An improved bound on the zero-error list-decoding capacity of the 4/3 channel

Marco Dalai
University of Brescia
Brescia, Italy
marco.dalai@unibs.it

Venkatesan Guruswami
Carnegie Mellon University
Pittsburgh, USA.
venkatg@cs.cmu.edu

Jaikumar Radhakrishnan
Tata Institute of Fundamental Research
Mumbai, India.
jaikumar@tifr.res.in

***Abstract*—We prove a new, improved upper bound on the size of codes $C \subseteq \{1,2,3,4\}^n$ with the property that every four distinct codewords in C have a coordinate where they all differ. Specifically, we show that such a code has size at most $2^{6n/19+o(n)}$, or equivalently has rate bounded by $6/19 \leq 0.3158$ (measured in bits). This improves the previous best upper bound of 0.3512 due to (Arikan 1994), which in turn improved the 0.375 bound that followed from general bounds for perfect hashing due to (Fredman and Komlós, 1984) and (Körner and Marton, 1988). The context for this problem is two-fold: zero-error list decoding capacity, where such codes give a way to communicate with no error on the "4/3 channel" when list-of-3 decoding is employed, and perfect hashing, where such codes give a perfect hash family of size n mapping C to $\{1,2,3,4\}$.**

I. Introduction

Shannon introduced the concept of zero error capacity of a discrete noisy channel [1], also referred to as the Shannon capacity of a graph. Such a channel can be modeled as a bipartite graph $H=(V,W,E)$, with V corresponding to channel inputs, W to channel outputs, where $(v,w) \in E$ if w can be received at the channel output when v is transmitted on the channel. One can associate a "confusability" graph $G=(V,E')$ with such a channel, where $(v_1,v_2) \in E'$ if there is a common output $w \in W$ such $(v_1,w),(v_2,w) \in E$, so that v_1, v_2 can be confused with each other. The zero error capacity of the channel is the largest asymptotic rate at which information can be transmitted with no error on the channel, in n independent uses of the channel for large n. This quantity is also called the Shannon capacity of the graph G, which is the limiting ratio of $(\log_2 \alpha(G^n))/n$ where $\alpha(G^n)$ is the size of the largest independent set in the n'th power G^n of G, where two n-tuples in V^n are adjacent if in every coordinate, they are either equal or adjacent in G. Lovász proved that the Shannon capacity of the 5-cycle, which is the smallest non-trivial case, is $\log_2 \sqrt{5}$ by introducing his influential Theta function [2].

In this work, we study the zero error capacity in the model of *list decoding*, for a basic channel whose Shannon capacity is trivially 0. The zero error *list decoding* capacity was introduced by Elias [3]. For a fixed L, the list-of-L zero error capacity of a channel H is the largest asymptotic rate at which one can communicate on the channel (over n independent uses for growing n) so that the decoder can pin down the correct message to one of at most L possibilities (in other words, the decoder can output L codewords which must include the transmitted one). More formally, for a channel $H=(V,W,E)$, a *code* $C \subseteq V^n$ is said to achieve zero error under list-of-L decoding if for every subset $\{c^{(1)}, c^{(2)}, \ldots, c^{(L+1)}\}$ of $L+1$ codewords of C, there is a coordinate i such that the symbols $c_i^{(1)}, c_i^{(2)}, \ldots, c_i^{(L+1)}$ don't share a common neighbor in W. Equivalently, C is an independent set in the $(L+1)$-uniform hypergraph defined on V^n where hyperedges correspond to tuples whose i'th symbols have a common neighbor for every i. (Note that the case $L=1$ corresponds to Shannon's zero error capacity.)

The smallest non-trivial case for zero error list decoding capacity is the *3/2 channel*, where $V=W=\{1,2,3\}$ and $(v,w) \in E$ iff $v \neq w$. Since every pair of input symbols can be confused with each other, the Shannon capacity of this channel is 0. However, there exists a code $C \subseteq \{1,2,3\}^n$ of rate R bounded away from 0 (i.e., of size 2^{Rn}) which permits list-of-2 decoding with no error on the 3/2 channel. The best known lower bound on R (to our knowledge) approaches $\frac{1}{4}\log_2 \frac{9}{5} \approx 0.212$ [4]. There is an easy upper bound of $\log_2(3/2)$ on the rate of such a code, which in fact holds for list-of-L decoding for any fixed L (or even $L \leq 2^{o(n)}$); the argument is just to take a random output sequence $w \in W^n$ and compute the expected fraction of codewords that are consistent with receiving w. As a side remark, we mention that the quantity $\log_2(3/2)$ equals the zero error capacity for list-of-2 decoding in the presence of noiseless feedback from the receiver to sender [3].

