An Upper Bound on the Zero-Error List-Coding Capacity

Erdal Arikan

Abstract—We present an upper bound on the zero-error list-coding capacity of discrete memoryless channels. Using this bound, we show that the list-3 capacity of the 4/3 channel is at most 0.3512 b, improving the best previous bound. The relation of the bound to earlier similar bounds, in particular, to Körner's graph-entropy bound, is discussed.

Index Terms—Zero-error capacity, list-coding, perfect-hashing, graph-entropy, Shannon capacity of graphs.

I. INTRODUCTION

In ordinary point to point communications, the communication system delivers to the destination a single estimate of the transmitted message. Such a system is said to be a zero-error system if the estimate is always correct. Zero-error systems of this type were first studied by Shannon [1], Elias [2] considered a more general type of system in which $L$ estimates $(L \text{ fixed})$ of the transmitted message are delivered to the destination and an error is said to occur if and only if all $L$ estimates are wrong.

The major problem of information-theoretic interest about such systems is to determine the zero-error list-$L$ capacity $C_L$, i.e., the highest possible rate of communication under the zero error list-$L$ condition. Unfortunately, no formula or algorithm is known for computing $C_L$. The aim of this correspondence is to give an upper bound on $C_L$.

We consider a system consisting of a finite discrete memoryless channel $K$ with input alphabet $I$, output alphabet $J$, and transition probability matrix $[P(j|i)]$, where $P(j|i)$ is the probability that output letter $j$ is received when input letter $i$ is transmitted. We write $P_N(y|x)$ to denote the probability that $y \in J^N$ is received when $x \in I^N$ is transmitted; since the channel is memoryless, $P_N(y|x) = \prod_{t=1}^{N} P(y_t|x_t)$.

A block code $\mathcal{C}$ is employed in the system, mapping $M$ messages into codewords $x(1), \ldots, x(M)$, with each codeword a sequence of length $N$ from $I$. When a codeword is transmitted through $K$, the receiver observes the channel output $y$, and generates the list $\mathcal{L}(y) = \{m : P_N(y|x(m)) > 0\}$ of all messages that may have been transmitted. $\mathcal{C}$ is called a list-$L$ code if, for each $y$, $\mathcal{L}(y)$ contains at most $L$ messages. Thus, for a list-$L$ code, the receiver can identify the transmitted message as one of at most $L$ alternatives.

In general, the codewords of a list-$L$ code do not have to be distinct. However, in a list-$L$ code at most $L - 1$ codewords can be identical to any given codeword. So, if we discard repeated codewords from a list-$L$ code, the size of the code is reduced at most by a factor of $1/L$. Since we shall be interested in asymptotic code rates for fixed $L$, there is no loss of generality in assuming, as we shall do henceforth, that all codewords in the codes under consideration are distinct. (This allows identification of codewords with messages and simplifies the notation considerably.)

The list-$L$ capacity of $K$ is defined by

$$C_L = \limsup_{N \to \infty} \frac{1}{N} \log M(N, L)$$

where $M(N, L)$ is the maximum possible size for a list-$L$ code of length $N$.

The upper bound on $C_L$ given in this correspondence is an extension of earlier bounds by Shannon [1], Elias [2], Fredman and Komlós [3], Körner [4], Körner and Marton [5], [6]. These bounds have in common the use of the information-theoretic mutual information function.

To obtain the basic mutual-information bound on $C_L$, consider the above system again. Let $\mathcal{C}$ be a list-$L$ code. Let $R = (1/N) \log M$ denote the rate of $\mathcal{C}$. Suppose a codeword $X$ is chosen equiprobably from $\mathcal{C}$ and transmitted through $K$. Let $Y$ denote the resulting channel output. Then, $NR - H(X) = H(Y|X) + I(X; Y) \leq \log L + I(X; Y)$, where the equalities follow from the definitions of entropy and mutual information functions (see, e.g., [8] for the definitions), and the inequality follows by noting that there are at most $L$ possibilities for $X$ when $Y$ is given. We may upper bound $I(X; Y)$ by $NC$ where $C$ is the ordinary Shannon capacity [8, p. 74] of $K$. Then, considering a sequence of list-$L$ codes with increasing block lengths and with rates approaching $C_L$, we obtain $C_L \leq C$.

