

Improved Upper Bound for the Size of a Trifferent Code

Siddharth Bhandari^{*}, Abhishek Khetan[†]

Abstract

A subset $\mathcal{C} \subseteq \{0, 1, 2\}^n$ is said to be a *trifferent* code (of block length n) if for every three distinct codewords $x, y, z \in \mathcal{C}$, there is a coordinate $i \in \{1, 2, \dots, n\}$ where they all differ, that is, $\{x(i), y(i), z(i)\}$ is same as $\{0, 1, 2\}$. Let $T(n)$ denote the size of the largest trifferent code of block length n . Understanding the asymptotic behavior of $T(n)$ is closely related to determining the zero-error capacity of the $(3/2)$ -channel defined by Elias [Eli88], and is a long-standing open problem in the area. Elias had shown that $T(n) \leq 2 \times (3/2)^n$ and prior to our work the best upper bound was $T(n) \leq 0.6937 \times (3/2)^n$ due to Kurz [Kur23]. We improve this bound to $T(n) \leq c \times n^{-2/5} \times (3/2)^n$ where c is an absolute constant.

1 Introduction

Let q be a positive integer and let $\Sigma = \{0, 1, 2, \dots, q-1\}$ be a finite alphabet. We use the notation $[n]$ to denote the set $\{1, 2, \dots, n\}$ when n is a positive integer.

Definition 1.1 (*q -perfect hash codes & Trifferent codes*). For positive integers $q \geq 2$ and n , a code $\mathcal{C} \subseteq \Sigma^n$ is said to be a *q -perfect hash code of block length n* if for any q distinct codewords x_1, x_2, \dots, x_q in \mathcal{C} we have a coordinate $i \in [n]$ such that $\{x_j(i) \mid 1 \leq j \leq q\} = \Sigma$, where $x(i)$ denotes the i^{th} coordinate of x . When $q = 3$, a *q -perfect hash code* is also referred to as a *trifferent code*.

We will write $T(q, n)$ to denote the maximum size a q -perfect hash code of block length n attains.

Understanding the asymptotics of $T(q, n)$ as n increases is an important question, both in information theory and computer science. In this paper we will focus solely on the case of $q = 3$. As mentioned in Definition 1.1, in this case a q -perfect hash code is popularly referred to as a ‘trifferent’ code and examining the growth of $T(3, n)$ is referred to as the ‘trifference problem’, a long-standing open problem that has garnered considerable attention. (See for instance the 2014 Shannon Lecture: ‘On The Mathematics of Distinguishable Difference’ by János Körner.) Since we concern ourselves only with the case of $q = 3$ in the remainder of the paper, we refer to $T(3, n)$ as $T(n)$. In a seminal work Elias [Eli88] showed that $T(n) \leq 2 \times (3/2)^n$. Prior to our work the best upper bound on $T(n)$ was $T(n) \leq 0.6937 \times (3/2)^n$ for $n \geq 10$, due to Kurz [Kur23]. We improve this bound for sufficiently large n in the following result.

Theorem 1.1 (Main theorem). *There exists a universal constant c with the following property. Let $\mathcal{C} \subseteq \{0, 1, 2\}^n$ be a trifferent code of block length n as defined in Definition 1.1. Then, $|\mathcal{C}| \leq c \times n^{-2/5} \times (3/2)^n$. Thus, $T(n) \leq c \times n^{-2/5} \times (3/2)^n$.*

Before delving into the trifference problem, we elucidate how q -perfect hash codes are connected to perfect hashing and how they simultaneously serve as error-correcting codes for a classical channel in information theory. Subsequently, we highlight notable findings in the estimation of $T(q, n)$ for the cases when $q > 3$.

^{*}Siddharth Bhandari is at the Toyota Technological Institute at Chicago. Email: siddharth@ttic.edu

[†]Abhishek Khetan is a post-doctoral researcher at the Tata Institute of Fundamental Research, CAM, Bangalore. Email: khetan21@tifrbng.res.in

General q -perfect Hash Codes

A q -perfect hash code $\mathcal{C} = \{x_1, x_2, \dots, x_s\}$ of block length n and size s is readily seen to be a family $\{h_1, \dots, h_n\}$ of n hash functions from $[s]$ to Σ where the i^{th} hash function h_i maps $j \in [s]$ to $x_j(i)$, i.e., the i^{th} coordinate of the codeword x_j . The family of hash functions $\{h_i\}$ has the property that any subset Q of size q of the domain $[s]$ is perfectly hashed by at least one of the hash functions h_i , i.e., $\{h_i(j) : j \in Q\} = \Sigma$. Alternatively, a q -perfect hash code, say \mathcal{C} as described above, can also be cast as a cover of the q -uniform complete hypergraph on the vertex set $[s]$, say $K_s(q)$, using n hypergraphs which are q -uniform and q -partite. Specifically, we think of each hash function h_i as a hypergraph H_i whose vertex set is $[s]$ and the edge set is $\{Q \subseteq [s] : |Q| = q \wedge \{h_i(j) \mid j \in Q\} = \Sigma\}$. See the excellent survey of Radhakrishnan [Rad01] for more details.

A q -perfect hash code also serves as an error-correcting code for a classical channel studied in zero-error information theory: the $q/(q-1)$ channel. The input and output alphabets of this channel are a set of q symbols, namely Σ ; when the channel receives the symbol $i \in \Sigma$ as input, the output symbol can be anything other than i itself. For the $q/(q-1)$ channel it is impossible to recover the message without error if the code has at least two codewords: in fact, no matter how large the block length, for every set of up to $q-1$ input codewords, one can construct an output word that is compatible with *all* of them. However, there exist codes with positive rate where on receiving an output word from the channel, one can narrow down the possibilities for the input message to a set of size at most $q-1$, that is, we can *list-decode* with lists of size $q-1$. Such codes are called $(q-1)$ -list-decoding codes for the $q/(q-1)$ channel. It is well known that a q -perfect hash code \mathcal{C} of block length n and size s is equivalent to a $(q-1)$ -list-decoding code for the $q/(q-1)$ channel with block length n (for instance see the introduction of Bhandari and Radhakrishnan [BR22]).

Definition 1.2 (Rate & Capacity). For positive integers $q \geq 2$ and n , let \mathcal{C} be a q -perfect hash code of block length n . Following Elias [Eli88], we define the **rate** of \mathcal{C} as $R_{\mathcal{C}} := \frac{1}{n} \log_2(|\mathcal{C}|/(q-1))$. We define the **capacity** as

$$\text{cap}(q) = \limsup_{n \rightarrow \infty} \frac{1}{n} \log_2 \frac{T(q, n)}{q-1}.$$

Remark 1.2. It is not known if ‘lim sup’ can be replaced by ‘lim’ in the definition of capacity; see [Ari94, Footnote 1].