The list-of-2 decoding setting for the 3/2 channel is completely equivalent to a question about perfect hash families. To achieve zero error with a list of size 2, the code $C \subseteq \{1,2,3\}^n$ should have the property that every triple of codewords has a coordinate where they all differ. The existence of such a code of cardinality N is thus equivalent to the existence of a perfect hash family of size n that maps a universe of size N to $\{1,2,3\}$ such that every three elements of the universe are mapped in a one-one fashion by at least one hash function. In this setting,

This research was supported by NSF grants CCF-1422045 and CCF-1563742, and by Italian Ministry of Education under grant PRIN 2015 D72F16000790001. The work was partly done while the authors were visiting the Simons Institute for the Theory of Computing at UC Berkeley, whose support is gratefully acknowledged.

we have a lower bound of $\log_{3/2} N$ on the size of such hash families, and it is a longstanding open problem to improve this. The bounds of Fredman and Komlós [5] and follow-ups (discussed further in Section II), give improvements for hashing into sets of size 4 and higher, but do not apply for hashing into $\{1,2,3\}$. It remains a major open question to improve the bound for the 3-element alphabet.

In this work, we address the perfect hashing problem into a range of size 4, or equivalently the zero error list-of-3 decoding capacity for the *4/3 channel* where $V=W=\{1,2,3,4\}$ and $(v,w)\in E$ iff $v\neq w$. For this channel, the zero error capacity is clearly 0 for list size 2, since any three input symbols share a common output symbol. Let us say that $C\subseteq\{1,2,3,4\}^n$ is a $4/3$ code if for every four distinct codewords $c^{(1)}$, $c^{(2)}$, $c^{(3)}$, and $c^{(4)}$ of C there is a coordinate $i\in\{1,2,\ldots,n\}$ for which $\{c_i^{(1)},c_i^{(2)},c_i^{(3)},c_i^{(4)}\}=\{1,2,3,4\}$. This is the exact criterion a code needs to meet in order to achieve zero error on the 4/3 channel with list-of-3 decoding. A simple random coding argument [4] shows the existence of $4/3$ codes of rate approaching $\frac{1}{3}\log_2\frac{32}{29}\approx 0.0473$ which is pretty small. The simple "random received word" argument mentioned above for the 3/2 channel shows an upper bound on capacity of $\log_2(4/3)$ in the case of 4/3 channel (this equals the zero error capacity with feedback).

In case of the 4/3 channel, an upper bound on capacity that is smaller than the simple $\log_2(4/3)\approx 0.415$ bound *is* known. The results of Fredman and Komlós on perfect hashing, when specialized to domain size 4, imply an upper bound of $3/8{=}0.375$ [5]. Körner and Marton [4] improved the Fredman-Komlós bounds using a hypergraph (as opposed to graph) covering approach, but did not get an improvement for alphabet size 4. Arikan improved the capacity upper bound for the $4/3$ channel to 0.3512 [6]. In this work, introducing some new ideas, we further improve the bound to $6/19<0.3158$:

Theorem 1: The size of a $4/3$ code $C\subseteq\{1,2,3,4\}^n$ satisfies $|C|\leq 2^{6n/19+o(n)}$.

In Section II, we discuss the techniques used in the earlier works [5], [6] and the novelty in our contribution, while in Section III we give the proof.

Notation. We call $\Sigma=\{1,2,3,4\}$ and, for general integer $n\geq 1$, $[n]=\{1,2,\ldots,n\}$. If $x\in\Sigma^n$ then x_i is the i-th component of x and, by extension, $x_{[k]}=(x_1,x_2,\ldots,x_k)$. All logarithms are to the base 2.

