Manuscript submitted to
AIMS' Journals
Volume **X**, Number **0X**, XX **200X**

doi:10.3934/xx.xx.xx.xx

pp. **X–XX**

CHANNEL DECOMPOSITION FOR MULTILEVEL CODES OVER MULTILEVEL AND PARTIAL ERASURE CHANNELS

Carolyn Mayer*

Department of Mathematics
University of Nebraska-Lincoln
Lincoln, Nebraska 68588–0130, USA

Kathryn Haymaker

Department of Mathematics & Statistics
Villanova University
Villanova, PA 19085, USA

Christine A. Kelley

Department of Mathematics
University of Nebraska-Lincoln
Lincoln, Nebraska 68588–0130, USA

(Communicated by the associate editor name)

Abstract. We introduce the Multilevel Erasure Channel (MEC) for binary extension field alphabets. The channel model is motivated by applications such as non-volatile multilevel read storage channels. Like the recently proposed q-ary partial erasure channel (QPEC), the MEC is designed to capture partial erasures. The partial erasures addressed by the MEC are determined by erasures at the bit level of the q-ary symbol representation. In this paper we derive the channel capacity of the MEC and give a multistage decoding scheme on the MEC using binary codes. We also present a low complexity multistage p-ary decoding strategy for codes on the QPEC when $q = p^k$. We show that for appropriately chosen component codes, capacity on the MEC and QPEC may be achieved.

1. Introduction.

Non-volatile memories are prevalent in current computer memory technologies such as flash memory, ferroelectric RAM, and magnetic storage systems. In some applications, symbols may be prone to error events in which some partial information can be obtained about the symbol at the receiving end. For example, both NAND flash memory and phase change memory are susceptible to retention errors and read errors [1, 2]. Cohen and Cassuto [3] introduced the q-ary Partial Erasure Channel (QPEC) to model partial erasure events, and provided a thorough analysis of iterative decoding of LDPC codes on the QPEC [3]. Their model is motivated by measurement channels, where information is read by assessing the level of charge in a cell or memory unit. Partial erasure events are characterized by incomplete read processes, such as when readouts of symbols from

2010 *Mathematics Subject Classification.* Primary: 94A40, 94B35; Secondary: 94A15.

Key words and phrases. Partial erasure channels, multilevel codes, multistage decoding, storage channels.

* Corresponding author: C. Mayer.

the channel terminate prematurely. Consequently, in their model, partially erased symbols are regarded as sets of a fixed cardinality M of possible symbols.

In this paper, we introduce another channel model that captures partial erasure events that may happen in practice. For example, in applications using q-ary symbols, typically from GF(2^k), the symbols may be stored or transmitted as binary k-tuples. This prompts the study of erasure events at the bit level. A symbol that has an erased bit (or bits) may also have partial information at the receiving end. To address these types of errors, we introduce the Multilevel Erasure Channel (MEC) model in which a partially erased symbol may belong to a set of size 2^j, where j ranges from 1 to k. Thus, a symbol that is erased has partial information available at the receiver, provided $j < k$. When $j = k$, the symbol is fully erased and may be any one of the q field elements. The MEC model may be applied, for example, when a erasure event is caused by a symbol readout terminating before completion.

Moreover, the MEC is tolerant to varying bit error probabilities that are inherent to the flash storage medium and possibly other storage media. Multilevel cell (MLC) flash and Triple-level cell (TLC) flash are two common settings, where in the MLC case, 4-ary symbols are stored as two bits, and in the TLC case, 8-ary symbols are stored as three bits. Moreover, in the storage architecture, each bit of a symbol is stored on a different "page," and the bits on each page are prone to different bit error probabilities. Thus, the channels characterizing bit transmission on each page are different. In [4], [5], the authors explain this phenomenon for MLC and TLC flash, and also examine the dependence among positions of error events. In [6], bit-interleaved schemes were analyzed to determine the effect on decoding performance. Our proposed MEC model is relevant for these applications in that it can allow for different bit error probabilities within each symbol. Codes for the MEC have similarities to codes designed with unequal error protection (e.g.[7, 8]).

Due to their simplicity, low complexity encoding/decoding, and capacity-achieving performance, Low-Density Parity-Check (LDPC) codes have become ubiquitous in modern technologies [10, 11, 12]. Low-density codes over the QPEC are also amenable to iterative and linear programming decoding [3], however, the analysis of the exact behavior of the decoder for the QPEC relates to the subset sum problem in group theory, and is a difficult problem [13]. Motivated by this, we examine the achievable rates of codes over the MEC and QPEC using multistage decoding. Imai and Hirakawa introduced multilevel coding in [14], leading to much work in the area of multilevel codes and multistage decoding [8, 15, 16]. In [17], a multilevel coding approach was proposed for magnetic storage applications.

While the idea of multilevel coding is not new, this paper gives the first application of multilevel coding to partial erasure channels. We determine that the MEC over GF(2^k) is equivalent to k independent BEC channels with transition probabilities given by the MEC. Prompted by this result, we examine the breakdown of the QPEC into knowledge per bit, and establish a multistage decoding scheme for codes on the QPEC over GF(2^k) using binary codes on the BEC. Although the BEC subchannels are no longer independent, we determine their erasure probabilities in terms of the parameters of the QPEC. When $M = 2$, we prove that capacity can be achieved by using binary codes optimized for channels with these erasure probabilities. Moreover, we show that for the QPEC with other values of M and q, multistage decoding may still be applied, but the subchannels may not be simple erasure channels.

This paper is organized as follows. Basic background and notation is given in Section 2. In Section 3 we present the multilevel erasure channel and derive the capacity of the channel. In Section 4 we briefly discuss coding techniques for LDPC codes on the MEC and QPEC. In Section 5, we analyze multistage decoding on the MEC and QPEC and determine when low complexity decoders may be used. Conclusions are given in Section 6.

2. **Preliminaries.**

We will focus on codes over nonbinary alphabets, particularly binary extension fields. Let α be a root of a primitive polynomial of degree k over GF(2). Since field elements can be represented as binary k-tuples or as powers of α, we will use the following mapping between these representations:

$$(a_{k-1}, a_{k-2}, \dots, a_1, a_0) \leftrightarrow a_{k-1}\alpha^{k-1} + \cdots + a_1\alpha^1 + a_k\alpha^0,$$

where $a_i \in \mathrm{GF}(2)$. When the context is clear, we will denote $(a_{k-1}, \dots, a_1, a_0)$ by $a_{k-1}a_{k-2}\dots a_1a_0$. We will also let $[n]$ denote the set of integers $\{1, 2, \dots, n\}$.

While there is no restriction on which types of codes may be used on the MEC, QPEC, or multilevel settings, we will assume low-density parity-check (LDPC) codes and focus on regular LDPC code ensembles [10]. An LDPC code over GF(q) is defined by a parity-check matrix H that is sparse in the number of nonzero entries, and is (d_v, d_c)-regular if there are d_v nonzero elements in each column and d_c nonzero elements in each row. A Tanner graph is a bipartite graph with a variable node (vertex) for each column of H, and a check node for each row of H. There is an edge from variable node v_i to check node c_l with label ψ precisely when $\psi \in \mathrm{GF}(q)$ is entry (l, i) in the parity-check matrix. The sparsity of H leads to a sparse graph representation, making these codes amenable to efficient graph-based iterative decoding algorithms [11].