Many significant improvements have been made in understanding $\text{cap}(q)$ and related quantities for $q > 3$. We list some of them below and refer the reader to the work of Bhandari and Radhakrishnan [BR22] for a more detailed survey. Fredman and Komlós’ seminal work [FK84] established $\text{cap}(q) \leq \exp(-B_1 q)$ for a constant $B_1 > 0$, independent of q . Guruswami and Riazanov [GR19] demonstrated the non-optimality of the Fredman-Komlós upper bound for $q \geq 4$ and provided explicit improvements for $q = 5, 6$. Costa and Dalai [CD20] resolved a conjecture by Guruswami and Riazanov, completing the explicit computation for improving the Fredman-Komlós bound across all q , and introduced an alternative method yielding substantial enhancements for $q = 5, 6$. For $q = 4$, Dalai, Guruswami, and Radhakrishnan [DGR20] improved the upper bound to $\text{cap}(4) \leq 6/19 \approx 0.3158$, surpassing Arikan’s previous bound of 0.3512 [Ari94], while Körner and Marton [KM88, eq (1.2)] established a lower bound of $\text{cap}(4) \geq (1/3) \lg(32/29) \approx 0.0473$. Additionally, Xing and Yuan [XY19] extended Körner and Marton’s concatenation technique, demonstrating improved lower bounds on capacity for $q = 4, 8$, all odd integers greater than 3 and less than 25, and sufficiently large q not congruent to 2 (mod 4).

The Trifference Problem

Despite receiving considerable attention, progress for the trifference problem has been relatively modest when compared to the situation for $q > 3$. Elias [Eli88] showed that $0.08 \approx \lg(3) - 1.5 \leq \text{cap}(3) \leq \lg(3) - 1 \approx 0.58$; Körner and Marton [KM88] improved the lower bound above to $0.212 \approx (1/4) \lg(9/5) \leq \text{cap}(3)$ via code concatenation. Under the further assumption of linearity, i.e., if we think of Σ as \mathbb{F}_3 and assume that the trifference code $\mathcal{C} \subseteq \mathbb{F}_3^n$ is a linear subspace of \mathbb{F}_3^n , some improvements have been obtained in the upper bound on $\text{cap}(3)$. Pohoata and Zakharov [PZ22] obtained $\text{linear-cap}(3) \leq (1/4 - \epsilon) \times \log_3(2) \approx 0.3962 - \epsilon$ for some absolute constant $\epsilon > 0$ following which Bishnoi, D’haeseleer, Gijswijt and Potukuchi [BDGP23] obtained $\text{linear-cap}(3) \leq (1/4.55) \times \log_3(2) \approx 0.3483$.

Notwithstanding the above results, the current best upper bound on $\text{cap}(3)$ for general trifferent codes remains the one given by Elias [Eli88], up to a constant factor. As such, there has been an impetus to view $T(n)$, the largest size of a trifferent code of block length n , with a more refined lens. Elias' upper bound and Körner and Marton's lower bound can be recast in terms of $T(n)$ as $T(n) \leq 2 \times (3/2)^n$ and $T(n) \geq (9/5)^{n/4}$ respectively. Recently, via a computer search for a large trifferent code of block length up to $n \leq 6$, combined with a number theoretic argument, Fiore, Gnutti and Polak [FGP22] showed that $T(n) \leq 1.09 \times (3/2)^n$ for $n \geq 12$. Even more recently Kurz [Kur23] extended the computer search for trifferent codes of block lengths up to $n \leq 7$ and obtained $T(n) \leq 0.6937 \times (3/2)^n$ for $n \geq 10$.

What makes studying $T(n)$ intriguing is the fact that the upper bound of $2 \times (3/2)^n$ is obtained via a relatively simple pruning argument (described below) which has proved difficult to improve. (See for instance the work of Costa and Dalai [CD21] as to the limits of the 'slice rank' method for the triffence problem, which, however was successful in bounding the largest size of a 3-AP free set in \mathbb{F}_3^n). Additionally, the lower bound of $(9/5)^{n/4}$ is obtained not via a purely random construction, but via concatenating a random outer code and an algebraic inner code (known as the Tetra code). The pruning argument for the upper bound of $2 \times (3/2)^n$ is as follows: let \mathcal{C} be a trifferent code of block length n and let $a_1 \in \{0, 1, 2\}$ be a least occurring symbol in the first coordinate of all the codewords. Then, let \mathcal{C}_1 be the code obtained by deleting those codewords from \mathcal{C} that have a_1 in the first coordinate. Observe that $|\mathcal{C}_1| \geq (2/3)|\mathcal{C}|$. Now since \mathcal{C} was trifferent, same is true for \mathcal{C}_1 . But any three distinct strings from \mathcal{C}_1 cannot exhibit the triffence property in the first coordinate and hence for any three distinct codewords x, y, z in \mathcal{C}_1 there must exist a coordinate $i > 1$ such that $\{x(i), y(i), z(i)\} = \{0, 1, 2\}$. Proceeding iteratively in this manner we let a_2 be a least occurring symbol in the second coordinate of codewords in \mathcal{C}_1 , and then obtain \mathcal{C}_2 from \mathcal{C}_1 by deleting those codewords which have a_2 in their second coordinate, and so on, till we obtain \mathcal{C}_n . Thus, $|\mathcal{C}_n| \geq (2/3)^n \times |\mathcal{C}|$. But observe that \mathcal{C}_n is a trifferent code where three distinct strings cannot exhibit the triffence property in *any* coordinate. Therefore $2 \geq |\mathcal{C}_n|$, which leads to $|\mathcal{C}| \leq 2 \times (3/2)^n$.

We restate our main result below, which is an improved upper bound on $T(n)$.

Theorem 1.1 (Main theorem). *There exists a universal constant c with the following property. Let $\mathcal{C} \subseteq \{0, 1, 2\}^n$ be a trifferent code of block length n as defined in Definition 1.1. Then, $|\mathcal{C}| \leq c \times n^{-2/5} \times (3/2)^n$. Thus, $T(n) \leq c \times n^{-2/5} \times (3/2)^n$.*

To prove Theorem 1.1 we first introduce a close variant of trifferent codes which we call 'bounded trifferent' codes.

