Optimal Coding for the Erasure Channel with Arbitrary Alphabet Size

Shervan Fashandi, Shahab Oveis Gharan and Amir K. Khandani
ECE Dept., University of Waterloo, Waterloo, ON, Canada, N2L3G1
email: {sfashand,shahab,khandani}@cst.uwaterloo.ca

Abstract

An erasure channel with an arbitrary alphabet size q is studied. It is proved that over any erasure channel (with or without memory), *Maximum Distance Separable* (MDS) codes achieve the minimum probability of error (assuming maximum likelihood decoding). Next, based on the performance of MDS codes, a lower-bound on the error probability of block codes over the erasure channel is derived. It is shown that this lower-bound (denoted by $\mathrm{L}_m(N, K, q)$) is valid whether or not an MDS code of size $[N, K]$ exists. For the case of a memoryless erasure channel with any arbitrary alphabet size, the exponential behavior of the lower-bound is studied. Finally, it is proved that both random codes and random linear codes have the same exponent as $\mathrm{L}_m(N, K, q)$ for the memoryless erasure channel of any arbitrary alphabet size and rates above the *critical rate*. In other words, considering rates above the critical rate, both random codes and random linear codes are exponentially optimal over the memoryless erasure channel[1].

I. Introduction

A. Motivation

Erasure channels, especially the ones with large alphabet sizes, have recently received significant attention in networking applications. Different erasure channel models (with strong memory) are adopted to study the performance of end-to-end connections over the Internet [1], [2]. In such models, each packet is seen as a $q = 2^b$-ary symbol where b is the packet length in bits. In this work, it is proved that over any erasure channel (with or without memory), *Maximum Distance Separable* (MDS) codes achieve the minimum probability of error. Moreover, a memoryless erasure channel with an arbitrarily large alphabet size is considered. The error probability over this channel (assuming maximum-likelihood decoding) for

[1]Financial support provided by Nortel and the corresponding matching funds by the Natural Sciences and Engineering Research Council of Canada (NSERC), and Ontario Centres of Excellence (OCE) are gratefully acknowledged.

MDS, random, and random linear codes are compared and shown to be exponentially identical for any alphabet size and any rate above the critical rate[2].

B. Related Works

Shannon [4] was the first who observed that the error probability for maximum likelihood decoding of a random code ($P_{E,ML}^{rand}$) can be upper-bounded by an exponentially decaying function with respect to the code block length N. This exponent is positive as long as the rate stays below the channel capacity, $R < C$. Following this result, tighter bounds were proposed in the literature [5]–[7]. For rates below the critical rate, modifications of random coding are proposed to achieve tighter bounds [8]. Interestingly, the exponential upper-bound on $P_{E,ML}^{rand}$ remains valid regardless of the alphabet size q, even in the case where q is larger than the block size N (e.g. see the steps of the proofs in [7]). There is also a lower-bound on the probability of error for any codebook which is known as the *sphere packing* bound [9]. For channels with a relatively small alphabet size ($q \ll N$), both the sphere packing lower-bound and the random coding upper-bound on the error probability are exponentially tight for rates above the critical rate [10]. However, the sphere packing bound is not tight if the alphabet size, q, is comparable to the coding block length N (noting the terms $o_1(N)$ and $o_2(N)$ in [9]). Hence, the exponential optimality of random codes is not necessarily valid for the case of q comparable with or larger than N.

Minimum distance, distance distribution, and error probability of random linear codes over the binary symmetric channel are discussed in [11], [12]. Pierce studies the asymptotic behavior of the minimum distance of binary random linear codes [11]. Error exponent of random linear codes over a binary symmetric channel is analyzed in [12]. Barg et al. also study the minimum distance and distance distribution of random linear codes and show that random linear codes have a better expurgated error exponent as compared to random codes for rates below the critical rate [12].

Maximum Distance Separable (MDS) [13] codes are optimum in the sense that they achieve the largest possible minimum distance, d_{min}, among all block codes of the same size. Indeed, any codeword in an MDS code of size $[N, K]$ can be successfully decoded from any subset of its coded symbols of size K or more. This property makes MDS codes suitable for use over erasure channels like the Internet [1], [2], [14]. However, the practical encoding-decoding algorithms for such codes have quadratic time complexity in terms of the code block length [15]. Theoretically, more efficient ($O\left(N \log^2 N\right)$) MDS codes can be constructed based on evaluating and interpolating polynomials over specially chosen finite fields using Discrete Fourier Transform [16]. However, in practice, these methods can not compete with the quadratic methods except for extremely large block sizes. Recently, a family of almost-MDS codes with low

[2]This is an extended version of the conference paper published in ISIT 2008 [3].

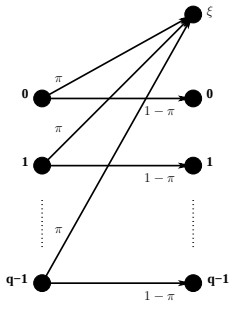


Fig. 1. Erasure memoryless channel model with the alphabet size q, probability of erasure π, and the erasure symbol ξ.

encoding-decoding complexity (linear in length) is proposed and shown to provide a practical alternative for coding over the erasure channels like the Internet [17]. In these codes, any subset of symbols of size $K(1+\epsilon)$ is sufficient to recover the original K symbols with high probability [17]. Fountain codes, based on the idea of almost-MDS codes with linear decoding complexity, are proposed for information multicasting to many users over an erasure channel [18], [19].

C. Contribution and Organization

In this work, an erasure channel with an arbitrarily large alphabet size, q, is studied. First, it is proved that MDS block codes offer the minimum probability of decoding error over any erasure channel (with or without memory). Then, based on the performance of MDS codes, a lower-bound on the error probability of block codes over the erasure channel is derived. This lower-bound, denoted by $\mathrm{L}_m(N,K,q)$, is valid whether or not an MDS code of size $[N,K]$ exists. For the case of a memoryless erasure channel, the exponential behavior of $\mathrm{L}_m(N,K,q)$ is studied. Finally, it is proved that both random codes and random linear codes have the same exponent as $\mathrm{L}_m(N,K,q)$ for the memoryless erasure channel of any arbitrary alphabet size as long as the rate is above the critical rate. More precisely, we prove that for rates above the critical rate, the error probability of the random codes and random linear codes can be upper-bounded as $O\left(N\mathrm{L}_m(N,K,q)\right)$ and $O\left(N^3\mathrm{L}_m(N,K,q)\right)$, respectively[3]. Interestingly, this result is valid whether or not q is comparable with N. Thus, our result is different from the known result on the exponential optimality of random codes which is based on the sphere packing bound [10].

The rest of this paper is organized as follows. In section II, the erasure channel model is introduced. Section III proves the optimality of MDS codes over any erasure channel. Error exponents of MDS codes, random codes, and random linear codes over a memoryless erasure channel are compared in section IV. Finally, section V concludes the paper.

[3]It should be noted that the normalized critical rate scales as $O(\frac{1}{q})$ to zero for large values of q.

II. Erasure Channel Model

The memoryless erasure channel studied in this work has the alphabet size q and the erasure probability π (see Fig. 1). The alphabet size q is assumed to be arbitrarily large, i.e., $q \gg 1$.

The described channel model occurs in many practical scenarios such as the Internet. From an end to end protocol's perspective, performance of the lower layers in the protocol stack can be modeled as a random *channel* called an *Internet channel*. Since each packet usually includes an internal error detection mechanism (for instance a Cyclic Redundancy Check), the Internet channel can be modeled as an erasure channel with packets as symbols [20]. If each packet contains b bits, the corresponding channel will have an alphabet size of $q = 2^b$ which is huge for typical packet sizes. Therefore, in practical networking applications, the block size is usually much smaller than the alphabet size. Algebraic computations over Galois fields $\mathbb{F}_q$ of such large cardinalities is now practically feasible with the increasing processing power of electronic circuits. Note that network coding schemes, recently proposed and applied for content distribution over large networks, have a higher computational complexity as they require the inverse matrix computation for square matrices of size $O(N)$ [21]–[27].

