# Upper bounds on the rate of linear $q$-ary $k$-hash codes

S. Della Fiore and M. Dalai

University of Salerno, University of Brescia

ISIT 2024, July

# Background

A $q$-ary code $C$ of length $n$ is a subset of $\{0, 1, \ldots, q-1\}^n$. Denote with $R = \frac{1}{n}\log_q |C|$ its rate.

**Definition ($(q, k)$-hash code)**

*A $q$-ary code $C$ is a $(q, k)$-hash code if for any $k$ distinct elements of $C$ we can find a coordinate in which they all differ.*

**Example ($(3, 3)$-hash code or trifferent code)**

$$\begin{matrix} \cdots & 0 & 0 & 0 & 0 & \cdots \\ \cdots & 0 & 1 & 2 & 1 & \cdots \\ \cdots & 0 & 2 & 1 & 2 & \cdots \\ \cdots & 1 & 0 & 2 & 2 & \cdots \end{matrix}$$

# Background

A $q$-ary code $C$ of length $n$ is a subset of $\{0, 1, \ldots, q-1\}^n$. Denote with $R = \frac{1}{n} \log_q |C|$ its rate.

**Definition ($(q, k)$-hash code)**

*A $q$-ary code $C$ is a $(q, k)$-hash code if for any $k$ distinct elements of $C$ we can find a coordinate in which they all differ.*

**Example ($(3, 3)$-hash code or trifferent code)**

$$\begin{array}{cccccc} \cdots & 0 & 0 & 0 & 0 & \cdots \\ \cdots & 0 & 1 & 2 & 1 & \cdots \\ \cdots & 0 & 2 & 1 & 2 & \cdots \\ \cdots & 1 & 0 & 2 & 2 & \cdots \end{array}$$

# Background

A $q$-ary code $C$ of length $n$ is a subset of $\{0, 1, \ldots, q-1\}^n$. Denote with $R = \frac{1}{n} \log_q |C|$ its rate.

**Definition ($(q,k)$-hash code)**

*A $q$-ary code $C$ is a $(q,k)$-hash code if for any $k$ distinct elements of $C$ we can find a coordinate in which they all differ.*

**Example ($(3,3)$-hash code or trifferent code)**

$$\begin{array}{cccccc} \cdots & 0 & \color{red}{0} & 0 & 0 & \cdots \\ \cdots & 0 & \color{red}{1} & 2 & 1 & \cdots \\ \cdots & 0 & \color{red}{2} & 1 & 2 & \cdots \\ \cdots & 1 & 0 & 2 & 2 & \cdots \end{array}$$

# Background

A $q$-ary code $C$ of length $n$ is a subset of $\{0,1,\dots,q-1\}^n$. Denote with $R = \frac{1}{n}\log_q |C|$ its rate.

**Definition ($(q,k)$-hash code)**

*A $q$-ary code $C$ is a $(q,k)$-hash code if for any $k$ distinct elements of $C$ we can find a coordinate in which they all differ.*

**Example ($(3,3)$-hash code or trifferent code)**

$$\begin{matrix} \cdots & 0 & 0 & 0 & 0 & \cdots \\ \cdots & 0 & 1 & 2 & 1 & \cdots \\ \cdots & 0 & 2 & 1 & 2 & \cdots \\ \cdots & 1 & 0 & 2 & 2 & \cdots \end{matrix}$$

# Background

A $q$-ary code $C$ of length $n$ is a subset of $\{0, 1, \dots, q-1\}^n$. Denote with $R = \frac{1}{n}\log_q |C|$ its rate.

**Definition ($(q, k)$-hash code)**

*A $q$-ary code $C$ is a $(q, k)$-hash code if for any $k$ distinct elements of $C$ we can find a coordinate in which they all differ.*

**Example ($(3, 3)$-hash code or trifferent code)**

$$
\begin{array}{cccccc}
\cdots & 0 & 0 & \color{red}{0} & 0 & \cdots \\
\cdots & 0 & 1 & 2 & 1 & \cdots \\
\cdots & 0 & 2 & \color{red}{1} & 2 & \cdots \\
\cdots & 1 & 0 & \color{red}{2} & 2 & \cdots
\end{array}
$$

# Background

A $q$-ary code $C$ of length $n$ is a subset of $\{0, 1, \ldots, q-1\}^n$. Denote with $R = \frac{1}{n} \log_q |C|$ its rate.

