3}$ (from the definition of the covering radius). If there exists 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 $m < \tilde{m}(n, \alpha, \epsilon, p)$, it follows from Theorem 4 that all other error words of larger Hamming weights have a smaller probability than the error word corresponding to the SMD decision. Hence the ML decoder will give the same output. ■

In Figs. 2 and 3, we plot $m^*(n, \epsilon, p)$ and $\tilde{m}(n, \alpha, \epsilon, p)$ versus the channel correlation coefficient ϵ (ranging from 0.03 to 1) of the QBC with parameter $\alpha = 1$ (hence the QBC reduces to the FMCC) and for $M = 1$ and $M = 2$, respectively, for different values of the BER p and the block length n . We notice from these figures that the condition of Theorem 5 is 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 when $\epsilon = 0$ (i.e., when the QBC becomes memoryless and reduces to the BSC), unsurprisingly all block codes satisfy it.

Corollary 2: Let \mathcal{C} be an (n, K, d_{min}) perfect code (in the classical sense) used over the QBC with $M = 1$ or $M = 2$ and parameters α, ϵ and p . If

$$r_{cov} = \left\lfloor \frac{d_{min} - 1}{2} \right\rfloor < \min \left\{ m^*(n, \epsilon, p), \frac{n}{2} \right\} \quad \text{for } M = 1,$$

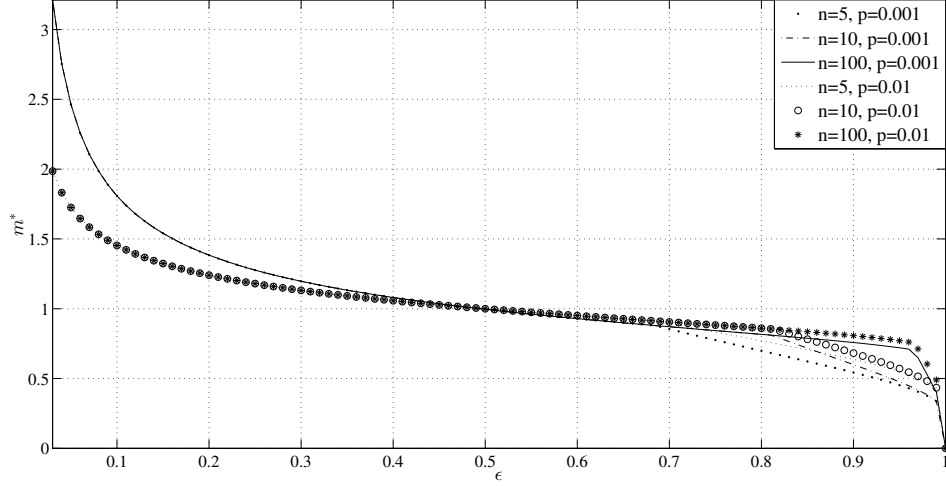


Fig. 2: Plot of $m^*(n, \epsilon, p)$ with respect to ϵ for different values of n and for $p = 0.001, 0.01$, QBC with $\alpha = 1$ and $M = 1$.

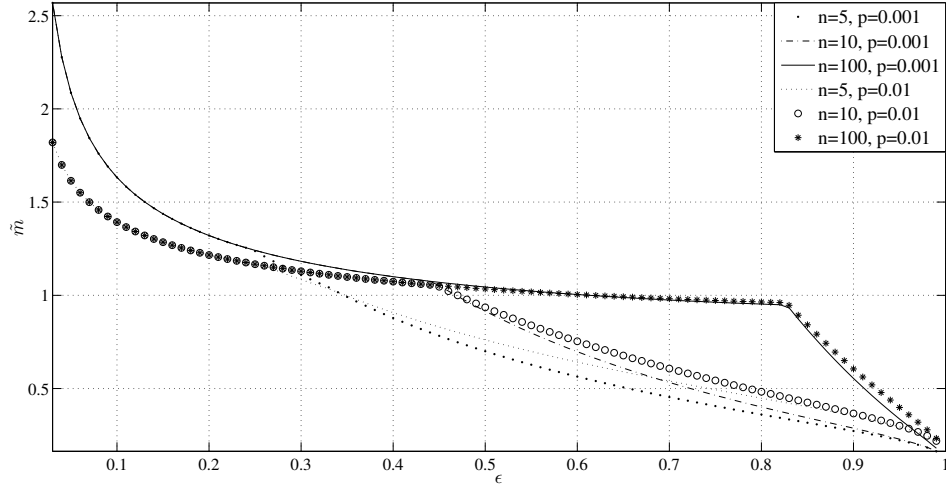


Fig. 3: Plot of $\tilde{m}(n, \alpha, \epsilon, p)$ with respect to ϵ for different values of n and for $p = 0.001, 0.01$, QBC with $\alpha = 1$, $M = 2$.