Note that all of the known MDS codes have alphabets of a large size (growing at least linearly with the block length N). Indeed, to have a feasible MDS code over a channel with the alphabet size q, the block size N should satisfy $N \leq q + 1$ in almost all cases (see Remark I in section III and [28]).

III. Optimality of MDS Codes over Erasure Channels

Maximum Distance Separable (MDS) codes are optimum in the sense of achieving the largest possible minimum distance, d_{min}, among all block codes of the same size [13]. The following Theorem shows that MDS codes are also optimum over any erasure channel in the sense of achieving the minimum probability of error. It is worth mentioning that practical erasure channels (like the Internet channel) typically demonstrate a strong dependency between the erasure events. However, in such scenarios, the erasure events are independent of the input symbols. The following definition includes the most general case of such erasure channels, with or without memory. Obviously, the memoryless and Markovian-modeled [1], [2] erasure channels are special cases of the following definition.

Definition I. An erasure channel is defined as the one which maps every input symbol to either itself or to an erasure symbol ξ. More accurately, an arbitrary channel (memoryless or with memory) with the input vector $\mathbf{x} \in \mathcal{X}^N$, $|\mathcal{X}| = q$, the output vector $\mathbf{y} \in (\mathcal{X} \cup \{\xi\})^N$, and the transition probability $p(\mathbf{y}|\mathbf{x})$ is defined to be erasure *iff* it satisfies the following conditions:

1) $p(y_j \notin \{x_j, \xi\} | x_j) = 0, \ \forall \ j$, where x_j and y_j denote the j'th elements of the vectors $\mathbf{x}$ and $\mathbf{y}$.

TABLE I

A TERNARY CODEBOOK WITH $N = 5$ AND $K = 3$, AND $d = 2$.

Information Symbols	Codewords
000	00000
001	00110
002	00220
010	01011
011	01121
012	01201
020	02022
021	02102
022	02212
100	10011
101	10121
102	10201
110	11022
111	11102
112	11212
120	12000
121	12110
122	12220
200	20022
201	20102
202	20212
210	21000
211	21110
212	21220
220	22011
221	22121
222	22201

2) Defining the erasure identifier vector $\mathbf{e}$ as

$$e_j \triangleq \begin{cases} 1 & y_j = \xi \\ 0 & \text{otherwise} \end{cases}$$

$p(\mathbf{e}|\mathbf{x}) = p(\mathbf{e})$, i.e. $\mathbf{e}$ is independent of $\mathbf{x}$.

Theorem I. A block code of size $[N, K]$ with equiprobable codewords over an arbitrary erasure channel (memoryless or with memory) has the minimum probability of error (assuming optimum, i.e., maximum likelihood decoding) among all block codes of the same size *if* that code is *Maximum Distance Separable* (MDS).

Proof. Consider any arbitrary $[N, K, d]$ codebook $\mathcal{C}$ with the q-ary codewords of length N, number of code-words q^K, and minimum distance d. The distance between two codewords is defined as the number of positions in which the corresponding symbols are different (Hamming distance). Assume a codeword $\mathbf{x} \in \mathcal{C}$ is transmitted and a vector $\mathbf{y} \in (\mathcal{X} \cup \{\xi\})^N$ is received. According to the definition of the erasure identifier vector, a unique erasure identifier vector $\mathbf{e}$ exists corresponding to the received vector $\mathbf{y}$. Let us define $w(\mathbf{e})$ and $\mathbb{P}\{\mathbf{e}\}$ as the Hamming weight and the probability of $\mathbf{e}$, correspondingly. According to the definition of the erasure channel, $\mathbf{e}$ and $\mathbf{x}$ are independent. Let us assume that $w(\mathbf{e}) = m$. Hence, the decoder decodes the transmitted codeword based on the $N - m$ correctly received symbols.

We say a codeword $\mathbf{c} \in \mathcal{C}$ *satisfies* a received vector $\mathbf{y}$ iff $\forall j, y_j \neq \xi \Rightarrow c_j = y_j$. It is easy to observe

TABLE II

BINNING OF THE CODEBOOK IN TABLE I FOR $\mathbf{e} = (0, 1, 0, 1, 1)$

Bin Index	$b_{\mathbf{e}}(i)$	Codewords in Bin i	Received Vector Corresponding to Bin i
1	3	00000 01011 02022	$0\xi0\xi\xi$
2	3	00110 01121 02102	$0\xi1\xi\xi$
3	3	00220 01201 02212	$0\xi2\xi\xi$
4	3	10011 11022 12000	$1\xi0\xi\xi$
5	3	10121 11102 12110	$1\xi1\xi\xi$
6	3	10201 11212 12220	$1\xi2\xi\xi$
7	3	20022 21000 22011	$2\xi0\xi\xi$
8	3	20102 21110 22121	$2\xi1\xi\xi$
9	3	20212 21220 22201	$2\xi2\xi\xi$

that receiving the vector $\mathbf{y}$, probability of the codeword $\mathbf{c} \in \mathcal{C}$ being transmitted is zero ($p(\mathbf{c}|\mathbf{y}) = 0$) if $\mathbf{c}$ does not satisfy $\mathbf{y}$. On the other hand, all $\mathbf{c}$'s that satisfy $\mathbf{y}$ are equiprobable to be transmitted given $\mathbf{y}$. Hence, the task of the maximum likelihood decoder is to randomly select one of the codewords which satisfy $\mathbf{y}$.

Notice that there exist q^{N-m} possible received vectors $\mathbf{y}$ whose erasure identifier vector equals $\mathbf{e}$. Now, based on the erasure identifier vector $\mathbf{e}$, we partition the codebook $\mathcal{C}$ into q^{N-m} bins. Each bin is associated with a possible received vector $\mathbf{y}$ (with erasure identifier vector $\mathbf{e}$) and contains all the codewords in $\mathcal{C}$ that satisfy $\mathbf{y}$. It is easy to verify that each codeword belongs to only one bin. Let us denote the number of codewords in the i'th bin by $b_{\mathbf{e}}(i)$ for $i = 1, ..., q^{N-m}$. As an example, consider the $[5, 3, 2]$ codebook defined in Table I over $\mathbb{F}_3$. Assume $\mathbf{x} = (0, 0, 0, 0, 0)$ is transmitted, and $\mathbf{y} = (0, \xi, 0, \xi, \xi)$ is received. Thus, $\mathbf{e} = (0, 1, 0, 1, 1)$ is the erasure identifier vector corresponding to $\mathbf{y}$. For this specific $\mathbf{e}$, $\mathcal{C}$ can be partitioned into 3^2 bins as shown in Table II. Hence, the Maximum Likelihood decoder should choose one of the 3 codewords in bin 1 randomly. Accordingly, the error probability of the decoder for this case is $\frac{2}{3}$.

Let us assume a bin with index i is associated with the received vector $\mathbf{y}$. If $b_{\mathbf{e}}(i) = 1$, the transmitted codeword $\mathbf{x}$ can be decoded with no ambiguity. Otherwise, the optimum decoder randomly selects one of the $b_{\mathbf{e}}(i) > 1$ codewords in the bin. Thus, given bin i, the probability of error is $1 - \frac{1}{b_{\mathbf{e}}(i)}$.