Definition 1.3 (r -bounded trifferent codes). *Let $\mathcal{C} \subseteq \{0, 1, 2\}^n$ be a trifferent code of block length n . For an integer $r \geq 0$, we call \mathcal{C} an r -bounded trifferent code if for all codewords $x \in \mathcal{C}$ we have that the number of 2's in x is r , i.e., $|\{i \in [n] : x(i) = 2\}| = r$. Further, for $n \geq r$ let $T_b(n, r)$ denote the maximum size an r -bounded trifferent code of block length n attains.*

In the remainder of the paper, when we talk about $T_b(n, r)$ it is to be understood that $n \geq r$ as an r -bounded trifferent code of block length $n < r$ has size 0. Note that $T_b(n, r) \leq 2 \times \binom{n}{r}$ as for a given subset of coordinates $S \subseteq [n]$, in any trifferent code of block length n there can be at most two codewords, say x and y , such that S is precisely the location of 2's for both x and y , i.e., $S = \{i \in [n] \mid x(i) = 2\} = \{i \in [n] \mid y(i) = 2\}$. If there were three, they would be a counter-example to the triffence property. Next, we prove a simple lemma relating $T(n)$ and $T_b(n, r)$, thus, highlighting the importance of studying r -bounded trifferent codes.

Lemma 1.3 (Size of trifferent codes in terms of r -bounded trifferent codes).

$$T(n) \leq 2^{(r-n)} \times \frac{T_b(n, r)}{\binom{n}{r}} \times 3^n$$

Proof. It will be convenient to think of $\Sigma := \{0, 1, 2\}$ as \mathbb{F}_3 . Let \mathcal{C} be a trifferent code of block length n . Let A_r denote the set of all those $x \in \Sigma^n$ such that x witnesses exactly r 2s. More precisely

$$A_r = \{x \in \mathbb{F}_3^n : |\{i \in [n] : x(i) = 2\}| = r\}.$$

For strings x and v in \mathbb{F}_3^n let $x + v$ denote addition in \mathbb{F}_3^n . Observe that the code $\mathcal{C} + v := \{x + v \mid x \in \mathcal{C}\}$ is also trifferent for any string $v \in \mathbb{F}_3^n$. Hence, $|\mathcal{C} + v \cap A_r| \leq T_b(n, r)$.

Let \mathbf{v} be a uniformly random element of \mathbb{F}_3^n . We write $\mathcal{C} + \mathbf{v}$ to denote the random subset $\mathcal{C} + v$ where v is picked according to the distribution of \mathbf{v} . Note that for any $x \in \mathbb{F}_3^n$ we have:

$$\mathbb{E}_{\mathbf{v}}[\mathbf{1}[x \in \mathcal{C} + \mathbf{v}]] = \mathbb{P}_{\mathbf{v}}(x - \mathbf{v} \in \mathcal{C}) = \frac{|\mathcal{C}|}{3^n}.$$

Therefore,

$$T_b(n, r) \geq \mathbb{E}_{\mathbf{v}}[|(\mathcal{C} + \mathbf{v}) \cap A_r|] = \sum_{x \in A_r} \mathbb{E}_{\mathbf{v}}[\mathbf{1}[x \in \mathcal{C} + \mathbf{v}]] = |A_r| \times \frac{|\mathcal{C}|}{3^n}.$$

Now using $|A_r| = \binom{n}{r} 2^{n-r}$, we get the desired result. \square

Remark 1.4. Even more generally, the above lemma shows that for any $S \subseteq \{0, 1, 2\}^n$ if $T_b(S)$ denotes that largest size of a trifferent code contained in S , then we have $T(n) \leq \frac{T_b(S)}{|S|} \times 3^n$. Hence, it might also be interesting to look at sets S which are not of the form $\{x \in \{0, 1, 2\}^n : \#2\text{'s in } x = r\}$ for some integer r .

In light of the above lemma it is natural to define the notion of an r -bounded density.

Definition 1.4 (r -bounded density). Recall that $T_b(n, r)$ denotes the maximum size an r -bounded trifferent code of block length n attains. The r -bounded density at block length n , denoted as $\rho_b(n, r)$, is defined as

$$\rho_b(n, r) = 2^{(r-n)} \times \frac{T_b(n, r)}{\binom{n}{r}}$$

Hence, [Lemma 1.3](#) can be cast as $T(n) \leq \rho_b(n, r) \times 3^n$ for all $r \geq 0$. The bound obtained using the pruning argument, that is $T(n) \leq 2 \times (3/2)^n$, can be obtained as an instantiation of [Lemma 1.3](#) with $r = 0$. In this case, it is clear that $T_b(n, r) = 2$ (there can be at most two codewords in a trifferent code if the symbol 2 does not appear in any codeword) and hence $\rho_b(n, r) = 2^{1-n}$. It turns out that $\rho_b(n, 1) = 2^{2-n}$ as we show that $T_b(n, 1) = 2n$ (see [Lemma 3.1](#)). This bound is worse than what we obtained on $\rho_b(n, 0)$. However, the situation improves when we consider $r \geq 2$. Our main contribution is showing that for $r = 3$, $\rho_b(n, r) \leq c \times n^{-2/5} \times 2^{-n}$ for some absolute constant c . More precisely, we have the following result.

Theorem 1.5 (Bounding r -bounded density). Let $T_b(n, r)$ and $\rho_b(n, r)$ be as defined in [Definitions 1.3](#) and [1.4](#) respectively. Then, we have constants c' and c such that

- a) $T_b(n, 2) \leq c' \times n^{5/3}$ and hence $\rho_b(n, 2) \leq c \times n^{-1/3} \times 2^{-n}$ and
- b) $T_b(n, 3) \leq c' \times n^{13/5}$ and hence $\rho_b(n, 3) \leq c \times n^{-2/5} \times 2^{-n}$.

We describe the proof idea of [Theorem 1.5](#) for the case when $r = 2$. We proceed via constructing a graph related to an r -bounded trifferent code, say \mathcal{C}_b , with block length n . This graph has roughly as many edges as $|\mathcal{C}_b|$. The crucial observation is that certain bipartite structures are forbidden in this graph. Then, an application of the famous Kővári–Sós–Turán (KST) theorem yields a bound on the number of edges in this graph which also serves as a bound on the size of \mathcal{C}_b . We give the details below.

