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 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 provide a self-contained proof that such codes have size at most $2^{6n/19+o(n)}$, that is, rate bounded asymptotically 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). Finally, using a combination of our approach with a simple idea which exploits powerful bounds on the minimum distance of codes in the Hamming space, we further improve the upper bound to 0.31477.**

I. Introduction

Shannon introduced the concept of zero-error capacity of a discrete noisy channel [1]. 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 always 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 in $W(H)$ 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 two input symbols can be confused with each other, the Shannon capacity of this channel is 0. However, there exist codes $C \subseteq \{1,2,3\}^n$ of rate R bounded away from 0 (i.e., of size 2^{Rn}) which permit 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 $\log_2(3/2)+o(1)$ on the rate of such codes, 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, we have a lower bound of $\log_{3/2}(N/2)$ on the size of such hash families, and it is a longstanding open problem to improve this. The bounds of

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.

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 every 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 x, y, z and t of C there is a coordinate $i\in\{1,2,\ldots,n\}$ for which $\{x_i,y_i,z_i,t_i\}=\{1,2,3,4\}$. 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 rather low. 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 the 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 [6] improved the capacity upper bound for the $4/3$ channel to 0.3512.

The main contribution of this work, Theorem 3 below, is a further improvement of the bound to $6/19<0.3158$. The proof of this result, which uses a delicate probabilistic combination of the Plotkin bound in coding theory and Hansel's lemma for covering the complete graph by bipartite graphs, is self-contained. Then, we show that the bound is not tight and, invoking some rather non-trivial results on the minimum distance of codes, we provide a slightly better bound of 0.31477.

We close the introduction with a side remark. In the last years, clever applications of the polynomial method in the breakthrough work of Croot, Lev and Pach [7] and follow-ups, have led to exponential improvements in size (or equivalently in the value of the associated "capacity") for several longstanding combinatorial problems, including 3-term arithmetic progression free sets in $\{0,1,2\}^n$ by Ellenberg and Gijswijt [8], and their generalization sunflower-free sets in $\{0,1,2,\ldots,D-1\}^n$ by Naslund and Sawin [9]. These are subsets A such that for every distinct triple $x,y,z\in A$ there is a coordinate i where *exactly two* of x_i,y_i,z_i are equal (compare this to the perfect hashing requirement of having x_i,y_i,z_i all be distinct). In each of these cases, the results give the first upper bounds on capacity that are bounded away from the trivial bound of 1 (or $\log_2 D$ when measured in bits, where D is the alphabet size). Recall that for a $3/2$ code, we already have a simple upper bound of $\log_2(3/2)$ on the capacity. A straightforward adaptation of the recent methods to the setting of $3/2$ codes seems not even to yield a bound better than 3^n. It is an interesting question if the new insights can also be exploited to improve the upper bounds on rate when the known upper bound on capacity is bounded away from 1.

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 of our main result. Finally, in Section IV we show that even this bound is not tight and we give a slight numerical improvement.

Notation. Let $\Sigma=\{1,2,3,4\}$ and, for general integer $n\geq 1$, let $[n]=\{1,2,\ldots,n\}$. 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 [10] has been rediscovered several times (see Krichevskii [11], Katona and Szemerédi [12], Pippenger [13], Fredman and Komlós [5], Körner and Marton [4]), and is a special case of the subadditivity property of Körner's graph entropy (see [14], [15]).

Lemma 1 (Hansel [10]): 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,

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

We provide a proof for the reader's convenience since it is very short and it is the only tool needed for our bound.

Proof: Call A_i and B_i the two parts of non-isolated vertices in G_i. For each i, randomly delete all the vertices in A_i or in B_i, independently, each with probability $1/2$. At most one vertex can remain at the end of this process since $\bigcup_{i\in I}E(G_i)=E(K_r)$. On the other hand, the probability that $v\in[r]$ is not deleted is 2^{-t_v} where t_v is the number of graphs G_i in which v is not isolated; so the expected number of vertices that survive is $\sum_{v\in[r]}2^{-t_v}$. So we have

$$\begin{aligned}1&\geq\sum_{v\in[r]}2^{-t_v}\\&\geq r2^{\frac{1}{r}\sum_{v\in[r]}-t_v}\\&=r2^{-\frac{1}{r}\sum_{v\in[r]}t_v}\end{aligned}$$

which gives the desired result since $\frac{1}{r}\sum_{v\in[r]}t_v=\sum_i\tau_i$ ∎

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 distinct 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 analysis differs 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] and Nilli [16] 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 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 [17], we can confirm using a computer supported calculation that a bound 0.3276 can be derived (see Section IV-A below for more details); 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. 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.