$$r_{cov} = \left\lfloor \frac{d_{min} - 1}{2} \right\rfloor < \min \left\{ \tilde{m}(n, \alpha, \epsilon, p), \frac{n}{3} \right\} \text{ for } M = 2,$$

then \mathcal{C} is a generalized perfect code for the corresponding channel and hence is optimal (under ML decoding) among all codes of the same length and dimension sent over the same channel.

Proof: Immediate from Lemma 2 and Theorems 4 and 5. ■

Remark 3: Note that we can not make a similar statement to the one in Corollary 2 for quasi-perfect codes, since two error words of the same weight are not guaranteed to have the same probability.

Remark 4: It should be pointed out that Corollary 2 directly implies the result of [15, Theorem 1] stating that every binary (classical) perfect of minimum Hamming distance three is optimal among all other codes of the same block length and dimension over any binary first-order Markov additive noise channel (which is identical to the QBC and FMCC with $M = 1$) with BER p and noise correlation coefficient

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

V. DISCUSSION

A. Numerical Validation and Listing of Short-Length Codes

We have verified the conditions of Theorems 3 and 5 established in the previous section by simulating over the IMCC and the QBC the performance of several short-length codes, including Hamming, Golay and Reed-Muller codes, as well as a perfect nonlinear code due to Vasil'ev [20], [21]. As expected, we observed that the probabilities of codeword error agree under the different decoding regimes within the parameters of the decoding equivalence conditions. Detailed numerical results can be found in [18].

It is also worth pointing out that the above (Hamming, Golay and Vasil'ev) perfect codes and other perfect codes satisfying Theorems 3 and 5 will perform optimally (in terms of yielding a minimal codeword error probability) among all binary block codes of identical size and block length when operated on the IMCC or the QBC under ML decoding.

We next determine in Table I the largest possible covering radii permitted by Theorems 2 and 5 for replacing ML decoding by MD decoding over the IMCC and by SMD decoding over the QBC with $M = 1$ (i.e., FMCC with $M = 1$) and QBC with $M = 2$ and $\alpha = 1$ for typical values of the channel parameters p and ϵ . It is observed that as the noise correlation coefficient ϵ increases, the class of codes for which the decoding equivalence holds gets smaller (becoming empty when $\epsilon = 0, 9$). In Table II, we list examples of codes with short block lengths

and covering radii that can be handily chosen according to the channel parameters of Table I. These codes can hence be useful for practical delay and complexity constrained systems, such as wireless sensor networks, when the systems' end-to-end binary channels are better approximated via the above channel models than the BSC.

BER p	ϵ			
	0.1	0.2	0.5	0.9
0.1	(7,2,2)	(3,1,1)	(0,0,1)	(0,0,0)
0.01	(8,1,1)	(3,1,1)	(0,0,1)	(0,0,0)
0.001	(8,1,1)	(3,1,1)	(0,0,1)	(0,0,0)

TABLE I: Largest possible covering radius values as specified by Theorems 2 and 5 for the IMCC, the QBC with $M = 1$ (i.e., the FMCC with $M = 1$) and the QBC with $M = 2$ and $\alpha = 1$ displayed as a triplet $(r_{cov}(\text{IMCC}), r_{cov}(\text{FMCC with } M = 1), r_{cov}(\text{QBC with } M = 2))$ for different values of the channel BER p and noise correlation coefficient ϵ .

r_{cov}	List of codes
1	Family of Hamming codes, Vasil'ev nonlinear perfect code.
2	$(23, 2^{12}, 7)$ Golay code, family of extended Hamming codes, $(8, 2^4, 4)$ Reed-Muller code.
3	$(24, 2^{12}, 8)$ extended Golay code, $(23, 2^{14}, 5)$ Wagner code, $(15, 2^7, 5)$ BCH code.
> 3	$(31, 2^{11}, 11)$ BCH code ($r_{cov} = 7$), $(16, 2^5, 8)$ Reed-Muller code ($r_{cov} = 6$).