Recall that each codeword in \mathcal{C}_b has exactly two 2's. Now, consider the graph $G_{\mathcal{C}_b}$ on the vertex set $[n]$ where for each codeword $x \in \mathcal{C}_b$ an edge $\{i, j\}$ with $i \neq j$ is added to $G_{\mathcal{C}_b}$ if $x(i) = x(j) = 2$, i.e., i and j are the locations of 2's in x . Note that an edge $\{i, j\}$ can be added by at most 2 codewords in \mathcal{C}_b . Hence, $G_{\mathcal{C}_b}$ has at least half as many edges as $|\mathcal{C}_b|$. Next, we show via the PHP and trifference property of \mathcal{C}_b that $K_{3,9}$ —the complete bipartite graph with the partite sets having sizes 3 and 9 respectively—is forbidden in $G_{\mathcal{C}_b}$. Applying the KST theorem yields an upper bound on the number of edges in $G_{\mathcal{C}_b}$ as $c'' \times n^{5/3}$ for some constant c'' . Hence, we obtain our desired bound on $|\mathcal{C}_b|$ and $T_b(n, 2)$.

The proof for when $r = 3$ proceeds along similar lines but we define the graph $G_{\mathcal{C}_b}$ more prudently. The detailed proof of [Theorem 1.5](#) appears in [Section 2](#).

Armed with [Theorem 1.5](#) and [Lemma 1.3](#) we easily obtain the proof of [Theorem 1.1](#).

Proof of Theorem 1.1. With $r = 3$ we have the following.

$$\begin{aligned} T(n) &\leq 2^{(r-n)} \times \frac{T_b(n, r)}{\binom{n}{r}} \times 3^n && \text{by Lemma 1.3} \\ &= \rho_b(n, r) \times 3^n \\ &\leq c \times n^{-2/5} \times 2^{-n} && \text{by Theorem 1.5.} \end{aligned}$$

giving the desired result. \square

From the above discussion it is tempting to analyze $\rho_b(n, r)$ when both r and n are growing with n growing much faster than r . As such, we define a notion of r -bounded deficit and sup-bounded deficit which serve to get a sense of the speed at which $\rho_b(n, r)$ decays.

Definition 1.5 (r -bounded deficit & sup-bounded deficit). For an integer $r \geq 1$, let $T_b(n, r)$ denote the maximum size an r -bounded trifferent code of block length n attains. Let $\Delta_r(n) = r - \frac{\log T_b(n, r)}{\log n}$. Then, the r -bounded deficit is defined as

$$\begin{aligned} \Delta_r &:= \limsup_{n \rightarrow \infty} \left(r - \frac{\log T_b(n, r)}{\log n} \right) \\ &= \limsup_{n \rightarrow \infty} \Delta_r(n). \end{aligned}$$

and the sup-bounded deficit is defined as

$$\Delta_\infty := \lim_{r \rightarrow \infty} \Delta_r.$$

Remark 1.6. 1. To be more precise we should have defined Δ_∞ as $\limsup_{r \rightarrow \infty} \Delta_r$. However, Corollary 1.6.1 shows that Δ_r is increasing in r .

2. If we unpack the above definition in terms of $T_b(n, r)$, then Lemma 1.3 yields that

$$T(n) \leq c_r \times n^{-\Delta_r(n)} \times (3/2)^n$$

where c_r is a constant depending only on r . Hence, studying Δ_r as r grows is helpful in proving better upper bounds on $T(n)$.

Since, $T_b(n, r)$ is at most $2 \times \binom{n}{r}$, the r -bounded deficit, Δ_r , is always non-negative. In fact, by Theorem 1.5 we have shown that $\Delta_2 \geq 1/3$ and $\Delta_3 \geq 2/5$. Via a simple application of the PHP we show that Δ_r is increasing in r .

Corollary 1.6.1. $\Delta_{r+1} \geq \Delta_r$ for any integer $r \geq 1$. Hence, for all integers $r \geq 3$, there are constants c_r and c'_r such that we have: $T_b(n, r) \leq c'_r \times n^{r-2/5}$ and hence $\rho_b(n, r) \leq c_r \times n^{2/5} \times 2^{-n}$.

Proof. Let $\mathcal{C}_b \subseteq \{0, 1, 2\}^n$ be an $r+1$ -bounded trifferent code of block length n . By double-counting we can find a coordinate $i \in [n]$ such that there is a subset $\mathcal{C}'_b \subseteq \mathcal{C}_b$ of size at least $n/r \times |\mathcal{C}_b|$ and every codeword of \mathcal{C}'_b has a 2 at coordinate i . Notice that \mathcal{C}'_b can be equivalently thought of as an r -bounded trifferent code with block length $n-1$ since the i -th coordinate is same (namely 2) for each codeword. Hence, $T_b(n, r+1) \leq n/r \times T_b(n-1, r)$ from where it follows that $\Delta_{r+1} \geq \Delta_r$. \square

Next, we turn our attention to establish upper bounds on Δ_r , i.e., prove lower bounds on $T_b(n, r)$. Our constructions are based on thinking about the codewords as point-line incidences in an appropriate finite-dimensional vector space over a finite field. See Section 3 for the detailed constructions.

Theorem 1.7 (Upper bounds on Δ_r). $T_b(n, 1) = 2n$ and hence $\Delta_1 = 0$. Also, $\Delta_3 \leq 3/2$, i.e., $T_b(n, 3) \geq c_1 \times n^{3/2}$ for some constant $c_1 > 0$: further, for every positive integer r , a power of 3, we have $\Delta_r \leq r - r^\alpha$ where $\alpha = 1 - \log_3(2) \approx 0.369$.

Further Questions

A few interesting questions which arise are as follows. Is $\Delta_\infty = \infty$? If yes, this immediately shows that for any constant d we have $T(n) \leq n^{-d} \times (3/2)^n$, i.e., the size of a family of trifferent codes of growing block lengths, say $n \rightarrow \infty$, decays faster than $(3/2)^n$ divided by any polynomial factor. A more nuanced understanding in terms of $\rho_b(n, r)$ with growing r might also lead to an improved upper bound on $\text{cap}(3)$, hence making progress on the long-standing open problem. Further, it is also interesting to understand Δ_r more precisely for small values of r such as 2 and 3: is $\Delta_2 = 1/2$?

Acknowledgements

We are highly indebted to Prof. Jaikumar Radhakrishnan for various insightful discussions and for painstakingly proofreading the manuscript. We thank Prof. Alexander Razborov for helpful discussions. We also acknowledge Aakash Bhowmick for helping us with experiments related to constructing trifferent codes. Lastly, we thank Prof. Nishant Chandgotia for his inputs and encouragement.

2 Upper bounds on the size of r -bounded trifferent codes

In this section we prove [Theorem 1.5](#). We restate the theorem below for convenience.