II. Background

The previous upper bounds on the rate of 4/3 codes (due to Fredman and Komlós [5] and Arikan [6]), as well as our new upper bound, can be based on an information theoretic inequality regarding graph covering. This inequality due to Hansel [7] has been rediscovered several times (see Krichevskii [8], Katona and Szemerédi [9], Pippenger [10], Fredman and Komlós [5], Körner and Marton [4]), and is a special case of the subadditivity property of Körner's graph entropy ([11], [12]).

Lemma 2 (Hansel [7]): Let K_r be the complete graph with vertex set $[r]$. Let I be a set of indices, and for $i\in I$, let G_i be a bipartite graph with vertex set $[r]$; let τ_i be the fraction of vertices in $[r]$ that appear non-isolated in G_i. Suppose $\bigcup_{i\in I}E(G_i)=E(K_r)$. Then,

$$\sum_{i\in I}\tau_i\geq\log_2 r \quad .$$

Let us recall how graph covering enters the discussion on 4/3 codes. Fix a 4/3 code $C\subseteq\Sigma^n$. Let x and x' be two codewords in C. Let $K^{x,x'}$ be the complete graph with vertex set $C\setminus\{x,x'\}$. For $m\in[n]$, let $G_m^{x,x'}$ be the graph with vertex set $C\setminus\{x,x'\}$ and edge set

$$E(G_m^{x,x'})=\{(y,y'):\{x_m,x'_m,y_m,y'_m\}=\Sigma\}\,.$$

It follows immediately from the definition of a 4/3 code that $\bigcup_{m\in[n]}G_m^{x,x'}=K^{x,x'}$; if we denote the fraction of non-isolated vertices in $G_m^{x,x'}$ by $\tau_m(x,x')$, then Hansel's lemma implies that

$$\sum_{m\in[n]}\tau_m(x,x')\geq\log(|C|-2). \tag{1}$$

To obtain a good upper bound on the rate of C, one would like to show that the left hand side of the above inequality is small. There are two ways in which $\tau_m(x,x')$ might be small for a choice of x and x': (1) if $x_m=x'_m$, then $\tau_m(x,x')=0$, so it is advantageous to pick x and x' that agree on a lot of coordinates; (2) if $x_m\neq x'_m$, then any codeword in $C\setminus\{x,x'\}$ that agrees with either x or x' in the m-th position will appear isolated in $G_m^{x,x'}$, so it is advantageous to pick x and x' that take the most popular values in the m-th coordinate.

Fredman and Komlós [5] and Arikan [6] exploit (1) in different ways, by devising different strategies for choosing x and x'. We review their approaches below, and pinpoint how our new analysis departs from theirs.

a) The Fredman-Komlós bound:: The approach of Fredman and Komlós [5] amounts to picking x and x' at random (without replacement) from C. It can be shown that for each m, $\mathbb{E}[\tau_m(x,x')]$ is at most $\frac{3}{8}(1+o(1))$. It then follows immediately from (1) that

$$|C|\leq 2^{\frac{3}{8}(1+o(1))n}.$$

In this approach, the two ways in which $\tau_m(x,x')$ can be made small are addressed simultaneously by the random choice of x and x'. By reducing the problem to hypergraph covering instead of graph covering, Körner and Marton [4] improve upon the Fredman-Komlós bound for perfect hashing for certain values of parameters; however, their method yields no improvement for 4/3 codes.

b) The Arikan bound:: Arikan's approach [6], on the other hand, places greater emphasis on ensuring that x and x' agree on many coordinates. Indeed, standard bounds in coding theory let us conclude that codes with non-trivial rate must have codewords that agree in significantly more coordinates than randomly chosen codewords. Arikan combines this insight with an ad hoc balancing argument that lets one bound $\tau_m(x,x')$ non-trivially even when $x_m \neq x'_m$. To obtain the best bound, one must balance parameters using the best results in the literature on rate versus distance (e.g., the Plotkin bound) for codes over $\{1,2,3,4\}$. Arikan [6] while using the Plotkin bound to derive the bound of 0.3512 for 4/3 codes , observes that it should be possible to derive better bounds using stronger trade-offs between rate and distance that are now available. In fact, combining Arikan's approach with one of the JPL (linear programming) bounds from Aaltonen [13], we can confirm using a computer supported calculation that a bound 0.3276 can be derived; perhaps, more complicated calculations can yield somewhat better bounds.