TABLE II: Example of codes with small covering radii.

B. Application to Syndrome Source Coding

In light of the duality between source and channel coding (e.g., see [22]–[25] and other works), the results that we derived in this work can be directly applied to syndrome source coding with or without side information at the receiver. Our results extend similar work in [24].

The syndrome source coding scheme without side information uses an $(n, 2^k, d_{min})$ *linear* code \mathcal{C} . The source encoder computes the syndrome \mathbf{s}_1^{n-k} of the source output \mathbf{e}_1^n and sends it over a noiseless communication channel. The source decoder outputs its estimate $\hat{\mathbf{e}}_1^n$ of the source output from its syndrome. If the n -bit source output is identically distributed to the n -bit error pattern generated by one of the channels that we considered in this work and if the *linear* code \mathcal{C} satisfies the conditions we presented for that channel, then we obtain the same equivalence relation between the ML and the MD (or SMD) decoders.

Alternatively, in the syndrome source coding scheme with side information at the decoder, the source gives two outputs Y_i and X_i , where the latter is only available at the decoder. Let $U_i = X_i \oplus Y_i$. In this scheme, the source encoder computes the syndrome \mathbf{z}_1^{n-k} of the n -bit source output \mathbf{y}_1^n and transmits it over a noiseless communication channel. The source decoder computes the syndrome of the second n -bit source output \mathbf{x}_1^n and adds it bitwise to \mathbf{z}_1^{n-k} (modulo-2). It can be easily proven that the result is the syndrome of $\mathbf{u}_1^n = \mathbf{x}_1^n \oplus \mathbf{y}_1^n$. Hence, the decoder computes its estimate $\hat{\mathbf{u}}_1^n$ of \mathbf{u}_1^n from its syndrome and outputs $\hat{\mathbf{y}}_1^n = \hat{\mathbf{u}}_1^n \oplus \mathbf{x}_1^n$. Similarly, if the \mathbf{U}_1^n is identically distributed to the n -bit error pattern of one of the channels that we considered in this work and if the *linear* code \mathcal{C} satisfies the condition we presented for that channel, then we obtain the same equivalence relation between the ML and the MD (or SMD) decoders.

VI. CONCLUSION

In this work, we presented sufficient conditions on general binary codes under which SMD and ML are equivalent over the QBC with $M = 1, 2$. We also established sufficient conditions under which classical perfect codes are optimal under ML decoding over these channels with memory. For the IMCC and the QBC with $M \geq n$, we provided both necessary and sufficient conditions on binary codes for which ML and MD are equivalent. As intuition suggests, it is generally observed that the class of block codes for which the MD-ML decoding equivalence holds gets larger as the channels' noise correlation decreases. Furthermore, we determined sufficient conditions under which classical perfect and quasi-perfect codes are optimal under ML decoding over these channels; these codes can hence be beneficial to deploy in delay and complexity constrained applications where the underlying binary-input binary-output end-to-end

channel is better represented by the above channel models than the traditional memoryless BSC.

ACKNOWLEDGEMENT

The Editor and the reviewers are sincerely thanked for their incisive comments which significantly improved the paper.

APPENDIX

Proof of Theorem 4: For the sake of brevity, we only include a proof for the case of the QBC with $M = 2$. Interested readers can find the full proof in [18, Theorems 5.1 and 6.4]. We present the following two lemmas without their proofs which can be found in [18, Lemmas 6.1 and 6.2].

Lemma 5: The error pattern \mathbf{z}_1^n of Hamming weight $0 < m < n$ where all zeros and ones are consecutive (e.g., $\mathbf{z}_1^n = 00\dots011\dots1$) is the most likely among all other error patterns of the same length and weight generated by the QBC with $M = 2$.

Lemma 6: Consider the error words of length n having a Hamming weight $0 < m < \frac{n}{3}$. The pattern $\mathbf{b}_1^n = (001001\dots00)$ is the least likely among all patterns of the same length and weight generated by the QBC with $M = 2$.