This bound may be tightened by observing that $C_L$ depends on the transition probabilities of $K$ only through the channel adjacency function $\phi_K$, defined as follows. For any $n \geq 1$ and $S \subset I^n$,

$$\phi_K(S) = \begin{cases} 1 & \text{if there exists } y \in J^n \text{ s.t. } P_N(y|x) > 0 \text{ for all } x \in S; \\ 0 & \text{otherwise.} \end{cases}$$

Thus, $\phi_K(S) = 1$ if and only if the sequences in $S$ are adjacent in the sense that there is a common channel output sequence reachable from all of them. (Note that, since $K$ is memoryless, $\phi_K$ is determined by its values on subsets of $I$.)

It is easy to see that $\mathcal{C}$ is a list-$L$ code for $K$ if and only if $\phi_K(S) = 0$ for each $S \subset \mathcal{C}$ with more than $L$ elements. Thus, if $K'$ is any other channel with the same input alphabet as $K$ and $\phi_{K'} \leq \phi_K$, then $C_L(K') \leq C_L(K')$. This observation leads to the Shannon–Elias bound [1], [2]:

$$C_L(K) \leq \min_{K': \phi_{K'} \leq \phi_K} C(K').$$

(1)

The bound (1) turns out to be rather weak in many examples, apparently because the channel output $Y$ (whichever admissible $K'$ is considered) carries more than enough information necessary to identify the transmitted $X$ as one of $L$ possible alternatives. That list-$L$ codes fail to achieve rates as high as $C$ (unlike codes designed for an average probability of error criterion) may be attributed to the rigid combinatorial constraints that they must satisfy.

A more general framework for obtaining bounds on $C_L$, which allows exploitation of the combinatorial constraints on the structure of list-$L$ codes, is to choose $K'$ from the class of multiinput channels with side information, as we shall do in the next section and as previously done (in a different notation) in the papers

1 It is not known if the lim sup can be replaced by lim for any $L \geq 2$. For $L = 1$, this is possible [1].
In general, by a b/l channel we mean a channel \( K \) with a b-letter input alphabet \( Z \) such that the considered in Application of the same bound to arbitrary b/l channels is closely related to perfect-hashing, which is a method of information storage and retrieval (cf. [7] for a general discussion of hashing). Kőrner and Marton over a b-letter alphabet \( k \)-separated if for every \( k \) tuple of fixed \( t, b, k, \) let \( N(t, b, k) \) be a list-L code for \( K \) with \( k \) separated. It can be seen that \( \log N(t, b, k) \) is equal to the list-(\( k - 1 \)) capacity \( C_{b,k-1} \) of a b/(k - 1) channel. Thus, the bound developed in Section II readily yields upper bounds on \( C_{b,k} \), and in some distances improves earlier such bounds, as demonstrated in Section III for \( (b, k) = (4, 4) \) and in [10] for several other \( (b, k) \).

II. THE NEW BOUND

Throughout this section, let \( K \) be the channel specified in Section I. To obtain a bound on \( C_b(K) \), we consider an alternative communication system with a discrete memoryless channel \( K' \) that has input alphabet \( I' \), output alphabet \( J' \), and transition probabilities \( \{ P(j' | i', h) \} \), \( j' \in J' \), \( i' \in I' \), \( h \in H \), where \( H \) is a random variable from the uniform distribution on \( H \). We assume that the \( h \) input of the channel is provided to the receiver in the system as side-information, i.e., when \( (i, h) \) is transmitted, the receiver observes \( h \) (in addition to the channel output \( j \)). The parameters \( m \) and \( k \) are arbitrary integers satisfying \( m \geq 1 \) and \( k \geq 0 \), respectively. Let \( \mathcal{F}_{m,k} \) denote the class of all such channels for fixed \( m, k \).