Obviously, the transmitted codeword $\mathbf{x}$ always lies in the bin associated with the received vector $\mathbf{y}$. Hence, $\mathbb{P}\{\text{bin } i \text{ selected}|\, \mathbf{e}\} = \sum_{\mathbf{x}\in \text{bin } i} \mathbb{P}\{\mathbf{x} \text{ transmitted}|\, \mathbf{e}\} \overset{(a)}{=} \frac{b_{\mathbf{e}}(i)}{q^K}$, where (a) follows from the facts that $\mathbf{x}$ and $\mathbf{e}$ are independent and all the codewords of $\mathcal{C}$ are equiprobable. Based on the above arguments, probability of decoding error for the maximum likelihood decoder of any codebook, $\mathcal{C}$, is equal to

$$\begin{aligned}
P_{E,ML}^{\mathcal{C}} &\overset{(a)}{=} \sum_{m=d}^{N} \sum_{\mathbf{e}:w(\mathbf{e})=m} \mathbb{P}\{\mathbf{e}\}\mathbb{P}\{\text{error}|\mathbf{e}\} \\
&= \sum_{m=d}^{N} \sum_{\mathbf{e}:w(\mathbf{e})=m} \mathbb{P}\{\mathbf{e}\} \sum_{i=1,\ b_{\mathbf{e}}(i)>0}^{q^{N-m}} \mathbb{P}\{\text{bin } i \text{ selected}|\, \mathbf{e}\}\, \mathbb{P}\{\text{error}\,|\mathbf{e}, \text{bin } i \text{ selected}\} \\
&= \sum_{m=d}^{N} \sum_{\mathbf{e}:w(\mathbf{e})=m} \mathbb{P}\{\mathbf{e}\} \sum_{i=1,\ b_{\mathbf{e}}(i)>0}^{q^{N-m}} \frac{b_{\mathbf{e}}(i)}{q^K} \left(1 - \frac{1}{b_{\mathbf{e}}(i)}\right) \\
&\overset{(b)}{=} \sum_{m=d}^{N} \sum_{\mathbf{e}:w(\mathbf{e})=m} \mathbb{P}\{\mathbf{e}\} \left(1 - \frac{b_{\mathbf{e}}^{+}}{q^K}\right) \\
&\overset{(c)}{\geq} \sum_{m=d}^{N} \sum_{\mathbf{e}:w(\mathbf{e})=m} \mathbb{P}\{\mathbf{e}\} \left(1 - \frac{\min\{q^K, q^{N-m}\}}{q^K}\right) \qquad (1)
\end{aligned}$$

where $b_{\mathbf{e}}^{+}$ indicates the number of bins containing one or more codewords. (a) follows from the fact that the transmitted codeword can be uniquely decoded if the number of erasures in the channel is less than the minimum distance of the codebook [29], and (b) follows from the fact that $\sum_{i=1}^{q^{N-m}} b_{\mathbf{e}}(i) = q^K$. (c) is true since $b_{\mathbf{e}}^{+}$ is less than both the total number of codewords and the number of bins.

According to (1), $P_{E,ML}^{\mathcal{C}}$ is minimized for a code-book $\mathcal{C}$ *if* two conditions are satisfied. First, the minimum distance of $\mathcal{C}$ should achieve the maximum possible value, i.e., $d = N - K + 1$. Second, we should have $b_{\mathbf{e}}^{+} = \min\{q^K, q^{N-m}\}$ for all possible erasure vectors $\mathbf{e}$ with any weight $d \leq m \leq N$. Any MDS code satisfies the first condition by definition. Moreover, it is easy to show that for any MDS code, we have $b_{\mathbf{e}}(i) = q^{K-N+m}$, for $N - K \leq m \leq N$. We first prove this for the case of $m = N - K$. Consider the bins of an MDS code for any arbitrary erasure pattern $\mathbf{e}, w(\mathbf{e}) = N - K$. From the fact that $d = N - K + 1$, we have $b_{\mathbf{e}}(i) \leq 1$. Knowing $\sum_{i=1}^{q^K} b_{\mathbf{e}}(i) = q^K$, it is concluded that each bin contains exactly one codeword. Therefore, there exists only one codeword which matches any K correctly received symbols. Now, consider any general erasure identifier vector $\mathbf{e}, w(\mathbf{e}) = m > N - K$. Let us choose an arbitrary $K - N + m$ erasure positions in $\mathbf{e}$. Also, consider a received vector $\mathbf{y}$ whose erasure identifier vector is $\mathbf{e}$. By assigning arbitrary values from $\mathcal{X}$ to the $K - N + m$ chosen erasure positions, $\mathbf{y}$ can be extended to another vector $\mathbf{y}'$ with K non-erasure symbols. According to the statement proved for $m = N - K$, there exists exactly one codeword in the MDS codebook that satisfies $\mathbf{y}'$ (and consequently $\mathbf{y}$). Hence, there exists q^{K-N+m} codewords in the MDS code that satisfy $\mathbf{y}$, i.e. we have $b_{\mathbf{e}}(i) = q^{K-N+m}$. This completes the proof ■

Remark I. Theorem I is valid for any N and $1 \leq K < N$. However, it does not guarantee the existence of an $[N, K]$ MDS code for all such values of N and K. In fact, a conjecture on MDS codes states that for every linear $[N, K]$ MDS code over the Galois field $\mathbb{F}_q$, if $1 < K < q$, then $N \leq q + 1$, except when q is even and $K = 3$ or $K = q - 1$, for which $N \leq q + 2$ [28].

Definition II. $\mathrm{L}_m(N, K, q)$ is defined as

$$\mathrm{L}_m(N, K, q) \triangleq \sum_{m=N-K+1}^{N} \sum_{\mathbf{e}:w(\mathbf{e})=m} \mathbb{P}\{\mathbf{e}\} \left(1 - \frac{q^{N-m}}{q^K}\right). \tag{2}$$

Now, based on the Singleton bound and the fact that $\min\{q^K, q^{N-m}\} = q^{N-m}$ for $m > N - K$, the inequality in (1) leads to

$$P_{E,ML}^{\mathcal{C}} \geq \mathrm{L}_m(N, K, q). \tag{3}$$

Thus, $\mathrm{L}_m(N, K, q)$ is a lower-bound on the probability of error for any codebook $\mathcal{C}$ of size $[N, K]$ over an erasure channel, whether an MDS code of that size exists or not. Noting the proof of Theorem I, the lower bound in (3) is tight for any MDS code, i.e.

$$P_{E,ML}^{MDS} = \mathrm{L}_m(N, K, q). \tag{4}$$

Corollary I. For $N \leq q + 1$, converse of Theorem I is also true if the following condition is satisfied

$$\forall \mathbf{e} \in \{0, 1\}^N : \ \mathbb{P}\{\mathbf{e}\} > 0 \tag{5}$$

Proof. For $N \leq q + 1$ and $1 \leq K < N$, we know that an MDS code of size $[N, K]$ does exist (an $[N, K]$ Reed-Solomon code can be constructed over $\mathbb{F}_q$, see [30]). Let us assume the converse of Theorem I is not true. Then, there should be a non-MDS codebook, $\mathcal{C}$, with the size $[N, K, d]$, $d < N - K + 1$, which achieves the minimum probability of error ($P_{E,ML}^{\mathcal{C}} = P_{E,ML}^{MDS}$). For any erasure vector $\mathbf{e}'$ with the weight $w(\mathbf{e}') = N - K$, we can write

$$\begin{aligned}
\mathbb{P}\{\mathbf{e}'\} \left(1 - \frac{b_{\mathbf{e}'}^+}{q^K}\right) &\overset{(a)}{\leq} \sum_{\mathbf{e}:w(\mathbf{e})=N-K} \mathbb{P}\{\mathbf{e}\} \left(1 - \frac{b_{\mathbf{e}}^+}{q^K}\right) \\
&\overset{(b)}{\leq} \sum_{m=d}^{N-K} \sum_{\mathbf{e}:w(\mathbf{e})=m} \mathbb{P}\{\mathbf{e}\} \left(1 - \frac{b_{\mathbf{e}}^+}{q^K}\right) \\
&\overset{(c)}{\leq} \sum_{m=d}^{N-K} \sum_{\mathbf{e}:w(\mathbf{e})=m} \mathbb{P}\{\mathbf{e}\} \left(1 - \frac{b_{\mathbf{e}}^+}{q^K}\right) \\
&\quad + \sum_{m=N-K+1}^{N} \sum_{\mathbf{e}:w(\mathbf{e})=m} \mathbb{P}\{\mathbf{e}\} \left(1 - \frac{b_{\mathbf{e}}^+}{q^K} - 1 + \frac{q^{N-m}}{q^K}\right) \\
&\overset{(d)}{=} P_{E,ML}^{\mathcal{C}} - P_{E,ML}^{MDS} \overset{(e)}{=} 0
\end{aligned} \tag{6}$$