Theorem 1.5 (Bounding r -bounded density). *Let $T_b(n, r)$ and $\rho_b(n, r)$ be as defined in [Definitions 1.3](#) and [1.4](#) respectively. Then, we have constants c' and c such that*

- a) $T_b(n, 2) \leq c' \times n^{5/3}$ and hence $\rho_b(n, 2) \leq c \times n^{-1/3} \times 2^{-n}$ and
- b) $T_b(n, 3) \leq c' \times n^{13/5}$ and hence $\rho_b(n, 3) \leq c \times n^{-2/5} \times 2^{-n}$.

To prove [Theorem 1.5](#) we will need to apply the famous result of Kővári, Sós and Turán, known popularly as the KST theorem, or rather a version of it due to Hyltén-Cavallius.

Theorem 2.1 (KST theorem due to Hyltén-Cavallius [[HC58](#)]). *The Zarankiewicz function $z(u, v; s, t)$ denotes the maximum possible number of edges in a bipartite graph $G = (U \cup V, E)$ for which $|U| = u$ and $|V| = v$, but which does not contain a subgraph of the form $K_{s,t}$ where s vertices come from U and t from V (here $K_{s,t}$ denotes the complete bipartite graph with s and t vertices in the two partite sets). Then,*

$$z(u, v; s, t) < (t-1)^{\frac{1}{s}}(u-s+1)v^{1-\frac{1}{s}} + (s-1)v.$$

Proof of [Theorem 1.5](#). We first proceed with the proof when $r = 2$. Let \mathcal{C}_b be an r -bounded trifferent code of block length n with $r = 2$. Thus, every codeword of \mathcal{C}_b has two 2's. As discussed previously, we construct a graph $G_{\mathcal{C}_b}$ using the code \mathcal{C}_b . The graph $G_{\mathcal{C}_b}$ has the vertex set $[n]$ and the set of edges E is

$$E = \{\{i, j\} \mid i \neq j \wedge \exists x \in \mathcal{C}_b : x(i) = x(j) = 2\}.$$

In other words, for each codeword $x \in \mathcal{C}_b$, an edge, say $x_e = \{i, j\}$, is added to $G_{\mathcal{C}_b}$, where i and j are the locations of 2's in x .

As we have argued previously, for any subset $S \subseteq [n]$ of coordinates there can be at most two codewords, say x and y , such that S is precisely the location of 2's for both x and y , i.e., $S = \{i \in [n] \mid x(i) = 2\} = \{i \in [n] \mid y(i) = 2\}$. (If there were three, they would contradict the trifference property.) This shows that an edge $\{i, j\}$ can be added by at most 2 codewords in \mathcal{C}_b . Hence,

$$|E| \geq |\mathcal{C}_b|/2,$$

i.e., there are at least half as many edges as $|\mathcal{C}_b|$. At the cost of another factor of $1/2$ on $|E|$ we can assume that $G_{\mathcal{C}_b}$ is bipartite of the form $(U \cup V, E')$, where $V = [n] \setminus U$ and $|E'| \geq |E|/2$. (A random equi-bipartition of $[n]$ will have this property on expectation.)

Next, we show that $K_{3,9}$ is forbidden in $G_{\mathcal{C}_b}$. To see this suppose there exist distinct i_1, i_2, i_3 in U and distinct j_1, j_2, \dots, j_9 in V such that subgraph induced by $G_{\mathcal{C}_b}$ on these vertices is $K_{3,9}$. We denote the edge $\{i_k, j_\ell\}$ by $e_{k,\ell}$. Further, let $x_{k,\ell}$ denote a codeword in \mathcal{C}_b corresponding to which the edge $e_{k,\ell}$ was

added to G_{C_b} : if there are two such codewords then choose one arbitrarily and fix it. By the PHP there is a subset $T \subseteq [9]$ of size at least 3 such that all codewords in $\{x_{1,\ell} \mid \ell \in T\}$ have the same value on the coordinates i_2 and i_3 , i.e., $|\{x_{1,\ell}(i_2) \mid \ell \in T\}| = 1$ and $|\{x_{1,\ell}(i_3) \mid \ell \in T\}| = 1$. Again, by the PHP we can find a subset $T' \subseteq T$ of size at least 2 such that all codewords in $\{x_{2,\ell} \mid \ell \in T'\}$ have the same value on the coordinate i_3 , i.e., $|\{x_{2,\ell}(i_3) \mid \ell \in T'\}| = 1$. WLOG let us assume that $T' = \{j_1, j_2\}$ (see Figure Fig. 1).

Then, by the PHP there must exist two vertices i_c and i_d (with $c < d$) in $\{i_1, i_2, i_3\}$ such that $x_{c,1}(j_2) = x_{d,1}(j_2)$. But then the three codewords $x_{c,1}$, $x_{d,1}$ and $x_{c,2}$ are a counter-example to the trifference of C_b . To see this let $w \in [n]$. If $w \notin \{i_c, i_d, j_1, j_2\}$, then none of the three codewords have the symbol 2 at w and hence $x_{c,1}$, $x_{d,1}$ and $x_{c,2}$ cannot witness trifference at w . Now if $w = i_c$, then both $x_{c,1}$ and $x_{c,2}$ have the symbol 2 at w ; if $w = i_d$, then, because of our definition of T' we have $x_{c,1}(w) = x_{c,2}(w)$; if $w = j_1$, then both $x_{c,1}$ and $x_{d,1}$ have the symbol 2 at w ; finally if $w = j_2$, then by our assumption $x_{c,1}(j_2) = x_{d,1}(j_2)$. So no matter what w is, some two of $x_{c,1}$, $x_{d,1}$ and $x_{c,2}$ agree on w , and hence these three codewords contradict the trifference property. Hence, G_{C_b} is $K_{3,9}$ -free.

Applying Theorem 2.1 with $u, v = n/2$ and $s = 3, t = 9$, yields $c'' \times n^{5/3}$ as an upper bound on the size of E' for some constant c'' . Hence, we obtain our desired bound on $|C_b|$ and $T_b(n, 2)$.

Next, we turn our attention to the case when $r = 3$. Again let C_b be an r -bounded trifferent code with $r = 3$ and block length n . Recall that each codeword in C_b now has three 2's. This time we construct a bi-partite graph $G_{C_b} = (U, V, E)$ with $U = [n]$, $V = \binom{[n]}{2}$ and edge set E described as follows. For each codeword $x \in C_b$ we add the edge $x_e = (i, \{j, k\})$ to E where $x(i) = x(j) = x(k) = 2$ and $i < j < k$. As argued previously, edge $e \in E$ can be added by at most 2 codewords in C_b : therefore, $|E| \geq 0.5 \times |C_b|$. Now, we will show that G_{C_b} is $K_{s=5, t=2^{21}}$ -free.