c) Our contribution:: We combine insights from the above approaches, but look deeper into how two codewords with small distance are obtained. In particular, we examine the standard argument that leads to the Plotkin bound more closely. This involves fixing a rich subcode of codewords with a common prefix and picking two distinct codewords (say, x and x') at random from this subcode. On the other hand, instead of concluding that this process on average yields codewords that agree on many coordinates, we directly estimate the expected contribution to the left hand side of (1), that is $\mathbb{E}[\tau_m(x,x')]$. It is crucial for our proof that we do not focus on one subcode but average over all of them. We need a technical balance condition on symbol frequencies in each codeword position in our formal justification that certain functions we encounter are concave. A simple calculation, similar to what Arikan also needed, can be used to justify this balance assumption. Our calculations do not require any non-trivial computation and are completely self-contained.

We anticipate, though this is by no means an easy extension, that further improvements to the bounds may be possible by combining our approach with ideas underlying the Elias bound in coding theory. (The Elias bound works by intersecting the code with Hamming balls rather than subsets with a common prefix used in the Plotkin bound.) We also hope that our ideas will renew interest into the many challenges in this area, including the most significant challenge of obtaining better upper bounds on the rate of $3/2$ codes.

III. Rate upper bound for 4/3 codes

Let us recap the definition of the central object of interest.

Definition 3: A code $C \subseteq \Sigma^n$ is said to be a $4/3$ code if for every subset of four distinct codewords $x,y,z,w \in C$, there exists a coordinate $i \in \{1,2,\ldots,n\}$ such that $\{x_i,y_i,z_i,w_i\} = \Sigma$.

In this section, we will prove our main theorem, restated below.

Theorem 4: Let $C \subseteq \Sigma^n$ be a $4/3$ code. Then $|C| \leq 2^{6n/19+o(n)}$.

We prove the above theorem in three steps. First, will prove the theorem under an assumption that no coordinate is very skewed in terms of the distribution of codeword symbols in that coordinate (Section III-A). For this we utilize a technical concavity result which we state and prove in Section III-C. A simple argument reduces the general case to the situation where there is no skewed coordinate (Section III-B).

A. The balanced case

For a code $C \subseteq \Sigma^n$ and $m \in [n] := \{1,2,\ldots,n\}$, let $\mathsf{f}_m \in \mathbb{R}^4$ be the frequency vector that records for each letter of the alphabet, the fraction of codewords in C that contain that letter in the m-th coordinate; that is, for $a \in \Sigma$,

$$\mathsf{f}_m[a] := \frac{1}{|C|}|\{x \in C : x_m = a\}|. \tag{2}$$

(Note we suppress the dependence on C in the notation f_m for notational simplicity.)

Lemma 5: Let $C \subseteq \Sigma^n$ be a $4/3$ code (for some $n \geq 4$). Suppose for all $m \in [n]$ and $a \in \Sigma$, we have $\mathsf{f}_m[a] \geq \frac{1}{6}$. Then, $|C| \leq 2^{6n/19+o(n)}$.

Proof: Let $M := |C| = 2^{R_0 n}$, $\ell = \lceil R_0 n/2 - \log n - 1 \rceil$, and $S = [n] \setminus [\ell]$. For each prefix $w \in \Sigma^\ell$, consider the subcode

$$C_w := \{z \in C : z_{[\ell]} = w\};$$

let $M_w := |C_w|$. Then, $C = \bigcup_w C_w$ and $M = \sum_w M_w$. We partition the set of prefixes into two sets:

$$\mathsf{Heavy} = \{w : M_w \geq n\}; \qquad \mathsf{Light} = \{w : M_w < n\}.$$