where (a), (b), and (c) follow from the fact that $b_{\mathbf{e}}^{+} \leq \min\{q^{N-m}, q^{K}\}$ if $w(\mathbf{e}) = m$. (d) and (e) are based on (1), (4), and the assumption that $P_{E,ML}^{\mathcal{C}} = P_{E,ML}^{MDS}$. Applying the fact that $b_{\mathbf{e}}^{+} \leq \min\{q^{N-m}, q^{K}\}$ for $w(\mathbf{e}) = N - K$, results in $\mathbb{P}\{\mathbf{e}'\} \left(1 - \frac{b_{\mathbf{e}'}^{+}}{q^{K}}\right) \geq 0$. Combining this result with (5) and (6), proves that $b_{\mathbf{e}'}^{+} = q^{K}$. Thus, we have $b_{\mathbf{e}'}(i) = 1$ for all $1 \leq i \leq q^{K}$ and any $\mathbf{e}'$ with the weight of $w(\mathbf{e}') = N - K$.

On the other hand, we know that the minimum distance of $\mathcal{C}$ is d. Thus, there exist two codewords $\mathbf{c}_1$ and $\mathbf{c}_2$ in $\mathcal{C}$ with the distance of d from each other. We define the vector $\mathbf{e}_{12}$ as follows

$$e_{12_j} \triangleq \begin{cases} 0 & \text{if } c_{1_j} = c_{2_j} \\ 1 & \text{otherwise} \end{cases} \tag{7}$$

where e_{12_j}, c_{1_j}, and c_{2_j} denote the j'th elements of $\mathbf{e}_{12}$, $\mathbf{c}_1$, and $\mathbf{c}_2$, respectively. Since we assumed that $\mathcal{C}$ is a non-MDS code, we have $w(\mathbf{e}_{12}) = d \leq N - K$. Then, we construct the binary vector $\mathbf{e}^{\star}$ by replacing $N - K - d$ zero elements in $\mathbf{e}_{12}$ with one. The positions of these replacements can be arbitrary. Again, it is obvious that $w(\mathbf{e}^{\star}) = N - K$.

Now, we define the received vector $\mathbf{y}^{\star}$ as

$$y_j^{\star} \triangleq \begin{cases} c_{1_j} & \text{if } e_j^{\star} = 0 \\ \xi & \text{otherwise.} \end{cases} \tag{8}$$

Based on (7) and (8), it can be seen that $y_j^{\star} = c_{1_j} = c_{2_j}$ if $e_j^{\star} = 0$. Thus, both $\mathbf{c}_1$ and $\mathbf{c}_2$ satisfy $\mathbf{y}^{\star}$. Therefore, in the binning corresponding to the erasure vector $\mathbf{e}^{\star}$, both $\mathbf{c}_1$ and $\mathbf{c}_2$ would be in the same bin (the one corresponding to $\mathbf{y}^{\star}$). However, we already know that $b_{\mathbf{e}^{\star}}(i) = 1$ for all $1 \leq i \leq q^{K}$ since $w(\mathbf{e}^{\star}) = N - K$. This contradiction proves the corollary ■

The memoryless erasure channel obviously satisfies the condition in (5). Combining Theorem I and Corollary I results in Corollary II.

Corollary II. A block code of size $[N, K]$ with equiprobable codewords over a memoryless erasure channel has the minimum probability of error (assuming optimum, i.e. maximum likelihood decoding) among all block codes of the same size *iff* that code is *Maximum Distance Separable* (MDS).

Remark II. In the proof of Theorem I, we do not assume any memoryless property for the erasure channel. Instead, the proof is based on the analysis of a block of transmitted symbols, regardless of the dependency between the erasures in the block. To prove the converse of Theorem I (Corollary I), the only extra condition we assume is the one in (5). This condition (that all erasure patterns can occur) is obviously much weaker than the memoryless condition. Therefore, converse of Theorem I is valid for many erasure channels with memory as long as they satisfy the condition in (5).

A. MDS codes with Suboptimal Decoding

In the proof of Theorem I, it is assumed that the received codewords are decoded based on maximum likelihood decoding which is optimum in this case. However, in many practical cases, MDS codes are

decoded by simpler decoders [30]. We consider the following suboptimal decoding method. If K or more symbols are received correctly, the decoder finds the transmitted codeword of length N uniquely with no error (by the binning method described in the proof of Theorem I or by an algebraic method as in [30]). In case more than $N-K$ symbols are erased, a decoding error is declared. Let $P_{E,sub}^{MDS}$ denote the probability of this event. $P_{E,sub}^{MDS}$ is obviously different from the decoding error probability of the maximum likelihood decoder denoted by $P_{E,ML}^{MDS}$. Theoretically, an optimum maximum likelihood decoder of an MDS code may still decode the original codeword correctly with a positive, but small probability, if it receives less than K symbols. More precisely, according to the proof of Theorem I, such a decoder is able to correctly decode an MDS code over $\mathbb{F}_q$ with the probability of $\frac{1}{q^i}$ after receiving $K-i$ correct symbols. Of course, for Galois fields with large cardinality, this probability is negligible. The relationship between $P_{E,sub}^{MDS}$ and $P_{E,ML}^{MDS}$ can be summarized as follows

$$\begin{aligned}
P_{E,sub}^{MDS} &= \sum_{m=N-K+1}^{N} \mathbb{P}\{m \text{ symbols erased}\} \\
&= \sum_{m=N-K+1}^{N} \sum_{\mathbf{e}:w(\mathbf{e})=m} \mathbb{P}\{\mathbf{e}\} \\
&= \sum_{m=N-K+1}^{N} \sum_{\mathbf{e}:w(\mathbf{e})=m} \mathbb{P}\{\mathbf{e}\} \left(1 - \frac{q^{N-m}}{q^K} + \frac{q^{N-m}}{q^K}\right) \\
&\overset{(a)}{=} P_{E,ML}^{MDS} + \sum_{m=N-K+1}^{N} \frac{1}{q^{K-N+m}} \sum_{\mathbf{e}:w(\mathbf{e})=m} \mathbb{P}\{\mathbf{e}\} \\
&\leq P_{E,ML}^{MDS} + \frac{1}{q} \sum_{m=N-K+1}^{N} \sum_{\mathbf{e}:w(\mathbf{e})=m} \mathbb{P}\{\mathbf{e}\} \\
&= P_{E,ML}^{MDS} + \frac{P_{E,sub}^{MDS}}{q}
\end{aligned} \tag{9}$$

where (a) follows from (4). Hence, $P_{E,ML}^{MDS}$ is bounded as

$$P_{E,sub}^{MDS}\left(1 - \frac{1}{q}\right) \leq P_{E,ML}^{MDS} \leq P_{E,sub}^{MDS}. \tag{10}$$

Note that the above inequality is valid whether the erasure channel is memoryless or has memory.

