TU Delft

Delft University of Technology

Document Version
Final published version

Licence
Dutch Copyright Act (Article 25fa)

Citation (APA)
Bishnoi, A., Kielak, B., Kovacs, B., Nagy, Z. L., Somlai, G., Vizer, M., & Zheng, Z. (2026). The generalized trifference problem. *IEEE Transactions on Information Theory*, *72*(5), 2907-2914. https://doi.org/10.1109/TIT.2026.3665439

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

The Generalized Trifference Problem

Anurag Bishnoi, Bartłomiej Kielak, Benedek Kovács, Zoltán Lóránt Nagy, Gábor Somlai, Máté Vizer, and Zeyu Zheng

***Abstract*—We study the problem of finding the largest number $T(n, m)$ of ternary vectors of length n such that for any three distinct vectors there are at least m coordinates where they pairwise differ. This problem is a special case of the perfect k-hashing problem in theoretical computer science, corresponding to the $k = 3$ case. For $m = 1$, we get the classical trifference problem which is wide open. We prove upper and lower bounds on $T(n, m)$ for various ranges of the parameter m and determine the phase transition threshold on $m = m(n)$ where $T(n, m)$ jumps from constant to exponential in n. By relating the linear version of this problem to a problem on blocking sets in finite geometry, we give explicit constructions and probabilistic lower bounds. We also compute the exact values of this function and its linear variation for small parameters. Moreover, we relate the trifference problem to the sunflower conjecture.**

***Index Terms*—k-hashing, minimal codes, trifference problem.**

I. Introduction

Let Σ denote a finite set of symbols, known as the alphabet, where we typically take $\Sigma = \{0, 1, 2\}$ or a finite field $\mathbb{F}_q$. We will refer to subsets of Σ^n as codes of length n over Σ. If $a, b, c \in \Sigma^n$ are three codewords, then we say that a, b and c *triffer* at position i if $\{a_i, b_i, c_i\} = \{0, 1, 2\}$. The well-known *trifference problem* refers to the determination of the largest size $T(n)$ of a ternary code $C \subseteq \{0, 1, 2\}^n$ in which for every three distinct codewords $a, b, c \in C$, there is at least one position where they triffer. This problem has important applications in computer science [1], [2] and cryptography [3], [4], and is closely related to many seemingly distant topics, such as strong blocking sets in finite projective spaces [5], minimal codes [5], [6], and covering codes [7]. Recently, the problem has gained significant interest, in part due to these connections, which has led to various improvements in the known bounds [5], [8], [9], [10], [11].

The trifference problem is a special case of the following perfect k-hashing problem. Let $k \geq 3$ be a fixed integer. A code C over an alphabet Σ is called a perfect k-hash code if for every k-tuple of distinct codewords in C, there exists a position where all these elements differ. Finding optimal upper bounds for the maximum size of perfect k-hash codes is a fundamental open problem in theoretical computer science. The cornerstone result is based on an elementary double counting argument due to Körner [12], and it states that $|C| \leq (k-1)\left(1 + \frac{1}{k-1}\right)^n$. This upper bound was improved by Fredman and Komlós [13], and later, by Arikan [14]. In recent years, there have been further improvements by Dalai et al. [15], and most recently by Guruswami and Riazanov [16] in the cases when $k > 3$. However, improving the trivial upper bound in the case $k = 3$ remained a notoriously difficult problem. Very recently, Bhandari and Khetan [9] obtained the following polynomial improvement on the upper bound,

$$T(n) \leq cn^{-2/5}\left(\frac{3}{2}\right)^n,$$

where c is an absolute constant. On the other hand, the best lower bound on $T(n)$ is $\left(\frac{9}{5}\right)^{n/4}$ due to Körner and Marton [1] and dates back to 1988 (see [5] for an alternate proof of the same bound).

In this paper, we study the following natural generalization of the trifference problem.

Definition 1: Suppose that $n \geq m \geq 1$ are integers. Let $T(n, m)$ be the largest size of a ternary code $C \subseteq \{0, 1, 2\}^n$ such that for any three distinct codewords x, y, z in C, we have at least m coordinates where they triffer. A code with such a property will be referred to as an *m-trifferent code*.

Recall that the *Hamming distance* $d(u, v)$ between two codewords u, v is the number of coordinates where u and v differ. The expression $A_q(n, d)$ denotes the maximum number of possible codewords in a q-ary block code of length n and minimum Hamming distance d, see [17]. The determination of (the asymptotic behavior of) $A_q(n, d)$ is one of the most central problems in coding theory. The determination of the maximum size of generalized trifferent codes $T(n, m)$ can be interpreted as the analogue of maximum size codes with given Hamming distance, where m corresponds to the 'minimum distance' in

Received 21 May 2025; revised 18 January 2026; accepted 2 February 2026. Date of publication 16 February 2026; date of current version 23 April 2026. This work was supported in part by the EKÖP-25 University Research Scholarship Program; in part by the University Excellence Fund of Eötvös Loránd University; in part by the Bolyai Scholarship of the Hungarian Academy of Sciences; in part by Országos Tudományos Kutatási Alapprogramok (OTKA) under Grant 138596, Grant FK 142993, and Grant STARTING 150576; in part by the Australian Research Council (ARC) Discovery Project under Grant DP250104965; in part by the EXCELLENCE-24 Combinatorics and Geometry of the National Research, Development and Innovation Office (NRDI) Fund under Project 151504; and in part by NSF under Grant DMS-2154063. An earlier version of this paper was presented in part at the 16th Emléktábla Workshop, Vác, Hungary, in 2024. *(Corresponding author: Máté Vizer.)*

Anurag Bishnoi is with Delft Institute of Applied Mathematics, TU Delft, 2628 Delft, The Netherlands (e-mail: a.bishnoi@tudelft.nl).

Bartłomiej Kielak was with the Faculty of Informatics, Masaryk University, 611 37 Brno, Czech Republic. He is now with he Institute of Mathematics, Leipzig University, 04109 Leipzig, Germany (e-mail: bkielak@alumni.uj.edu.pl).

Benedek Kovács and Zoltán Lóránt Nagy are with the ELTE Linear Hypergraphs Research Group, Eötvös Loránd University, 1117 Budapest, Hungary (e-mail: benoke98@student.elte.hu; zoltan.lorant.nagy@ttk.elte.hu).

Gábor Somlai is with the Algebra and Number Theory Department, Eötvös Loránd University, 1117 Budapest, Hungary, and also with The University of Melbourne, Parkville, VIC 3010, Australia (e-mail: gabor.somlai@ttk.elte.hu).

Máté Vizer is with Budapest University of Technology and Economics, 1117 Budapest, Hungary (e-mail: vizermate@gmail.com).

Zeyu Zheng is with the Carnegie Mellon University, Pittsburgh, PA 15213 USA (e-mail: zeyuzhen@andrew.cmu.edu).

Communicated by X. Zhang, Associate Editor for Coding and Decoding.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2026.3665439.

Digital Object Identifier 10.1109/TIT.2026.3665439

0018-9448 © 2026 IEEE. All rights reserved, including rights for text and data mining, and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: TU Delft Library. Downloaded on May 06,2026 at 12:18:01 UTC from IEEE Xplore. Restrictions apply.

this case, thus highlighting its relevance. This generalization has been studied by Bassalygo et al. in [18] under the name of k-hash codes with distance d_k over an alphabet of size q. Our problem corresponds to the special case of $q = k = 3$ and $d_3 = m$.

Our main contribution is to give both lower and upper bounds on $T(n,m)$ for various ranges of m. We first show that there is a phase transition around the value $m \approx \frac{2}{9}n$, which is analogous to the phase transition of error-correcting codes. This was already stated implicitly in [18] for the general problem, but we also include a proof, and we significantly improve their bounds for $q = k = 3$.