Introduced by Cohen and Cassuto in [3], the q-ary partial erasure channel (QPEC) is a generalization of the BEC. Given a q-ary input, the QPEC outputs the original input with probability $1 - \varepsilon$, or a partial erasure with probability ε. A partial erasure is a set containing the original input as well as $M - 1$ other q-ary symbols. Each of the $\binom{q-1}{M-1}$ possible erasures for a symbol occur with equal probability. An example of the QPEC for $q = 2^2$ and $M = 2$ is given in Figure 1. The capacity of the QPEC over GF(q) with partial erasures of size M is $1 - \varepsilon \log_q M$ [3].

3. **The Multilevel erasure channel.**

3.1. **Channel Model.**

A 2^k-ary symbol x sent across the multilevel erasure channel (MEC) with erasure probability ε will either be received without error with probability $1 - \varepsilon$, or an erasure event will occur with probability ε. An erasure event may consist of any combination of bits in the binary representation of x being erased. For $i = 1, \dots, k$, we will use γ_i to denote the probability that bit i is erased. For each $j = 1, \dots, k$, there are $\binom{k}{j}$ sets of symbols whose binary representations differ from the binary representation of x in at most j selected bits. Using similar notation as in [3], we define the super-symbol $?_x^B$ to be the set of 2^k-ary symbols differing from x in at most the bits given by index set $B \subseteq [k]$.

For a 2^k-ary symbol x in which each bit has the same probability, γ, of being erased, the transition probability of the MEC is given by

$$\Pr(Y = y \mid X = x) = \begin{cases} 1 - \varepsilon & y = \{x\} \\ \gamma^j (1 - \gamma)^{k-j} & y = ?_x^B \text{ with } |B| = j, \end{cases}$$

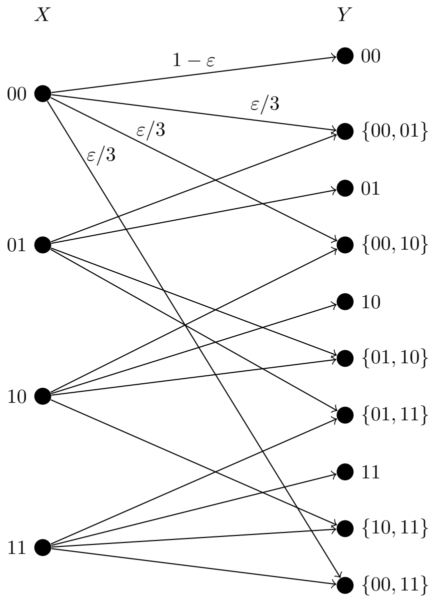

FIGURE 1. 4-ary partial erasure channel (QPEC) with $M = 2$. Here, the binary representation of each 4-ary symbol is shown.

for $j = 1, \ldots, k$, and where

$$\varepsilon = \sum_{j=1}^{k} \binom{k}{j} \left(\gamma^j (1-\gamma)^{k-j}\right) = 1 - (1-\gamma)^k.$$

Figure 2 is an example of the MEC for $q = 2^2$.

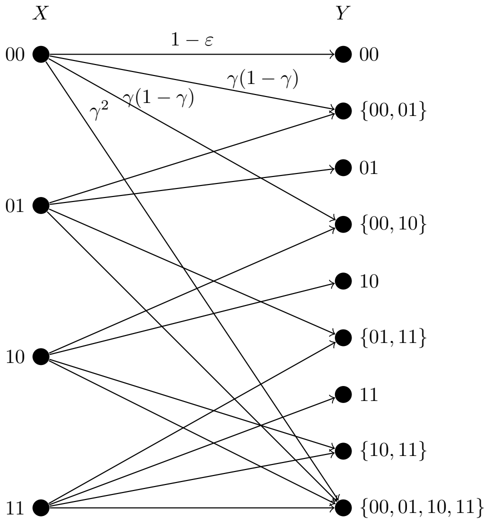

FIGURE 2. 4-ary multilevel erasure channel with erasure probability ε and bit error probability γ.

Example 1. For an 8-ary symbol in which each bit has the same probability, γ, of being erased, the transition probability of the MEC is given by

$$\Pr(Y = y \mid X = x) = \begin{cases} 1-\varepsilon & y = \{x\} \\ \gamma(1-\gamma)^2 & y = ?^B_x \text{ with } |B| = 1 \\ \gamma^2(1-\gamma) & y = ?^B_x \text{ with } |B| = 2 \\ \gamma^3 & y = ?^B_x \text{ with } |B| = 3, \end{cases}$$

where

$$\varepsilon = \sum_{j=1}^{3} \binom{3}{j} \left(\gamma^j (1-\gamma)^{3-j}\right) = 1 - (1-\gamma)^3.$$

For example, assume $x = (0,0,0)$ is transmitted corresponding to the symbol $0 \in \mathrm{GF}(8)$. Then the possible output sets and their transition probabilities are

$$\Pr(Y = y \mid X = x) = \begin{cases} (1-\gamma)^3 & y = \{0\} \\ \gamma(1-\gamma)^2 & y = \{0,1\}, \{0,\alpha\}, \\ & \quad \text{or } \{0,\alpha^2\} \\ \gamma^2(1-\gamma) & y = \{0,1,\alpha,1+\alpha\}, \\ & \{0,1,\alpha^2,1+\alpha^2\}, \\ & \quad \text{or } \{0,\alpha,\alpha^2,\alpha+\alpha^2\} \\ \gamma^3 & y = \mathrm{GF}(8) \end{cases} .$$

□

The MEC over $GF(2^k)$ that has the same erasure probability, γ, for each bit in the 2^k-ary symbol will be referred to as the *constrained* MEC. The general case of the MEC where bit i of the 2^k-ary symbol has erasure probability γ_i, for $i = 1, 2, \ldots, k$, will be referred to as the *unconstrained* MEC. We will now consider the unconstrained MEC case.

For $q = 2^k$, the transition probability of the unconstrained MEC is given by

$$\Pr(Y = y \mid X = x) = \begin{cases} 1-\varepsilon & y = \{x\} \\ \prod_{i\in B} \gamma_i \prod_{i \notin B} (1-\gamma_i) & y = ?^B_x, \\ & B \subseteq [k] \end{cases}$$

where

$$\varepsilon = \sum_{B \subseteq [k]} \left(\prod_{i \in B} \gamma_i \prod_{i \notin B} (1-\gamma_i) \right).$$

Example 2. Assume again that $x = (0,0,0)$ is transmitted corresponding to the symbol $0 \in \mathrm{GF}(8)$. Then the possible output sets and their transition probabilities are

$$\Pr(Y = y \mid X = x) =$$

$$\begin{cases} (1-\gamma_1)(1-\gamma_2)(1-\gamma_3) & y=\{0\} \\ \gamma_1(1-\gamma_2)(1-\gamma_3) & y=\{0,1\} \\ (1-\gamma_1)\gamma_2(1-\gamma_3) & y=\{0,\alpha\} \\ (1-\gamma_1)(1-\gamma_2)\gamma_3 & y=\{0,\alpha^2\} \\ \gamma_1\gamma_2(1-\gamma_3) & y=\{0,1,\alpha,1+\alpha\} \\ \gamma_1(1-\gamma_2)\gamma_3 & y=\{0,1,\alpha^2,1+\alpha^2\} \\ (1-\gamma_1)\gamma_2\gamma_3 & y=\{0,\alpha,\alpha^2,\alpha+\alpha^2\} \\ \gamma_1\gamma_2\gamma_3 & y=\mathrm{GF}(8) \end{cases}.$$

Remark 1. We introduce the multilevel erasure channel for binary extension fields for practical applications, but note that it can be extended in the obvious way to fields of order $q = p^k$ for any prime $p > 2$, or more general alphabets in which symbols can be represented as words using a subalphabet.