Similar to $\mathrm{L}_m(N,K,q)$, $\mathrm{L}_s(N,K,q)$ can be defined as the lower-bound on the probability of error for any codebook of size $[N,K]$ with suboptimal decoder, whether an MDS code of that size exists or not. In other words, $\mathrm{L}_s(N,K,q)$ is defined as

$$\mathrm{L}_s(N,K,q) \triangleq \sum_{m=N-K+1}^{N} \sum_{\mathbf{e}:w(\mathbf{e})=m} \mathbb{P}\{\mathbf{e}\}. \tag{11}$$

In case an MDS code of size $[N, K]$ does exist, we have $P_{E,sub}^{MDS} = \mathrm{L}_s(N, K, q)$. Now the inequality in (10) can be written as

$$\mathrm{L}_s(N, K, q)\left(1 - \frac{1}{q}\right) \leq \mathrm{L}_m(N, K, q) \leq \mathrm{L}_s(N, K, q). \tag{12}$$

IV. Error Exponents of MDS, Random, and Random Linear Codes

A. Error Exponent of MDS Codes over a Memoryless Erasure Channel

Consider a block code of size $[N, K]$ over the memoryless erasure channel of Fig. 1. Let $\alpha \triangleq \frac{N-K}{N}$ denote the coding overhead. For a q-ary $[N, K]$ code, the rate per symbol, R, is equal to

$$R = \frac{K}{N}\log q = (1 - \alpha)\log q. \tag{13}$$

In a block code of length N, the number of lost symbols would be $\sum_{i=1}^{N} e_i$ where e_i is defined in Definition I of section III. Thus, assuming the suboptimal decoder of subsection III-A, $\mathrm{L}_s(N, K, q)$ can be written as

$$\mathrm{L}_s(N, K, q) = \mathbb{P}\left\{\frac{1}{N}\sum_{i=1}^{N} e_i > \alpha\right\} = \sum_{i=0}^{K-1} P_i \tag{14}$$

where P_i denotes the probability that i symbols are received correctly. As stated in subsection III-A, $\mathrm{L}_s(N, K, q)$ would be equal to $P_{E,sub}^{MDS}$ if an MDS code of size $[N, K]$ exists. Since the channel is assumed to be memoryless in this section, e_i's are i.i.d random variables with Bernoulli distribution. Thus, we have

$$P_i = (1 - \pi)^i \pi^{N-i}\binom{N}{i}. \tag{15}$$

It is easy to see that

$$\frac{P_i}{P_{i-1}} = \frac{(N - i + 1)(1 - \pi)}{i\pi} > 1 \quad \text{for } i = 1, \cdots, K - 1 \tag{16}$$

if $\alpha = \frac{N-K}{N} \geq \pi$. According to equation (13), the condition $\alpha \geq \pi$ can be rewritten as $R \leq (1 - \pi)\log q = C$ where C is the capacity of the memoryless erasure channel. Therefore, the summation terms in equation (14) are always increasing, and the largest term is the last one. Now, we can bound $\mathrm{L}_s(N, K, q)$ as

$$P_{K-1} \leq \mathrm{L}_s(N, K, q) \leq KP_{K-1}. \tag{17}$$

The term $\binom{N}{K-1}$ in P_{K-1} can be bounded using the fact that for any $N > K > 0$, we have [31]

$$\frac{1}{N+1}e^{NH\left(\frac{K}{N}\right)} \leq \binom{N}{K} \leq e^{NH\left(\frac{K}{N}\right)} \tag{18}$$

where the entropy, $H\left(\frac{K}{N}\right)$, is computed in nats. Thus, using (15), (17), and (18), $\mathrm{L}_s(N, K, q)$ is bounded as

$$\frac{\pi(1-\alpha)Ne^{-Nu(\alpha)}}{(1-\pi)(N+1)(\alpha N+1)} \leq P_{K-1} \leq \mathrm{L}_s(N, K, q) \leq KP_{K-1} \leq \frac{\pi(1-\alpha)^2N^2e^{-Nu(\alpha)}}{(1-\pi)(\alpha N+1)} \tag{19}$$

where $u(\alpha)$ is defined as

$$u(\alpha) \triangleq \begin{cases} 0 & \text{for } \alpha \leq \pi \\ \alpha \log\left(\dfrac{\alpha(1-\pi)}{\pi(1-\alpha)}\right) & \\ -\log\left(\dfrac{1-\pi}{1-\alpha}\right) & \text{for } \pi \leq \alpha \leq 1. \end{cases} \tag{20}$$

with the $\log$ functions computed in the Neperian base. $u(\alpha)$ can be interpreted as the error exponent of MDS codes over a memoryless erasure channel if an MDS code of size $[N, K]$ exists. Otherwise, $u(\alpha)$ is the exponent of the lower-bound on the probability of error for any codebook of size $[N, K]$ with the suboptimal decoder of subsection III-A, $\mathrm{L}_s(N, K, q)$. Using equation (13), $u()$ can be expressed in terms of R instead of α

$$u(R) \triangleq \begin{cases} 0 & \text{for } 1-\pi \leq r \\ -r \log \dfrac{(1-\pi)(1-r)}{r\pi} & \\ -\log \dfrac{\pi}{1-r} & \text{for } 0 < r \leq 1-\pi \end{cases} \tag{21}$$

where $r \triangleq \dfrac{R}{\log q}$ is the normalized rate.

B. Random Coding Error Exponent of a Memoryless Erasure Channel

It is interesting to compare the error exponent in (21) with the random coding error exponent as described in [7]. This exponent, $E_r(R)$, can be written as

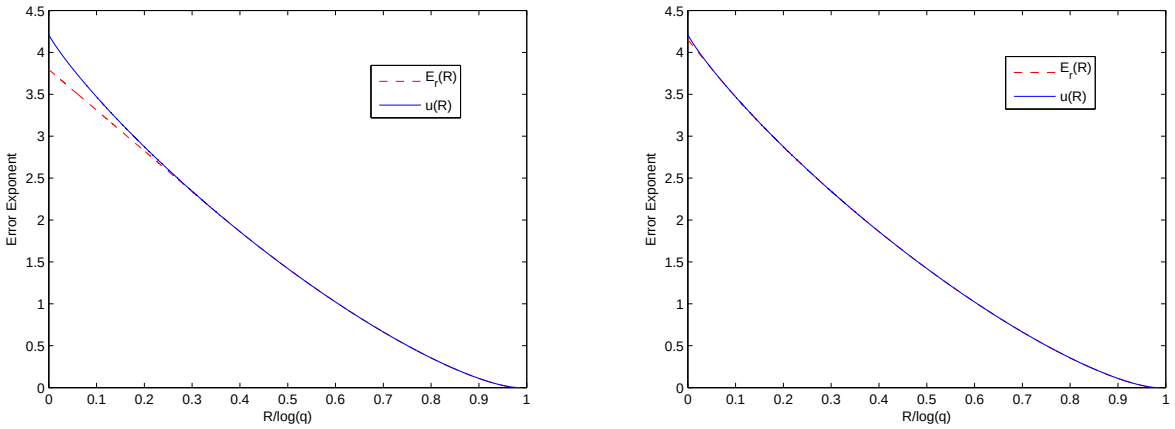
$$E_r(R) = \max_{0 \leq \rho \leq 1} \left\{ -\rho R + \max_{\mathbf{Q}} E_0(\rho, \mathbf{Q}) \right\} \tag{22}$$

where $\mathbf{Q}$ is the input distribution, and $E_0(\rho, \mathbf{Q})$ equals

$$E_0(\rho, \mathbf{Q}) = -\log \left(\sum_{j=0}^{q} \left[\sum_{k=0}^{q-1} Q(k) P(j|k)^{\frac{1}{1+\rho}} \right]^{1+\rho} \right). \tag{23}$$

Due to the symmetry of the channel transition probabilities, the uniform distribution maximizes (22) over all possible input distributions. Therefore, $E_0(\rho, \mathbf{Q})$ can be simplified as

$$E_0(\rho, \mathbf{Q}) = -\log \left(\frac{1-\pi}{q^\rho} + \pi \right). \tag{24}$$



(a) (b)

Fig. 2. Error exponents of random coding ($E_r(R)$) and MDS coding ($u(R)$) for a memoryless erasure channel with $\pi = 0.015$, and (a): $q = 128$, (b): $q = 1024$.