Theorem 1: Suppose that $\varepsilon > 0$ is a real number. If $m > \left(\frac{2}{9} + \varepsilon\right) n$, then $T(n,m)$ is bounded above by a constant depending only on ε. On the other hand, if $m < \left(\frac{2}{9} - \varepsilon\right) n$, then $T(n,m)$ is bounded from below by an exponential function $(1+\delta)^n$, where $\delta > 0$ depends only on ε.

Before stating our other main results, let us first state some elementary observations on the function in view.

Proposition 1:

(i) $T(n,n) = 3$ for every $n \geq 1$.
(ii) $T(n,m) \leq T(n-1,m-1)$ for every $n \geq m \geq 2$.
(iii) $T(n,m) \geq T(n/m,1)$ for every $n \geq m \geq 1$ with $m \mid n$.
(iv) $\frac{2}{3}T(n,m) \leq T(n-1,m)$ for every $n > m \geq 1$.

Proof: (*i*) is straightforward. To see (*ii*), observe that if we erase a coordinate of an m-trifferent code, we obtain an $(m-1)$-trifferent code. To prove (*iii*), we take the m-fold repetition of each codeword in a 1-trifferent code of length n/m, giving an m-trifferent code of length n. Finally for (*iv*), if we fix a coordinate i in an m-trifferent code and take all the codewords whose i-th coordinate is not equal to j for an arbitrarily chosen j, we again obtain an m-trifferent code. If we then erase the i-th coordinate, the resulting code is still m-trifferent. Taking j to be the least frequent value appearing at coordinate i, we get the desired bound. □

Using the upper bound on $T(n) = T(n,1)$ from [9], Proposition 1 (ii) applied $m-1$ times immediately gives the following.

Corollary 1: $T(n,m) \leq c(n-m+1)^{-2/5}\left(\frac{3}{2}\right)^{n-m}$ holds, where c is an absolute constant.

In [18], it was shown that the elementary recursive bound can be improved to give

$$T(n,m) \leq m\left(\frac{3}{2}\right)^{n-\frac{9}{2}(m-1)}. \tag{1}$$

We provide significant improvements on these upper bounds in various ranges of values for m.

Theorem 2: Suppose that $m \leq \frac{2}{9}n$. Then

$$T(n,m) \leq \left(\frac{3}{2}\right)^n \frac{3m \cdot 2^{\lceil 2m/3 \rceil}}{\binom{n}{\lceil 2m/3 \rceil - 1}}.$$

Note that for constant m this bound gives a polynomial improvement over Corollary 1 and Equation (1), whereas for $m = \Theta(n)$ it gives an exponential improvement. For small values of m, we obtain the following bounds:

Theorem 3:

$$T(n+1,3) \leq T(n,2) \leq \left(\frac{3}{2}\right)^n \frac{10}{n},$$

$$T(n,4) \leq \left(\frac{3}{2}\right)^n \frac{44}{n^2},$$

$$T(n,5) \leq \left(\frac{3}{2}\right)^n \frac{232}{n^3}.$$

If m is of order $\Theta(n)$, the bounds of Theorem 2 can be further improved and provide asymptotic improvements to (1):

Theorem 4 (Refined upper bound when m is large): Let $m = \lceil \lambda n \rceil$ for some $\lambda \in \left(0, \frac{2}{9}\right)$, and $\hat{x}$ be the smallest real root of the cubic polynomial $\lambda - \frac{3}{2}x(1-x)^2$.

Then

$$T(n,m) \leq \left(\frac{3}{2}\right)^n \frac{C}{2^{-\hat{x}n}\binom{n}{\lceil \hat{x}n \rceil - 1}},$$

with a suitable constant $C = C(\lambda)$.

As far as lower bounds are concerned, we prove the following in the same range of parameters.

Theorem 5 (Lower bound): For a fixed value $0 < \lambda < \frac{2}{9}$, let $m = \lfloor \lambda n \rfloor$. Then there exists a constant $c > 0$ such that

$$T(n,m) \geq c\left(\frac{9}{2}\lambda\right)^{\frac{1}{2}\lambda n}\left(\frac{9}{7}(1-\lambda)\right)^{\frac{1}{2}(1-\lambda)n}$$

for n large enough.

The paper is organized as follows. In Section II we introduce the necessary notation. We present the proofs of Theorems 1, 5, 2, 3 and 4 in Section III. Then we study the linear version of the problem in Section IV. We describe the finite geometric analogues of m-trifferent linear codes and their relation to minimal codes. Using probabilistic constructions, we obtain lower bounds on linear m-trifferent codes $T_L(n,m)$. Finally, in Section V, we present computational results on $T(n,m)$ and $T_L(n,m)$, and discuss the relation between the sunflower conjecture and the trifference problem.

II. Notation and Preliminaries

For a codeword $c \in \Sigma^n$, $\sigma(c)$ denotes the *support* of the codeword, i.e., the set of indices where c is non-zero. The notation $[n]$ denotes the integer interval $\{x \in \mathbb{Z} : 1 \leq x \leq n\}$.

Definition 2: A linear code $C \subseteq \mathbb{F}_q^n$ is called *minimal* if for any two linearly independent $c, c' \in C$, we have $\sigma(c) \not\subseteq \sigma(c')$. A linear code $C \subseteq \mathbb{F}_q^n$ is called *m-minimal* if after the deletion of any $m-1$ coordinates, all codewords remain distinct and the remaining code is still minimal.

Logarithms in this paper are taken base 2.

Definition 3: The *binary entropy function* $H(p)$ is defined as

$$H(p) = -p\log p - (1-p)\log(1-p)$$

for $0 \leq p \leq 1$.

We will use the following well-known bound on binomial coefficients:

Lemma 1:

$$\frac{1}{n+1}2^{nH\left(\frac{k}{n}\right)} \leq \binom{n}{k} \leq 2^{nH\left(\frac{k}{n}\right)}.$$

For a refined bound, see [19, Lemma 17.5.1, p. 666.]

The *relative entropy* or Kullback-Liebler divergence is defined as

$$D_{\mathrm{KL}}(P \parallel Q) = \sum_{x \in \mathcal{X}} P(x)\log\left(\frac{P(x)}{Q(x)}\right)$$

Authorized licensed use limited to: TU Delft Library. Downloaded on May 06,2026 at 12:18:01 UTC from IEEE Xplore. Restrictions apply.

for a pair of discrete probability distributions P and Q defined on the same sample space $\mathcal{X}$. In the special case when P and Q correspond to the distribution of indicator variables with appearance probability p and p', we use the simpler notation $H(p,p')$, which simplifies to

$$H(p,p') = p\log\frac{p}{p'} + (1-p)\log\frac{1-p}{1-p'}.$$

Definition 4: For any $S \subseteq \{0,1,2\}^n$, we define $T_S(n,m)$ to be the size of the largest m-trifferent set contained in S.

The lemma below follows the ideas of the proof in [9] for the similar bound on $T_S(n)$. This statement will enable us to give upper bounds on $T(n,m)$ via bounding the maximum size of m-trifferent sets contained in certain subsets of the space Σ^n.

Lemma 2: For any subset $S \subseteq \{0,1,2\}^n$,

$$\frac{T(n,m)}{3^n} \leq \frac{T_S(n,m)}{|S|}.$$

Proof: Let C be an m-trifferent code in $\mathbb{F}_3^n$, identified with $\{0,1,2\}^n$. For a vector $v \in \mathbb{F}_3^n$, let $S_v = S + v = \{x+v : x \in S\}$. We double count the size of the set $\{(x,v) : x \in C \cap S_v\}$. Since $x \in C$ is contained in S_v if and only if $v = x - s$ for some $s \in S$, each element of C is contained in $|S|$ many S_v's. Therefore,

$$\sum_{v \in \mathbb{F}_3^n} |C \cap S_v| = |C||S|.$$

By averaging, there must exist a vector u such that $|C \cap S_u| \geq |C||S|/3^n$. We know that $C \cap S_u$ is an m-trifferent set inside S_u, and since S_u is just a translate of S, $|C \cap S_u| \leq T_S(n,m)$, giving us the inequality. □

Notice that the proof of the previous lemma uses the fact that the set of subsets of $\mathbb{F}_3^n$ that triffer is closed under translation.