3.2. **Capacity.** Let C be a channel with input X, output Y, and let $p_x := Pr(X = x)$ be the input distribution to the channel where $x \in (\mathrm{GF}(2))^k$. By definition, the channel capacity, *cap*(C) is given by

$$cap(C) = \max_{\{p_x\}} I(X;Y) = \max_{\{p_x\}}(H(Y) - H(Y|X))$$

where $I(X;Y)$ is the mutual information between X and Y, $H(Y)$ is the entropy of Y, and $H(Y|X)$ is the conditional entropy of Y given X.

A channel is said to be *uniformly dispersive* if the multiset

$$\mathcal{A}(x) := \{P_{Y|X}(y_1 \mid x), \ldots, P_{Y|X}(y_t \mid x)\}$$

is identical for each input symbol x [18]. We first show that the constrained MEC has this property, which can be expected since the bit-erasure probabilities are symbol-independent.

Lemma 3.1. *The constrained MEC is uniformly dispersive.*

Proof. For any input x on the constrained MEC, there are $\binom{k}{j}$ possible outputs of size 2^j for each $j = 0, \ldots, k$. For each output y of size 2^j, $P_{Y|X}(y \mid x) = \gamma^j(1-\gamma)^{k-j}$. Moreover, an output set of size 2^j is obtained by choosing which bits to vary in $\binom{k}{j}$ ways and choosing the remaining bits in 2^{k-j} ways. Thus, there are a total of $\sum_{j=0}^{k}\binom{k}{j}2^{k-j} = 3^k$ possible channel outputs. Of these 3^k outputs, $3^k - \sum_{j=0}^{k}\binom{k}{j} = 3^k - 2^k$ have probability 0 of occurring given a fixed input x. Therefore,

$$\mathcal{A}(x) := \{\underbrace{0, \ldots, 0}_{3^k - 2^k \text{ times}}\} \cup \bigcup_{j=0}^{k} \{\underbrace{\gamma^j(1-\gamma)^{k-j}, \ldots, \gamma^j(1-\gamma)^{k-j}}_{\binom{k}{j} \text{ times}}\}$$

regardless of x, as desired. □

Since the constrained MEC is a uniformly dispersive channel, it is easy to show that for any input distribution,

$$H(Y|X) = -\sum_{j=0}^{k}\binom{k}{j}\gamma^j(1-\gamma)^{k-j}\log(\gamma^j(1-\gamma)^{k-j}).$$

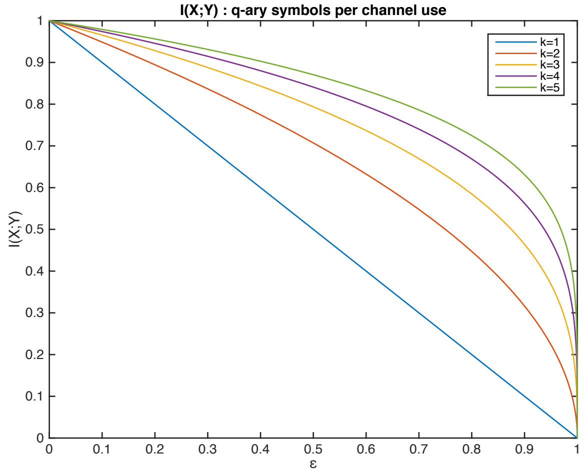

FIGURE 3. $I(X;Y)$ as a function of ε, given a uniform input distribution.

Thus, to determine capacity, it is enough to maximize $H(Y)$. For the uniform input distribution $\{p_x\}$ where $Pr(X=x)=1/q$ for each $x \in \mathrm{GF}(q)$, we have

$$H(Y) = -\sum_{j=0}^{k} \binom{k}{j} \gamma^j (1-\gamma)^{k-j} \log(2^{j-k}\gamma^j(1-\gamma)^{k-j}).$$

Given the fact that $1-\varepsilon = (1-\gamma)^k$, it follows that given a uniform input distribution and $q=2^k$,

$$I(X;Y) = k(1-\varepsilon)^{1/k}$$

measured in bits per channel use, or

$$I(X;Y) = k\log_q(2)(1-\varepsilon)^{1/k} = (1-\varepsilon)^{1/k} \cdot \log_q(2^k) = (1-\varepsilon)^{1/k}$$

measured in q-ary symbols per channel use.

Theorem 3.2. *The capacity on the constrained multilevel erasure channel is*

$$cap(C\text{-}MEC) = (1-\varepsilon)^{1/k}.$$

A proof of Theorem 3.2 is given in Section 5. Additionally, a proof using a numerical search for $k \le 2$ is included in the appendix. Observe that when $k=1$, the channel is simply the Binary Erasure Channel (BEC) with $q=2$ and capacity $1-\varepsilon$ as expected.

4. **Coding Techniques for the MEC.** In this section, we discuss two approaches for LDPC code design on the MEC. The first is based on a q-ary error correcting code and q-ary iterative decoder, and the second uses multilevel coding with multistage decoding.

In the first approach, an (N,K) q-ary LDPC code is used where the code symbols are transmitted across the MEC, and the receiver uses an iterative decoder on the received code symbols with partial erasures. Several iterative decoders have been proposed for q-ary LDPC codes, including an efficient FFT-based implementation, variations of the min-sum decoder, and others. In addition, Cohen and Cassuto

[3] give an iterative decoding algorithm for LDPC codes over GF(q) on the QPEC. Specifically, they incorporate the partial erasures into the standard sum-product algorithm.

Channel information on the MEC takes the form of subsets of GF(2^k) of cardinality 2^j, for some $j \in \{0, 1, \ldots, k\}$. If a variable node is assigned a subset of size one, no error has occured. In the QPEC iterative decoding algorithm, check node calculations involve taking Minkowski sums of subsets of GF(q), and variable node calculations are performed by taking the intersection of the incoming subsets [3]. Decoding succeeds when the intersection at each variable node is a set with cardinality one. Moreover, the codeword estimate that results when all variable node subsets have size one is guaranteed to be the original codeword. The decoder fails if the size of an erasure subset at one or more variable nodes never reduces to one. In the case of the QPEC, density evolution may be approximated by tracking the probability that message sets have cardinality $m = 1, 2, \ldots, q$ [3].

In the second approach, multilevel codes [14] may be used. In this paper, we focus on a detailed analysis of the multilevel coding with multistage decoding approach for the MEC, rather than analyzing iterative decoding of q-ary codes on the channel. Using this approach to obtain an overall (N, K) q-ary code for the MEC, we first consider the case where $q = 2^k$. (The case for $q = p^k$, for prime p, is a natural generalization of the $p = 2$ case.) Observe that N q-ary symbols can be represented using Nk bits and K q-ary symbols can be represented using Kk bits. Thus, the Kk information bits are subdivided into k groups with $\ell_1, \ell_2, \ldots, \ell_k$ information bits in each group, respectively.