A block code \( \mathcal{C}' \) of length \( N \) for a channel \( K' \in \mathcal{F}_{m,k} \) is a subset of \( I^N \), the set of \( t \) tuples over \( I^N \). We write the codewords of such a code in the form \( (x, z) = (x_1, x_2, \ldots, x_m, z_1, \ldots, z_k) \), where \( x_1, \ldots, x_m \in I^m \), \( r = 1, \ldots, m \), \( s = 1, \ldots, k \). The sequence \( x \) is transmitted via the \( r \)th input, and \( z \) via the \( s \)th input of \( K' \). When a codeword \( (x, z) \) is sent, the receiver observes the channel output \( y \) and the side-information \( z \), and produces the list \( \mathcal{L}(y, z) = \{(x', z') \in \mathcal{C}' : P_y(x', z') > 0\} \) of all possible codewords that may have been transmitted. \( \mathcal{C}' \) is called a list-L code for \( K' \) if \( \mathcal{L}(y, z) \) contains not more than \( L \) elements for every possible \( y \) and \( z \).

We introduce some notation before proceeding. Let \( T \) be a set of \( m \) tuples over \( I^N \). Let \( \mathcal{C} \) be a list over \( I^N \). We use the notation \( \phi_c(T|x) \) as a shorthand for \( \phi_c(S) \) where \( S = T \times \{x\} \). We write \( |T| \) to denote the set of all words in \( I^N \) that appear \( c \) times as coordinates of \( m \) tuples in \( T \). More precisely, if the elements of \( T \) are denoted \( x_n = (x_{n1}, \ldots, x_{nm}) \), \( x_n' \in I^N \), \( u = 1, \ldots, m \), \( v = 1, \ldots, n \), then \( |T| \) is the set of all such \( x_n \). We write \( [z] \) to denote \( (z_1, \ldots, z_k) \), the set of coordinates of \( z \). For any finite set \( S, \{ S \} \) denotes the cardinality of \( S \).

For any set \( U \subseteq I^N \) and any \( z \in I^N \), we define \( \mathcal{F}_{m,k}(U, z) \) as the set of all \( K' \in \mathcal{F}_{m,k} \) such that, for any \( T \subseteq U \) with \( |T| \geq 2 \), \( \phi_c(T|x) \leq \phi_c(T) \cup \{z\} \). Note that \( \mathcal{F}_{m,k}(U, z) \) is nonempty, always containing the trivial channel \( K' \) whose output identically equals its input.

**Lemma 1:** Let \( \mathcal{C}' \) be a list-L code for \( K, \mathcal{C}' \), any subset of \( \mathcal{C}' \), and \( z \) any point in \( \mathcal{C}' \). Then, \( \mathcal{C}' = \mathcal{C}' \times \{z\} = \{x, z : x \in \mathcal{C}'\} \) is a list-L code for \( K' \in \mathcal{F}_{m,k}(\mathcal{C}', z) \).

**Proof:** \( \mathcal{C}' \times \{z\} \) is a list-L code for \( K' \) if (and only if) \( \phi_c(T|x) = \phi_c(T) \) for every \( T \subseteq \mathcal{C}' \), with \( |T| \geq 2 \). Suppose a proof by contradiction, that there exists \( T \subseteq \mathcal{C}' \) such that \( |T| \geq 2 \) and \( \phi_c(T|x) > \phi_c(T) \cup \{z\} \). Also, \( |T| \leq |\mathcal{C}'| \), since \( K' \in \mathcal{F}_{m,k}(\mathcal{C}', z) \). But \( S \) is a subset of \( \mathcal{C}' \), a list-L code for \( K' \); so, \( \phi_c(S) = 1 \) implies \( |S| \leq L \). Also, \( |T| \leq |\mathcal{C}'| \), since \( T \) is a set of \( m \) tuples over \( S \). Thus, \( |T| \leq L^m \), a contradiction, and the proof is complete.

Let \( \mathcal{C}', \mathcal{C}' \), \( K' \) be as in the hypothesis of the lemma. Let \( X \) denote a random variable from the equiprobable distribution on \( \mathcal{C}' \), and \( Y \) the output of \( K' \) when \( (X, z) \) is transmitted. That is, suppose that \( P_x(z) = 1/|\mathcal{C}'| \) for \( z \in \mathcal{C}' \), and \( P_{y|x}(y) = P_{y|x}(y|x, z) \), where \( P_{y|x}(y|x, z) \) is the transition probability for \( K' \). Then, we have

\[
\log|\mathcal{C}'| = H(X) = H(X|Z) = H(X|Y) + I(X; Y|Z) \leq L^m + I(X; Y|Z) \tag{2}
\]

where the second equality follows by the independence of \( X \) and \( Z \) and the inequality by Lemma 1.