Define $\mathbf{a}_1^n(t)$ and $\mathbf{b}_1^n(t)$ to be the following n -bit patterns of Hamming weight t :

$$\begin{aligned} \mathbf{a}_1^n(t) &= (000\dots111) && \text{for } 0 < t < n, \\ \mathbf{b}_1^n(t) &= (001001\dots000) && \text{for } 0 < t < \frac{n}{3}. \end{aligned}$$

In other words, $\mathbf{a}_1^n(t)$ contains t consecutive ones at the end of its sequence (with $0 < t < n$) and $\mathbf{b}_1^n(t)$ consists of the pattern 001 repeated t times followed by $n - t$ zeros (with $0 < t < n/3$).

We first prove the first direction (\implies): Consider the following three cases:

- Case 1: $m = 0$

In this case, \mathbf{z}_1^n is the all-zero error pattern. From Lemma 1, $P(\mathbf{Z}_1^n = \mathbf{0}^n) > P(\mathbf{Z}_1^n = \bar{\mathbf{z}}_1^n)$.

- Case 2: $0 < m < \frac{n}{3}$ and $1 \leq i < n - m - 1$

$$\frac{P(\mathbf{Z}_1^n = \mathbf{z}_1^n)}{P(\mathbf{Z}_1^n = \bar{\mathbf{z}}_1^n)} \geq \frac{\min_{\mathbf{z}_1^n \in \mathbf{F}_2^n: w_H(\mathbf{z}_1^n) = m} P(\mathbf{Z}_1^n = \mathbf{z}_1^n)}{\max_{\mathbf{z}_1^n \in \mathbf{F}_2^n: w_H(\mathbf{z}_1^n) = m+i} P(\mathbf{Z}_1^n = \mathbf{z}_1^n)}$$

$$\begin{aligned}
&= \frac{P(\mathbf{Z}_1^n = \mathbf{b}_1^n(m))}{P(\mathbf{Z}_1^n = \mathbf{a}_1^n(m+i))} \\
&= \left[\frac{\frac{\alpha\epsilon}{1-\epsilon} + 1 - p}{1-p} \right]^m \left[\frac{\frac{(1+\alpha)\epsilon}{1-\epsilon} + p}{\frac{\epsilon}{1-\epsilon} + p} \right]^{-(m+i-2)} \left[\frac{\frac{\epsilon}{1-\epsilon} + p}{p} \right]^{-(m+i-1)} \\
&\quad \left[\frac{1-p}{p} \right]^i \left[\frac{\frac{(1+\alpha)\epsilon}{1-\epsilon} + 1 - p}{\frac{\epsilon}{1-\epsilon} + 1 - p} \right]^{i-2m} \left[\frac{\frac{\epsilon}{1-\epsilon} + 1 - p}{1-p} \right]^{i-m}
\end{aligned}$$

where the first equality follows from Lemmas 5 and 6. Thus,

$$\begin{aligned}
\frac{P(\mathbf{Z}_1^n = \mathbf{z}_1^n)}{P(\mathbf{Z}_1^n = \bar{\mathbf{z}}_1^n)} &\geq \left\{ \frac{p\left(\frac{\alpha\epsilon}{1-\epsilon} + 1 - p\right)\left(\frac{\epsilon}{1-\epsilon} + 1 - p\right)}{\left[\frac{(1+\alpha)\epsilon}{1-\epsilon} + p\right] \left[\frac{(1+\alpha)\epsilon}{1-\epsilon} + 1 - p\right]^2} \right\}^m \\
&\quad \frac{\left[\frac{(1+\alpha)\epsilon}{1-\epsilon} + 1 - p\right]^i \left[\frac{(1+\alpha)\epsilon}{1-\epsilon} + p\right]^2}{\left[\frac{(1+\alpha)\epsilon}{1-\epsilon} + p\right] p \left(\frac{\epsilon}{1-\epsilon} + p\right)} \\
&\geq \left\{ \frac{p\left(\frac{\alpha\epsilon}{1-\epsilon} + 1 - p\right)\left(\frac{\epsilon}{1-\epsilon} + 1 - p\right)}{\underbrace{\left[\frac{(1+\alpha)\epsilon}{1-\epsilon} + p\right] \left[\frac{(1+\alpha)\epsilon}{1-\epsilon} + 1 - p\right]^2}_{\leq 1}} \right\}^m \\
&\quad \frac{\left[\frac{(1+\alpha)\epsilon}{1-\epsilon} + 1 - p\right] \left[\frac{(1+\alpha)\epsilon}{1-\epsilon} + p\right]}{p \left(\frac{\epsilon}{1-\epsilon} + p\right)} \tag{8} \\
&> 1. \tag{9}
\end{aligned}$$