Let $C^+ = \bigcup_{w \in \mathsf{Heavy}} C_w$, and $C^- = C \setminus C^+$. We have,

$$|C^-| \leq \sum_{w \in \mathsf{Light}} M_w \leq \sum_{w \in \mathsf{Light}} n \leq 4^\ell n \leq 4^{R_0 n/2 - \frac{1}{2}\log n} = |C|/n,$$

and therefore, for a random z uniformly distributed over C,

$$\Pr[z \in C^+] \geq 1 - \frac{1}{n}.$$

Let x and x' be two random codewords in C^+ generated as follows. First pick x uniformly at random from C^+; let $w = x_{[\ell]}$. Next, pick x' uniformly from $C_w \setminus \{x\}$ (which is non-empty). With this (random) choice of x and x' consider the bipartite graph $G_m^{x,x'}$ with vertex set $C \setminus \{x,x'\}$ and edge set $\{(y,y') : \{x_m, x'_m, y_m, y'_m\} = \Sigma\}$. Since C is a 4/3 code, we have

$$\bigcup_{m \in S} G_m^{x,x'} = K^{x,x'},$$

and the situation is ripe for using Hansel's lemma. The fraction of non-isolated vertices in $G_m^{x,x'}$ is precisely

$$\tau_m(x,x') := \left(\frac{|C|}{|C|-2}\right)(1 - \mathsf{f}_m[x_m] - \mathsf{f}_m[x'_m])\,\mathbf{1}\{x_m \neq x'_m\}, \tag{3}$$

where $\mathbf{1}\{x[m]\neq x'[m]\}$ is the indicator random variable for the event $x[m]\neq x'[m]$. By (1) we have

$$\log_2(M-2)\leq \sum_{m\in S}\tau_m(x,x').$$

Taking expectations over the choices of (x,x'), we obtain

$$\log_2(M-2)\leq \sum_{m\in S}\mathbb{E}[\tau_m(x,x')]. \tag{4}$$

We will estimate each term of the sum separately.

Claim 1: For each $m\in S$, we have

$$\mathbb{E}[\tau_m(x,x')]\leq \left(\frac{3}{8}\right)(1+o(1))\ . \tag{5}$$

Let us first assume this claim and complete the proof of the lemma. We have from (4) and the above claim that

$$\begin{aligned}\frac{\log_2(M-2)}{1+o(1)}&\leq (n-\ell)\left(\frac{3}{8}\right)\\ &\leq \left(n-\frac{R_0 n}{2}+\log(2n)\right)\left(\frac{3}{8}\right)\\ &\leq n\left(1-\frac{R_0}{2}\right)\left(\frac{3}{8}\right)+\log(2n)\end{aligned}$$

Recalling that $M=|C|=2^{R_0 n}$, the above implies that

$$R_0\leq \frac{3}{8}\left(1-\frac{R_0}{2}\right)+o(1)\ ,$$

This yields $R_0\leq \frac{6}{19}+o(1)$, as desired.

We still need to establish Claim 1.

a) Proof of Claim 1: For $m\in S$, let $\mathsf{f}_{m|w}$ be the frequency vector of the m-th coordinate in the subcode C_w. Note that $\mathbb{E}_W[\mathsf{f}_{m|W}]=\mathsf{f}_m$ if W is the random prefix $W=z_{[\ell]}$ induced by a z taken uniformly at random from C. Fix m. Now, for each $w\in\mathsf{Heavy}$, taking expectations over x,x' in (3), we obtain,

$$\begin{aligned}&\mathbb{E}[\tau_m(x,x')\,|\,x\in C_w]\leq\\ &\frac{|C|}{|C|-2}\cdot\frac{n}{n-1}\sum_{(a,b):a\neq b}\mathsf{f}_{m|w}(a)\mathsf{f}_{m|w}(b)(1-\mathsf{f}_m(a)-\mathsf{f}_m(b)),\end{aligned}$$

where the adjustment by the $\frac{n}{n-1}$ factor arises because x,x' are sampled without replacement from C_w, and $|C_w|\geq n$ for $w\in\mathsf{Heavy}$.