The information bits within each group are encoded to a length N binary codeword using a component code that is specific for each group. In particular, group i uses a component code $\mathcal{C}_i$ that represents an (N, ℓ_i) binary code. Thus each group is encoded into N bits. The first encoded bits from each of the k groups are combined to form the first q-ary symbol for transmission on the MEC, where the bit from group i forms the i-th coordinate of the binary representation of the q-ary symbol. Similarly, the second encoded bit from each of the groups are combined to form the second q-ary symbol, and so on. Thus, a length N q-ary codeword is obtained via k binary codewords of length N. From the above, we have $Kk = \ell_1 + \cdots + \ell_k$, and the overall code rate of the multi-level code is $R = \frac{K}{N} = \frac{\ell_1+\ell_2+\cdots+\ell_k}{Nk} = \frac{R_1+R_2+\cdots+R_k}{k}$, where $R_i = \frac{\ell_i}{N}$ is the code rate of the i-th binary component code $\mathcal{C}_i$.

Recall that for $x \in$ GF(2^k), we will denote its binary representation by $x = (x_1, x_2, \ldots, x_k)$. Thus, in the above scheme, the code $\mathcal{C}_i$ represents the code over the i-th coordinate of the binary representation of the input alphabet, for $i = 1, 2, \ldots, k$. To obtain an information-theoretic perspective of the multilevel code and a corresponding multistage decoding scheme, let X_i be a random variable representing the i-th coordinate, for $i = 1, \ldots, k$. We now discuss techniques for designing these codes so that the capacity of the MEC is achieved.

First, observe that the random variable X representing the input to the channel has a one-to-one correspondence with the vector $(X_1, \ldots, X_k)$. Thus, the mutual information between the input X and the output (or, received word) Y is

$$I(X; Y) = I((X_1, X_2, \ldots, X_k); Y)$$

Using the chain rule of mutual information, we can write the following:

$$I((X_1, X_2, \ldots, X_k); Y) = I(X_1; Y) + I(X_2; Y|X_1)+$$

$$I(X_3;Y|X_1,X_2)+\cdots+I(X_k;Y|X_1,\ldots,X_{k-1}).$$

This can be interpreted as follows. The decoder first decodes Y using the code $\mathcal{C}_1$, and obtains an estimate for X_1. Using this knowledge and the received word, the decoder then decodes using $\mathcal{C}_2$ to obtain an estimate for X_2, etc. In this multistage decoding scheme, as the stages of decoding progress, the channels typically improve due to greater side information. Intuitively, it makes sense to assign the strongest code to the weakest channel (i.e., the channel with the largest erasure probability). The code $\mathcal{C}_1$ will have the lowest code rate R_1, and in general, the codes will be organized so that $R_1 \le R_2 \le \cdots \le R_k$ where R_i is the rate of code $\mathcal{C}_i$. (However, note that it is not always the case that $I(X_i;Y|X_1,\ldots,X_{i-1}) \ge I(X_j;Y|X_1,\ldots,X_{j-1})$ for $i \le j$, as these conditional mutual informations depend on the choice of the signal set in the channel.) To achieve capacity via multistage decoding, the code rates should be chosen so that $R_i = I(X_i;Y|X_1,\ldots,X_{i-1})$ for each $i = 1,\ldots,k$ [14, 15].

Remark 2. Density evolution may be performed on each component code of the multilevel scheme on the MEC, assuming multistage decoding.

The decoding complexity of the multilevel scheme is much smaller than that of the q-ary LDPC scheme. Typically, for a blocklength N q-ary LDPC code with average column weight d, the decoding complexity of an efficient message-passing implementation is in the order of $O(Ndq^2)$ [9]. The decoding complexity of a binary LDPC code of blocklength N and average column degree d is in the order of $O(Nd)$. The multilevel scheme with k binary LDPC component codes requires k binary decoders. Thus, the overall decoding complexity for such a scheme is $O(kNd)$ as opposed to a single q-ary LDPC coding scheme that requires $O(Nd2^{2k})$ operations per decoding iteration. Furthermore, in terms of practical implementation, if the same type of error-correction code (i.e. LDPC/BCH etc.) is used for each component code of the k parallel channels, then the same hardware may be used for each level and the hardware may be programmed with parameters appropriate for each code/level.

5. Channel decomposition of the MEC and QPEC.

In this section we consider multistage decoding of multilevel codes designed for the MEC and QPEC. We examine how the subchannels break down and determine cases when the subchannels are simple erasure channels.

5.1. Multistage decoding on the MEC.

We now calculate the effective erasure probabilities $\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_k$ for each subchannel assuming an unconstrained multilevel erasure channel C with erasure probability ε and bit error probabilities $\gamma_1,\gamma_2,\ldots,\gamma_k$.

Theorem 5.1. *For the MEC channel over* $\mathrm{GF}(2^k)$ *with erasure probability* ε *and bit erasure probabilities* $\gamma_1,\gamma_2,\ldots,\gamma_k$ *for* $X_1,\ldots,X_k$*, respectively,*

$$I(X_i;Y|X_1,\ldots,X_{i-1}) = I(X_i;Y),$$

for $i = 1,2,\ldots,k$ *and where* $X_0 = \{\}$*. That is, the MEC can be decomposed into* k *independent binary erasure channels (BECs), each with erasure probability* γ_i*, and the overall capacity of the MEC is* $I(X;Y) = \sum_{i=1}^{k} I(X_i;Y) = \sum_{i=1}^{k}(1-\gamma_i)$ *bits/channel-use.*

Proof. We show that the channels representing $I(X_i;Y)$ and $I(X_i;Y \mid X_1,\dots,X_{i-1})$ each have transition probability γ_i. To see this, consider sending a 2^k-ary symbol across the channel. The probability of receiving an output with uncertainty in the i-th bit is then

$$\gamma_i \sum_{B\subseteq[k]\setminus i}\left(\prod_{j\in B}\gamma_j \prod_{j\in([k]\setminus i)\setminus B}(1-\gamma_j)\right)=\gamma_i,$$

where the sum is over all subsets that do not include the i-th bit. Now suppose that a 2^k-ary symbol is sent across the channel and bits $1,\dots,i-1$ are known. The probability of receiving an output with uncertainty in the i-th bit is then

$$\gamma_i \sum_{B\subseteq[k]\setminus[i]}\left(\prod_{j\in B}\gamma_j \prod_{j\in([k]\setminus[i])\setminus B}(1-\gamma_j)\right)=\gamma_i,$$

where the sum is over all subsets excluding the first i positions. Therefore $I(X_i;Y) = I(X_i;Y \mid X_1,\dots,X_{i-1})$. □

Note that in the special case that γ is the same for each of the k bits, the same code may be applied to each bit or subchannel. We now use Theorem 5.1 to prove Theorem 3.2.

Proof of Theorem 3.2. *The capacity of the constrained multilevel erasure channel is*

$$cap(C\text{-}MEC) = (1-\varepsilon)^{1/k},$$

2^k*-ary symbols/channel use.*

Proof. Theorem 5.1 shows that $I(X;Y)=\sum_{i=1}^{k}(1-\gamma_i)$ bits/channel-use can be acheived using multi-level codes on the MEC. In the case of the constrained MEC, $\gamma_i=\gamma$ for all i, so

$$I(X;Y)=k-k\gamma=k(1-\gamma)$$

bits/channel use. Recall that $1-\gamma=(1-\varepsilon)^{1/k}$. Thus in symbols/channel use we have

$$I(X;Y)=(1-\gamma)=(1-\varepsilon)^{1/k}.$$

We now show that for any input distribution $\{p_x\}$ of X where $p_x = Pr(X=x)$, $I(X;Y)\le k-k\gamma$ in bits/channel use.