**Definition ($(q, k)$-hash code)**

*A $q$-ary code $C$ is a $(q, k)$-hash code if for any $k$ distinct elements of $C$ we can find a coordinate in which they all differ.*

**Example ($(3, 3)$-hash code or trifferent code)**

$$\begin{matrix} \cdots & 0 & 0 & 0 & 0 & \cdots \\ \cdots & 0 & \color{red}{1} & 2 & 1 & \cdots \\ \cdots & 0 & \color{red}{2} & 1 & 2 & \cdots \\ \cdots & 1 & \color{red}{0} & 2 & 2 & \cdots \end{matrix}$$

# Some bounds from the literature

Let $R$ be the largest asymptotic rate of $(q,k)$-hash codes.

**Theorem (Körner and Marton 1988)**

$$R \leq \min_{0 \leq j \leq k-2} \frac{q^{\underline{j+1}}}{q^{j+1}} \log_q \left( \frac{q-j}{k-j-1} \right) + o(1)\,.$$

For $q$ sufficiently larger than $k$ one can obtain a better bound.

**Theorem (Mehlhorn, Blackburn and Wild (1984, 1998))**

$$R \leq \frac{1}{k-1} + o(1)\,.$$

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

[2] S. R. Blackburn and P. R. Wild, "Optimal linear perfect hash families," Journal of Combinatorial Theory, Series A, vol. 83, no. 2, pp. 233–250, 1998.

# Some bounds from the literature

Let $R$ be the largest asymptotic rate of $(q,k)$-hash codes.

**Theorem (Körner and Marton 1988)**

$$R \leq \min_{0 \leq j \leq k-2} \frac{q^{\underline{j+1}}}{q^{j+1}} \log_q \left( \frac{q-j}{k-j-1} \right) + o(1)\,.$$

For $q$ sufficiently larger than $k$ one can obtain a better bound.

**Theorem (Mehlhorn, Blackburn and Wild (1984, 1998))**

$$R \leq \frac{1}{k-1} + o(1)\,.$$

---

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

[2] S. R. Blackburn and P. R. Wild, "Optimal linear perfect hash families," Journal of Combinatorial Theory, Series A, vol. 83, no. 2, pp. 233–250, 1998.

# The case $q = k = 3$

Let $C$ be a $(3,3)$-hash code (trifferent code) of maximum size

**Theorem (Körner and Marton 1988)**

$$\left(\frac{9}{5}\right)^{n/4+o(1)} \leq |C| \leq 2\left(\frac{3}{2}\right)^n$$

Improvements on the upper bound multiplicative constant

**Theorem (Kurz 2024)**

$$|C| \leq 0.6937 \cdot \left(\frac{3}{2}\right)^n \textit{ for } n \geq 10$$

Improvements on the upper bound polynomial factor

**Theorem (Bhandhari and Keta 2024)**

$$|C| \leq c \cdot n^{-2/5} \cdot \left(\frac{3}{2}\right)^n \textit{ for some constant } c$$

# The case $q = k = 3$

Let $C$ be a $(3,3)$-hash code (trifferent code) of maximum size

**Theorem (Körner and Marton 1988)**

$$\left(\frac{9}{5}\right)^{n/4+o(1)} \leq |C| \leq 2\left(\frac{3}{2}\right)^n$$

Improvements on the upper bound multiplicative constant

**Theorem (Kurz 2024)**

$$|C| \leq 0.6937 \cdot \left(\frac{3}{2}\right)^n \textit{ for } n \geq 10$$

Improvements on the upper bound polynomial factor

**Theorem (Bhandhari and Keta 2024)**

$$|C| \leq c \cdot n^{-2/5} \cdot \left(\frac{3}{2}\right)^n \textit{ for some constant } c$$

# The case $q = k = 3$

Let $C$ be a $(3,3)$-hash code (trifferent code) of maximum size

**Theorem (Körner and Marton 1988)**

$$\left(\frac{9}{5}\right)^{n/4+o(1)} \leq |C| \leq 2\left(\frac{3}{2}\right)^{n}$$

Improvements on the upper bound multiplicative constant

**Theorem (Kurz 2024)**

$$|C| \leq 0.6937 \cdot \left(\frac{3}{2}\right)^{n} \quad \textit{for } n \geq 10$$

Improvements on the upper bound polynomial factor

**Theorem (Bhandhari and Keta 2024)**

$$|C| \leq c \cdot n^{-2/5} \cdot \left(\frac{3}{2}\right)^{n} \quad \textit{for some constant } c$$

# Linear trifferent codes

Let $C$ be a linear trifferent code in $\mathbb{F}_3^n$. Using connections between minimal codes and blocking sets we have the following results.