III. Proof of the Lower and Upper Bounds

We start by proving the lower bound on $T(n,m)$.

Theorem 6: Suppose that $1 \leq m < \frac{2}{9}n$. Then $T(n,m) \geq \frac{2\sqrt{2}}{3} \cdot 2^{nh/2} - 1$, where $h = H\left(\frac{m}{n}, \frac{2}{9}\right)$.

Proof: We use the alteration method. For a positive integer t fixed later on, choose codewords $c_1, c_2, \ldots, c_t \in \{0,1,2\}^n$ independently and uniformly at random, and let $C_0 = \{c_1, c_2, \ldots, c_t\}$. A triple of codewords $\{c_x, c_y, c_z\}$ is called *bad* if they triffer at less than m positions. Let B denote the number of bad triples in C_0. Clearly $\mathbb{E}[B] = \binom{t}{3} p_b$ where $p_b = \mathbb{P}[\{c_1, c_2, c_3\} \text{ is bad}]$. Create a new code C from C_0 by deleting one of the codewords from each bad triple. Then, C will be m-trifferent, and $|C| \geq t - B$. Therefore, $T(n,m) \geq \mathbb{E}[|C|] \geq t - \binom{t}{3} p_b$.

If c_1, c_2 and c_3 are uniformly random elements of $\{0,1,2\}^n$, then for each $1 \leq i \leq n$, the probability that the codewords triffer at position i is $\frac{3!}{3^3} = \frac{2}{9}$. Therefore, if X denotes the number of positions where the three codewords triffer, then X follows the binomial distribution $\mathrm{Bin}\left(n, \frac{2}{9}\right)$. By the well-known Chernoff bound, $p_b \leq \mathbb{P}[X \leq m] \leq 2^{-nh}$, where $h = H\left(\frac{m}{n}, \frac{2}{9}\right)$ satisfies $0 < h < \log\frac{9}{7}$.

Thus, if we choose $t = \lfloor \sqrt{2} \cdot 2^{nh/2} \rfloor$, then we get

$$T(n,m) \geq t - \binom{t}{3} p_b \geq t - \frac{1}{6} t^3 p_b$$

$$\geq \sqrt{2} \cdot 2^{nh/2} - 1 - \frac{\sqrt{2}}{3} \cdot 2^{3nh/2} 2^{-nh}$$

$$= \frac{2\sqrt{2}}{3} \cdot 2^{nh/2} - 1.$$

□

Remark 1:

(i) In the case where m is a function of n with $m = o(n)$ (for example $m \geq 1$ is a constant) and $n \to \infty$, we get that for some $c > 0$, $T(n,m) \geq c\left(\frac{3}{\sqrt{7}}\right)^n$ for all n. Note that $\frac{3}{\sqrt{7}} \approx 1.1339$.

(ii) If $m = \lfloor \lambda n \rfloor$ for a fixed value $0 < \lambda < \frac{2}{9}$, and $n \to \infty$, we get that for some $c > 0$, $T(n,m) \geq c \cdot 2^{\frac{1}{2}H(\lambda, \frac{2}{9})n} = cb(\lambda)^n$, where

$$b(\lambda) = 2^{\frac{1}{2}H(\lambda,\frac{2}{9})} = \left(\frac{9}{2}\lambda\right)^{\frac{1}{2}\lambda} \left(\frac{9}{7}(1-\lambda)\right)^{\frac{1}{2}(1-\lambda)}$$

is a strictly decreasing function for $\lambda \in \left[0, \frac{2}{9}\right]$ with $b(0) = \frac{3}{\sqrt{7}}$ and $b\left(\frac{2}{9}\right) = 1$.

This completes the proof of Theorem 5, which also implies Theorem 1 when $m < \left(\frac{2}{9} - \varepsilon\right)n$. Next, we complete the proof of Theorem 1 by considering the case $m > \left(\frac{2}{9} + \varepsilon\right)n$. Note that the following result is also a consequence of Theorem 4 in [18] and it is similar to the proof of the Plotkin bound [20].

Theorem 7: For any fixed $\frac{2}{9} < \lambda \leq 1$, there exists a constant $c(\lambda)$ such that $T(n, \lceil \lambda n \rceil) \leq c(\lambda)$ for all n.

Proof: Let $C = \{c_1, c_2, \ldots, c_T\} \subseteq \{0,1,2\}^n$ be a code for which any three codewords triffer in at least λn positions.

Let K denote the number of quadruples (x,y,z,i) such that:

- $1 \leq x < y < z \leq T$ and $1 \leq i \leq n$ are integers,
- and codewords c_x, c_y and c_z triffer at position i.

We will bound K from below and above. Firstly, we can see that for every triple (x,y,z) with $x < y < z$, we have at least λn positions where the three codewords triffer, so $K \geq \binom{T}{3} \cdot \lambda n$.

On the other hand, for a fixed value $1 \leq i \leq n$, let us put an upper bound on the number of triples triffering at position i. If the number of codewords in C having i-th coordinate 0, 1 and 2 is r_i, s_i and t_i respectively, then the following hold:

- $r_i, s_i, t_i \geq 0$,
- $r_i + s_i + t_i = T$,
- the number of triples triffering at position i is $r_i s_i t_i$.

By the AM-GM inequality, $r_i s_i t_i \leq \left(\frac{r_i+s_i+t_i}{3}\right)^3 = \frac{1}{27}T^3$. Therefore,

$$K = \sum_{i=1}^{n} r_i s_i t_i \leq \frac{1}{27} n T^3.$$

This means that $\binom{T}{3} \cdot \lambda n \leq \frac{1}{27} n T^3$, giving

$$\frac{T(T-1)(T-2)}{T^3}\lambda \leq \frac{2}{9}.$$

Since $\frac{T(T-1)(T-2)}{T^3} \to 1$ as $T \to \infty$ and $\frac{2}{9}/\lambda < 1$, we conclude that T is bounded by a constant depending only on λ. □

We continue by proving upper bounds on $T(n,m)$. First, we present a consequence of the threshold value $\lambda = \frac{2}{9}$ and of Proposition 1. This is similar to Theorem 5 of [18] and the Singleton bound on codes.

Theorem 8: Let $0 < \alpha < \frac{2}{9}$ and $0 < \varepsilon < \frac{7}{9}$ be real numbers. Then

$$T\left(n, \left\lceil \left(\frac{2}{9} - \alpha\right) n \right\rceil\right) \leq c \cdot \left(\frac{3}{2}\right)^{4.5(\alpha+\varepsilon)n}$$

Authorized licensed use limited to: TU Delft Library. Downloaded on May 06,2026 at 12:18:01 UTC from IEEE Xplore. Restrictions apply.

for some constant $c = c(\varepsilon)$.

Proof: For $m = \left\lceil \left(\frac{2}{9} - \alpha\right) n \right\rceil$, Proposition 1 (iv) can be applied t times iteratively, where

$$t = \left\lceil n - \frac{1}{\frac{2}{9} + \varepsilon} m \right\rceil.$$

We hence get $T(n,m) \le \left(\frac{3}{2}\right)^t T(n-t,m)$ and it follows from the choice of t that $m \ge \left(\frac{2}{9} + \varepsilon\right)(n-t)$, so by Theorem 7 we have that $T(n-t,m)$ is at most a constant $c' = c'(\varepsilon)$. Altogether,

$$T\left(n, \left\lceil \left(\frac{2}{9} - \alpha\right) n \right\rceil\right) \le c' \cdot \left(\frac{3}{2}\right)^{n - \frac{1}{\frac{2}{9}+\varepsilon} m + 1} \le c \cdot \left(\frac{3}{2}\right)^{n\left(1 - \frac{2/9 - \alpha}{2/9 + \varepsilon}\right)} \le c \cdot \left(\frac{3}{2}\right)^{\frac{9}{2}(\alpha+\varepsilon)n},$$

where $c(\varepsilon) = \frac{3}{2} c'(\varepsilon)$. □

We now look at the other regime where m is small. We first prove a better upper bound than $T(n,2) \le T(n,1) \le cn^{-2/5}(3/2)^n$.