For probability vectors $f,g\in\mathbb{R}^4$, let

$$\phi(f,g):=\sum_{(i,j):i\neq j}f[i]f[j](1-g[i]-g[j]). \tag{6}$$

We thus have, for $w\in\mathsf{Heavy}$,

$$\mathbb{E}[\tau_m(x,x')\,|\,x\in C_w]\leq\left(\frac{|C|}{|C|-2}\right)\left(\frac{n}{n-1}\right)\phi(\mathsf{f}_{m|w},\mathsf{f}_m). \tag{7}$$

Let W be the random variable equal to $z_{[\ell]}\in\Sigma^\ell$ for a random z uniformly distributed over C and chosen independently of x (note that unlike x, which is picked from C^+, z is picked from the full code C). Taking expectations over W in (7), and conditioning on $W\in\mathsf{Heavy}$, we have

$$\begin{aligned}&\mathbb{E}_{W,x,x'}[\tau_m(x,x')\,|\,x\in C_W\ \wedge\ W\in\mathsf{Heavy}]\leq\\ &\left(\frac{|C|}{|C|-2}\right)\left(\frac{n}{n-1}\right)\mathbb{E}_W[\phi(\mathsf{f}_{m|W},\mathsf{f}_m)\,|\,W\in\mathsf{Heavy}]\ .\end{aligned} \tag{8}$$

Now note that the left hand side of (8) is simply $\mathbb{E}_{x,x'}[\tau_m(x,x')]$, so we have

$$\begin{aligned}&\mathbb{E}[\tau_m(x,x')]\\ &\leq\left(\frac{|C|}{|C|-2}\right)\left(\frac{n}{n-1}\right)\mathbb{E}_W[\phi(\mathsf{f}_{m|W},\mathsf{f}_m)\,|\,W\in\mathsf{Heavy}]\ .\end{aligned} \tag{9}$$

Now, using (9)

$$\begin{aligned}&\mathbb{E}_W[\phi(\mathsf{f}_{m|W},\mathsf{f}_m)]\\ &\quad\geq\Pr[W\in\mathsf{Heavy}]\cdot\mathbb{E}_W[\phi(\mathsf{f}_{m|W},\mathsf{f}_m)\,|\,W\in\mathsf{Heavy}]\\ &\quad\geq\Pr[z\in C^+]\cdot\left(\frac{|C|-2}{|C|}\right)\left(\frac{n-1}{n}\right)\mathbb{E}[\tau_m(x,x')]\,.\end{aligned}$$

As $\Pr[z\in C^+]\geq 1-1/n$, we have

$$\begin{aligned}\mathbb{E}[\tau_m(x,x')]&\leq\left(\frac{|C|}{|C|-2}\right)\left(\frac{n}{n-1}\right)^2\mathbb{E}_W[\phi(\mathsf{f}_{m|W},\mathsf{f}_m)]\\ &\leq\frac{3}{8}(1+o(1))\qquad,\end{aligned}$$

where the last inequality follows from Lemma 7, which we state and prove in Section III-C. In our application, we set $\mathsf{f}_w\leftarrow\mathsf{f}_{m|w}$ and $\mathsf{f}\leftarrow\mathsf{f}_m$; note $\mathbb{E}_W[\mathsf{f}_{m|W}]=\mathsf{f}_m$. This completes the proof of our claim and the lemma. ■

Remark 1: There is a technical reason for choosing x,x' to be uniformly distributed over C^+ while W to be over all prefixes (i.e., $\mathsf{Heavy}\cup\mathsf{Light}$), instead of just removing C^- and only consider the subcode C^+. Indeed, removing C^- would introduce a modification of the frequencies f_m and hence the assumption that $\mathsf{f}_m[a]\geq 1/6\,,\forall a$ would not hold anymore for the subcode. On the other hand, assuming balanced frequencies with some safety margin on C, say $\mathsf{f}_m[a]\geq 1/6+1/n$ on C (as to ensure $\mathsf{f}_m[a]\geq 1/6$ on C^+) only moves the technicality to how we deal with the balancing assumption in Section III-B.

Remark 2: We point out that, despite the same coefficient $3/8$ which appears, Claim 1 is not equivalent to the bound devised by Fredman and Komlós because our x and x' are constrained to have common prefix $x_{[\ell]}=x'_{[\ell]}$, while they are picked without replacement from the whole code C in their approach.

B. Wrapping it up in general

We now remove the restriction that the codeword symbol frequencies are balanced[1].

[1]A similar argument appears in [6, Lemma 4].

