Improved Lower Bounds for 3-Query Matching Vector Codes

Divesh Aggarwal
Centre for Quantum Technologies, National University of Singapore, Singapore

Pranjal Dutta
National University of Singapore, Singapore

Zeyong Li
Centre for Quantum Technologies, National University of Singapore, Singapore

Maciej Obremski
National University of Singapore, Singapore

Sidhant Saraogi
Georgetown University, Washington, DC, USA

Abstract

A Matching Vector (**MV**) family modulo a positive integer $m \geq 2$ is a pair of ordered lists $\mathcal{U} = (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_K)$ and $\mathcal{V} = (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_K)$ where $\boldsymbol{u}_i, \boldsymbol{v}_j \in \mathbb{Z}_m^n$ with the following property: for any $i \in [K]$, the inner product $\langle \boldsymbol{u}_i, \boldsymbol{v}_i \rangle = 0 \pmod{m}$, and for any $i \neq j$, $\langle \boldsymbol{u}_i, \boldsymbol{v}_j \rangle \neq 0 \pmod{m}$. An **MV** family is called *r-restricted* if inner products $\langle u_i, v_j \rangle$, for all i, j, take *at most* r different values. The r-restricted **MV** families are extremely important since the only known construction of constant-query subexponential locally decodable codes (LDCs) are based on them. Such LDCs constructed via matching vector families are called *matching vector codes*. Let $\mathbf{MV}(m, n)$ (respectively $\mathbf{MV}(m, n, r)$) denote the *largest* K such that there exists an **MV** family (respectively r-restricted **MV** family) of size K in $\mathbb{Z}_m^n$. Such a **MV** family can be transformed in a black-box manner to a *good* r-query locally decodable code taking messages of length K to codewords of length $N = m^n$.

For small prime m, an *almost* tight bound $\mathbf{MV}(m, n) \leq O(m^{n/2})$ was first shown by Dvir, Gopalan, Yekhanin (FOCS'10, SICOMP'11), while for general m, the same paper established an upper bound of $O(m^{n-1+o_m(1)})$, with $o_m(1)$ denoting a function that goes to zero when m grows. For any arbitrary constant $r \geq 3$ and composite m, the best upper bound till date on $\mathbf{MV}(m, n, r)$ is $O(m^{n/2})$, is due to Bhowmick, Dvir and Lovett (STOC'13, SICOMP'14).In a breakthrough work, Alrabiah, Guruswami, Kothari and Manohar (STOC'23) implicitly improve this bound for 3-restricted families to $\mathbf{MV}(m, n, 3) \leq O(m^{n/3})$.

In this work, we present an upper bound for $r = 3$ where $\mathbf{MV}(m, n, 3) \leq m^{n/6+O(\log n)}$, and as a result, any 3-query matching vector code must have codeword length of $N \geq K^{6-o(1)}$.

2012 ACM Subject Classification Theory of computation → Error-correcting codes

Keywords and phrases Locally Decodable Codes, Matching Vector Families

Digital Object Identifier 10.4230/LIPIcs.ITCS.2025.2

Related Version *Previous Version*: `https://eccc.weizmann.ac.il/report/2024/061/`

Funding *Divesh Aggarwal*: Funded by the bridging grant at Centre for Quantum Technologies titled "Quantum algorithms, complexity, and communication".
Pranjal Dutta: Funded by NRF Singapore under the project "Computational Hardness of Lattice Problems and Implications" [NRF-NRFI09-0005].
Zeyong Li: Funded by NRF Singapore under the project "Computational Hardness of Lattice Problems and Implications" [NRF-NRFI09-0005].
Maciej Obremski: Funded by MOE tier 2 grant T2EP20223-0024, Breaking the Box – On Security of Cryptographic Devices.
Sidhant Saraogi: Funded in part by the National Science Foundation CAREER award to Justin Thaler (grant CCF-1845125).

16th Innovations in Theoretical Computer Science Conference (ITCS 2025).
Editor: Raghu Meka; Article No. 2; pp. 2:1–2:19
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Acknowledgements The authors would like to thank Sivakanth Gopi and Anup Rao for sharing a proof of Theorem 21. They would also like to thanks anonymous reviewers for their helpful comments.