Theorem 9:

$$T(n,2) \le \frac{10}{n} \left(\frac{3}{2}\right)^n.$$

Proof: We use the setup of Lemma 2 for the set $S = \{x \in \{0,1,2\}^n : \#\{i : x_i = 2\} = 1\}$. Let $C' \subseteq S$ be a 2-trifferent code. Observe that if $x_i = y_i = 2$ for some $x, y \in S$ and $i \in [n]$, then for any $z \in S$ the triple (x,y,z) would triffer in at most one position. Hence, for any $i \in [n]$ there exists at most one $x \in C'$ such that $x_i = 2$.

Fix some $t \in C'$ and assume that $t_1 = 2$. The remaining codewords in C' start with 0 or 1 so if $|C'| \ge 6$, then by pigeonhole principle there exist three codewords $x, y, z \in C'$ such that $x_1 = y_1 = z_1$. Without loss of generality, $t = (2, *, \cdots, *)$, $x = (0, 2, *, \cdots, *)$, $y = (0, *, 2, *, \cdots, *)$, and $z = (0, *, *, 2, *, \cdots, *)$, where each * is either a 0 or 1. By considering the triples t, x, y and t, x, z, we get $y_2 = z_2$, and from the triples t, x, y and t, y, z we get $x_3 = z_3$. Therefore, $\{x_i, y_i, z_i\} = \{0,1,2\}$ is only possible at $i = 4$, so x, y and z triffer in at most 1 position, contradicting the 2-trifference property.

Hence $|C'| \le 5$, and by Lemma 2 we get

$$T(n,2) \le \frac{5 \cdot 3^n}{n \cdot 2^{n-1}} = \frac{10}{n} \left(\frac{3}{2}\right)^n.$$

□

Before proving the main upper bound (Theorem 4), we present an upper bound in the spirit of the bound on $T(n,1)$ given by Bhandari and Khetan [9], using extremal graph theoretic tools for an auxiliary graph assigned to the trifferent code.

Theorem 10:

$$T(n, 2\ell+1) \le \left(\frac{3}{2}\right)^n \frac{\binom{n}{\ell}}{\binom{n}{2\ell}} \cdot 2(2\ell+1).$$

Proof: We now use the setup of Lemma 2 for $m = 2\ell + 1$ and $S = \{x \in \{0,1,2\}^n : \#\{i : x_i = 2\} = 2\ell\}$. We wish to bound $T_S(n, 2\ell+1)$ from above. To do this, we will define a bipartite graph G on the two vertex classes U and V, where U and V are disjoint sets with size $|U| = |V| = \binom{n}{\ell}$, both of whose elements are labelled with the ℓ-element subsets of $[n]$. Fix any $(2\ell+1)$-trifferent code $C' \subseteq S$. For each codeword $x \in C'$, there exists a set of indices $I \subseteq [n]$ of size 2ℓ such that $x_i = 2$ for each $i \in I$ and $x_i \ne 2$ otherwise. For each partition $I = I_1 \cup I_2$ with $|I_1| = |I_2| = \ell$, we add an edge (I_1, I_2) to G, where $I_1 \in U$ and $I_2 \in V$. This way, we add $\binom{2\ell}{\ell}$ edges for each $x \in C'$. Moreover, since C' is $(2\ell+1)$-trifferent, each edge corresponds to at most one codeword from C'. Therefore, $T_S(n, 2\ell+1)$ is at most $\frac{|E(G)|}{\binom{2\ell}{\ell}}$.

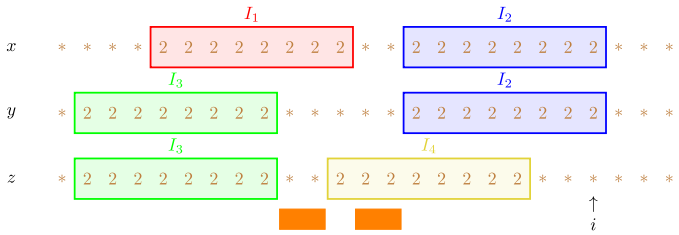


Fig. 1. Illustration of the structure of x, y, z in the proof of Theorem 10, where each * indicates a digit 0 or 1. The codewords may only triffer at positions where exactly one of them has a 2, as indicated in orange. Such positions must lie in the symmetric difference of I_1 and I_4.

We claim that G does not contain a P_4 (a four-vertex path). Assume it is not the case and (I_1, I_2), (I_3, I_2), and (I_3, I_4) are edges of some P_4 (where $I_1, I_3 \in U$ and $I_2, I_4 \in V$) and these edges correspond to some codewords x, y, $z \in C'$. Fig. 1 shows an example for the possible structures of x, y, z. Since $I_1 \cup I_2 \ne I_3 \cup I_2$, we must have $x \ne y$; analogously, $y \ne z$. Moreover, since $I_2 \ne I_4$ and both are of the same size and their union is disjoint from I_3, there exists $i \in I_2 \setminus (I_3 \cup I_4)$. In particular, $x_i = 2 \ne z_i$, which implies $x \ne z$, hence x, y, z are pairwise different. Since x, y, z are codewords from a $(2\ell+1)$-trifferent code, there must exist a set $J \subseteq [n]$ of size at least $2\ell + 1$ such that $\{x_i, y_i, z_i\} = \{0,1,2\}$ for each $i \in J$. However, for every such $i \in J$, exactly one of x_i, y_i, z_i is 2, so by the structure of the three codewords, J must be a subset of the symmetric difference of I_1 and I_4 (see Fig. 1), which consists of at most 2ℓ elements. This gives a contradiction.

Since G is P_4-free (and also triangle-free by definition), it must be a star forest, hence it contains at most $|V(G)| - 1 < 2\binom{n}{\ell}$ edges. This implies $T_S(n, 2\ell+1) \le \frac{2}{\binom{2\ell}{\ell}} \cdot \binom{n}{\ell}$ and

$$T(n, 2\ell+1) \le 3^n \cdot \frac{T_S(n, 2\ell+1)}{|S|} \le 3^n \cdot \frac{\frac{2}{\binom{2\ell}{\ell}} \cdot \binom{n}{\ell}}{\binom{n}{2\ell} 2^{n-2\ell}} \le \left(\frac{3}{2}\right)^n \frac{\binom{n}{\ell}}{\binom{n}{2\ell}} \frac{\cdot 2^{2\ell+1}}{\binom{2\ell}{\ell}} \le \left(\frac{3}{2}\right)^n \frac{\binom{n}{\ell}}{\binom{n}{2\ell}} \cdot 2(2\ell+1).$$

□

Remark 1: We used Lemma 1 here in the last step. However, it is well known that $\frac{2^{2\ell}}{\binom{2\ell}{\ell}} = (1 + o(1))\sqrt{\pi\ell}$ holds. Thus, the above argument also shows that

$$T(n, 2\ell+1) \le (1 + o(1)) \left(\frac{3}{2}\right)^n \frac{\binom{n}{\ell}}{\binom{n}{2\ell}} \cdot \sqrt{\pi\ell}.$$

Next, we prove Theorems 2 and 4 and compare them with the upper and lower bounds above. The proof is similar in flavor to the proof of Elias-Bassalygo bound [21] for binary codes.

Authorized licensed use limited to: TU Delft Library. Downloaded on May 06,2026 at 12:18:01 UTC from IEEE Xplore. Restrictions apply.