Write the MEC output Y as a vector $(b_1,\dots,b_k)$, where b_i takes the value of the i-th bit if it is known, and $b_i = ?$ otherwise. For example, the output set $\{00,01\}$ will be written as $(0,?)$. Next, define a random vector $E=(a_1,\dots,a_k)$ corresponding to Y as follows. Let $a_i=1$ (resp., $a_i=0$) if the i-th position of the bit representation of Y is erased (resp., not erased). Thus E may be regarded as a random variable representing an outcome of the MEC. Moreover, $H(Y) = H(Y,E) = H(E)+H(Y|E)$ by the chain rule of entropy, where the first equality follows from the fact that E is a function of Y.

Observe that

$$Pr[E=(a_1,\dots,a_k) \text{ and } a_j=1 \text{ for exactly } j \text{ indices}] = \gamma^j(1-\gamma)^{k-j}$$

So

$$H(E)=-\sum_{j=0}^{k}\binom{k}{j}\gamma^i(1-\gamma)^{k-j}\log(\gamma^j(1-\gamma)^{k-j}),$$

which can be shown to equal $\sum_x p_x H(Y|X=x) = H(Y|X)$.

Since $I(X;Y) = H(Y) - H(Y|X)$, we have

$$I(X;Y) = H(E) + H(Y|E) - H(Y|X) = H(E) + H(Y|E) - H(E) = H(Y|E).$$

Write $y = (b_1, \ldots, b_k)$ where b_i is 0, 1, or erased. Note that $Pr(Y = y|E = a) = 0$ if the positions in Y that are erased are not identical to the set $\{i | a_i = 1 \in a\}$.

We have

$$\begin{aligned} H(Y|E) &= \sum_{a \in \mathrm{GF}(2)^k} Pr(E = a) H(Y|E = a) \\ &= (1-\gamma)^k H(Y|E = a^0) + \sum_{a^1 \in \mathrm{GF}(2)^k} \gamma(1-\gamma)^{k-1} H(Y|E = a^1) \\ &+ \sum_{a^2 \in \mathrm{GF}(2)^k} \gamma^2 (1-\gamma)^{k-2} H(Y|E = a^2) + \ldots + \gamma^k H(Y|E = a^k), \end{aligned}$$

where a^i denotes a vector in $\mathrm{GF}(2)^k$ with weight i for $i = 0, \ldots, k$.

Observe that each $H(Y|E = a^j)$ is at most $k - j$ since j components of Y are erased and there are 2^{k-j} possible values of Y conditioned on $E = a^j$. So the maximum entropy for $H(Y|E = a^j)$ is at most $\log(2^{k-j}) = k - j$. Thus,

$$\begin{aligned} H(Y|E) &\leq k(1-\gamma)^k + (k-1)\binom{k}{1}\gamma(1-\gamma)^{k-1} + \cdots + (k-k)\binom{k}{k}\gamma^k \\ &= \sum_{j=0}^{k} (k-j)\binom{k}{j}\gamma^j(1-\gamma)^{k-j} \\ &= k(1-\gamma)\sum_{j=0}^{k-1}\binom{k-1}{j}\gamma^j(1-\gamma)^{k-1-j} = k(1-\gamma), \end{aligned}$$

where the last equality is from the binomial identity.

This proves that $I(X;Y) \leq k - k\gamma$. Since we have already shown that the mutual information $I(X;Y) = k - k\gamma$ is achievable using multi-level codes and multi-stage decoding, the capacity of the MEC is indeed $k - k\gamma$ bits/channel-use (or, $(1-\epsilon)^{1/k}$ symbols/channel-use).

□

Example 3. Recall the 4-ary MEC from Figure 2. By using the transition probabilities in Figure 2 and assuming equally likely inputs, the effective channel for X_1 is a binary erasure channel with erasure probability γ, shown in Figure 3. To see this, observe that if $X_1 = 0$, the outputs $\{00, 10\}, \{00, 01, 10, 11\}, \{01, 11\}$ from input symbols $\{00\}, \{01\}$ lead to uncertainty in the value of X_1 at the output. The probability of receiving these outputs is

$$\frac{1}{4}\gamma(1-\gamma) + \frac{2}{4}\gamma^2 + \frac{1}{4}\gamma(1-\gamma) = \frac{1}{2}\gamma.$$

Conditioning on $X_1 = 0$, the probability is $\frac{(\gamma/2)}{\frac{1}{2}} = \gamma$.

Without any side knowledge of X_1, the variable X_2 would see an effective channel that is a BEC with erasure probability γ.

In fact, the effective erasure probability for the BEC for X_2 conditioned on the value of X_1 is also γ, so X_2 can be decoded independent of X_1. The subchannel for X_2 assuming that $X_1 = 0$ is shown in Figure 5.

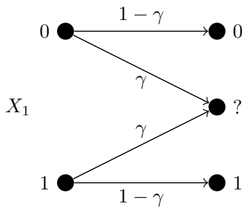

FIGURE 4. Subchannel for X_1 on 4-ary multilevel erasure channel with parameters ε, γ.

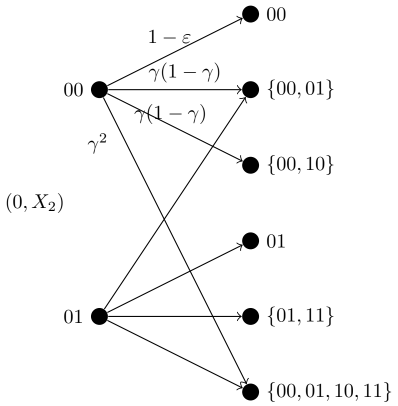

FIGURE 5. Subchannel for X_2 on 4-ary multilevel erasure channel with parameters ε, γ.

5.2. **Multistage decoding on the QPEC.** While it is perhaps not surprising in hindsight that the MEC can be decomposed into independent binary erasure channels and decoded bitwise in parallel, we show in this subsection that the QPEC over GF(2^k) may also be decomposed into subchannels and decoded via multistage decoding. When $M = 2$, these subchannels are binary erasure channels, and capacity may be achieved by optimizing the component codes with respect to each channel's capacity. While the subchannels are not independent, it is notable that codes over the 2^k-ary QPEC with $M = 2$ may be decoded using standard efficient decoders over the BECs, as an alternative to q-ary decoding over the QPEC. For other values of M and q, the QPEC decomposes into simpler channels but they are not necessarily p-ary erasure channels or BECs, for $q = p^k$ or $q = 2^k$, respectively.

We first calculate the effective erasure probabilities $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k$ for each subchannel assuming a 2^k-ary partial erasure channel C with erasure probability ε for different values of k and $M = 2$ when multistage decoding is performed.

Example 4. Consider the QPEC channel with partial erasure probability ε, $q = 4$, and $M = 2$. Recall the 4-ary QPEC from Figure 1. The mutual information $I(X;Y)$ for this channel can be written as $I(X_1, X_2; Y) = I(X_1; Y) + I(X_2; Y|X_1)$. By using the transition probabilities in Figure 1 and assuming equally likely inputs, the effective channel for X_1 is a binary erasure channel with erasure probability $\frac{2}{3}\varepsilon$, shown in Figure 6. To see this, observe that if $X_1 = 0$, the outputs $\{00, 10\}, \{01, 10\}, \{01, 11\}$ and $\{00, 11\}$ from input symbols $\{00\}, \{01\}$ lead to uncertainty in the value of X_1

at the output. The probability of receiving these outputs is $4(\frac{\varepsilon}{3})$. Conditioning on $X_1 = 0$, the erasure probability is $\varepsilon_1 = \frac{1}{2}(4(\frac{\varepsilon}{3})) = \frac{2}{3}\varepsilon$.