1 Introduction

1.1 Locally Decodable Codes

A code $C : \Sigma^K \to \Sigma^N$ is said to be (r, δ, ε)-locally decodable if for every $i \in [K]$, the i-th coordinate of the message $\boldsymbol{x}_i \in \Sigma$ can be recovered with probability at least $1 - \varepsilon$ by a randomized decoding procedure that makes only r queries to the codeword, i.e., reads the codeword at r positions even if the codeword is corrupted on up to δN locations.

Locally decodable codes were formally introduced in [19] but had already been studied informally in the context of the PCP theorem [3, 4]. Locally decodable codes have found applications in many areas of computer science such as worst-case to average-case reductions, private information retrieval, secure multi-party computation, derandomization, matrix rigidity, data structures, and fault-tolerant computation. See [25] for a detailed survey.

A central research question is to understand the largest possible message length K (as a function of N) that can be achieved by a r-query locally decodable code. For the simplest non-trivial setting of $r = 2$, we have a nearly complete understanding – the Hadamard code achieves $K = \log_2 N$, and it was shown in [20, 12] that this is the best possible up to a constant factor, i.e., $K = O(\log_2 N)$.

Much less is understood about the dependence between K and N for the number of queries $r \geq 3$. In particular, the only known constructions for r-query locally decodable codes for a constant $r \geq 3$ are based on families of matching vector codes [24, 10, 9]. These constructions achieve $K = (\log N)^{\Omega(\log \log N)}$. On the other hand, the known upper bounds on K are very far from what is achieved by these constructions. In a series of works [19, 20, 22, 23, 6], it has been shown that $K = N^{1-2/(r+1)} \operatorname{polylog} N$, when r is odd, and $K = N^{1-2/r} \operatorname{polylog} N$, when r is even. For the special case where $r = 3$, it was shown that $K = O(N^{1/3} \log^2 N)$ in a celebrated recent result [2].

1.2 Matching Vector Families and Application to Locally Decodable Codes

A matching vector (**MV**) family is a combinatorial object that has been used in different contexts such as Ramsey graphs and weak representation of OR polynomials[11, 17, 1, 13], but the most prominent application of matching vector families is in the construction of constant-query locally decodable codes [9, 25]. A matching vector family is defined by two ordered lists $\mathcal{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_K)$ and $\mathcal{V} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K)$ where for all $i \in [K]$, $\boldsymbol{u}_i, \boldsymbol{v}_i$ are in $\mathbb{Z}_m^n$ for some positive integers $m, n > 1$. An S-matching vector family can be defined as follows.

▶ **Definition 1** (Matching Vector Family). *Let $S \subseteq \mathbb{Z}_m \setminus \{0\}$. An S-matching vector family is a set of vectors $\mathcal{U} = \{u_1, \ldots, u_K\}$, $\mathcal{V} = \{v_1, \ldots, v_K\}$, $\mathcal{U}, \mathcal{V} \subseteq \mathbb{Z}_m^n$ such that:*

- $\forall i \in [K], \langle u_i, v_i \rangle = 0$.
- $\forall i \neq j,\ i, j \in [K], \langle u_i, v_j \rangle \in S$.

An S-matching vector family is often called $(|S| + 1)$-restricted matching vector family, and such a family gives a locally decodable code for messages of block length K, codeword length $N = m^n$, and number of queries $r = |S| + 1$. See Theorem 15 ([25]) for the construction of a locally decodable code from a matching vector family.

Finally, we denote $\mathbf{MV}(m, n)$ (respectively $\mathbf{MV}(m, n, r)$), as the *largest* K such that there exists an **MV** family (respectively r-restricted **MV** family) of size K in $\mathbb{Z}_m^n$.

1.3 Upper Bounds for Matching Vector Families

Since the only known approach that has led to constructions of constant-query locally decodable codes is via constructing a large matching vector family, it is natural to ask what is the largest K as a function of m, n for which there is a r-restricted matching vector family of size K. This in particular will give an upper bound for the rate of a r-query locally decodable code via matching vector codes, i.e., any r-query locally decodable code achieving a better rate must be via some novel approach that does not require a matching vector family.