Proof: [Proof of Theorem 2 and 4] In the setup of Lemma 2, suppose that S consists of the codewords with exactly $k < \frac{1}{3}n$ many 2's, where k is an integer to be fixed later. Let $C = \{c_1, c_2, \ldots, c_T\} \subseteq S$ be an m-trifferent code. To bound $|C| = T$, we double count the number of quadruples

$$K = |\{(x, y, z, i) : 1 \le x < y < z \le T,\ i \in [n],\\ c_x, c_y \text{ and } c_z \text{ triffer at position } i\}|$$

along the lines of the proof of Theorem 7. Then $K \ge m\binom{T}{3}$. Let r_i, s_i, t_i denote the number of codewords in C having i-th coordinate 0, 1, and 2, respectively. Then $\sum_{i=1}^{n} r_i s_i t_i = K$ and $\sum_{i=1}^{n} t_i = kT$. Note that

$$m\binom{T}{3} \le \sum_{i=1}^{n} r_i s_i t_i \le \sum_{i=1}^{n} t_i \left(\frac{T - t_i}{2}\right)^2 \\ \le n\frac{kT}{n}\left(\frac{T - \frac{kT}{n}}{2}\right)^2 \le \frac{kT^3}{4}. \tag{2}$$

Here, the use of Jensen's inequality in the step $\sum_{i=1}^{n} t_i \left(\frac{T-t_i}{2}\right)^2 \le n\frac{kT}{n}\left(\frac{T-\frac{kT}{n}}{2}\right)^2$ is justified only in the case when $t_i \le \frac{2}{3}T$ for every i, as the function $x(T-x)^2$ is concave in $\left[0, \frac{2}{3}T\right]$. However, the maximum value of the left hand side over $(t_i) \in \mathbb{Z} \cap [0, T]$ cannot be taken when there exists an index i for which $t_i > \frac{2}{3}T$. Indeed, the sum $\sum_{i=1}^{n} t_i = kT$ implies that there exists an index j with $t_j < \frac{1}{3}T$ in view of the assumption on k. But then altering $t_i \to t_i - 1$, $t_j \to t_j + 1$ for this index pair, we would obtain an increment in the sum, as

$$(t_i - 1)\left(\frac{T - t_i + 1}{2}\right)^2 + (t_j + 1)\left(\frac{T - t_j - 1}{2}\right)^2 \\ - t_i\left(\frac{T - t_i}{2}\right)^2 - t_j\left(\frac{T - t_j}{2}\right)^2 \\ = \frac{1}{4}(4T - 3\ t_i - 3\ t_j)(-1 + t_i - t_j) > 0,$$

since both factors $(4T - 3\ t_i - 3\ t_j) = ((3T - 3\ t_i) + (T - 3\ t_j))$ and $(-1 + t_i - t_j)$ are positive.

Now let us simplify the inequality (2): $m\binom{T}{3} \le \frac{kT^3}{4}$. By expanding the formula, we obtain

$$(2m - 3k)T^2 - 6mT + 4m \le 0. \tag{3}$$

Plugging in $k = \lceil 2m/3 \rceil - 1$ gives us the bound

$$T \le \begin{cases} 6m & \text{if } m \equiv 2 \pmod 3 \\ 3m & \text{if } m \equiv 1 \pmod 3 \\ 2m & \text{if } m \equiv 0 \pmod 3. \end{cases}$$

Hence, if m is of form $m = 3\ell + r$ for $r \in \{0, 1, 2\}$, then Lemma 2 implies

$$T(n, m) \le 3^n \cdot \frac{T_S(n, m)}{\binom{n}{k} \cdot 2^{n-k}} \le \begin{cases} \left(\frac{3}{2}\right)^n \frac{6(3\ell+2)2^{2\ell+1}}{\binom{n}{2\ell+1}} & \text{if } r = 2 \\ \left(\frac{3}{2}\right)^n \frac{3(3\ell+1)2^{2\ell}}{\binom{n}{2\ell}} & \text{if } r = 1 \\ \left(\frac{3}{2}\right)^n \frac{6\ell 2^{2\ell-1}}{\binom{n}{2\ell-1}} & \text{if } r = 0, \end{cases} \tag{4}$$

as claimed in Theorem 2. Note that the same argument can be slightly refined, as Inequality (3) implies a slightly better upper bound on T. Indeed, for small values of m, we get the exact bounds presented (m), k, and T, as shown at the bottom of the page. These bounds on T together with Inequality (4), and the fact that T is an integer, give the upper bounds of Theorem 3.

Now we refine the argument to derive Theorem 4. Suppose now that $m = \lceil \lambda n \rceil$ for some $0 < \lambda < \frac{2}{9}$. Let $\hat{x}$ be the smallest real root of $\lambda = \frac{3}{2}x(1-x)^2$. From Inequality (2), we then get

$$m\binom{T}{3} \le kT\left(\frac{T - \frac{kT}{n}}{2}\right)^2 = \frac{T^3}{4}k\left(1 - \frac{k}{n}\right)^2.$$

Take $k = \mu n := \lceil \hat{x}n \rceil - 1$. Then we get

$$\lambda(T-1)(T-2) \le \frac{3T^2}{2}\mu(1-\mu)^2,$$

which is equivalent to

$$\left(\lambda - \frac{3}{2}\mu(1-\mu)^2\right)T^2 \le \lambda(3T - 2).$$

Consequently, $T < \frac{3\lambda}{\lambda - \frac{3}{2}\mu(1-\mu)^2}$, provided that $\lambda - \frac{3}{2}\mu(1-\mu)^2$ is positive, which holds in view of the choice of μ. Then, similarly to the previous case, we obtain that

$$T(n, m) \le 3^n \cdot \frac{T_S(n, m)}{\binom{n}{k} \cdot 2^{n-k}} \le \left(\frac{3}{2}\right)^n \frac{C}{2^{-\hat{x}n}\binom{n}{\mu n}},$$

with a suitable constant $C = C(\lambda)$.

Here, note that $\hat{x}/\lambda = \hat{x}(\lambda)\lambda^{-1}$ is monotone increasing in $\lambda \in (0, 2/9)$ for fixed n, with $\hat{x}/\lambda \ge 2/3$ on the whole interval while $\hat{x}n/m \to \frac{3}{2}$ as $\lambda \to 2/9$ and $n \to \infty$. On the other hand, μ is a good approximation of $\hat{x}$ as $|\mu - \hat{x}| \le n^{-1}$. □

Remark 2: In Fig. 2, we compare the lower bound obtained from the probabilistic construction in Remark 1, with the former upper bound and three new upper bounds from Corollary 1 and Theorems 4 and 10, for a fixed $0 \le \lambda \le \frac{2}{9}$. For each λ, the values $b(\lambda)$ are plotted where $T(n, \lfloor \lambda n \rfloor) \ge (b(\lambda)(1 - o(1)))^n$ for lower bounds and $T(n, \lceil \lambda n \rceil) \le (b(\lambda)(1 + o(1)))^n$ for upper bounds. Because of Theorem 7 we know that $b(\lambda) = 1$ is the correct value for $\frac{2}{9} < \lambda \le 1$.

IV. The Linear Version

The equivalence between linear trifferent codes and minimal codes was first showed in [5]. Here, we establish an analogous equivalence between linear m-trifferent codes and m-minimal codes, which allows us to leverage the rich theory of minimal codes to study linear trifferent codes.

Definition 5 (m-minimal codes): A linear code $C \subseteq \mathbb{F}_q^n$ is called *m-minimal* if after the deletion of any $m-1$ coordinates, all codewords of C remain distinct and the remaining code is still minimal.

Lemma 3: A linear code $C \subseteq \mathbb{F}_3^n$ is m-trifferent if and only if it is m-minimal.