We have inequality in (8) since we set $i = 1$ (indeed, the term raised to the power i is greater than 1 and hence is increasing in i where $i \geq 1$). Inequality (9) is a result of the condition $m < \tilde{m}_1(\alpha, \epsilon, p)$.

- Case 3: $0 < m < \frac{n}{3}$ and $n - m - 1 \leq i \leq n - m$

In this case, $t_{000}(\bar{\mathbf{z}}_1^n) = t_{000}(\bar{\mathbf{z}}_1^n) = 0$. We have

$$\frac{P(\mathbf{Z}_1^n = \mathbf{z}_1^n)}{P(\mathbf{Z}_1^n = \bar{\mathbf{z}}_1^n)} \geq \frac{P(\mathbf{Z}_1^n = \mathbf{b}_1^n(m))}{P(\mathbf{Z}_1^n = \mathbf{a}_1^n(m+i))}$$

$$\begin{aligned}
&= \left[\frac{(1+\alpha)\frac{\epsilon}{1-\epsilon} + 1 - p}{1 - p} \right]^n \left[\frac{1 - p}{(1+\alpha)\frac{\epsilon}{1-\epsilon} + p} \right]^i \\
&\quad \left[\frac{p \left(\frac{\epsilon}{1-\epsilon} + 1 - p \right) \left(\frac{\alpha\epsilon}{1-\epsilon} + 1 - p \right) (1 - p)}{\left(\frac{(1+\alpha)\epsilon}{1-\epsilon} + 1 - p \right)^3 \left[\frac{(1+\alpha)\epsilon}{1-\epsilon} + p \right]} \right]^m \\
&\quad \frac{\left[\frac{(1+\alpha)\epsilon}{1-\epsilon} + p \right]^2 (1 - p) \left(\frac{\epsilon}{1-\epsilon} + 1 - p \right)}{p \left(\frac{\epsilon}{1-\epsilon} + p \right) \left[\frac{(1+\alpha)\epsilon}{1-\epsilon} + 1 - p \right]^2} \\
&> 1.
\end{aligned}$$

The last inequality is a result of the condition $m < \tilde{m}_2(n, \alpha, \epsilon, p)$.

We now prove the other direction (\Leftarrow):

- Assume $m \geq \tilde{m}_1(\alpha, \epsilon, p)$: In the proof of Case 2, all the inequalities except the last one can be met with equality by choosing the error patterns as follows: $\mathbf{z}_1^n = \mathbf{b}_1^n(m)$ and $\bar{\mathbf{z}}_1^n = \mathbf{a}_1^n(m+1)$. Under the assumption that $m \geq \tilde{m}_1(\alpha, \epsilon, p)$, we get:

$$\frac{P(\mathbf{Z}_1^n = \mathbf{z}_1^n)}{P(\mathbf{Z}_1^n = \bar{\mathbf{z}}_1^n)} \leq 1.$$

Therefore, we proved that there exist at least two words \mathbf{z}_1^n and $\bar{\mathbf{z}}_1^n$ satisfying:

- $w_H(\mathbf{z}_1^n) = m$, where $0 < m \leq \frac{n}{2}$
- $w_H(\bar{\mathbf{z}}_1^n) = m + i$, where $1 \leq i \leq n - m$

such that:

$$\frac{P(\mathbf{Z}_1^n = \mathbf{z}_1^n)}{P(\mathbf{Z}_1^n = \bar{\mathbf{z}}_1^n)} \leq 1.$$

- Assume $m \geq \tilde{m}_2(n, \alpha, \epsilon, p)$: The proof follows a similar reasoning as above, only this time we choose $\bar{\mathbf{z}}_1^n$ to be the all-one error word (while \mathbf{z}_1^n is unchanged).