Suppose not, and that there exist vertices i_1, i_2, \dots, i_5 in U and S_1, S_2, \dots, S_t (note that each S_i is a 2-subset of coordinates) in V such that G_{C_b} induces a $K_{5,t}$ on these vertices. We denote the edge (i_k, S_ℓ) by $e_{k,\ell}$ and the codeword corresponding to it with $x_{k,\ell}$: if there are two such codewords then choose one arbitrarily. By the PHP there is a subset $T_1 \subseteq [t]$ of size at least $t/2^4$ such that all codewords in $\{x_{1,\ell} \mid \ell \in T_1\}$ have the same value on the coordinates i_2, i_3, i_4, i_5 , i.e., $|\{x_{1,\ell}(i_2) \mid \ell \in T_1\}| = 1$ and so on for i_3, i_4, i_5 . Again, by the PHP there is a subset $T_2 \subseteq T_1$ of size at least $|T_1|/2^4 = t/2^8$ such that all codewords in $\{x_{2,\ell} \mid \ell \in T_2\}$ have the same value on the coordinates i_1, i_3, i_4, i_5 . Continuing this way for the remaining coordinates i_3, i_4, i_5 , we obtain a subset T_5 of size at least $t/2^{20} = 2$: WLOG say $T_5 = \{S_1, S_2\}$, thus, for any $c, d \in [5]$ we have $x_{c,1}(i_d) = x_{c,2}(i_d)$. Further, let $S = S_2 \setminus S_1$, then $|S| \leq 2$. By the PHP there exists $c, d \in [5]$ with $c < d$ such that $x_{c,1}(S) = x_{d,1}(S)$ where $x(S)$ denotes the tuple $(x(i) \mid i \in S)$. However, now the three codewords $x_{c,1}$, $x_{c,2}$ and $x_{d,1}$ are a counter-example to the trifference of C_b . This is because for any coordinate $w \in [n]$ if w is not i_c, i_d or in S_1, S_2 , then none of the three codewords have the symbol 2 at w : further, if $w = i_c$, then both $x_{c,1}$ and $x_{c,2}$ have the symbol 2 at w ; if $w = i_d$, then, because of our definition of T_5 we have $x_{c,1}(w) = x_{c,2}(w)$; if $w \in S_1$, then both $x_{c,1}$ and $x_{d,1}$ have the symbol 2 at w ; finally if $w = S_2$, then by our assumption $x_{c,1}(S_2) = x_{d,1}(S_2)$. Hence, G_{C_b} is $K_{5,2^{21}}$ -free.

Applying Theorem 2.1 with $u = n$ and $v = \binom{n}{2}$ and $s = 5, t = 2^{21}$, yields a bound on the number of edges in E as $c'' \times n^{3-2/5} = c'' \times n^{13/5}$ for some constant c'' . Hence, we obtain our desired bound on $|C_b|$ and $T_b(n, 3)$. \square

Remark 2.2. We can improve the (huge) constant 2^{21} appearing in the above proof to some extent by following a strategy similar to the one employed in the case when $r = 2$. However, it is easier to work with the current numbers and this doesn't seem to hurt the bounds of Theorem 2.1 beyond a constant factor.

3 Lower Bounds on the size of r -bounded trifferent codes

In this section we will prove Theorem 1.7 which we restate below for convenience.

Theorem 1.7 (Upper bounds on Δ_r). $T_b(n, 1) = 2n$ and hence $\Delta_1 = 0$. Also, $\Delta_3 \leq 3/2$, i.e., $T_b(n, 3) \geq c_1 \times n^{3/2}$ for some constant $c_1 > 0$: further, for every positive integer r , a power of 3, we have $\Delta_r \leq r - r^\alpha$ where $\alpha = 1 - \log_3(2) \approx 0.369$.

We will first focus on the case of $r = 1$.