(m)	2	3	4	5	6
k	1	1	2	3	3
$T \le$	$2(3+\sqrt{7}) \approx 11.29$	$3+\sqrt{5} \approx 5.24$	$2(3+\sqrt{7}) \approx 11.29$	$15+\sqrt{205} \approx 29.32$	$2(3+\sqrt{7}) \approx 11.29$

Authorized licensed use limited to: TU Delft Library. Downloaded on May 06,2026 at 12:18:01 UTC from IEEE Xplore. Restrictions apply.

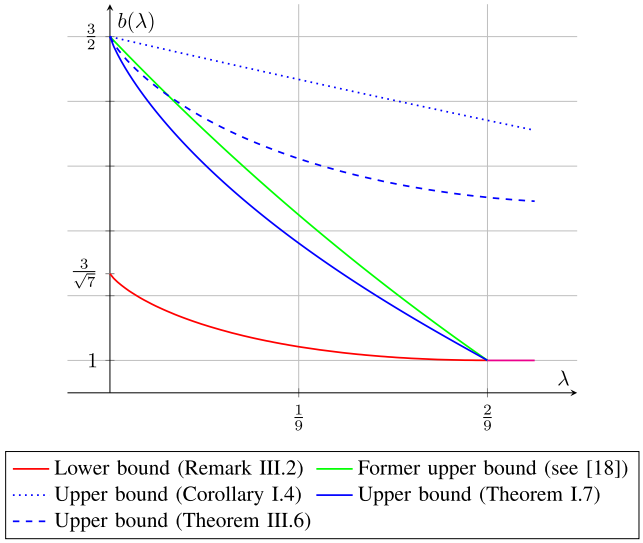


Fig. 2. Graph indicating various bounds of the form $(b(\lambda)(1 \pm o(1)))^n$ on $T(n, \lambda n)$.

Proof: Suppose C is m-trifferent but not m-minimal. Then the codewords of C must still be distinct after deleting any set of $m-1$ coordinates: otherwise we would have two codewords which differ in at most $m-1$ positions, so together with an arbitrary third codeword they would violate m-trifference. So there must exist linearly independent codewords x and y such that after deleting some set of $m-1$ coordinates, $\sigma(x) \subseteq \sigma(y)$. Consider the codeword $z = -x$, which must be in C by linearity. Let i be a coordinate (after the deletion of $m-1$ coordinates) where x, y, z triffer. One of x_i, y_i, z_i has to be zero. Note that $y_i = 0$ implies $z_i = x_i = 0$, while $x_i = 0$ if and only if $z_i = 0$. Therefore, the triple $\{x, y, z\}$ does not triffer at position i, a contradiction.

For the backward implication, suppose C is m-minimal but not m-trifferent. Let x, y, z be three distinct codewords and let S denote the coordinates where x, y, z triffer. Suppose $|S| \leq m-1$. We extend S to $\widetilde{S}$, a subset of $[n]$ with exactly $m-1$ coordinates. Without loss of generality, assume $z = 0$ by replacing x, y, z with $x-z, y-z, 0$. Consider $w = x+y$. For any coordinate i outside S, $\{x_i, y_i\} \neq \{1, 2\}$. Therefore, $\sigma(w) \setminus \widetilde{S} = (\sigma(x) \setminus \widetilde{S}) \cup (\sigma(y) \setminus \widetilde{S})$. Since x and y remain distinct and nonzero after deleting the coordinates in $\widetilde{S}$ (C is m-minimal), we have $(\sigma(x) \setminus \widetilde{S}) \neq (\sigma(y) \setminus \widetilde{S})$, which implies that either $(\sigma(x) \setminus \widetilde{S}) \subsetneq (\sigma(w) \setminus \widetilde{S})$ or $(\sigma(y) \setminus \widetilde{S}) \subsetneq (\sigma(w) \setminus \widetilde{S})$, contradicting m-minimality. □

A. The Upper Bound

Let $\mathrm{PG}(k-1, q)$ denote the projective space $(\mathbb{F}_q^k \setminus \{0\})/\sim$, where $u \sim v$ if $u = \lambda v$.

Definition 6: Let $m \geq 1$ be an integer. A set of points S on a hyperplane H is a spanning set *with strength* m, if S remains a spanning set after removing any $m-1$ of its points.

As introduced in [22], a *strong blocking set* $B \subseteq \mathrm{PG}(k-1, q)$ is a set of points that intersects every hyperplane H in a spanning set. See also [5], [7], [23].

Definition 7: Let $m \geq 1$ be an integer. A set of points B in $\mathrm{PG}(k-1, q)$ is a strong blocking set *with strength* m if B remains a strong blocking set after removing any $m-1$ of its points.

It has been shown in [24] that C is a minimal code of dimension k in $\mathbb{F}_q^n$ with a generator matrix having no two linearly dependent columns if and only if the columns of its generator matrix form a strong blocking set in $\mathrm{PG}(k-1, q)$. We now provide a generalization of this statement.

Lemma 4: C is an m-minimal code of dimension k in $\mathbb{F}_q^n$ with a generator matrix having no two linearly dependent columns if and only if the columns of its generator matrix form a strong blocking set S with strength m in $\mathrm{PG}(k-1, q)$.

Proof: A code has the m-minimal property if and only if after every deletion of $m-1$ columns it is a minimal code. In other words, the strong blocking set formed by the columns of the generator matrix is a strong blocking set even after removing any set of $m-1$ elements. □

In [5], it was also shown that a set of points $\mathcal{L} \subseteq \mathrm{PG}(k-1, q)$ is a strong blocking set if and only if $B = \cup_{\ell \in \mathcal{L}} \ell \subseteq \mathbb{F}_q^k$ is an affine 2-blocking set, i.e., if B meets every $(k-2)$-dimensional affine subspace of $\mathbb{F}_q^k$ in at least one point. Hence we have the following corollary.

Corollary 2: Let C be a linear code of dimension k in $\mathbb{F}_3^n$ with generator matrix G. Suppose G has no two linearly dependent columns. Then C is an m-trifferent code if and only if the union of 1-dimensional subspaces of $\mathbb{F}_3^k$ spanned by the columns of G meet every affine subspace of dimension $(k-2)$, not passing through the origin, in at least m points. □

Let $T_L(n, m)$ be the largest size of a linear m-trifferent code of length n.

Theorem 11: For some absolute constant $c > 0$, $T_L(n, m) \leq c \cdot 3^{\frac{n-m}{4.55}}$.

Proof: Let $k = \log_3 T_L(n, m)$ be the dimension of the code. Then we have a set of n points in $\mathrm{PG}(k-1, 3)$ such that every $n-m+1$ element subset of it is a strong blocking set. Since a strong blocking set in $\mathrm{PG}(k-1, 3)$ has size at least $4.55(k-1)$ (see [5]), we get that $n \geq 4.55(k-1) + m - 1$, giving the statement. □

B. The Lower Bound

From the analysis above, to lower bound the rate of $T_L(n, m)$, we only need to upper bound the size of S. We use a probabilistic approach to construct a strong blocking set S with strength m in the spirit of the bound of Héger and Nagy on the length of minimal codes [7].

Theorem 12: In $\mathrm{PG}(k-1, 3)$, there exists a strong blocking set with strength m of size at most $43k+18m-93$. In particular, when k is sufficiently large and $m = o(k)$, the bound can be improved to $10k + 5m - 21$.

Proof: We uniformly choose t lines $\ell_1, \ell_2, \ldots \ell_t$ from $\mathrm{PG}(k-1, 3)$. Consider the probability that $\mathcal{B} = \bigcup_{i=1}^{t} \ell_i$ fails to be a strong blocking set with strength m in $\mathrm{PG}(k-1, 3)$. This happens if and only if there exists a hyperplane H such that, after the removal of some $m-1$ points from $\mathcal{B}$ to obtain a modified set $\mathcal{B}^*$, the span of $H \cap \mathcal{B}^*$ is strictly contained in H, i.e., $\dim\langle H \cap \mathcal{B}^* \rangle < k-2$. In particular, there exists a $(k-3)$-dimensional projective subspace H' such that apart from at most $m-1$ lines, $H' \cap \ell_i$ is non-empty. This happens with probability at most

$$\begin{bmatrix} k \\ k-2 \end{bmatrix}_3 \cdot \mathbb{P}(\ell_1 \cap H' \neq \emptyset)^{t-m+1} \cdot \binom{t}{m-1},$$

Authorized licensed use limited to: TU Delft Library. Downloaded on May 06,2026 at 12:18:01 UTC from IEEE Xplore. Restrictions apply.

which gives an upper bound on the probability that $\mathcal{B}$ fails to be a strong blocking set.