III. Rate upper bound for 4/3 codes

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

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

In this section, we present and prove our main result, stated in the following theorem.

Theorem 3: As n grows unbounded, $4/3$ codes $C \subseteq \Sigma^n$ have size $|C| \leq 2^{6n/19 + o(n)}$.

We prove the above theorem in three steps. First, we 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 $m \in [n]$, 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 4: 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 $M := |C| = 2^{R_0 n}$ and $\ell = \lceil R_0 n/2 - \log n - 1 \rceil$. 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,

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

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 because $|C_w| \geq n \geq 4$.). 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

$$\bigcup_{m \in [n] \setminus [\ell]} G_m^{x,x'} = K^{x,x'},$$

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

$$\log_2(M - 2) \leq \sum_{m \in [n] \setminus [\ell]} \tau_m(x, x').$$

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 $m \in [n] \setminus [\ell]$, 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 (5) that

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

Since $M = |C| = 2^{R_0 n}$, 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.

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) we obtain

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

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 5, which we state and prove in Section III-C. Lemma 5 is stated in terms of probability vectors f_w and f. To obtain the above conclusions, set $\mathsf{f}_w \leftarrow \mathsf{f}_{m|w}$ and $\mathsf{f} \leftarrow \mathsf{f}_m$. 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 [5] 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 the earlier approach.

B. Proof of Theorem 3

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

Fix a code C of sufficiently large length n. We will use Lemma 4. 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 4 to C_t and conclude that $|C_t| \leq 2^{(6/19)(n-t)+o(n)}$. Then, using (10), we obtain

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

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

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 5: 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 $\mathsf{f} = (A, B, C, D)$ (which we treat as a vector in $\mathbb{R}^4$). Let $\Delta_w := \mathsf{f}_w - \mathsf{f} = (\alpha_w, \beta_w, \gamma_w, \delta_w)$. 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,

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

Since, no component of $\mathsf{f} = (A, B, C, D)$ 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 $\mathsf{f} = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$, 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 $\phi(\mathsf{f}, \mathsf{f})$.) ∎

IV. Improvement of the Bound

In this section we show that Theorem 3 is not tight by providing the following slight explicit improvement.

Theorem 6: For n large enough, any $4/3$ code $C \subseteq \Sigma^n$ has rate bounded above by 0.31478.

Despite the improvement being small, it should be noted that this new result uses the linear programming bound on the minimum Hamming distance of codes [17] as a black-box. Hence, unlike Theorem 3, it is far from being self-contained. We first give a qualitative explanation and then compute the resulting bound.

A. Qualitative Analysis

We first observe that our bound $R \leq 6/19$ corresponds to the bound that would be obtained in Arikan's paper [6] if the result of his Lemma 4 was replaced by the *assumption* that the code is maximally balanced, that is, if it is assumed that the frequency vectors f_m all equal $(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4})+o(1)$. Under such an assumption, we always have $\tau_m(x,x') = \frac{1}{2}\mathbf{1}\{x_m \neq x'_m\}+o(1)$, and hence

$$\sum_{m\in[n]} \tau_m(x,x') = \frac{1}{2} d_{\mathrm{H}}(x,x') + o(n), \tag{14}$$

where $d_{\mathrm{H}}(x,x')$ is the Hamming distance between the codewords x and x'. The bound of equation (1) thus reduces to

$$\log_2 |C| \leq \frac{1}{2} d_{\mathrm{H}}(x,x') + o(n). \tag{15}$$

Plotkin's bound states that in a code C of rate R_0 there exist distinct codewords x, x' at distance $d_H(x,x') \leq \frac{3}{4}(1-R_0/2)$, which when used in the above equation gives back the bound $R_0 \leq 6/19 + o(1)$. So, we may say that the analysis in the previous section achieves the goal of removing the *assumption* that the coordinates are uniform using an averaging approach over equation (1).