**Theorem (Pohata and Zakharov 2022)**

*For some* $\epsilon > 0$

$$|C| \leq 3^{(1/4-\epsilon)n} \approx 1.3161^n$$

**Theorem (Bishnoi et al. 2024)**

$$(9/5)^{n/4+o(1)} \leq |C| \leq 3^{n/4.5516+o(n)} \approx 1.2731^n$$

[3] C. Pohoata and D. Zakharov, "On the trifference problem for linear codes," IEEE Transactions on Information Theory, vol. 68, no. 11, pp. 7096–7099, 2022.

[4] A. Bishnoi, J. D'haeseleer, D. Gijswijt and A. Potukuchi, "Blocking sets, minimal codes and trifferent codes," Journal of the London Mathematical Society, 109(6), 2024.

# Linear trifferent codes

Let $C$ be a linear trifferent code in $\mathbb{F}_3^n$. Using connections between minimal codes and blocking sets we have the following results.

**Theorem (Pohata and Zakharov 2022)**

*For some* $\epsilon > 0$

$$|C| \leq 3^{(1/4-\epsilon)n} \approx 1.3161^n$$

**Theorem (Bishnoi et al. 2024)**

$$(9/5)^{n/4+o(1)} \leq |C| \leq 3^{n/4.5516+o(n)} \approx 1.2731^n$$

[3] C. Pohoata and D. Zakharov, "On the trifference problem for linear codes," IEEE Transactions on Information Theory, vol. 68, no. 11, pp. 7096–7099, 2022.

[4] A. Bishnoi, J. D'haeseleer, D. Gijswijt and A. Potukuchi, "Blocking sets, minimal codes and trifferent codes," Journal of the London Mathematical Society, 109(6), 2024.

# Our contribution

- Rederivation of the best upper bounds for $q = k = 3$
- Generalization to the case $q \geq k \geq 3$

**Remark**

*When $q$ is small compared to $k > 3$ no linear $k$-hash codes of dimension $2$ exist. Blackburn and Wild proved that this holds for $q \leq 2k - 4$.*

# Our contribution

- Rederivation of the best upper bounds for $q = k = 3$
- Generalization to the case $q \geq k \geq 3$

**Remark**

*When $q$ is small compared to $k > 3$ no linear $k$-hash codes of dimension $2$ exist. Blackburn and Wild proved that this holds for $q \leq 2k - 4$.*

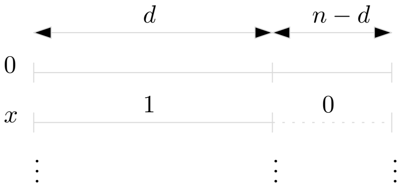
# Our contribution for $q = k = 3$

The main tool that we used is the Jamison's bound.

**Theorem (Jamison 1977)**

*Let $q \geq 3$ be a prime power, and let $\mathcal{H}$ be a set of hyperplanes in $\mathbb{F}_q^m$ whose union is $\mathbb{F}_q^m \setminus \{0\}$. Then $|\mathcal{H}| \geq (q-1)m$.*

- Let $C$ be a linear trifferent code in $\mathbb{F}_3^n$
- Let $G$ be its generator $m \times n$ matrix
- Let $d$ be the minimum hamming distance of $C$




where $x = uG$ is a codeword of weight $d$. WLOG we can assume it has 1 in the first $d$ coordinates and 0 in the others (reordering and rescaling).

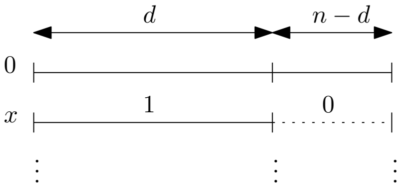
# Our contribution for $q = k = 3$

The main tool that we used is the Jamison's bound.