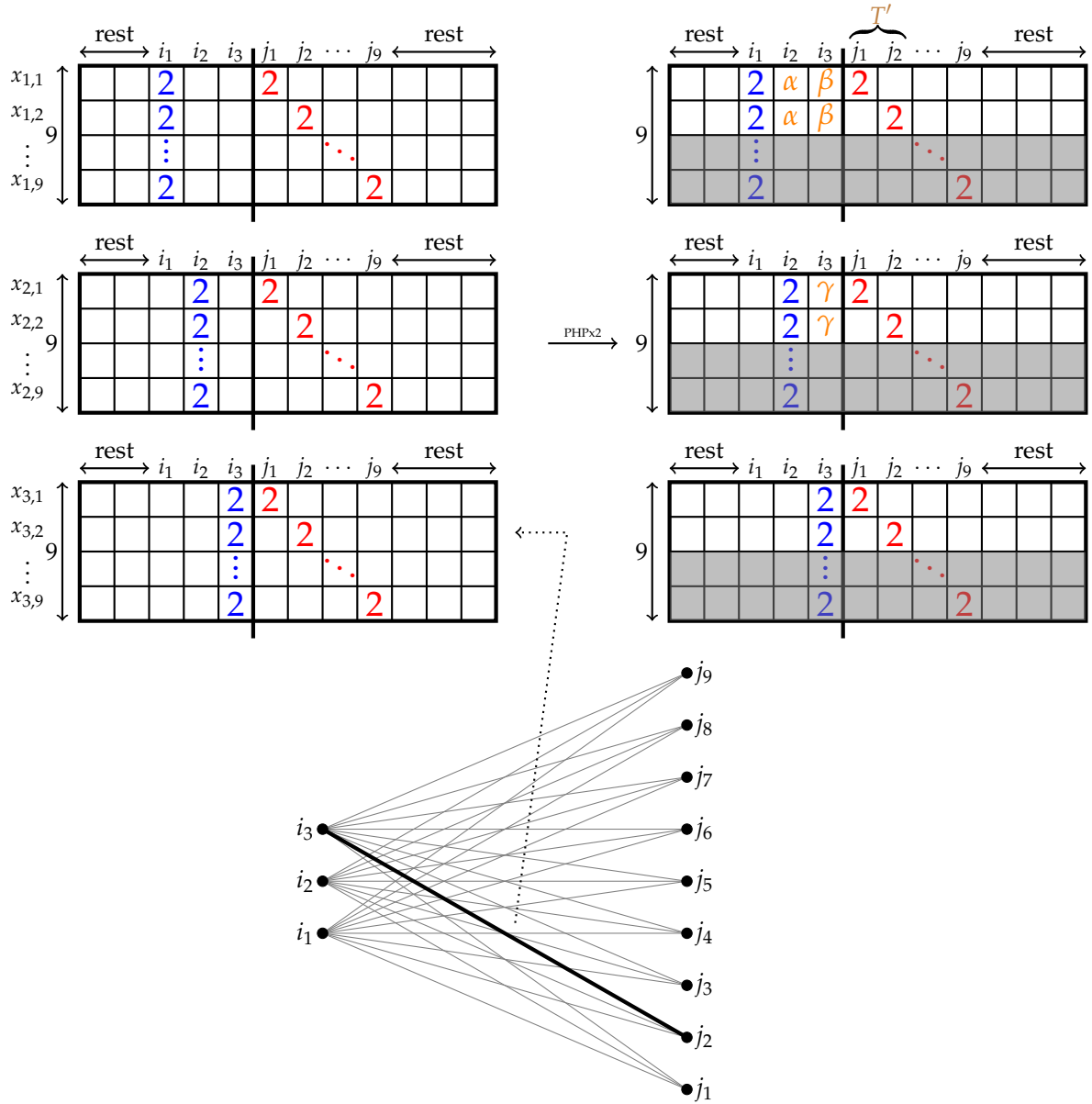


Figure 1: Illustrating the proof for $r = 2$. The collection of codewords before the application of the PHP is displayed in the left column of the figure: they correspond to the edges of the $K_{3,9}$ supposed to exist for the sake of contradiction. For example, the codeword $x_{3,2}$ corresponds to the highlighted edge $\{i_3, j_2\}$, and the first block of codewords on the left corresponds to the edges incident to i_1 . After applying the PHP we are left with the non-grayed codewords in the right column corresponding to the set $T' = \{j_1, j_2\}$. Eventually, we find three codewords from the non-grayed ones which do not exhibit the trifference property.

Lemma 3.1 (Maximum size of trifferent codes where each codeword has one 2). *For each integer $n \geq 1$ we have $T_b(n, 1) = 2n$.*

Proof. As we have argued previously, in any trifferent code \mathcal{C} , for any subset $S \subseteq [n]$ of coordinates there can be at most two codewords, say x and y , such that S is precisely the location of 2's for both x and y , i.e., $S = \{i \in [n] \mid x(i) = 2\} = \{i \in [n] \mid y(i) = 2\}$. Hence, $T_b(n, 1) \leq 2n$. Next, we construct an r -bounded trifferent code ($r = 1$) with block length n and size $2n$. Let $A_1 = \{x \in \{0, 1, 2\}^n \mid \#2\text{'s in } x = 1\}$. Further, for each $i \in [n]$, let $u_i, v_i \in \{0, 1, 2\}^n$ be defined as

$$u_i(j) = \begin{cases} 2 & \text{if } j = i, \\ 1 & \text{if } i < j, \\ 0 & \text{otherwise,} \end{cases} \quad v_i(j) = \begin{cases} 2 & \text{if } j = i, \\ 1 & \text{if } i > j, \\ 0 & \text{otherwise.} \end{cases}$$

Consider the code $\mathcal{C}_b \subseteq A_1$ defined as $\cup_{i \in [n]} \{u_i, v_i\}$. Clearly, $|\mathcal{C}_b| = 2n$. We claim the \mathcal{C}_b is a trifferent code. Consider any three different codewords x, y and z : there are two cases of interest to check (other cases can be reduced to these): (a) $x = u_i, y = v_i, z = u_j$ or v_j where $i \neq j$, in which case we have $\{x(j), y(j), z(j)\} = \{0, 1, 2\}$ and (b) $x = u_i, y = u_j, z = u_k$ or v_k where $i < j < k$, in which case either $\{x(j), y(j), z(j)\} = \{0, 1, 2\}$ or $\{x(i), y(i), z(i)\} = \{0, 1, 2\}$ respectively. \square

Now, we turn to the general case when r is a power of 3.

Lemma 3.2 (Maximum size of trifferent codes where each codeword has 3^t many 2's). *Let $t \geq 0$ be an integer and let $r = 3^t$. Suppose that for all positive integers $n \geq r$ we have $T_b(n, r) \geq c_t \times n^{(3/2)^t}$ for some constant $c_t > 0$ depending only on t . Then, for all positive integers $n \geq 3r$ we have $T_b(n, 3r) = c_{t+1} \times n^{(3/2)^{t+1}}$ for some constant $c_{t+1} > 0$ depending only on $t + 1$.*

Proof. Let \mathbb{F}_q be a large enough finite field. Let $\mathcal{P} = \mathbb{F}_q^2$ and \mathcal{L} be the set of all the affine lines in \mathbb{F}_q^2 . Thus,

$$|\mathcal{P}| = q^2, \quad |\mathcal{L}| = q^2 + q.$$

Let $\mathcal{S} = \{(p, \ell) \in \mathcal{P} \times \mathcal{L} \mid p \in \ell\}$ and note that $|\mathcal{S}| = q$ and $|\mathcal{L}| = q^3 + q^2$. For each line $\ell \in \mathcal{L}$ choose a permutation σ_ℓ of the points of ℓ such that σ_ℓ has no fixed points. Further, let $f : \mathcal{S} \rightarrow \mathcal{P}$ be defined as $f(p, \ell) = \sigma_\ell(p)$. Notice that $f(p, \ell)$ is a point on ℓ but is different from p . Let $\varphi : \mathcal{P} \rightarrow \{0, 1, 2\}^n$ and $\psi : \mathcal{L} \rightarrow \{0, 1, 2\}^n$ be injective functions whose images are 3^t -bounded trifferent codes where n is the smallest integer such that $c_t \times n^{(3/2)^t} \geq q^2 + q$. The maps φ and ψ exist by the inductive hypothesis for t .

Define a map $\theta : \mathcal{S} \rightarrow \{0, 1, 2\}^n$ as $\theta(p, \ell) = \varphi(f(p, \ell))$. Finally, define $\tau : \mathcal{S} \rightarrow \{0, 1, 2\}^n \times \{0, 1, 2\}^n$ as

$$\tau(p, \ell) = (\varphi(p), \psi(\ell), \theta(p, \ell)).$$

We claim that the image of τ , denoted by $\tau(\mathcal{S})$, is a $3r = 3^{t+1}$ -bounded trifferent code.