Now, by inspection of the proof, we notice that in order for the bound of Theorem 3 to be tight, we would need the second inequality in (12) to be tight, when f is replaced by f_m, for an overwhelming fractions of $m \in S$, that is, we would need for almost all $m \in S$

$$\sum_{(i,j):i\neq j} \mathsf{f}_m[i]\mathsf{f}_m[j](1-\mathsf{f}_m[i]-\mathsf{f}_m[j]) = \frac{3}{8} + o(1).$$

For this to hold, f_m must equal $(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4}) + o(1)$; so we reach the conclusion that a necessary condition for our bound to be tight is that $\mathsf{f}_m = (\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4}) + o(1)$ for a fraction $1 - o(1)$ of the coordinates in the suffix set S. But clearly we can permute the coordinates of our code at will and the procedure of the previous section still applies. So a necessary condition for tightness is that actually $\mathsf{f}_m = (\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4}) + o(1)$ holds for a fraction $1 - o(1)$ of *all* coordinates. However, we saw that under this latter assumption equation (1) reduces to equation (15). Since the condition $R_0 \leq 6/19$ is obtained from (15) using the Plotkin bound on $d_{\mathrm{H}}(x,x')$, and since there are strictly better bounds on the minimum distance of codes, we conclude that the bound of Theorem (3) is certainly not tight.

As a side remark, we note that using the first linear programming bound on the Hamming distance [17], condition (15) gives $R_0 \leq 0.2845$. Of course, this bound holds only under the assumption of uniform coordinate distributions and not in general, but it points to the limitations of this approach. The refinement of Arikan's bound 0.3276 mentioned in Section II is obtained by relaxing the condition on f_m as follows. One first assumes f_m has all components larger than a constant μ, deducing that $\tau_m(x,x') \leq (1-2\mu)\mathbf{1}\{x_m \neq x'_m\}$. The obvious extension of (15) used with the linear programming bound on d_{H} then gives an upper bound on R_0 as a function of μ, say $R(\mu)$. Finally, the parameter μ is lowered to the minimum value $\bar{\mu}$ below which single coordinates would not "support" a rate equal to $R(\bar{\mu})$; that is, if a $4/3$ code C with rate $R(\bar{\mu}) = \frac{1}{n}\log|C|$ has in some coordinate m a symbol with frequency μ smaller than $\bar{\mu}$, then removing that coordinate and all codewords which take that value in that coordinate would leave a new code of length $n-1$ and rate $\frac{1}{n-1}\log(|C|(1-\mu)) > R(\bar{\mu})$. Thus $\bar{\mu}$ is the solution of the equation $\mu = 1 - 2^{-R(\mu)}$. This is essentially the meaning of [6, Lemma 4], which corresponds to what we did in III-B. The resulting bound $R(\mu^*)$ is then unconditional, and we numerically evaluated it as $R(\mu^*) = 0.3276$.

To summarize, the above discussion leads to the conclusion that

- either for a fraction $(1-o(1))$ of the coordinates we have $\mathsf{f}_m = (\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4}) + o(1)$, in which case we can use equation (15) with the linear programming bound on the Hamming distance, deducing $R_0 \leq 0.2845$,
- or there exist $\epsilon, \delta > 0$ such that for a fraction δ of the coordinates $\min_a \mathsf{f}_m[a] \leq 1/4 - \epsilon$, in which case there exists a $\gamma > 0$ such that $R_0 \leq 6/19 - \gamma$.

So, in any case there is a $\gamma > 0$ such that $R_0 \leq 6/19 - \gamma$.

B. Quantitative Analysis

We now turn the previous discussion into an explicit quantitative bound on R_0. The idea is to exploit the dichotomy between balanced/unbalanced coordinates, where in the first case we take advantage of the minimum distance, and in the second case we refine our original argument.

For each $m \in [n]$, let $\mu_m = \min_a \mathsf{f}_m(a)$, and assume without loss of generality that $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n$. We can also assume that $\mu_n \geq 1/6$, since coordinates with $\mu_m < 1/6$ support rates not larger than $\log_2(6/5) = 0.2630$ (see Section III-B), which is even smaller than our target upper bound.

Again let $\ell = \lceil R_0 n/2 - \log n - 1 \rceil$, and let C^- and C^+ as in Section III-A. Proceeding as in the proof of Theorem 3 and taking an intermediate step in Lemma 5 (i.e., the first

inequality in (12)), we reach the conclusion that

$$\frac{nR_0}{1+o(1)} \leq \sum_{m=\ell+1}^{n} \phi(\mathsf{f}_m, \mathsf{f}_m). \tag{16}$$

Since the quantity $\phi(\mathsf{f}_m, \mathsf{f}_m)$ is not decreased by substituting two components with their average, using $(\mu_m, \frac{1-\mu_m}{3}, \frac{1-\mu_m}{3}, \frac{1-\mu_m}{3})$ in place of f_m we find the upper bound

$$\phi(\mathsf{f}_m, \mathsf{f}_m) \leq \frac{6}{27}(1-\mu_m)^2(8\mu_m+1).$$

So finally we have the bound

$$\frac{nR_0}{1+o(1)} \leq \sum_{m=\ell+1}^{n} \frac{6}{27}(1-\mu_m)^2(8\mu_m+1). \tag{17}$$