Note that for H' with $\dim H' = k-3$,

$$\begin{aligned}\mathbb{P}(\ell_1 \cap H' \neq \emptyset) &= \frac{\begin{bmatrix}k-2\\2\end{bmatrix}_3 + \frac{1}{3}\cdot\begin{bmatrix}k-2\\1\end{bmatrix}_3\left(\begin{bmatrix}k\\1\end{bmatrix}_3 - \begin{bmatrix}k-2\\1\end{bmatrix}_3\right)}{\begin{bmatrix}k\\2\end{bmatrix}_3}\\ &= \frac{\frac{(3^{k-2}-1)(3^{k-3}-1)}{16} + \frac{1}{3}\cdot\frac{3^{k-2}-1}{2}\left(\frac{3^k-1}{2} - \frac{3^{k-2}-1}{2}\right)}{\frac{(3^k-1)(3^{k-1}-1)}{16}}\\ &= \frac{(3^{k-2}-1)(3^{k-3}-1) + 32\cdot 3^{2k-5} - 32\cdot 3^{k-3}}{(3^k-1)(3^{k-1}-1)}\\ &= \frac{11\cdot 3^{2k-4} - 4\cdot 3^{k-1} + 1}{27\cdot 3^{2k-4} - 4\cdot 3^{k-1} + 1}\\ &< \frac{11}{27}.\end{aligned}$$

Also we can upper bound $\begin{bmatrix}k\\k-2\end{bmatrix}_3$ by $3^{2k-1}/16$. Hence, $\mathcal{B}$ is a blocking set with strength m with positive probability if

$$\frac{3^{2k-1}}{16}\cdot\left(\frac{11}{27}\right)^{t-m+1}\cdot\binom{t}{m-1} < 1.$$

Taking logarithms and relaxing, above inequality holds if

$$\begin{aligned}&(2k-1)\log 3 - 4\\ &+ (t-m+1)\log\left(\frac{11}{27}\right) + t\cdot H\left(\frac{m-1}{t}\right) < 0. \qquad (5)\end{aligned}$$

A further relaxation of the entropy term yields the simpler sufficient condition

$$(2k-1)\log 3 - 4 + (t-m+1)\log\left(\frac{11}{27}\right) + t < 0. \qquad (6)$$

Rearranging this inequality gives

$$t > -\frac{2\log 3}{\log\left(\frac{22}{27}\right)}k + \frac{\log\left(\frac{11}{27}\right)}{\log\left(\frac{22}{27}\right)}m + \frac{\log 3 + 4 - \log\left(\frac{11}{27}\right)}{\log\left(\frac{22}{27}\right)}. \qquad (7)$$

Hence, one may choose

$$t = \lceil 10.729k + 4.385m - 23.287\rceil.$$

Since each line in $\mathrm{PG}(k-1,3)$ contains 4 points,

$$|\mathcal{B}| \leq 4t \leq 43k + 18m - 93.$$

We now assume k is sufficiently large and $m = o(k)$. From equation (7) we see that t grows at least linearly with k, hence $m = o(t)$. Consequently, the entropy term satisfies $H\left(\frac{m-1}{t}\right) \to 0$. A sufficient condition for inequality (6) becomes for some $\epsilon > 0$,

$$\begin{aligned}&\left(\log\left(\frac{11}{27}\right) + \epsilon\right)t\\ &< -2\log 3k + \log\left(\frac{11}{27}\right)m + \log 3 + 4 - \log\left(\frac{11}{27}\right).\end{aligned}$$

We may choose

$$t = \lceil 2.447k + 1.001m - 5.310\rceil,$$

which yields

$$|\mathcal{B}| \leq 4t \leq 10k + 5m - 21.$$

□

Corollary 3: For any integer $m < 2n/9$, we have

$$T_L(n,m) \geq 3^{\lfloor\frac{n-18m+93}{43}\rfloor}.$$

Moreover, when $m = o(n)$, for sufficiently large n, this bound improves to

$$T_L(n,m) \geq 3^{\left\lfloor\frac{n-5m+21}{10}\right\rfloor}.$$

Proof: Let $C \subseteq \mathbb{F}_3^n$ be a linear code of dimension k. Its generator matrix defines a set of n points in $\mathrm{PG}(k-1,3)$. As established earlier, C is m-trifferent if and only if these points form a strong blocking set $\mathcal{B}$ with strength m.

By Theorem 12, there exists a strong blocking set $\mathcal{B} \subseteq \mathrm{PG}(k-1,3)$ with strength m satisfying $|\mathcal{B}| \leq 43k + 18m - 93$. Thus, given $n \geq 43k + 18m - 93$, one can select a collection of exactly a set of n columns for a generator matrix that yields an m-trifferent code of dimension k. Since a linear code of dimension k has 3^k codewords, it follows that $T_L(n,m) \geq 3^k$. Rewriting $n \geq 43k + 18m - 93$ as

$$k \leq \frac{n - 18m + 93}{43}$$

shows that the choice

$$k = \left\lfloor\frac{n - 18m + 93}{43}\right\rfloor$$

is valid, and we deduce that

$$T_L(n,m) \geq 3^{\lfloor\frac{n-18m+93}{43}\rfloor}.$$

For the regime $m = o(n)$, note that the above argument shows k grows linearly in n. Hence, we may apply the refined bound in Theorem 12 and the same argument gives

$$T_L(n,m) \geq 3^{\lfloor\frac{n-5m+23}{10}\rfloor}$$

for sufficiently large n. □

V. Concluding Remarks

In this section, we discuss some related results to the m-trifference problem and present computational results as well.

A. Computational Results and Explicit Constructions

Corollary 2 leads to an easy ILP formulation of the problem. Namely, we take a binary variable x_i for the i-th 1-dimensional subspace V_i of $\mathbb{F}_3^k$, and for every $(k-2)$-dimensional affine subspace S, we add the constraint $\sum_{i:V_i\leq S} x_i \geq m$. Using Gurobi [25], we have computed the following optimal values of these m-fold affine blocking sets.

(k,m)	(3, 2)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)	(4, 7)
n	12	16	24	28	30	35	38

Via Corollary 2, this table implies that $T_L(n,2) = 3^3$ for all $12 \leq n \leq 15$, and $T_L(16,2) = 3^4$.

For the non-linear case, direct computer computations give the following values: $T(4,2) = 4$, $T(5,2) = T(6,2) = 6$. Also, $T(4,3) = T(5,3) = 3$, $T(6,3) = T(7,3) = 4$.

Authorized licensed use limited to: TU Delft Library. Downloaded on May 06,2026 at 12:18:01 UTC from IEEE Xplore. Restrictions apply.

B. Explicit Constructions

Large linear trifferent codes were constructed via random methods [5], so it was a relevant goal to obtain good explicit constructions as well. Such codes were presented by Alon et al. [23] using expander graphs. A similar approach can be applied in the general trifference problem as well.

For a graph G, let $\iota(G) = \min_{S \subseteq V(G)}\{|S| + \kappa(G - S)\}$, where $\kappa(H)$ is the largest size of a connected component in H.

Let C be an infinite family of $[n, Rn, \delta n]_3$ codes, and let G be an infinite family of d-regular graphs on n vertices such that $\iota(G) \geq (1-\delta)n + m - 1$. Then by the construction in [23] and the Lemma above, we get a linear m-trifferent code of length $(1+d)n$ and dimension $k = Rn$.