Inequality (2) can be used to obtain upper bounds on the size \( M \) of \( \mathcal{C}' \) by choosing particular forms for \( \mathcal{C}' \). For example, setting \( \mathcal{C}' = \mathcal{C}' \) yields \( H(X) = m \log M \). Another possibility, which has yielded better results in applications, is to set \( \mathcal{C}' = \mathcal{C}' \) where \( \mathcal{C}' = \mathcal{C}' \) and \( Y \) is distinct. Then, \( H(X) = m \log M \) where \( M \in \mathcal{F}_{m,k}(\mathcal{C}', m - i) \). The rest of the paper will be based on this latter choice with the further restriction that \( z \in \mathcal{C}' \). The result thus far can be summarized as follows.

**Proposition 1:** The size \( M \) of any list-L code \( \mathcal{C}' \) for a discrete memoryless channel \( K \) satisfies, for any \( k \geq 0, m \geq 1 \)

\[
\log M \leq m \log L + \min_{K' \in \mathcal{F}_{m,k}(\mathcal{C}', \mathcal{Z})} \min_{z \in \mathcal{C}' \setminus \mathcal{Z}} H(I(X; Y|Z)) \tag{3}
\]

where \( X \) is a random variable from the uniform distribution on \( \mathcal{C}' \) and \( P_{y|x}(y|x) = P_{y|x}(y|x, z) \) with \( P_{y|x}(y|x, z) \) the transition probability for \( K' \).

Inequality (3) represents the general form of the bound proposed in this correspondence. An equivalent bound is implicit in Kőrner's work [4]. The bound (3) is not amenable to computation due to its involved structure. In actual calculations, one finds it necessary to make the range of minimization over \( K' \) independent of \( \mathcal{C}' \). Such a simplified form of the bound is

\[
\log M \leq m \log L + \min_{K' \in \mathcal{F}_{m,k}(\mathcal{C}', \mathcal{Z})} \min_{z \in \mathcal{C}' \setminus \mathcal{Z}} H(I(X; Y|Z)) \tag{4}
\]

where \( \mathcal{Z} \) is a random variable from an arbitrary probability distribution on \( \mathcal{C}' \). By choosing the distribution of \( \mathcal{Z} \) suitably, the bound (5) may be computed relatively easily in specific instances. For example, in [6], the bound (5) was applied to \( L \)-uniform channels with \( Z \) from the uniform distribution on \( \mathcal{Z} \).
I work is the demonstration of this fact. For the 4/3 channel IEEE bound on its list-3 capacity that improves all previous bounds, in (3), we derive a bound on its list-3 capacity that improves all previous bounds, in particular, the bound (5) with Z uniform.

To end this section, let us note that the Shannon–Elias bound (1) is a special case of (4) with m = 1, k = 0. Let us also note that, due to the memoryless property of the channels involved, the term $\sum_{x(\ast)} P_{\mathcal{X}}(x(\ast))$ can be upper bounded by $S_{\mathcal{X}}(\ast)$. Thus, $H(Y) = H(Y|I) + I(X; Y|I)$.

III. The 4/3 Channel

In this section, we consider a 4/3 channel $K$, and apply the bound (4) to show that its list-3 capacity satisfies $C_3 \leq 0.35212$. This improves the best previous bound $C_3 \leq 0.385$, which was obtained by applying (5) with $m = 1$, $k = 0$, and $Z$ uniform. This demonstrates that choosing the random variable $Z$ in (5) from a nonuniform distribution in particular, concentrating it on a single point as in (6), may yield better bounds, as might be expected. In the following, all rates will be in bits and all logarithms to base 2.