Theorem 6: For all large enough n, if $C \subseteq \Sigma^n$ is a $4/3$ code, then $|C| \leq 2^{6n/19+o(n)}$

Proof: Fix a code C. We will use Lemma 5. For that, we must first ensure that the frequency vector for each coordinate is not too skewed. We ask if there is a coordinate $m \in [n]$ and $a \in \Sigma$ such that $\mathsf{f}_m[a] < \frac{1}{6}$. If there is such a coordinate m, we create a new code by deleting all codewords $x \in C$ for which $x_m = a$, and shortening the remaining codewords to the indices in $[n] \setminus \{m\}$. By repeating this process, starting with $C_0 = C$, we obtain codes $C_0, C_1, \ldots$, where $C_i \subseteq \Sigma^{n-i}$ and $|C_i| \geq (5/6)|C_{i-1}|$. Suppose the process stops after completing t steps at which point C_t is obtained. (If the process stops without completing the first step, then $t = 0$.) Then,

$$\begin{aligned} |C_0| &\leq \left(\frac{6}{5}\right)^t |C_t| \qquad (10) \\ &\leq \left(\frac{6}{5}\right)^t 4^{n-t} \\ &= 2^{2n - t(2 - \log_2(6/5))} \leq 2^{2n - 1.736t}. \end{aligned}$$

If $t \geq 0.99n$, then this gives $|C_0| \leq 2^{0.281n} \leq 2^{6n/19}$ and our claim holds. On the other hand, if $t < 0.99n$, then we may apply Lemma 5 to C_t and conclude that $|C_t| \leq 2^{(6/19)(n-t)+o(n)}$. Then, using (10), we obtain

$$\begin{aligned} |C_0| &\leq \left(\frac{6}{5}\right)^t 2^{(6/19)(n-t)+o(n)} \\ &\leq 2^{(6/19)n - [(6/19) - \log_2(6/5)]t + o(n)}. \qquad (11) \end{aligned}$$

The right hand side is at most $2^{(6/19)n+o(n)}$ because the coefficient of t is negative (since $\log_2(6/5) < 6/19$). ∎

C. Concavity of ϕ function

Recall the definition of $\phi(f, g)$ given in (6) for probability vectors $f, g \in \mathbb{R}^4$. We now establish the following concavity result that was used in the proof of Claim 1 above.

Lemma 7: Let W be a random variable taking values in a set $\mathcal{W}$. For each $w \in \mathcal{W}$, let $\mathsf{f}_w \in \mathbb{R}^4$ be a probability vector. Suppose $\mathsf{f} := \mathbb{E}_W[\mathsf{f}_w]$ is such that $\min_a \mathsf{f}[a] \geq \frac{1}{6}$. Then,

$$\mathbb{E}_W[\phi(\mathsf{f}_W, \mathsf{f})] \leq \phi(\mathsf{f}, \mathsf{f}) \leq \frac{3}{8}. \qquad (12)$$

Proof: Let $\mathsf{f} = \langle A, B, C, D \rangle$ (which we treat as a vector). Let $\Delta_w := \mathsf{f}_w - \mathsf{f} = \langle \alpha_w, \beta_w, \gamma_w, \delta_w \rangle$. Then Δ_w satisfies the following two conditions.

$$\begin{aligned} \mathbb{E}_w[\Delta_w] &= \mathbb{E}_w[\mathsf{f}_w - \mathsf{f}] \\ &= \mathbb{E}_w[\mathsf{f}_w] - \mathsf{f} = \mathbf{0}; \\ \mathbf{1} \cdot \Delta_w &= 0, \qquad (13) \end{aligned}$$

where $\mathbf{0} = (0,0,0,0)$ and $\mathbf{1} = (1,1,1,1)$. Let

$$M := (m_{ij} : i, j \in \Sigma) = \begin{pmatrix} 0 & C+D & B+D & B+C \\ C+D & 0 & A+D & A+C \\ B+D & A+D & 0 & A+B \\ B+C & A+C & A+B & 0 \end{pmatrix}.$$

Note that the off-diagonal entries $m_{ij} = 1 - \mathsf{f}[i] - \mathsf{f}[j]$. Then,

$$\begin{aligned} \phi(\mathsf{f}_w, \mathsf{f}) &= \mathsf{f}_w M \mathsf{f}_w^t \\ &= (\mathsf{f} + \Delta_w) M (\mathsf{f} + \Delta_w)^t \\ &= \phi(\mathsf{f}, \mathsf{f}) + \Delta_w M \ \mathsf{f}^t + \mathsf{f} \ M \Delta_w^t + \Delta_w M \Delta_w^t \end{aligned}$$