We now also consider a bound on R_0 which comes from using the minimum Hamming distance. In particular, let $d_{\min}(R)$ be the largest minimum distance of any code of rate R. Then, by taking x, x' in (1) at minimum distance, we find

$$\begin{aligned} nR_0 + o(n) &\leq \sum_m \tau_m(x, x') \\ &\leq \sum_{m=n-d_{\min}(R_0)+1}^{n} (1-2\mu_m). \end{aligned}$$

If we call $d_{\mathrm{LP}}(R)$ the linear programming upper bound on $d_{\min}(R)$, we have

$$nR_0 \leq \sum_{m=n-d_{\mathrm{LP}}(R_0)+1}^{n} (1-2\mu_m) + o(n). \tag{18}$$

The combination of (17) and (18) is a quantitative version of the qualitative discussion in Section IV-A. For any sequence μ we can upper bound R_0 by taking the best of (17) and (18) then we can bound R_0 unconditionally by taking the worst case sequence μ_m. Let μ_m^* be any such worst case sequence and R_0^* be the corresponding upper bound on R_0.

First note that if $\mu_m^* < 1/4$ for $m \leq n - d_{\mathrm{LP}}(R_0^*)$, then by replacing it with $\mu_m' = 1/4$ the minimum distance bound in (18) is not affected while the bound on R_0 from (17) will in general become larger. Indeed, since $\ell = \lceil R_0 n/2 - \log n - 1 \rceil < n - d_{\mathrm{LP}}(R_0^*)$, the terms of the sum in (17) with indices m in the range $\ell < m \leq n - d_{\mathrm{LP}}(R_0^*)$ are maximized at $\mu_m = 1/4$. So, the sequence with μ_m' is at least as bad as the original one, and we can thus assume that $\mu_m^* = 1/4$ for $m \leq n - d_{\mathrm{LP}}(R_0^*)$. Finally, we see that we can assume μ_m^* is constant for $m \geq n - d_{\mathrm{LP}}(R_0^*)+1$. In fact, if this is not the case, by replacing those μ_m^* with their average, again the bound in (18) will not change, while the bound in (17) will increase, since the summand is a concave function of μ_m.

So, we conclude that we can assume that there is a constant μ^* such that

$$\mu_m^* = \begin{cases} 1/4 & m \leq n - d_{\mathrm{LP}}(R_0^*), \\ \mu^* & m > n - d_{\mathrm{LP}}(R_0^*). \end{cases} \tag{19}$$

Then, the minimum distance bound (18) reads

$$nR_0^* \leq (1-2\mu^*) d_{\mathrm{LP}}(R_0^*) + o(n) \tag{20}$$

while (17) becomes

$$\begin{aligned} nR_0^* \leq \frac{3}{8}(n - d_{\mathrm{LP}}(R_0^*) - nR_0^*/2) \\ + \frac{6}{27}(1-\mu^*)^2(8\mu^*+1)\, d_{\mathrm{LP}}(R_0^*) + o(n). \end{aligned} \tag{21}$$

We can then plug in Aaltonen's q-ary extension of the linear programming bound [17], which can be stated as $d_{\mathrm{LP}}(R) = n\delta + o(n)$ where $\delta \leq (q-1)/q$ satisfies the condition (in our case $q=4$)

$$R = H_q\left(\frac{q-1-(q-2)\delta - 2\sqrt{(q-1)\delta 1-\delta}}{q}\right), \tag{22}$$

H_q being the q-ary entropy function

$$H_q(t) = t\log(q-1) - t\log t - (1-t)\log(1-t). \tag{23}$$

Solving numerically for μ^* we experimentally determine $\mu^* = 0.217165$ and $R_0^* = 0.31477$.

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] E. Croot, V. Lev, and P. Pach, "Progression-free sets in Z_4^n are exponentially small," *ArXiv e-prints*, May 2016.

[8] J. S. Ellenberg and D. Gijswijt, "On large subsets of F_q^n with no three-term arithmetic progression," *ArXiv e-prints*, May 2016.

[9] E. Naslund and W. F. Sawin, "Upper bounds for sunflower-free sets," *ArXiv e-prints*, Jun. 2016.

[10] 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.

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

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

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

[14] 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.

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

[16] A. Nilli, "Perfect hashing and probability," *Combinatorics, Probability and Computing*, vol. 3, pp. 407–409, 1994.

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