**Theorem (Jamison 1977)**

*Let $q \geq 3$ be a prime power, and let $\mathcal{H}$ be a set of hyperplanes in $\mathbb{F}_q^m$ whose union is $\mathbb{F}_q^m \setminus \{0\}$. Then $|\mathcal{H}| \geq (q-1)m$.*

- Let $C$ be a linear trifferent code in $\mathbb{F}_3^n$
- Let $G$ be its generator $m \times n$ matrix
- Let $d$ be the minimum hamming distance of $C$




where $x = uG$ is a codeword of weight $d$. WLOG we can assume it has $1$ in the first $d$ coordinates and $0$ in the others (reordering and rescaling).

# Our contribution for $q = k = 3$

Since $C$ is trifferent, any codeword different from $0$ and $x$ must have at least a $2$ in the first $d$ coordinates.

Let $g_i$ be the $i$-th column of $G$. Then the following $d$ affine subspaces (hyperplanes) in $\mathbb{F}_q^m$

$$H_i = \{v \in \mathbb{F}_q^m \mid v \cdot g_i = 2\} \text{ for } i = 1, \ldots, d$$

cover all the points in $\mathbb{F}_q^m$ except for $0$ and $u$.

Adding another hyperplane $H_{d+1} = \{v \in \mathbb{F}_q^m \mid v \cdot g_1 = 1\}$ we cover also $u$.

# Our contribution for $q = k = 3$

Since $C$ is trifferent, any codeword different from $0$ and $x$ must have at least a $2$ in the first $d$ coordinates.

Let $g_i$ be the $i$-th column of $G$. Then the following $d$ affine subspaces (hyperplanes) in $\mathbb{F}_q^m$

$$H_i = \{v \in \mathbb{F}_q^m \mid v \cdot g_i = 2\} \text{ for } i = 1, \dots, d$$

cover all the points in $\mathbb{F}_q^m$ except for $0$ and $u$.

Adding another hyperplane $H_{d+1} = \{v \in \mathbb{F}_q^m \mid v \cdot g_1 = 1\}$ we cover also $u$.

# Our contribution for $q = k = 3$

Since $C$ is trifferent, any codeword different from $0$ and $x$ must have at least a $2$ in the first $d$ coordinates.

Let $g_i$ be the $i$-th column of $G$. Then the following $d$ affine subspaces (hyperplanes) in $\mathbb{F}_q^m$

$$H_i = \{v \in \mathbb{F}_q^m \mid v \cdot g_i = 2\} \text{ for } i = 1, \ldots, d$$

cover all the points in $\mathbb{F}_q^m$ except for $0$ and $u$.

Adding another hyperplane $H_{d+1} = \{v \in \mathbb{F}_q^m \mid v \cdot g_1 = 1\}$ we cover also $u$.

# Our contribution for $q = k = 3$

Hence by Jamison's bound we have $d + 1 \geq 2m$, that in terms of rates and relative minimum distance $\delta = d/n$

$$R \leq \delta/2 + o(1)$$

Using the Plotkin bound $\delta \leq 2/3(1 - R) + o(1)$, one obtains:

$$R \leq 1/4 + o(1)$$

Since the Plotkin bound is not tight at positive rates, one could use the linear programming bound for 3-ary codes to get

$$R \leq 1/4.5516 + o(1)$$

**Note**: this procedure can be seen as an application of a method presented by Calderbank et al. in 1993.

[5] A. Calderbank et al., "The sperner capacity of linear and nonlinear codes for the cyclic triangle," Journal of Algebraic Combinatorics: An International Journal, 1993

# Our contribution for $q = k = 3$

Hence by Jamison's bound we have $d + 1 \geq 2m$, that in terms of rates and relative minimum distance $\delta = d/n$

$$R \leq \delta/2 + o(1)$$

Using the Plotkin bound $\delta \leq 2/3(1 - R) + o(1)$, one obtains:

$$R \leq 1/4 + o(1)$$

Since the Plotkin bound is not tight at positive rates, one could use the linear programming bound for 3-ary codes to get

$$R \leq 1/4.5516 + o(1)$$

**Note**: this procedure can be seen as an application of a method presented by Calderbank et al. in 1993.

[5] A. Calderbank et al., "The sperner capacity of linear and nonlinear codes for the cyclic triangle," Journal of Algebraic Combinatorics: An International Journal, 1993