Note that constant degree expanders can be used to explicitly construct d-regular graphs G on n vertices with $\iota(G) \geq \left(1 - \frac{2\sqrt{d}}{d+\sqrt{d}-o(1)}\right)n$, c.f. [23].

C. Relation to the Sunflower Conjecture

Consider a code $C \subseteq \{0,1,2\}^n$ such that any three distinct codewords x, y, z triffer in at least 1 coordinate. Now consider the multiset

$$F(C) = \{\sigma(x) : x \in C\}$$

where each x is considered once. The elements of the multiset are in $\mathcal{P}([n])$. If three distinct codewords x, y, z triffer at position i then i appears in precisely two of the sets $\sigma(x), \sigma(y), \sigma(z)$. Observe that in $F(C)$, each word must have multiplicity at most 2. This is because if we had $\sigma(x) = \sigma(y) = \sigma(z)$ for distinct x, y, z then for every i, i appears in 0 or 3 of the three sets, and so the three words cannot triffer at any position.

A set system $\mathcal{F}$ consisting of $k \geq 3$ sets is called a *k-sunflower* if any two sets in the system have the same intersection, and a set system is called k-sunflower-free if no k distinct elements of it form a k-sunflower. A 3-sunflower-free set-system is just called *sunflower-free*. Naslund and Sawin [26] showed that a sunflower-free family $\mathcal{F}$ of subsets of $[n]$ must satisfy $|\mathcal{F}| \leq \left(\frac{3}{2^{2/3}}\right)^{n(1+o(1))}$.

Consider the ground set of $F(C)$, denoting it by $\mathcal{F}$. Still in $\mathcal{F}$, it is true that there cannot be three sets X, Y, Z such that for every $i \in [n]$, i appears in either 0, 1 or 3 of the sets X, Y, Z. We can see that three sets satisfy this property precisely when they form a 3-sunflower, therefore $\mathcal{F}$ must be a sunflower-free family, yielding $|C| \leq 2|\mathcal{F}| \leq \left(\frac{3}{2^{2/3}}\right)^{n(1+o(1))}$.

Here $\frac{3}{2^{2/3}} \approx 1.8899$. However, we already know that $T(n,1) \leq Cn^{-2/5}\left(\frac{3}{2}\right)^n$ from [9], so this does not give an improvement.

Finally, let us mention that while this work focused on the generalization of the trifference problem, which was proven to be the most difficult one among q-hash codes, our methods are applicable for the q-hashing case as well.

Acknowledgment

The authors thank the referees for their valuable suggestions. Bartłomiej Kielak thanks the Alexander von Humboldt Foundation in the framework of the Alexander von Humboldt Professorship of Daniel Kráľ endowed by the Federal Ministry of Education and Research.

References

[1] J. Korner and K. Marton, "New bounds for perfect hashing via information theory," *Eur. J. Combinatorics*, vol. 9, no. 6, pp. 523–530, Nov. 1988.
[2] H. Wang and C. Xing, "Explicit constructions of perfect hash families from algebraic curves over finite fields," *J. Combinat. Theory, A*, vol. 93, no. 1, pp. 112–124, Jan. 2001.
[3] A. Barg, G. R. Blakley, G. Kabatiansky, and C. Tavernier, "Robust parent-identifying codes," in *Proc. IEEE Inf. Theory Workshop*, Aug. 2010, pp. 1–4.
[4] G. Zaverucha, "Hash families and cover-free families with cryptographic applications," Ph.D. dissertation, David R. Cheriton School Comput. Sci., Univ. Waterloo, Waterloo, ON, Canada, 2010.
[5] A. Bishnoi, J. D'haeseleer, D. Gijswijt, and A. Potukuchi, "Blocking sets, minimal codes and trifferent codes," *J. London Math. Soc.*, vol. 109, no. 6, p. 12938, Jun. 2024.
[6] G. N. Alfarano, M. Borello, A. Neri, and A. Ravagnani, "Three combinatorial perspectives on minimal codes," *SIAM J. Discrete Math.*, vol. 36, no. 1, pp. 461–489, Mar. 2022.
[7] T. Héger and Z. L. Nagy, "Short minimal codes and covering codes via strong blocking sets in projective spaces," *IEEE Trans. Inf. Theory*, vol. 68, no. 2, pp. 881–890, Feb. 2022.
[8] S. Kurz, "Trifferent codes with small lengths," *Examples Counterexamples*, vol. 5, Jun. 2024, Art. no. 100139.
[9] S. Bhandari and A. Khetan, "Improved upper bound for the size of a trifferent code," *Combinatorica*, vol. 45, no. 1, p. 2, Feb. 2025.
[10] C. Pohoata and D. Zakharov, "On the trifference problem for linear codes," *IEEE Trans. Inf. Theory*, vol. 68, no. 11, pp. 7096–7099, Nov. 2022.
[11] S. Della Fiore and M. Dalai, "Bounds on k-hash distances and rates of linear codes," *IEEE Trans. Inf. Theory*, vol. 72, no. 2, pp. 865–875, Feb. 2026.
[12] J. Körner, "Coding of an information source having ambiguous alphabet and the entropy of graphs," in *Proc. 6th Prague Conf. Inf. Theory*, 1973, pp. 411–425.
[13] M. L. Fredman and J. Komlós, "On the size of separating systems and families of perfect hash functions," *SIAM J. Algebr. Discrete Methods*, vol. 5, no. 1, pp. 61–68, Mar. 1984.
[14] E. Arikan, "An upper bound on the zero-error list-coding capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1237–1240, Jul. 1994.
[15] M. Dalai, V. Guruswami, and J. Radhakrishnan, "An improved bound on the zero-error list-decoding capacity of the 4/3 channel," *IEEE Trans. Inf. Theory*, vol. 66, no. 2, pp. 749–756, Feb. 2020.
[16] V. Guruswami and A. Riazanov, "Beating Fredman-Komlós for perfect k-hashing," *J. Combinat. Theory, A*, vol. 188, May 2022, Art. no. 105580.
[17] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, vol. 16. Amsterdam, The Netherlands: Elsevier, 1977.
[18] L. A. Bassalygo, M. Burmester, A. Dyachkov, and G. Kabatianski, "Hash codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 1997, p. 174.
[19] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 1991.
[20] M. Plotkin, "Binary codes with specified minimum distance," *IEEE Trans. Inf. Theory*, vol. IT-6, no. 4, pp. 445–450, Sep. 1960.
[21] L. A. Bassalygo, "New upper bounds for error correcting codes," *Problemy Peredachi Informatsii*, vol. 1, no. 4, pp. 32–35, 1965. [Online]. Available: https://www.mathnet.ru/eng/ppi762
[22] A. A. Davydov, M. Giulietti, S. Marcugini, and F. Pambianco, "Linear nonbinary covering codes and saturating sets in projective spaces," *Adv. Math. Commun.*, vol. 5, no. 1, pp. 119–147, 2011.
[23] N. Alon, A. Bishnoi, S. Das, and A. Neri, "Strong blocking sets and minimal codes from expander graphs," *Trans. Amer. Math. Soc.*, vol. 377, no. 8, pp. 5389–5410, 2024.
[24] C. Tang, Y. Qiu, Q. Liao, and Z. Zhou, "Full characterization of minimal linear codes as cutting blocking sets," *IEEE Trans. Inf. Theory*, vol. 67, no. 6, pp. 3690–3700, Jun. 2021.
[25] Gurobi Optimization, LLC. (2024). *Gurobi Optimizer Reference Manual*. [Online]. Available: https://www.gurobi.com
[26] E. Naslund and W. Sawin, "Upper bounds for sunflower-free sets," *Forum Math., Sigma*, vol. 5, p. e15, Jan. 2017.

Authorized licensed use limited to: TU Delft Library. Downloaded on May 06,2026 at 12:18:01 UTC from IEEE Xplore. Restrictions apply.