This problem was first studied in [9], where the authors showed that when m is a large prime, $\mathbf{MV}(m,n) \leq O(m^{n/2})$. On the other hand, the same paper established a general upper bound of $O(m^{n-1+o_m(1)})$, with $o_m(1)$ denoting a function that goes to zero when a composite m grows. In [8], the authors improved this bound to $\text{poly}(m) \cdot m^{.625n}$, for a very special case when $m = pq$, for distinct primes p and q such that $p \approx q$, and $n \ll m$. On the other hand, when $m \ll n$ is a prime, it can be shown via linear algebraic methods that $\mathbf{MV}(m,n) \leq O(n^{m-1})$ [5].

Bhowmick, Dvir and Lovett [7] showed how to extend the results of [9] to the case of m being a composite integer. They showed that $\mathbf{MV}(m,n,r) \leq r^{O(r \log r)} m^{n/2}$, i.e., for constant query codes, they showed that $K = O(\sqrt{N})$.

Additionally, in the same paper, Bhowmick, Dvir and Lovett also showed that under the (recently proved) polynomial Freiman Ruzsa (PFR) theorem [16], for m constant, $\mathbf{MV}(m,n) \leq m^{O(n/\log n)}$.

1.4 Our Result

The main contribution of the paper is to give a better upper bound for 3-restricted matching vector families.

▶ **Theorem 2.** *Let $n, m \geq 2$ be positive integers with $n > 10m^3$. Suppose m has atleast two distinct prime factors, then*

$$\mathbf{MV}(m,n,3) \ \leq \ m^{\frac{n+2\log(n+1)}{2q} + 2m^2},$$

where q is the second-smallest *prime dividing m.*

We already know a polynomial bound for prime powers due to [15] which we state below and prove for completeness in Section 2.4.

▶ **Theorem 3.** *Suppose $n, m \geq 2$ be positive integers where m is a constant prime power. Then $\mathbf{MV}(m,n,3) \leq (2emn)^{m-1}$.*

This allows us to conclude an upper bound for arbitrary m:

▶ Remark 4. When m has at least one factor which is not 2, that immediately gives a $m^{n/6+O(\log n)}$ upper bound, while if the constant m is a power of 2, the bound is just $\text{poly}(n)$, as mentioned above.

Finally, we can also derive a lower bound for the 3-query LDCs derived from matching vector families:

▶ **Corollary 5.** *Suppose $C : \Sigma^K \to \Sigma^N$ is a 3-query locally decodable code constructed from a 3-restricted matching vector family in a black-box manner (as in Theorem 15), then $N \geq K^{6-o_K(1)}$ where $o_K(1)$ vanishes as K increases.*

1.5 Concurrent Work

Concurrent to our work, the Polynomial Freiman Ruzsa conjecture was proven by Gowers, Green, Manners and Tao [16] for all abelian groups with bounded torsion! As a corollary (via Theorem 6.4 of [7]), one could conclude that

$$\mathbf{MV}(m,n) \le 2^{1200c(m)m^{6\log m}n/\log n},$$

where $c(m)$ is some explicit constant depending only on m. Infact, if we plug in the bound of [16] to Lemma 6.3 from [7], we can conclude that $c(m) \ge m^3$. Note that this bound is non-trivial (i.e., less than m^n) only when $n \ge 2^{\frac{1200\cdot 6^{3+6\log 6}}{\log 6}} > 2^{2^{56}}$ (plugging in $m = 6$) and beats our bound only when $n \ge 2^{2^{58}}$.

While this result supersedes ours in many parameter regimes, the techniques and approaches are very different. And our approach might have the potential of breaking the $2^{n/\log n}$ barrier. (See discussions below.)

1.6 Our Techniques

In the discussion below, we want to prove upper bounds for a matching vector family $\mathcal{U}, \mathcal{V}$ of size K over $\mathbb{Z}_m^n$ for some positive integers m and n.