Since $\mathbb{E}_w[\Delta_w] = 0$, when we take expectations over w, the two middle terms drop out. Thus,

$$\mathbb{E}_W[\phi(\mathsf{f}_W, \mathsf{f})] = \phi(\mathsf{f}, \mathsf{f}) + \mathbb{E}_W[\Delta_W M \Delta_W^t].$$

To justify our claim we show that the second term $\Delta_w M \Delta_w^t$ is never positive. Let J be the 4×4 all 1's matrix, and F be the diagonal matrix with $F_{ii} = \mathsf{f}[i]$. Then,

$$M = J - FJ - JF - (I - 2F).$$

By (13), $\Delta_w J \Delta_w^t,\ \Delta_w F J \Delta_w^t,\ \Delta_w J F \Delta_w^t = 0$; thus,

$$\begin{aligned} \Delta_w M \Delta_w^t &= -\Delta_w (I - 2F) \Delta_w^t \\ &= -\left[(1-2A)\alpha_w^2 + (1-2B)\beta_w^2 \right. \\ &\quad \left. + (1-2C)\gamma_w^2 + (1-2D)\delta_w^2 \right]. \end{aligned}$$

Since, no component of $\mathsf{f} = \langle A, B, C, D \rangle$ exceeds $\frac{1}{2}$ (because all coordinates of f are at least $\frac{1}{6}$), the right hand side is never positive. This establishes the first inequality in (12).

To establish the second inequality, we check that $\phi(\mathsf{f}, \mathsf{f})$ takes its maximum value when $\mathsf{f} = \langle \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \rangle$, that is, the maximum is $\frac{3}{8}$. (Indeed, if some two components of f are not equal, replacing them both by their average will not reduce $\phi(\mathsf{f}, \mathsf{f})$.) ∎

REFERENCES

[1] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Information Theory*, vol. 2, no. 3, pp. 8–19, 1956.

[2] L. Lovász, "On the shannon capacity of a graph," *IEEE Trans. Information Theory*, vol. 25, no. 1, pp. 1–7, 1979.

[3] P. Elias, "Zero error capacity under list decoding," *IEEE Trans. Information Theory*, vol. 34, no. 5, pp. 1070–1074, 1988.

[4] J. Körner and K. Marton, "New bounds for perfect hashing via information theory," *European Journal of Combinatorics*, vol. 9, pp. 523–530, 1988.

[5] M. Fredman and J. Komlós, "On the size of separating systems and perfect hash functions," *SIAM J. Alg. Disc. Meth.*, vol. 5, pp. 61–68, 1984.

[6] E. Arikan, "An upper bound on the zero-error list-coding capacity," *IEEE Trans. Information Theory*, vol. 40, no. 4, pp. 1237–1240, 1994.

[7] G. Hansel, "Nombre minimal de contacts de fermature nécessaires pour réaliser une fonction booléenne symétrique de n variables," *C. R. Acad. Sci. Paris*, pp. 6037–6040, 1964.

[8] R. E. Krichevskii, "Complexity of contact circuits realizing a function of logical algebra," *Sov. Phys. Dokl.*, vol. 8, pp. 770–772, 1964.

[9] G. Katona and E. Szemerédi, "On a problem of graph theory," *Studia Sci. Math. Hungarica*, vol. 2, pp. 23–28, 1967.

[10] N. Pippenger, "An information-theoretic method in combinatorial theory," *Journal of Combinatorial Theory (A)*, vol. 23, pp. 99–104, 1977.

[11] J. Körner, "Coding of an information source having ambiguous alphabet and the entropy of graphs," in *Trans. 6th Prague Conference on Inform. Theory*, 1973, pp. 411–425.

[12] ——, "Fredman–Komlós bounds and information theory," *SIAM Journal on Algebraic Discrete Methods*, vol. 7, no. 4, pp. 560–570, 1986.

[13] M. Aaltonen, "A new upper bound on nonbinary block codes," *Discrete Mathematics*, vol. 83, no. 2, pp. 139–160, 1990.