# Our contribution for $q = k = 3$

Hence by Jamison's bound we have $d + 1 \geq 2m$, that in terms of rates and relative minimum distance $\delta = d/n$

$$R \leq \delta/2 + o(1)$$

Using the Plotkin bound $\delta \leq 2/3(1 - R) + o(1)$, one obtains:

$$R \leq 1/4 + o(1)$$

Since the Plotkin bound is not tight at positive rates, one could use the linear programming bound for 3-ary codes to get

$$R \leq 1/4.5516 + o(1)$$

**Note**: this procedure can be seen as an application of a method presented by Calderbank et al. in 1993.

[5] A. Calderbank et al., "The sperner capacity of linear and nonlinear codes for the cyclic triangle," Journal of Algebraic Combinatorics: An International Journal, 1993

# General case $q \geq k \geq 3$

The idea is to iterate the procedure for $q = k = 3$.

**Theorem (Bruen 1997)**

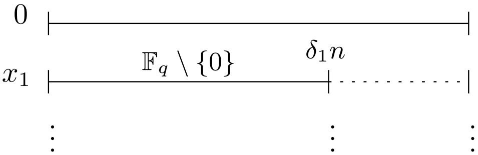
*Let $\mathcal{H}$ be a multiset of hyperplanes in $\mathbb{F}_q^m$. If no hyperplane in $\mathcal{H}$ contains $0$ and each point in $\mathbb{F}_q^m \setminus \{0\}$ is covered by at least $t$ hyperplanes in $\mathcal{H}$, then*

$$|\mathcal{H}| \geq (m + t - 1)(q - 1)\,.$$

**Note**: for $t = 1$ we get the classical Jamison's bound

# General case $q \geq k \geq 3$ – Sketch of the proof

- Let $C$ be a linear $k$-hash code in $\mathbb{F}_q^n$
- Let $x_1$ be a codeword of minimum weight $d = \delta_1 n$

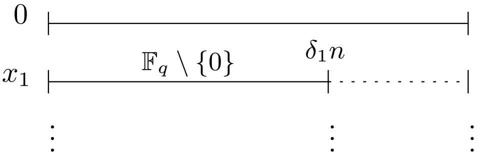



Any set of $k$ codewords that contains $0$ and $x_1$ cannot be $k$-hashed in the last $n - d$ coordinates.

We consider the punctured code $C_{[d]}$, that is clearly linear and by the $k$-hash property we have $|C_{[d]}| = |C|$ (injectivity).

# General case $q \geq k \geq 3$ – Sketch of the proof

- Let $C$ be a linear $k$-hash code in $\mathbb{F}_q^n$
- Let $x_1$ be a codeword of minimum weight $d = \delta_1 n$

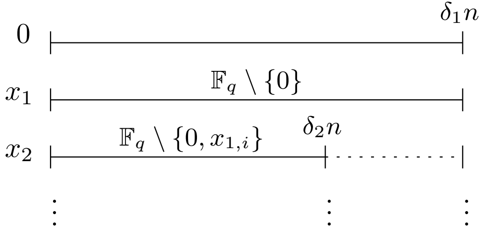



Any set of $k$ codewords that contains $0$ and $x_1$ cannot be $k$-hashed in the last $n - d$ coordinates.

We consider the punctured code $C_{[d]}$, that is clearly linear and by the $k$-hash property we have $|C_{[d]}| = |C|$ (injectivity).

# General case $q \geq k \geq 3$ – Sketch of the proof

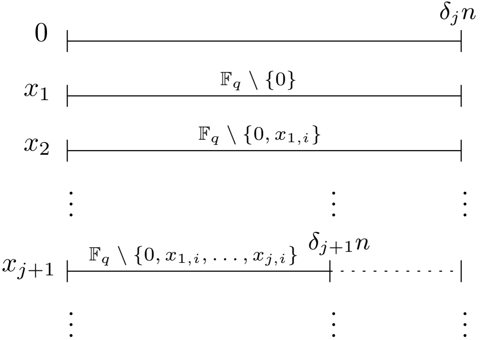
- Choose $\delta_2 \in [0,1]$ such that $R > \frac{q-2}{q-1}\delta_1 - \delta_2 + o(1)$
- By Bruen theorem, there exists $x_2$ that is linearly independent of $x_1$ such that $\{0, x_1, x_2\}$ is 3-hashed in at most $\delta_2 n$ coordinates