Techniques from [7]. We begin by describing Bhowmick, Dvir, and Lovett's ([7]) approach towards getting an upper bound of $m^{n/2}$. They observe that if we consider the random variables $\boldsymbol{U}, \boldsymbol{V}$ as sampled independently and randomly from $\mathcal{U} = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_K\}$ and $\mathcal{V} = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_K\}$, respectively, then we have that $\Pr[\langle \boldsymbol{U}, \boldsymbol{V}\rangle = 0] = \frac{1}{K}$, and hence $\langle \boldsymbol{U}, \boldsymbol{V}\rangle$ mod m is *far* from uniform.

If m is a prime, then standard properties of the inner product two-source extractor imply that $\mathbf{H}_\infty(\boldsymbol{U}) + \mathbf{H}_\infty(\boldsymbol{V}) = 2\log K$ is not much bigger than $n \log m$, else the inner product extractor's output would be very close to uniform. This straightaway yields the desired bound of $K \lessapprox m^{n/2}$.

For composite m, this approach does not work directly, since $\mathbb{Z}_m$ is *not* a field. In particular, say for $m = 6$, it is possible to have two random variables distributed uniformly over $\{0, 2, 4\}^n$ such that their inner product is *far* from uniform modulo 6, but the sum of the min-entropies is $2n \log_2 3 \gg n \log 6$. The authors overcome this by showing that this is essentially the *only* bad case, i.e., it can be attributed to some "bad" factor $s \mid m$ (in the example above, $s = 2$). Hence, with some careful inductive argument on the modulus m/s, one can still force out the fundamental observation: $\mathbf{H}_\infty(\boldsymbol{U}) + \mathbf{H}_\infty(\boldsymbol{V}) = 2\log K$, which cannot be too large due to the biased behavior of $\langle \boldsymbol{U}, \boldsymbol{V}\rangle$, showing the desired upper bound on K.

Going beyond $\boldsymbol{m^{n/2}}$. A natural attempt towards improving the upper bound for K would be to consider the inner product $\langle \boldsymbol{U}^\dagger, \boldsymbol{V}^\dagger\rangle$ where $\boldsymbol{U}^\dagger = \boldsymbol{U}_1 + \cdots + \boldsymbol{U}_\ell$ for some integer ℓ with each $\boldsymbol{U}_i$ sampled uniformly and independently from $\mathcal{U}$ (and $\boldsymbol{V}^\dagger$ similarly defined), and hope for two nice events:

(a) $\mathbf{H}_\infty(\boldsymbol{U}^\dagger) + \mathbf{H}_\infty(\boldsymbol{V}^\dagger) \approx 2\ell \log K \gg 2\log K$;

(b) $\langle \boldsymbol{U}^\dagger, \boldsymbol{V}^\dagger\rangle$ remains far from uniform just like $\langle \boldsymbol{U}, \boldsymbol{V}\rangle$. And hence $\mathbf{H}_\infty(\boldsymbol{U}^\dagger) + \mathbf{H}_\infty(\boldsymbol{V}^\dagger) \lessapprox m \log n$.

In other words, we are trying to boost the lower bound on min-entropy by adding ℓ independent samples of $\boldsymbol{U}$, and at the same time maintain the upper bound on min-entropy. And if the above is true, one would obtain a substantial improvement on the upper bound (e.g. $K \le m^{n/4}$ even for the smallest non-trivial $\ell = 2$).

In the rest of this proof overview, we will walk through how we show the two events above are true for any $\{\alpha,\beta\}$-matching vector family. For simplicity, let $m = pq$ for distinct primes p and q, with $q > p$. And one could further assume that $\alpha = 0 \pmod p, \alpha \neq 0 \pmod q$, and $\beta = 0 \pmod q, \beta \neq 0 \pmod p$; see Lemma 25 for more details.

Event (a)

In order to show that $\mathbf{H}_\infty(\boldsymbol{U}^\dagger) \approx \ell \log K$, it is sufficient to show that $\boldsymbol{U}^\dagger = \sum_{i=1}^{\ell} \boldsymbol{U}_i$ *does not* generate too many collisions.