■

REFERENCES

- [1] S. B. Wicker, *Error Control Systems for Digital Communication and Storage*, Prentice-Hall Inc., 1995.
- [2] R. G. Gallager, *Information Theory and Reliable Communication*, 1968.

- [3] P. Sadeghi, R. A. Kennedy, P. B. Rapajic, and R. Shams, "Finite-State Markov Modeling of Fading Channels," *IEEE Signal Processing Magazine*, vol. 25, no. 5, pp. 57-80, Sep. 2008.
- [4] Y. Q. Shi, X. M. Zhang, Z.-C. Ni and N. Ansari, "Interleaving for combating bursts of errors," *IEEE Circuits and Systems Magazine*, vol. 4, no. 1, pp. 29-42, First Quarter, 2004.
- [5] M. Zorzi, "Some results on error control for burst-error channels under delay constraints," *IEEE Trans. Veh. Technol.*, vol. 50, pp. 12-24, Jan. 2001.
- [6] R. L. Dobrushin and M. S. Pinsker, "Memory increases transmission capacity", *Probl. Pered. Inform.*, vol. 5, no. 1, pp. 94-95, 1969.
- [7] 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.
- [8] F. Alajaji and T. Fuja, "A communication channel modeled on contagion," *IEEE Trans. Inform. Theory*, vol. 40, pp. 2035-2041, Nov. 1994.
- [9] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, 1981.
- [10] J. Kieffer, "Epsilon-capacity for a class of nonergodic channels," in *Proc. IEEE Int. Symp. Inform. Theory*, Seattle, July 2006, pp. 1268-1271.
- [11] 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.
- [12] C. Pimentel and F. Alajaji, "Packet-based modeling of Reed-Solomon block coded correlated fading channels via a Markov finite queue model," *IEEE Trans. Veh. Technol.*, vol. 58, pp. 3124-3136, Sep. 2009.
- [13] 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.
- [14] A. Badr, A. Khisti, W.-T. Tan, and J. Apostolopoulos, "Streaming codes with partial recovery over channels with burst and isolated erasures," *IEEE J. Sel. Topics Signal Processing*, to appear (available at <http://www.comm.utoronto.ca/~akhisti/double.pdf>).
- [15] M. Hamada, "Near-optimality of subcodes of Hamming codes on the two-state Markovian additive channel," *IEICE Trans. Fundam. Electronics, Commun. Comp. Sciences*, vol. E84-A, pp. 2383-2388, Oct. 2001.
- [16] 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.
- [17] 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*, vol. 12, no. 9, pp. 678-680, Sep. 2008.
- [18] G. Azar, *On the Optimality of the Hamming Metric for Decoding Block Codes over Binary Additive Noise Channels*, M.A.Sc. Thesis, Department of Mathematics and Statistics, Queen's University, July 2013 (available at www.mast.queensu.ca/~web/Papers/azar-msc13.pdf).
- [19] 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.
- [20] J. L. Vasil’ev, “On nongroup close-packed codes,” *P robl. Kibernet.*, vol. 8, pp. 375-378, 1962.
- [21] Z. Wang, M. G. Karpovsky and K. J. Kulikowski, “Replacing linear Hamming codes by robust nonlinear codes results in a reliability improvement of memories,” in *Proc. IEEE/IFIP Int. Conf. Dependable Systems & Networks*, 2009, pp. 514-523.
- [22] T. Ancheta, “Syndrome source coding and its universal generalization,” *IEEE Trans. Inform. Theory*, vol. 22, no. 4, pp. 432-436, July 1976.
- [23] P. Mitran and J. Bajcsy, “Turbo source coding: A noise-robust approach to data compression,” in *Proc. IEEE Data Compression Conf.*, Snowbird, UT, Apr. 2002, p. 465.
- [24] X. Wu, *Syndrome Source Coding for Lossless Data Compression Based on Linear Block Codes*, M.Sc. Thesis, Dept. Mathematics and Statistics, Queen’s University, Sep. 2008.
- [25] A. D. Wyner, “Recent results in Shannon theory,” *IEEE Trans. Inform. Theory*, vol. 20, Jan. 1974.