# General case $q \geq k \geq 3$ – Sketch of the proof

Then we can iterate the procedure. At iteration $j$ we have $j$ linearly independent codewords $x_1, \ldots, x_j$ in $C_{[\delta_j n]}$

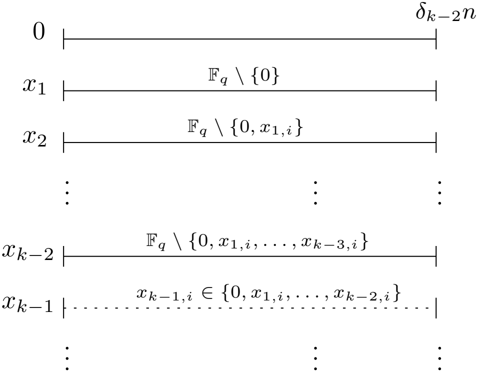
- Choose $\delta_{j+1} \in [0,1]$ such that $R > \frac{q-j-1}{q-1}\delta_j - \delta_{j+1} + o(1)$
- By Bruen theorem, there exists $x_{j+1}$ that is lin. ind. of $x_1, \ldots, x_j$ s. t. $\{0, x_1, \ldots, x_{j+1}\}$ is $(j+1)$-hashed in at most $\delta_{j+1}n$ coordinates



# General case $q \geq k \geq 3$ – Sketch of the proof

After $k-3$ iterations

- if $R > \frac{q-k+1}{q-1}\delta_{k-2} + o(1)$ we find one last lin. ind. codeword $x_{k-1}$




Hence the set $\{0, x_1, \ldots, x_{k-1}\}$ is not $k$-hashed.

# General case $q \geq k \geq 3$ – Sketch of the proof

Then we have a recursive sequence of conditions on $R$ that with initialization $\delta_1 = \delta$ take us to formulate the following theorem.

**Theorem**

*Let $C$ be a linear $k$-hash code in $\mathbb{F}_q^n$ of rate $R$ and relative distance $\delta$. Then*

$$R \leq \frac{\delta}{\sum_{i=1}^{k-2} \frac{(q-1)^i}{(q-2)^{\underline{i}}}} + o(1)$$

*where* $(q-2)^{\underline{i}} = (q-2)(q-3)\cdots(q-i-1)$.

**Note**: it can be seen that using the singleton bound $\delta \leq 1 - R$ we already improve the bound $1/(k-1)$ due to Mehlhorn for every $q \geq k$.

# Some numerical results

As done for the case $q = k = 3$, we can obtain bounds that depend only on $q$ and $k$ by using well-known bounds on the relative distance $\delta$ of a code.

| $q$ | Plotkin | LP | Körner and Marton |
|---|---|---|---|
| 3 | $1/4 = 0.25$ | $0.2198$ | $0.3691$ |
| 4 | $1/3 = 0.\overline{3}$ | $0.3000$ | $1/2 = 0.5$ |
| 5 | $3/8 = 0.375$ | $0.3441$ | $0.5694$ |
| 7 | $5/12 = 0.41\overline{6}$ | $0.3928$ | $0.6438$ |
| 8 | $3/7 = 0.\overline{428571}$ | $0.4080$ | $2/3 = 0.\overline{6}$ |
| 9 | $7/16 = 0.4375$ | $0.4200$ | $0.6846$ |
| 11 | $9/20 = 0.45$ | $0.4373$ | $0.7110$ |
| 13 | $11/24 = 0.458\overline{3}$ | $0.4497$ | $0.7298$ |
| 16 | $7/15 = 0.4\overline{6}$ | $0.4628$ | $3/4 = 0.75$ |
| $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ |
| 64 | $31/63 = 0.\overline{492063}$ | $0.5119$ | $5/6 = 0.8\overline{3}$ |

Table: Upper bounds on the rate of linear 3-hash codes in $\mathbb{F}_q^n$ for a prime power $q \in [3, 64]$. All numbers are rounded upwards.