Omitting most technical details, we show that if $\boldsymbol{u}_1 + \ldots + \boldsymbol{u}_\ell = \boldsymbol{u}'_1 + \ldots + \boldsymbol{u}'_\ell$ (i.e. a collision), then it must be the case that $\langle \boldsymbol{u}'_i, \boldsymbol{v}_1 \rangle = \alpha$ for all $1 \leq i \leq \ell \leq q$, where $\boldsymbol{v}_1$ is the matching vector for $\boldsymbol{u}_1$. This is due to the heavy constraint that – (1) any inner product can only take values from $\{0,\alpha,\beta\}$, and (2) inner product with $\boldsymbol{v}_1$ gives an expression of the form $x_1\alpha + x_2\beta$, on the LHS, while $y_1\alpha + y_2\beta$, on the RHS, for nonnegative integers x_i, y_i such that $x_1 + x_2 \leq \ell - 1$. and $y_1 + y_2 \leq \ell$. (Remember $\langle \boldsymbol{u}_1, \boldsymbol{v}_1 \rangle = 0$.)

This anomaly (that there is plausibly 1 less term in the LHS) is sufficient to conclude that if there are too many collisions, one can easily extract a large $\{\alpha\}$-matching vector family, which would contradict a known linear upper bound, since a simple pigeon-hole principle argument shows that $\mathbf{MV}(pq, n, 2) \leq n + 1$; see Lemma 18.

In the actual proof, for technical reasons, we work with a slightly different distribution $\boldsymbol{U}^\star$, which adds up ℓ *distinct* random vectors from $\mathcal{U}$. In other words, one may view $\boldsymbol{U}^\dagger$ as sampling *with replacement* and $\boldsymbol{U}^\star$ as sampling *without replacement* from $\mathcal{U}$. These two distributions are so close to each other that we are able to substitute one with the other without incurring much loss in the final argument. Moreover, this distinction also helps us to argue that the sum of the coefficients in LHS is *exactly* $\ell - 1$, while it is *exactly* ℓ in RHS. For the detailed proof on lower bounding the min-entropy, see Section 3.2.

Event (b)

To show that $\langle \boldsymbol{U}^\dagger, \boldsymbol{V}^\dagger \rangle$ is far from uniform, we directly examine its bias $\left|\mathbb{E}[\omega^{\langle \boldsymbol{U}^\dagger, \boldsymbol{V}^\dagger \rangle}]\right|$ in terms of Fourier coefficients (where ω is a primitive m-th root of unity). Via standard Fourier-analytic techniques, and using the "almost" multiplicative property of the bias (Lemma 11), we can *almost* show that

$$\left|\mathbb{E}[\omega^{\langle \boldsymbol{U}^\dagger, \boldsymbol{V}^\dagger \rangle}]\right| \gtrapprox \left|\mathbb{E}[\omega^{\langle \boldsymbol{U}, \boldsymbol{V} \rangle}]\right|^{\ell^2} .$$

That is, adding ℓ independent samples of $\boldsymbol{U}$ (and $\boldsymbol{V}$) will *not* smooth out the inner product for not too large ℓ (see Lemmas 11 and 13).

On the other hand, since $\langle \boldsymbol{U}, \boldsymbol{V} \rangle$ takes only 3 possible values with 0 showing up with negligible probability $\frac{1}{K}$, we can show that $\left|\mathbb{E}[\omega^{\langle \boldsymbol{U}, \boldsymbol{V} \rangle}]\right|$ has to be large! And this is essentially all we need. For the detailed proof on upper bounding the min-entropy, see Section 3.1.

Combining all of the above and choosing $\ell = q$, we get $2q \log K \lessapprox \mathbf{H}_\infty(\boldsymbol{U}^\star) + \mathbf{H}_\infty(\boldsymbol{V}^\star) \lessapprox n \log m$ which gives us the desired upper bound of $K \lessapprox m^{n/2q}$. For a general m one can always reduce it to working with $m = p^s q^t$, for primes p and q; see Lemmas 24 and 25. And for $m = p^s q^t$, the argument is similar to the above.