The combinatorial property characterizing list-3 codes for a 4/3 channel is that for any four distinct codewords $x_1, x_2, x_3, x_4$, there exists a coordinate $n$ such that $x_1_1, x_2_1, x_3_1, x_4_1$ are distinct. To obtain a bound on $C_3$, we employ the method of Section II with a set of parameters $K'$ from $S$. Thus, the inputs of $K'$ are of the form $(i_1, h_1, h_2) \in I^3$, where $I$ denotes the input alphabet of $K$, and the inputs $h_1, h_2$ are provided as side-information at the channel output. We specify the output alphabet of $K'$ as $\cup (e)$ where $e$ is a symbol not contained in $I$, and its transition probabilities as follows:

$$P(j_i, h_1, h_2) = \begin{cases} \delta_j & \text{if } h_1 = h_2, \\ 1/2 & \text{if } h_1 \neq h_2, \\ 1 & \text{if } (i_1, h_1, h_2, j) = I. \end{cases}$$

Lemma 2: $K'$ specified above belongs to $\mathcal{X}_2^3$.

Proof: Let $\mathcal{X}'$ be an arbitrary list-3 code for $K$, and $z = (z_1, z_2)$ an arbitrary point in $\mathcal{X}'$. We must show that, for every $T \subseteq \mathcal{X}'$ with $|T| \geq 2$, $\phi_K(T) \leq \phi_K(S)$, where $S = \{0, 1\} \cup I$. We only need consider $T$ for which $\phi_K(S) = 0$. Any such $T$ contains at least two codewords $x_1, x_2$ such that $x_1_1, x_2_1, x_2$ are distinct. So, by the defining property of list-3 codes, there exists a coordinate $n$ such that $x_1_1, x_2_1, x_1_2, x_2_2$ are distinct. Hence, by the way $K'$ has been specified, $\phi_K((x_1, x_2, k_1), x_1_2, x_2_2) = 0$. This implies $\phi_K(T) = 0$ (since $(x_1, x_2)$ is a subset of $T$), completing the proof.

Henceforth fix $\mathcal{X}'$ as a list-3 code for $K$ and $z = (z_1, z_2)$ as a point in $\mathcal{X}'$. Let $N$ be the length, $M$ the size, $R$ the rate of $\mathcal{X}'$. Let $X$ be a random variable equiprobable on $\mathcal{X}'$, and $Y$ the random variable observed at the output of $K$ when $(X, z)$ is transmitted. Thus, $P_{X|Y}(x|y) = (1/M)P_{X}(x|z)$, for $x \in \mathcal{X}'$.

(4), the rate of $\mathcal{X}'$ satisfies

$$NR \leq \log 3 + I(X; Y').$$

In the rest of this section, we develop an upper bound on $I(X; Y)$.

For any two sequences $u_1, u_2$ of equal length, let $d(u_1, u_2)$ denote the number of coordinates $n$ such that $u_1_n \neq u_2_n$, the Hamming distance. Likewise, for any three sequences $u_1, u_2, u_3$ of equal length, let $d(u_1, u_2, u_3)$ denote the number of coordinates $n$ such that $u_1_n, u_2_n, u_3_n$ are distinct.

Lemma 3: If $I(X; Y) \leq \sum_{z \in \mathcal{Z}} P_{X}(z_1)P_{X}(z_2)\sum_{z_1, z_2} I_{\mathcal{X}}(z_1, z_2)$, $I_{\mathcal{X}}(z_1, z_2)$ is the empirical distribution of the nth coordinate of the codewords in $\mathcal{X}'$, i.e.,

$$Q_n(i) = \frac{1}{M} \sum_{x \in \mathcal{X}'} \delta_{x_n}(i).$$

Proof: For coordinates $n$ where $z_{1_n} = z_{2_n}$, we have $Y_n = e$. For $z_{1_n} \neq z_{2_n}$, $Y_n$ can take one of at most two values. So, the number of possible values of $Y$ is at most $2^d(1, z_1, z_2)$. This gives $H(Y) \leq d(1, z_1, z_2)$. On the other hand, for each $x, y$ we have either $P_{X}(y|x, z) = 0$ or $P_{X}(y|x, z) = 2^{-d(1, x, z)}$.

Thus, $H(Y|x, z) = d(1, z_1, z_2) - \sum_{z_1, z_2} I_{\mathcal{X}}(z_1, z_2)d(z_1, z_2)$.

Since $H(Y|x, z) = H(Y|z) - I(Y; z)$, the inequality follows. The proof is complete by noting that $d(x, z_1, z_2) = \sum_{x(\ast)} I_{\mathcal{X}}(x_1, z_1, z_2)$ and $\sum_{z_1, z_2} I_{\mathcal{X}}(z_1, z_2) = 1 - Q_n(z_1) - Q_n(z_2)$ for $z_1, z_2$.