Solving the maximization in (22), gives us $E_r(R)$ as

$$E_r(R) = \begin{cases} -\log \dfrac{1-\pi+\pi q}{q} - r \log q \\ \qquad\qquad \text{for } 0 \le r \le \dfrac{R_c}{\log q} \\ -r \log \dfrac{(1-\pi)(1-r)}{r\pi} - \log \dfrac{\pi}{1-r} \\ \qquad\qquad \text{for } \dfrac{R_c}{\log q} \le r \le 1-\pi \end{cases} \tag{25}$$

where $r \triangleq \frac{R}{\log q}$, and $R_c \triangleq \frac{1-\pi}{1-\pi+\pi q} \log q$ are the normalized and the critical rates, respectively.

Comparing (21) and (25), we observe that MDS codes and random codes perform exponentially the same for rates between the critical rate and the capacity. However, for the region below the critical rate, where the error exponent of the random code decays linearly with R, MDS codes achieve a larger error exponent. It is worth noting that this interval is negligible for large alphabet sizes.

Figure 2 depicts the error exponents of random codes and MDS codes for the alphabet sizes of $q = 128$ and $q = 1024$ over an erasure channel with $\pi = 0.015$. Comparing Fig. 2(a) and Fig. 2(b), it is seen that the region where MDS codes outperform random codes ($R < R_c$) becomes very small even for moderate values of alphabet size ($q = 1024$).

C. Random Linear Coding Error Exponent of a Memoryless Erasure Channel

Maximum likelihood decoding of random codes generally has exponential complexity in terms of the block length (N). Random linear codes, on the other hand, have the advantage of polynomial decoding

complexity (assuming maximum likelihood decoding) over any arbitrary erasure channel [32]. In a linear codebook of size $[N, K]$, any codeword $\mathbf{c}$ can be written as $\mathbf{c} = \mathbf{bG}$, where $\mathbf{b}$ is a row vector of length K, indicating the information symbols, and $\mathbf{G}$ is the generator matrix of size $K \times N$. In the case of a random linear code, every element in $\mathbf{G}$ is generated independently according to a distribution $\mathbf{Q}$ [11], [12]. For a memoryless erasure channel, due to the symmetry of the channel transition probabilities, the uniform distribution is applied to generate $\mathbf{G}$.

Here, we describe a suboptimal decoder with polynomial complexity for decoding of linear block codes over erasure channels. This decoder is a slightly modified version of the optimum (maximum likelihood) decoder in [32]. In case that less than K symbols are received correctly, a decoding error is declared. When K or more correct symbols are received, the decoder determines the information vector $\mathbf{b}$ (and the transmitted codeword $\mathbf{c}$) by constructing a new matrix called the *reduced generator matrix*, $\tilde{\mathbf{G}}$. $\tilde{\mathbf{G}}$ consists of the columns in $\mathbf{G}$ whose corresponding symbols are received correctly. Thus, if the erasure identifier vector $\mathbf{e}$ has the weight of $w(\mathbf{e}) = m \leq N - K$, $\tilde{\mathbf{G}}$ would have the size of $K \times (N - m)$. Then, the decoder computes the row or column rank of $\tilde{\mathbf{G}}$. If this rank is less than K, a decoding error is declared. In case the rank is equal to K, the information symbol vector can be decoded uniquely by solving $\mathbf{b}\tilde{\mathbf{G}} = \tilde{\mathbf{y}}$. In this case, $\tilde{\mathbf{y}}$ is the *reduced received vector* consisting of the correctly received symbols only. In the described decoder, suboptimality arises only from the cases that less than K symbols are received correctly. In such cases, although an optimal (maximum likelihood) decoder can not reconstruct the transmitted codeword uniquely, it may be able to find it by making a random selection among a set of likely codewords [32]. However, the described decoder here just declares an error.

Using the described suboptimal decoder, the probability of error is the probability that the rank of $\tilde{\mathbf{G}}$ is less than K. Thus, the probability of error conditioned on an erasure vector of weight $w(\mathbf{e}) = m$ can be written as [33]

$$\mathbb{P}\left\{\text{error}|w(\mathbf{e}) = m\right\} = 1 - \prod_{i=N-m-K+1}^{N-m} \left(1 - \frac{1}{q^i}\right). \tag{26}$$

We bound the above probability as

$$\begin{aligned} \mathbb{P}\left\{\text{error}|w(\mathbf{e}) = m\right\} &\overset{(a)}{\leq} 1 - \left(1 - \frac{1}{q^{N-m-K+1}}\right)^K \\ &\overset{(b)}{\leq} \frac{K}{q^{N-m-K+1}} \end{aligned} \tag{27}$$

where (a) follows from the fact that $\left(1 - \frac{1}{q^{N-m-K+1}}\right) \leq \left(1 - \frac{1}{q^i}\right)$ for all $N - m - K + 1 \leq i \leq N - m$. (b) follows from Bernoulli's inequality [34] and the assumption that $w(\mathbf{e}) = m \leq N - K$. The total

probability of error is written as

$$\begin{aligned} P_{E,sub}^{lin} &= \sum_{i=0}^{K-1} P_i + \sum_{i=K}^{N} P_i \ \mathbb{P}\left\{\text{error}|w(\mathbf{e}) = N-i\right\} \\ &\overset{(a)}{\leq} \sum_{i=0}^{K-1} P_i + \sum_{i=K}^{N} \frac{KP_i}{q^{i-K+1}} \\ &\overset{(b)}{\leq} \sum_{i=0}^{K-2} P_i + A_{K-1} + K\sum_{i=K}^{N} A_i \end{aligned} \tag{28}$$

where P_i denotes the probability that i symbols are received correctly as defined in subsection IV-A, and $A_i = \frac{P_i}{q^{i-K+1}}$. (a) follows from (27), and (b) is true since $A_{K-1} = P_{K-1}$.

We define i_0 as $i_0 = \frac{(N+1)(1-\pi)}{1-\pi+q\pi}$. Of course, i_0 is not necessarily an integer. The rest of the analysis is divided to two cases, based on whether $i_0 \leq K$ or vice versa.

Case I: $i_0 \leq K$. Following equation (13), this condition is equivalent to $R_c\left(1+\frac{1}{N}\right) \leq R$ where $R_c = \frac{1-\pi}{1-\pi+\pi q}\log q$ is the critical rate as in (25). Now, similar to equation (16), we can write

$$\frac{A_i}{A_{i-1}} = \frac{(N-i+1)(1-\pi)}{qi\pi} \leq 1 \quad \text{for } i = K, \cdots, N. \tag{29}$$

Thus, A_i's are decreasing, and $A_{K-1} \geq A_i$ for $K \leq i \leq N$. This means that we can upper-bound the term $\sum_{i=K}^{N} A_i$ in (28) as

$$\sum_{i=K}^{N} A_i \leq (N-K+1)A_{K-1} = (N-K+1)P_{K-1}. \tag{30}$$

Then, we write

$$\begin{aligned} P_{E,sub}^{lin} &\overset{(a)}{\leq} \sum_{i=0}^{K-1} P_i + K(N-K+1)A_{K-1} \\ &\overset{(b)}{\leq} (N-K+2)KP_{K-1} \\ &\overset{(c)}{\leq} \frac{\pi K^2(N-K+2)}{(1-\pi)(N-K+1)} e^{-NE_r\left(\frac{K}{N}\log q\right)} \\ &\overset{(d)}{=} \frac{\pi N^2 r^2(N-Nr+2)}{(1-\pi)(N-Nr+1)} e^{-NE_r(R)} \end{aligned} \tag{31}$$

where $r = \frac{R}{\log q}$ is the normalized rate as in (25). (a) follows from (28) and (30). (b) is based on (16) and the facts that $\max\limits_{0\leq i\leq K-1} P_i = P_{K-1}$ and $\sum_{i=0}^{K-1} P_i \leq KP_{K-1}$. (c) results from (15) and (18), and (d) from (13).