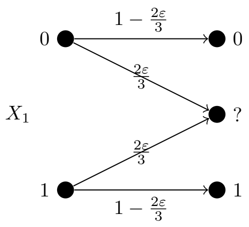

Figure 6. Subchannel for X_1 on 4-ary partial erasure channel (QPEC) with erasure probability ε.

Now suppose without loss of generality that the bit X_1 is known to equal 0. Figure 7 shows the 4-ary subchannel seen by bit X_2. We can observe that X_2 conditioned on X_1 is a BEC with probability of erasure derived using the outputs in Figure 7 that create an uncertainty on the value of X_2. Assuming $X_2 = 0$ for example, only the output $\{00, 01\}$ leads to uncertainty. The probability this output occurs when $X_2 = 0$ is just $\frac{\varepsilon}{3}$. Thus, the effective erasure probability for the binary erasure channel seen by X_2 conditioned on X_1 is $\frac{\varepsilon}{3}$. Observe that the sum of the mutual information for the subchannels equals the capacity of the QPEC.

□

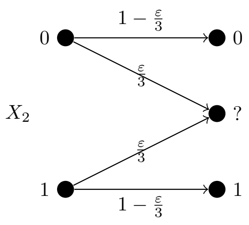

Figure 7. Subchannel for X_2 on 4-ary partial erasure channel (QPEC) with erasure probability ε.

In general, we have the following result.

Theorem 5.2. *For the QPEC with erasure probability ε, $q = 2^k$ and $M = 2$, the mutual information for the subchannels with multistage decoding is given by*

$$I(X_i; Y|X_1, X_2, \ldots, X_{i-1}) = 1 - \varepsilon_i,$$

where $\varepsilon_i = \frac{2^{k-i}}{2^k-1}\varepsilon$ for $i = 1, \ldots, k$. Thus, each subchannel is a binary erasure channel with erasure probability ε_i. Moreover, multistage decoding of binary codes optimized for each subchannel achieves the capacity on the QPEC.

Proof. Note that for the QPEC with $q = 2^k$ and $M = 2$, the probability of any specified partial erasure set is $\frac{\varepsilon}{2^k-1}$. For any given input symbol, there are a total of 2^{k-i} possible erasure sets with bits $1, \ldots, i-1$ known and uncertainty in the i-th bit. To see this, note that once $i-1$ bits are fixed, there are $k-1-(i-1) = k-i$ bits other than the (erased) i-th bit for which to choose values.

The probability of receiving an output with uncertainty in the i-th bit when bits $1, \ldots, i-1$ are known is thus

$$\varepsilon_i := \frac{2^{k-i}}{2^k - 1}\varepsilon$$

for each $i = 1, \ldots, k$. Thus $I(X_i; Y|X_1, X_2, \ldots, X_{i-1}) = 1 - \varepsilon_i$.

To see that capacity on the QPEC is achieved, note that

$$\sum_{i=1}^{k}(1-\varepsilon_i) = k - \sum_{i=1}^{k}\frac{2^{k-i}\varepsilon}{2^k-1} = k - \varepsilon = k\left(1 - \frac{\varepsilon}{k}\right).$$

which is $\left(1 - \frac{\varepsilon}{k}\right) = (1 - \varepsilon \log_q(M))$ 2^k-ary symbols per channel use.

□

The decomposition of the QPEC into simple erasure subchannels does not happen for general $p, M > 2$. However, the next theorem shows that when $q = p^k$, the k-th subchannel does have a simple mutual information expression. When $p = 2$, this coincides with the capacity of a BEC with transition probability in terms of ε.

Theorem 5.3. *For the QPEC with erasure probability ε, $q = p^k$ and any M, the last (k-th) subchannel has mutual information*

$$I(X_k; Y \mid X_1, \ldots, X_{k-1}) = (1-\varepsilon) + \frac{\varepsilon}{\binom{q-1}{M-1}} \sum_{t=0}^{\min\{M-1, p-1\}} \binom{q-p}{M-t-1}\binom{p-1}{t}\log_p\left(\frac{p}{t+1}\right).$$

A proof of Theorem 5.3 is given in the appendix.

Corollary 1. *For the QPEC with erasure probability ε, $q = 2^k$ and $M > 2$, multistage decoding may not decompose entirely into simple BECs. However, the last (k-th) subchannel is a simple BEC with transition probability*

$$\varepsilon_k = \left(\frac{M-1}{2^k-1}\right)\varepsilon.$$

Example 5. For the QPEC with erasure probability ε, $q = 2^2$ and $M = 3$,

$$I(X_1; Y) = 1 + \frac{2}{3}\varepsilon - \varepsilon \log_2 3 \text{ bits per channel use},$$

and

$$I(X_2; Y \mid X_1) = 1 - \frac{2}{3}\varepsilon \text{ bits per channel use}.$$

Observe that the sum is $2 - \varepsilon \log_2 3$ bits per channel use. This is equal to $(1 - \varepsilon \log_4(3)) = (1 - \varepsilon \log_q(M))$ symbols per channel use, which is the capacity of the QPEC. Thus, if binary codes are used for X_1 and X_2 that have these rates, the capacity of the QPEC is achieved.

When q is a perfect square, the mutual information of the subchannels may also be calculated in terms of ε and p, as Theorem 5.4 demonstrates. Moreover, one may observe that the subchannels are not simple erasure channels.

Theorem 5.4. *For the QPEC with ε, $q = p^2$ and any $M \geq 2$, the subchannel mutual information is given by:*

$$I(X_1;Y) = (1-\varepsilon)+$$

$$\frac{\varepsilon}{\binom{p^2-1}{M-1}} \sum_{t=0}^{\min\{p-1,M-1\}} \binom{p-1}{t}\binom{p^2-p}{M-1-t} \log_p\left(\frac{p(t+1)}{M}\right)$$

$$I(X_2;Y|X_1) = (1-\varepsilon)+$$

$$\frac{\varepsilon}{\binom{p^2-1}{M-1}} \sum_{t=0}^{\min\{p-1,M-1\}} \binom{p-1}{t}\binom{p^2-p}{M-1-t} \log_p\left(\frac{p}{t+1}\right).$$

A proof of Theorem 5.4 is given in the Appendix.

6. **Conclusion.**

We introduced the multilevel erasure channel to model partial erasures that may occur in modern storage applications and derived the capacity of this channel. We also showed that multilevel coding may be applied on the MEC and the QPEC, and that such codes may be decoded using multistage decoding, allowing them to be decomposed into simpler binary or p-ary component codes. In addition, by choosing the rates of the component codes appropriately, the capacity of the MEC or QPEC can be achieved using low complexity decoders. We are currently analyzing the subchannels in the decomposition of the QPEC which are not simple erasure channels. In particular, decoding on these channels, while easier than on the full channel, may still involve subset sums. Finally, we comment that by considering the p-ary representation of a q-ary symbol, the results for the MEC may also be extended to $q = p^k$ for $p > 2$.

REFERENCES

[1] Y. Cai, E. F. Haratsch, O. Mutlu, and K. Mai, Error patterns in MLC NAND flash memory: Measurement, characterization, and analysis, *Design, Automation and Test in Europe Conference Exhibition (DATE)*, March 2012.