Lemma 3 and (6) give the following constraint on the rate and composition of $\mathcal{X}'$:

$$NR \leq \log 3 + \sum_{n=1}^{N} (1 - Q_n(z_1) - Q_n(z_2))d(z_1, z_2).$$

(7)

To obtain a tight bound on $C_3$, using (7), we need to show that $\mathcal{X}'$ can be chosen with rate close to $C_3$ and with $Q_n(i)$ not too small for any $n, i$.

Lemma 4: Given any $\epsilon > 0$, there exist list-3 codes (for the 4/3 channel) of arbitrarily large lengths, with rates $\geq C_3 - \epsilon$, and for which $Q_n(i) \geq 1 - 2^{-(C_3-2\epsilon)}$ for all $n, i$.

Proof: For any $\epsilon > 0$, there exists a finite integer $N_0$ such that every list-3 code with length $N \geq N_0$ has rate $\geq C_3 + \epsilon$, this follows from the definition of $C_3$. Fix $\epsilon > 0$, and consider a list-3 code $\mathcal{X}'$ with rate $R \geq C_3 - \epsilon$ and length $N > 3N_0$. The existence of such a code for arbitrarily large $N$ is also guaranteed by the definition of $C_3$.

If there exist $n, i$ such that $Q_n(i) < 1 - 2^{-(C_3-2\epsilon)}$, consider the subcode $\mathcal{X}' = \{x \in \mathcal{X}' : x_1 = i\}$. $\mathcal{X}'$ is a list-3 code (any subcode of $\mathcal{X}'$ is a list-3 code) with $M \geq |M| - 1 - Q_n(i) \geq 2^{-N(C_3-2\epsilon)}$ codewords, where $M = 2^N$ is the number of codewords in $\mathcal{X}'$. Let $\mathcal{X}_1$ be the code obtained by deleting the nth coordinate of each codeword in $\mathcal{X}'$. $\mathcal{X}_1$ has length $N - 1$, and it is easy to see that it is also a list-3 code for the 4/3 channel. Thus, $\mathcal{X}_1$ has M1 codewords and rate $R_1 = [1/(N-1)]\log M_1 > C_3 - \epsilon + \epsilon/(N-1)$. Since $R_1 > C_3 - \epsilon$, we may iterate the above procedure with $\mathcal{X}_1$ in place of $\mathcal{X}'$. At that end the kth round, we shall have a code $\mathcal{X}_k$ with length $N_k = N - k$, number of codewords $M_k > M_k^{2^{-k(C_3-2\epsilon)}}$, and rate $R_k > C_3 - \epsilon + k\epsilon/(N - k)$. If this process could continue for more than $2N/3$ rounds, at round $k = [2N/3]$ we would have a code with length $N_k = N - k > N/3$, rate $R_k > C_3 - \epsilon$, and for which $Q_n(i) \geq 1 - 2^{-(C_3-2\epsilon)}$ for all $n = 1, \ldots, N_k$ and all $i \in I$. Since $N/3$ can be arbitrarily large, this completes the proof.
Proposition 2: The list-3 capacity of the 4/3 channel satisfies $C_3 \leq 0.3512$ b.

Proof: Let $\epsilon > 0$ be arbitrary and consider a list-3 code $\mathcal{C}'$ with rate $R \geq C_3 - \epsilon$ and $\mathcal{Q}(i) \geq 1 - 2^{-\left(C_3 - 2\epsilon\right)}$ for all $n, i$. By Lemma 4, such a code exists and its length $N$ can be assumed arbitrarily large. Substituting the parameters for this code into $C$, (71), we obtain

$$N(C_3 - \epsilon) \leq \log 3 + (2 - 2^{-\left(C_3 - 2\epsilon\right)}) - 1) d(z_1, z_2).$$

Let $d(\mathcal{C}') = \min \{d(z_1, z_2) : z_1, z_2 \in \mathcal{C}', z_1 \neq z_2\}$ and let

$$\delta(R) = \limsup \{d(\mathcal{C}')/N : N \to \infty\}.$$