Case II: $K < i_0$. Following equation (13), this condition is equivalent to $R < R_c\left(1+\frac{1}{N}\right)$. Computing the ratio $\frac{A_i}{A_{i-1}}$ as in (29), it can be seen that $\frac{A_i}{A_{i-1}} \leq 1$ for $i_0 \leq i$, and $\frac{A_i}{A_{i-1}} > 1$ for $i_0 > i$. Thus, the series of $\{A_i\}_{i=K-1}^{N}$ achieves its maximum at $i^\star = \lfloor i_0 \rfloor \geq K$, and we can upper-bound the term $\sum_{i=K}^{N} A_i$ in (28) as

$$\sum_{i=K}^{N} A_i \leq (N-K+1)A_{i^\star}. \tag{32}$$

Then, we have

$$\begin{aligned}
P_{E,sub}^{lin} &\overset{(a)}{\leq} \sum_{i=0}^{K-1} P_i + K(N-K+1)A_{i^\star} \\
&\overset{(b)}{\leq} KP_{K-1} + K(N-K+1)A_{i^\star} \\
&\overset{(c)}{\leq} K(N-K+2)A_{i^\star} \\
&\overset{(d)}{\leq} (N-K+2)K\exp\left(-N\left\{\frac{i^\star}{N}\log\frac{\frac{i^\star}{N}q\pi}{\left(1-\frac{i^\star}{N}\right)(1-\pi)} - \log\frac{\pi}{1-\frac{i^\star}{N}} - \frac{K}{N}\log q\right\}\right) \\
&\leq (N-K+2)K\exp\left(-N\left\{\frac{i_0-1}{N}\log\frac{\frac{i_0-1}{N}q\pi}{\left(1-\frac{i_0-1}{N}\right)(1-\pi)} - \log\frac{\pi}{1-\frac{i_0}{N}} - \frac{K}{N}\log q\right\}\right) \\
&\overset{(e)}{=} (N-rN+2)Nre^{-Nv(R,N)} \qquad (33)
\end{aligned}$$

where $v(R,N)$ is defined as

$$\begin{aligned}
v(R,N) &\triangleq \frac{N(1-\pi)-\pi q}{N(1-\pi+\pi q)}\log\frac{N(1-\pi)-\pi q}{(N+1)(1-\pi)} - \log\frac{\pi N(1-\pi+\pi q)}{N\pi q-1+\pi} - R \\
&= -\log\frac{1-\pi+\pi q}{q} - R + q\,O\left(\frac{1}{N}\right) \\
&= E_r(R) + q\,O\left(\frac{1}{N}\right). \qquad (34)
\end{aligned}$$

In (33), (a) follows from (28) and (32). (b) results from (16) and $\max\limits_{0\leq i\leq K-1} P_i = P_{K-1}$. (c) follows from the facts that $A_{i^\star}$ is the maximum of the series $\{A_i\}_{i=K-1}^{N}$, and $P_{K-1} = A_{K-1} \leq A_{i^\star}$. (d) is based on (18) and can be derived similar to (19). (e) follows from (13), and noting that $i_0 = \frac{(N+1)(1-\pi)}{1-\pi+q\pi}$.

(31) and (33) upper-bound $P_{E,sub}^{lin}$ for $R_c\left(1+\frac{1}{N}\right) \leq R$ and $R < R_c\left(1+\frac{1}{N}\right)$, respectively. Combining these two upper-bounds results in

$$P_{E,sub}^{lin} \leq \begin{cases} (N-rN+2)Nre^{-N\left(E_r(R)+q\,O\left(\frac{1}{N}\right)\right)} & \text{for } R < R_c\left(1+\frac{1}{N}\right) \\ \frac{\pi N^2r^2(N-Nr+2)}{(1-\pi)(N-Nr+1)}e^{-NE_r(R)} & \text{for } R \geq R_c\left(1+\frac{1}{N}\right) \end{cases} \qquad (35)$$

D. Exponential Optimality of Random Coding and Random Linear Coding

Using the sphere packing bound, it is already shown that random coding is exponentially optimal for the rates above the critical rate over channels with relatively small alphabet sizes ($q \ll N$) [9], [10]. In other words, we know that

$$P_{E,ML}^{rand} \doteq e^{-NE_r(R)} \qquad (36)$$

where the notation $\doteq$ means $\lim_{N->\infty} -\frac{\log P_{E,ML}^{rand}}{N} = E_r(R)$. However, the sphere packing bound is not tight for the channels whose alphabet size, q, is comparable to the block length. Here, based on Theorem I and the results of subsections IV-A, IV-B, and IV-C, we prove the exponential optimality of random coding and random linear coding over the erasure channels for all alphabet sizes (whether or not q is comparable with N).

The average decoding error probability for an ensemble of random codebooks with the maximum-likelihood decoding can be upper bounded as

$$P_{E,ML}^{rand} \overset{(a)}{\leq} e^{-NE_r(R)} \overset{(b)}{=} e^{-Nu(R)} \tag{37}$$

where (a) follows from [7], and (b) is valid only for rates above the critical rate according to (21) and (25). The similar upper-bound for $P_{E,sub}^{lin}$ is given in (31).

We can also lower bound $P_{E,ML}^{rand}$ and $P_{E,sub}^{lin}$ as

$$\begin{aligned} P_{E,ML}^{rand} &\overset{(a)}{\geq} \mathrm{L}_m(N,K,q) \\ &\overset{(b)}{\geq} \left(1-\frac{1}{q}\right)\mathrm{L}_s(N,K,q) \\ &\overset{(c)}{\geq} \frac{\left(1-\frac{1}{q}\right)\pi r N e^{-Nu(R)}}{(1-\pi)\,(N+1)\,((1-r)\,N+1)} \end{aligned} \tag{38}$$

where (a) follows from Theorem I and (3), (b) from inequality (12), and (c) from inequality (19). The inequality in (38) remains valid if $P_{E,ML}^{rand}$ is replaced by $P_{E,sub}^{lin}$. Now, we prove the main Theorem of this section.

Theorem II. Consider a memoryless erasure channel with an arbitrary alphabet size q and the erasure probability π. Random codes and random linear codes are exponentially optimal over such channel as long as the normalized rate, r, is above the normalized critical rate, $r_c \triangleq \frac{1-\pi}{1-\pi+\pi q}$. More precisely, let $\mathrm{L}_m(N,K,q)$, $P_{E,ML}^{rand}$, and $P_{E,sub}^{lin}$ denote the lower-bound on the error probability of block codes of size $[N,K]$ over $\mathbb{F}_q$ (defined in Definition II), error probability of random codes (with ML decoder), and error probability of random linear codes (with the suboptimal decoder), respectively. As long as $r \geq r_c$, we have

$$\begin{aligned} P_{E,ML}^{rand} &= O\left(N\mathrm{L}_m(N,K,q)\right) \\ P_{E,sub}^{lin} &= O\left(N^3\mathrm{L}_m(N,K,q)\right) \end{aligned} \tag{39}$$

Proof. Combining (37) and (38) guarantees that both the upper-bound and the lower-bound on $P_{E,ML}^{rand}$ are exponentially tight, and the decaying exponent of $P_{E,ML}^{rand}$ versus N is indeed $u(R)$. Combining (31)

and (38) proves the same result about the exponent of $P_{E,sub}^{lin}$ versus N. Moreover, we can write

$$\begin{aligned} \mathrm{L}_m(N,K,q) &\overset{(a)}{\leq} P_{E,ML}^{rand} \overset{(b)}{\leq} c_1 N \mathrm{L}_m(N,K,q) \\ \mathrm{L}_m(N,K,q) &\overset{(a)}{\leq} P_{E,sub}^{lin} \overset{(c)}{\leq} c_2 N^3 \mathrm{L}_m(N,K,q) \end{aligned} \tag{40}$$

where $c_1 = \frac{(1-\pi)(1+\frac{1}{N})(1-r+\frac{1}{N})}{\left(1-\frac{1}{q}\right)\pi r}$ and $c_2 = \frac{\pi r(1+\frac{1}{N})(1-r+\frac{2}{N})}{\left(1-\frac{1}{q}\right)}$. Here, (a) follows from Theorem I and (3), and (b) results from inequalities (37) and (38). (c) is based on (12), (19), and (33). (40) completes the proof ■

It should be noted that in (40), the term $\mathrm{L}_m(N,K,q)$ can be replaced by $P_{E,ML}^{MDS}$ if an MDS code of size $[N,K]$ exists over $\mathbb{F}_q$. In that case, since the coefficients of $P_{E,ML}^{MDS}$ in (40) do not include any exponential terms, it can be concluded that for rates above the critical rate, both random codes and random linear codes perform exponentially the same as MDS codes, which are already shown to be optimum.