Towards this end, let $e_1 = (p_1, \ell_1), e_2 = (p_2, \ell_2)$ and $e_3 = (p_3, \ell_3)$ be three pairwise distinct elements of \mathcal{S} and consider their encodings $\tau(p_1, \ell_1), \tau(p_2, \ell_2)$ and $\tau(p_3, \ell_3)$. If p_1, p_2 and p_3 are pairwise distinct, then $\varphi(p_1), \varphi(p_2)$ and $\varphi(p_3)$ will exhibit the trifference property as the image of φ is a trifferent code by construction. Similarly, if ℓ_1, ℓ_2 and ℓ_3 are pairwise distinct then $\psi(\ell_1), \psi(\ell_2)$ and $\psi(\ell_3)$ will exhibit the trifference property. So we may WLOG assume that $p_1 = p_2 = p$ (say) and $\ell_2 = \ell_3 = \ell$ (say). Write p' in place of p_3 and ℓ' in place of ℓ_1 . See Fig. 2. We show that $\theta(p, \ell'), \theta(p, \ell)$ and $\theta(p', \ell)$ exhibit the trifference property to finish the proof of the claim.

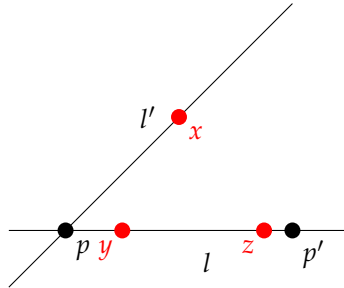


Figure 2: The case when there are only two points and two lines.

Note that $p \neq p'$ and $\ell \neq \ell'$ for otherwise e_1, e_2 and e_3 would not be pairwise distinct. Let $f(p, \ell') = \sigma_{\ell'}(p) = x$, $f(p, \ell) = \sigma_{\ell}(p) = y$ and $f(p', \ell) = \sigma_{\ell}(p') = z$. We observe that x, y and z are pairwise distinct. Indeed, the points y and z lie on ℓ and are different from p : hence, $x \neq y$ and $x \neq z$ as x lies on ℓ' and the only common point between ℓ and ℓ' is p . Further, $y \neq z$ because σ_{ℓ} is permutation of the points of ℓ .

Therefore $\varphi(x), \varphi(y)$ and $\varphi(z)$ exhibit the trifferent property. But

$$\theta(e_1) = \varphi(x), \quad \theta(e_2) = \varphi(y), \quad \theta(e_3) = \varphi(z)$$

showing that $\tau(\mathcal{S})$ is a trifferent code. It is also clear that the number of 2's in each codeword of $\tau(\mathcal{S})$ is $3r$ and the block length of the code is $3n$. Finally, since τ is injective, the size of the code is $|\mathcal{S}| = q^3 + q^2$.

Setting n to be the smallest integer larger than $\left(\frac{q^2+q}{c_t}\right)^{(2/3)^t}$, and $n' = 3n$, we obtain a $3r$ -bounded trifferent code $\tau(\mathcal{S})$ with size at least $c_{t+1} \times n'^{(3/2)^{t+1}}$. (For smaller values of $n' > 3r$ we can just adjust the constant c_{t+1} to make hypothesis true.) \square

Proof of Theorem 1.7. Lemma 3.1 proves the first claim. For the next claim: $T_b(n, 3^t) \geq c_t \times n^{(3/2)^t}$ is obtained directly from Lemma 3.2 with the base case of $t = 0$ being Lemma 3.1. Setting $r = 3^t$ gives $(3/2)^t = r^\alpha$ with $\alpha = 1 - \log_3(2) \approx 0.369$. Hence, $\Delta_r \leq r - r^\alpha$. \square

References

- [Ari94] Erdal Arikan. An upper bound on the zero-error list-coding capacity. *IEEE Trans. Inform. Theory*, 40(4):1237–1240, 1994. 2
- [BDGP23] Anurag Bishnoi, Jozefien D’haeseleer, Dion Gijswijt, and Aditya Potukuchi. Blocking sets, minimal codes and trifferent codes, 2023. 2
- [BR22] Siddharth Bhandari and Jaikumar Radhakrishnan. Bounds on the zero-error list-decoding capacity of the $q/(q-1)$ channel. *IEEE Transactions on Information Theory*, 68(1):238–247, 2022. 2
- [CD20] Simone Costa and Marco Dalai. New bounds for perfect k -hashing. (manuscript), 2020. 2
- [CD21] Simone Costa and Marco Dalai. A gap in the slice rank of k -tensors. *Journal of Combinatorial Theory, Series A*, 177:105335, 2021. 3
- [DGR20] Marco Dalai, Venkatesan Guruswami, and Jaikumar Radhakrishnan. An improved bound on the zero-error list-decoding capacity of the $4/3$ channel. *IEEE Trans. Inform. Theory*, 66(2):749–756, 2020. (Preliminary version in *IEEE International Symposium on Information Theory (ISIT)*, 2017). 2
- [Eli88] Peter Elias. Zero error capacity under list decoding. *IEEE Trans. Inform. Theory*, 34(5):1070–1074, 1988. 1, 2, 3
- [FGP22] Stefano Della Fiore, Alessandro Gnutti, and Sven Polak. The maximum cardinality of trifferent codes with lengths 5 and 6. *Examples and Counterexamples*, 2:100051, 2022. 3
- [FK84] Michael L. Fredman and János Komlós. On the size of separating systems and families of perfect hash functions. *SIAM Journal on Algebraic Discrete Methods*, 5(1):61–68, 1984. 2
- [GR19] Venkatesan Guruswami and Andrii Riazanov. Beating Fredman–Komlós for perfect k -hashing. In Christel Baier, Ioannis Chatzigiannakis, Paola Flocchini, and Stefano Leonardi, editors, *Proc. 46th International Colloq. of Automata, Languages and Programming (ICALP)*, volume 132 of *LIPIcs*, pages 92:1–92:14. Schloss Dagstuhl, 2019. 2
- [HC58] C. Hyltén-Cavallius. On a combinatorial problem. *Colloquium Mathematicum*, 6:59–65, 1958. As cited by Bollobás (2004). 6

- [KM88] János Körner and Katalin Marton. New bounds for perfect hashing via information theory. *European J. Combin.*, 9(6):523–530, 1988. [2](#)
- [Kur23] Sascha Kurz. Trifferent codes with small lengths, 2023. [1](#), [3](#)
- [PZ22] Cosmin Pohoata and Dmitriy Zakharov. On the trifference problem for linear codes. *IEEE Transactions on Information Theory*, 68(11):7096–7099, 2022. [2](#)
- [Rad01] Jaikumar Radhakrishnan. Entropy and counting. 2001. [2](#)
- [XY19] Chaoping Xing and Chen Yuan. Beating the probabilistic lower bound on perfect hashing. (manuscript), 2019. [2](#)