[2] Z. Zhang, W. Xiao, N. Park, and D. J. Lilja, Memory module-level testing and error behaviors for phase change memory, *IEEE 30th Int. Conf. on Computer Design (ICCD)*, Oct. 2012.

[3] R. Cohen and Y. Cassuto, Iterative Decoding of LDPC Codes Over the q -Ary Partial Erasure Channel, *IEEE Transactions on Information Theory*, **62** (2016), 2658–2672.

[4] R. Gabrys, E. Yaakobi, L. Grupp, S. Swanson, and L. Dolecek, Tackling Intracell Variability in TLC Flash Through Tensor Product Codes, *Proc. of IEEE International Symposium on Information Theory*, Cambridge, MA, July 2012.

[5] R. Gabrys, E. Yaakobi, and L. Dolecek, Graded Bit-Error-Correcting Codes With Applications to Flash Memory, *IEEE Transactions on Information Theory*, **59** (2013), 2315–2327.

[6] K. Haymaker and C. A. Kelley, Structured Bit-Interleaved LDPC Codes for MLC Flash Memory, *IEEE Journal on Selected Areas in Communications*, **32** (2014), 870–879.

[7] S. Borade, B. Nakiboğlu, and L. Zheng, Unequal Error Protection: An Information-Theoretic Perspective, *IEEE Transactions on Information Theory*, **55** (2009), 5511–5539.

[8] A. R. Calderbank and N. Seshadri, Multilevel codes for unequal error protection, *IEEE Transactions on Information Theory*, **39** (1993), 1234–1248.

[9] D. Declercq and M. Fossorier, Decoding algorithms for nonbinary LDPC codes over GF(q), *IEEE Transactions on Communications*, **55** (2007), 633–643.

[10] R. G. Gallager, *Low Density Parity Check Codes*, MIT Press, 1963.

[11] R. M. Tanner, A recursive approach to low complexity codes, *IEEE Transactions on Information Theory*, **27** (1981), 533–547.

[12] T. Richardson, A. Shokrollahi, and R. Urbanke, Design of capacity-approaching irregular low-density parity-check codes, *IEEE Transactions on Information Theory*, **47** (2001), 619–637.

[13] T. Tao and V. H. Vu, *Additive Combinatorics*, Cambridge University Press, 2006.

[14] H. Imai and S. Hirakawa, A new multilevel coding method using error-correcting codes, *IEEE Transactions on Information Theory*, **23** (1977), 371–377.

[15] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, Multilevel codes: Theoretical concepts and practical design rules, *IEEE Transactions on Information Theory*, **45** (1999), 1361–1391.

[16] J. Huber, U. Wachsmann, and R. Fischer, Coded modulation by multilevel-codes: Overview and state of the art, in *ITG-Fachberichte Conf. Rec.*, Aachen, Germany, March 1998.

[17] K. A. S. Abdel-Ghaffar and M. Hassner, Multilevel error-control codes for data storage channels, *IEEE Transactions on Information Theory*, **37** (1991), 735–741.

[18] M. Moser and P. Chen, *A Student's Guide to Coding and Information Theory*, Cambridge: Cambridge UP, 2012.

[19] T. J. Richardson and R. L. Urbanke, The capacity of low-density parity-check codes under message passing decoding, *IEEE Transactions on Information Theory*, **47** (2001), 599–618.

7. Appendix.

7.1. Alternative proof of Theorem 3.2 for $k \leq 2$.

The capacity on the constrained multilevel erasure channel is

$$cap(C\text{-}MEC) = (1-\varepsilon)^{1/k}$$

when $k = 1, 2$.

Proof. Observe that when $k = 1$, the channel is simply the Binary Erasure Channel (BEC) with $q = 2$ and capacity $1 - \varepsilon$ as expected.

Let $P(\cdot)$ be a capacity achieving input distribution for the constrained MEC. We show the proof for $k = 2$.

For each input $x_i \in \mathrm{GF}(4)$, let Y_i be the set of possible outputs. Then

$$I(x_i; Y) = \sum_{y \in Y_i} P(y \mid x_i) \log\left(\frac{P(y \mid x_i)}{P(y)}\right)$$

$$= (1-\varepsilon)\log(1-\varepsilon) + 2(1-\gamma)\gamma\log((1-\gamma)\gamma)$$

$$+\gamma^2\log(\gamma^2) - \sum_{y \in Y_i}\left(P(y \mid x_i)\log(P(y))\right).$$

For $x_i = 0$,

$$\sum_{y \in Y_0}\left(P(y \mid 0)\log(P(y))\right) = (1-\varepsilon)\log(P(0))$$

$$+2\gamma(1-\gamma)\log(\gamma(1-\gamma)) + \gamma(1-\gamma)\log(P(0)+P(1))$$

$$+\gamma(1-\gamma)\log(P(0)+P(\alpha)) + \gamma^2\log(P(\{0,1,\alpha,\alpha+1\}))$$

and the sum is similar for each $x_i \in \mathrm{GF}(4)$. By the Karush-Kuhn-Tucker (KKT) conditions [18], a capacity achieving input distribution for a channel with capacity C must satisfy $I(x;Y) = C$ for all x with $Pr(x) > 0$. If each $\sum_{y \in Y_i}\left(P(y \mid x_i)\log(P(y))\right)$

is equal, then we have the following system of equations.

$$\begin{aligned}
&\log(P(0)) + \frac{\gamma}{(1-\gamma)}\log(P(0)+P(\alpha))\\
&\qquad = \log(P(1)) + \frac{\gamma}{(1-\gamma)}\log(P(1)+P(\alpha+1))\\
&\log(P(0)) + \frac{\gamma}{(1-\gamma)}\log(P(0)+P(1))\\
&\qquad = \log(P(\alpha))) + \frac{\gamma}{(1-\gamma)}\log(P(\alpha)+P(\alpha+1))\\
&\log(P(1)) + \frac{\gamma}{(1-\gamma)}\log(P(0)+P(1))\\
&\qquad = \log(P(\alpha+1)) + \frac{\gamma}{(1-\gamma)}\log(P(\alpha)+P(\alpha+1))\\
&\log(P(\alpha)) + \frac{\gamma}{(1-\gamma)}\log(P(0)+P(\alpha))\\
&\qquad = \log(P(\alpha+1)) + \frac{\gamma}{(1-\gamma)}\log(P(1)+P(\alpha+1)).
\end{aligned}$$

Assuming each input symbol occurs with positive probability, we have verified through a numerical search that the uniform input distribution is the unique distribution satisfying the system of equations. Therefore for $k = 2$, $cap(\text{MEC}) = k\log_{2^k}(2)(1-\varepsilon)^{1/k}$ as claimed.