V. Conclusion

It is proved that *Maximum Distance Separable* (MDS) codes achieve the minimum probability of error over any erasure channel (with or without memory). Moreover, $\mathrm{L}_m(N,K,q)$ is introduced as a lower-bound on the error probability of block codes over the erasure channel. Finally, for the memoryless erasure channel and rates above the critical rate, the error exponents of random codes and random linear codes are studied and shown to be equal to the exponent of $\mathrm{L}_m(N,K,q)$. This result is proved to be valid whether or not the alphabet size is comparable with the codebook block length, N.

Acknowledgments

The authors would like to thank Dr. Muriel Medard and Dr. Amin Shokrollahi for their helpful comments and fruitful suggestions to improve this work.

References

[1] W. T. Tan and A. Zakhor, "Video Multicast Using Layered FEC and Scalable Compression," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 3, pp. 373–386, 2001.

[2] L. Dairaine, L. Lancerica, J. Lacan, and J. Fimes, "Content-Access QoS in Peer-to-Peer Networks Using a Fast MDS Erasure Code," *Elsevier Computer Communications*, vol. 28, no. 15, pp. 1778–1790, 2005.

[3] S. Fashandi, S. Oveisgharan, and A.K. Khandani, "Coding over an Erasure Channel with a Large Alphabet Size," in *IEEE International Symposium on Information Theory*, 2008, pp. 1053–1057.

[4] C. E. Shannon, "A Mathematical Theory of Communications," *Bell Systems Technical Journal*, vol. 27, pp. 379–423,623–656, 1948.

[5] P. Elias, "Coding for Noisy Channels," *IRE Convention Record*, vol. 4, pp. 37–46, 1955.

[6] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to Error Probability for Coding on Discrete Memoryless Channels," *Information and Control*, vol. 10, pp. 65–103,522–552, 1967.

[7] R. G. Gallager, *Information Theory and Reliable Communication*, 1st ed. New York, NY, USA: John Wiley & Sons, 1968, pp. 135–144.

[8] G. Forney, "Exponential Error Bounds for Erasure, List, and Decision Feedback Schemes," *IEEE Transactions on Information Theory*, vol. 14, no. 2, pp. 206–220, 1968.

[9] R. G. Gallager, *Information Theory and Reliable Communication*, 1st ed. New York, NY, USA: John Wiley & Sons, 1968, pp. 157–158.

[10] R. Gallager, "The Random Coding Bound is Tight for the Average Code," *IEEE Transactions on Information Theory*, vol. 19, no. 2, pp. 244–246, 1973.

[11] J Pierce, "Limit Distribution of the Minimum Distance of Random Linear Codes," *IEEE Transactions on Information Theory*, vol. 13, no. 4, pp. 595– 599, 1967.

[12] A. Barg and G. D. Forney, "Random codes: Minimum Distances and Error Exponents," *IEEE Transactions on Information Theory*, vol. 48, no. 9, pp. 2568– 2573, 2006.

[13] Ron M. Roth, *Introduction to Coding Theory*, 1st ed. Cambridge University Press, 2006, pp. 333–351.

[14] X. H. Peng, "Erasure-control Coding for Distributed Networks," *IEE Proceedings on Communications*, vol. 152, pp. 1075 – 1080, 2005.

[15] N. Alon, J. Edmonds, and M. Luby, "Linear Time Erasure Codes with Nearly Optimal Recovery," in *IEEE Symposium on Foundations of Computer Science, Proc. IEEE Vol. 3*, 1995, pp. 512–519.

[16] J. Justesen , "On the Complexity of Decoding Reed-Solomon Codes," *IEEE transactions on information theory*, vol. 22, no. 2, pp. 237– 238, 1993.

[17] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman, "Efficient Erasure Correcting Codes," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 569–584, 2001.

[18] M. G. Luby, "LT Codes," in *IEEE Symposium on the Foundations of Computer Science (FOCS)*, 2002, pp. 271–280.

[19] A. Shokrollahi, "Raptor Codes," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2551–2567, 2006.

[20] S. Fashandi, S. Oveisgharan, and A.K. Khandani, "Path Diversity in Packet Switched Networks: Performance Analysis and Rate Allocation," in *IEEE Global Telecommunications Conference, GLOBECOM '07*, 2007, pp. 1840–1844.

[21] R. Koetter and M. Medard , "An Algebraic Approach to Network Coding," *IEEE transactions on Networking*, vol. 11, no. 5, pp. 782–795, 2003.

[22] T. Ho, R. Koetter, M. Medard, D. R. Karger, and M. Effros, "The Benefits of Coding over Routing in a Randomized Setting," in *IEEE International Symposium on Information Theory*, 2003, p. 442.

[23] P. A. Chou, Y. Wu, and K. Jain, "Practical Network Coding ," in *51st Allerton Conference on Communication, Control and Computing*, 2003.

[24] C. Gkantsidis and P. R. Rodriguez, "Network coding for large scale content distribution," in *IEEE INFOCOM, Proc. IEEE Vol. 4*, 2005, pp. 2235–2245.

[25] C. Gkantsidis, J. Miller, and P. Rodriguez, "Comprehensive View of a Live Network Coding P2P System," in *ACM SIGCOMM Conference on Internet Measurement*, 2006, pp. 177 – 188.

[26] T. Ho, M. Medard, R. Koetter, D. R. Karger, M. Effros, Shi Jun, and B. Leong, "A Random Linear Network Coding Approach to Multicast," *IEEE Transactions on Information Theory*, vol. 52, no. 10, pp. 4413 – 4430, 2006.

[27] S. Jaggi, M. Langberg, S. Katti, T. Ho, D. Katabi, and M. Medard, "Resilient Network Coding in the Presence of Byzantine Adversaries," in *IEEE International Conference on Computer Communications (INFOCOM)*, 2007, pp. 616 – 624.

[28] J. L. Walker, "A New Approach to the Main Conjecture on Algebraic-geometric MDS Codes," *Journal of Designs, Codes and Cryptography*, vol. 9, no. 1, pp. 115– 120, 1996.

[29] Ron M. Roth, *Introduction to Coding Theory*, 1st ed. Cambridge University Press, 2006, pp. 16–17.

[30] ——, *Introduction to Coding Theory*, 1st ed. Cambridge University Press, 2006, pp. 183–204.

[31] T. Cover and J. Thomas, *Elements of Information Theory*, 1st ed. New York: Wiley, 2006, pp. 284–285.

[32] I. I. Dumer and P. G. Farrell, “Erasure Correction Performance of Linear Block Codes,” *Springer Lecture Notes in Computer Science, Algebraic Coding*, vol. 781, pp. 316–326, 1994.

[33] F. Didier, “A New Upper Bound on the Block Error Probability After Decoding Over the Erasure Channel,” *IEEE Transactions on Information Theory*, vol. 52, no. 10, pp. 4496–4503, 2006.

[34] D. S. Mitrinovic and P. M. Vaic, *Analytic Inequalities*, 1st ed. Springer-Verlag, 1970.