For $q = 2^k$, it can be shown that with the uniform input distribution, $I(x_i; Y) = I(x_j; Y)$ for each $x_i, x_j \in \text{GF}(2^k)$. □

7.2. **Proof of Theorem 5.3.** *For the QPEC with erasure probability ε, $q = p^k$ and any M, the last (k-th) subchannel has mutual information*

$$I(X_k; Y \mid X_1, \ldots, X_{k-1}) = (1-\varepsilon) + \frac{\varepsilon}{\binom{q-1}{M-1}} \sum_{t=0}^{\min\{M-1,p-1\}} \binom{q-p}{M-t-1}\binom{p-1}{t}\log_p\left(\frac{p}{t+1}\right).$$

Proof. Define the symbol $?_{\mathbf{0}}$ to be the set of all M-sets containing the all-zeros vector.

$$\begin{aligned}
&I(X_k; Y \mid X_1, \ldots, X_{k-1})\\
&\quad = \sum_y P(y \mid 0, \ldots, 0)\log_p\left(\frac{p \cdot P(y \mid 0, \ldots, 0)}{\sum_{x_k'} P(y \mid 0, \ldots, 0, x_k')}\right),
\end{aligned}$$

where the above equality exploits the symmetry across the values of $X_1, \ldots, X_k$ in the mutual information calculation.

$$\begin{aligned}
&I(X_k; Y \mid X_1, \ldots, X_{k-1})\\
&\quad =(1-\varepsilon) + \frac{\varepsilon}{\binom{q-1}{M-1}} \sum_{y \in ?_{\mathbf{0}}} \log_p\left(\frac{p \cdot \frac{\varepsilon}{\binom{q-1}{M-1}}}{\sum_{x_k'} P(y \mid 0, \ldots, 0, x_k')}\right)
\end{aligned}$$

Note that there are $\binom{q-1-(p-1)}{M-1-t}\binom{p-1}{t}$ erasure sets of $\mathbf{0}$ containing exactly t symbols of the form $0\cdots 0x_k'$, with $x_k' \in \{1,\dots,p-1\}$. There are $t+1$ input symbols of the form $0\cdots 0\tilde{x_k}$ with $\tilde{x}_k \in \{0,\dots,p-1\}$ for which each such erasure set is possible. Therefore

$$\begin{aligned}I(X_k;Y \mid X_1,\dots,X_{k-1}) =&(1-\varepsilon)+\\ &\frac{\varepsilon}{\binom{q-1}{M-1}}\sum_{t=0}^{\min\{M-1,p-1\}}\binom{q-p}{M-t-1}\binom{p-1}{t}\log_p\left(\frac{p\cdot\frac{\varepsilon}{\binom{q-1}{M-1}}}{(t+1)\frac{\varepsilon}{\binom{q-1}{M-1}}}\right)\\ =&(1-\varepsilon)+\\ &\frac{\varepsilon}{\binom{q-1}{M-1}}\sum_{t=0}^{\min\{M-1,p-1\}}\binom{q-p}{M-t-1}\binom{p-1}{t}\log_p\left(\frac{p}{t+1}\right).\end{aligned}$$

□

7.3. **Proof of Theorem 5.4.** *For the QPEC with ε, $q=p^2$ and any $M\geq 2$, the subchannel mutual information is given by:*

$$\begin{aligned}I(X_1;Y) &= (1-\varepsilon)+\\ &\frac{\varepsilon}{\binom{p^2-1}{M-1}}\sum_{t=0}^{\min\{p-1,M-1\}}\binom{p-1}{t}\binom{p^2-p}{M-1-t}\log_p\left(\frac{p(t+1)}{M}\right)\end{aligned}$$

$$\begin{aligned}I(X_2;Y|X_1) &= (1-\varepsilon)+\\ &\frac{\varepsilon}{\binom{p^2-1}{M-1}}\sum_{t=0}^{\min\{p-1,M-1\}}\binom{p-1}{t}\binom{p^2-p}{M-1-t}\log_p\left(\frac{p}{t+1}\right).\end{aligned}$$

Proof.

$$\begin{aligned}&I(X_1;Y)\\ &=\sum_{x_1,x_2}\sum_y P(y|x_1,x_2)P(x_1)P(x_2)\log_2\left(\frac{\sum_{x_2'}P(y|x_1,x_2')P(x_2')}{\sum_{x_1'x_2'}P(y|x_1'x_2')P(x_1')P(x_2')}\right)\\ &=\sum_y P(y|0,0)\log_p\left(\frac{p\sum_{x_2'}P(y|0,x_2')}{\sum_{x_1'x_2'}P(y|x_1'x_2')}\right),\end{aligned}$$

where the above equality exploits the symmetry across the values of X_1, X_2 in the mutual information calculation. Thus

$$\begin{aligned}
&I(X_1;Y)\\
=&(1-\varepsilon)\log_p\left(\frac{p(1-\varepsilon)}{(1-\varepsilon)}\right)+\sum_{y\in ?_{\mathbf{0}}}\frac{\varepsilon}{\binom{p^2-1}{M-1}}\log_p\left(\frac{p\sum_{x_2'}P(y|0,x_2')}{\sum_{x_1'x_2'}P(y|x_1'x_2')}\right)\\
=&(1-\varepsilon)+\frac{\varepsilon}{\binom{p^2-1}{M-1}}\sum_{y\in ?_{\mathbf{0}}}\log_p\left(\frac{p\sum_{x_2'}P(y|0,x_2')}{\sum_{x_1'x_2'}P(y|x_1'x_2')}\right)\\
&= (1-\varepsilon)+\frac{\varepsilon}{\binom{p^2-1}{M-1}}\sum_{t=0}^{\min\{p-1,M-1\}}\binom{p^{2-1}-1}{t}\binom{p^2-p^{2-1}}{M-1-t}\log_p\left(\frac{p\frac{t+1}{\binom{p^2-1}{M-1}}\varepsilon}{M\frac{\varepsilon}{\binom{p^2-1}{M-1}}}\right)\\
=&(1-\varepsilon)+\frac{\varepsilon}{\binom{p^2-1}{M-1}}\sum_{t=0}^{\min\{p-1,M-1\}}\binom{p-1}{t}\binom{p^2-p}{M-1-t}\log_p\left(\frac{p(t+1)}{M}\right)
\end{aligned}$$

and

$$\begin{aligned}
&I(X_2;Y|X_1)\\
&=\sum_{x_1,x_2}\sum_y P(y|x_1,x_2)P(x_1)P(x_2)\log_p\left(\frac{P(y|x_1,x_2)}{\sum_{x_2'}P(y|x_1x_2')P(x_2')}\right)\\
&=\sum_y P(y|0,0)\log_p\left(\frac{pP(y\mid 0,0)}{\sum_{x_2'}P(y|0x_2')}\right)\\
&=(1-\varepsilon)\log_p\left(\frac{p(1-\varepsilon)}{(1-\varepsilon)}\right)+\sum_{y\in ?_{\mathbf{0}}}\frac{\varepsilon}{\binom{p^2-1}{M-1}}\log_p\left(\frac{\frac{p\varepsilon}{\binom{p^2-1}{M-1}}}{\sum_{x_2'}P(y|0x_2')}\right)\\
&=(1-\varepsilon)+\frac{\varepsilon}{\binom{p^2-1}{M-1}}\sum_{t=0}^{\min\{p-1,M-1\}}\binom{p-1}{t}\binom{p^2-p}{M-1-t}\log_p\left(\frac{\frac{p\varepsilon}{\binom{p^2-1}{M-1}}}{(t+1)\frac{\varepsilon}{\binom{p^2-1}{M-1}}}\right)\\
&=(1-\varepsilon)+\frac{\varepsilon}{\binom{p^2-1}{M-1}}\sum_{t=0}^{\min\{p-1,M-1\}}\binom{p-1}{t}\binom{p^2-p}{M-1-t}\log_p\left(\frac{p}{t+1}\right).
\end{aligned}$$

□

Received xxxx 20xx; revised xxxx 20xx.

E-mail address: cmayer@huskers.unl.edu
E-mail address: kathryn.haymaker@villanova.edu
E-mail address: ckelley2@unl.edu