Lastly, it is worth noting that via a more direct analysis on the min-entropy of carefully chosen random variables, we are able to avoid the inductive argument in [7], which was tailored for composite m.

1.7 Conclusions and Open Questions

Our approach can be broadly summarized as considering $\boldsymbol{U}^\dagger := \boldsymbol{U}_1 + \cdots + \boldsymbol{U}_\ell$, and similarly $\boldsymbol{V}^\dagger := \boldsymbol{V}_1 + \cdots + \boldsymbol{V}_\ell$, and then proving that

1. $\omega^{\langle \boldsymbol{U}^\dagger, \boldsymbol{V}^\dagger \rangle}$ is far from uniform, where ω is the m-th root of unity, and
2. $\mathbf{H}_\infty(\boldsymbol{U}^\dagger) \approx \ell \cdot \mathbf{H}_\infty(\boldsymbol{U})$ for $\ell = q$.

The first item implies that $\mathbf{H}_\infty(\boldsymbol{U}^\dagger) + \mathbf{H}_\infty(\boldsymbol{V}^\dagger)$ is not much larger than $m \log n$, and this combined with the second item implies that $\mathbf{H}_\infty(\boldsymbol{U}) = \mathbf{H}_\infty(\boldsymbol{V})$ is small, giving an upper bound on K.

We prove these two items for $\ell = q$, and for r-restricted families for $r = 3$. Notice that the approach is much more general, and it is just a shortcoming of our proof techniques that doesn't allow us to go beyond $\ell > q$, or $r > 3$. In particular, for $r = 3$, the statement (1) above holds for $\ell \gg q$, and we leave it as an open question whether one can prove that for $\ell \gg q$, $\mathbf{H}_\infty(\boldsymbol{U}^\dagger) \gg q\mathbf{H}_\infty(\boldsymbol{U})$. This will immediately imply a better upper bound on K for 3-query matching vector codes.

Similarly, we leave it as an open question whether the same approach can be extended to a larger number of queries.

2 Preliminaries

2.1 Random Variable and Entropy

▶ **Definition 6** (Min-Entropy). *Let X be a random variable over a set $\mathcal{X}$, the Min-Entropy is defined as:*

$$\mathbf{H}_\infty(X) = \min_{x \in \mathcal{X}} -\log(\Pr[X = x]) \ .$$

▶ **Definition 7** (Collision Entropy). *Let X and Y be independent and identically distributed random variables, the Collision Entropy is defined as:*

$$\mathbf{H}_2(X) = -\log \Pr[X = Y] \ .$$

Alternatively,

$$\mathbf{H}_2(X) = -\log \sum_x (\Pr[X = x])^2 .$$

▶ **Observation 8** (Relation Between Collision and Min-Entropy). *For any random variable X, it holds that*

$$\mathbf{H}_2(X) \geq \mathbf{H}_\infty(X) \ .$$

▶ **Lemma 9.** *Let X and Y be independent random variables of the same domain $\mathcal{X}$ closed under addition. It holds that*

$$\mathbf{H}_\infty(X) \leq \mathbf{H}_\infty(X + Y) \ .$$

Proof. For any $z \in \mathcal{X}$, we have:

$$\begin{aligned}
\Pr[X + Y = z] &= \sum_{y \in \mathcal{X}} \Pr[X = z - y | Y = y] \Pr[Y = y] \\
&= \sum_{y \in \mathcal{X}} \Pr[X = z - y] \Pr[Y = y] \\
&\leq \sum_{y \in \mathcal{X}} (\max_x \Pr[X = x]) \Pr[Y = y] \\
&\leq \max_x \Pr[X = x] \ .
\end{aligned}$$

Therefore,

$$\mathbf{H}_\infty(X) \;=\; \log \frac{1}{\max_x \Pr[X=x]} \;\leq\; \log \frac{1}{\max_z \Pr[X+Y=z]} \;=\; \mathbf{H}_\infty(X+Y)\,,$$

as desired. ◀

Let ω be a primitive root of unity of order m. Let X and Y be two probability distributions over $\mathbb{Z}_m^n$. The bias of the output distribution wrt ω, denoted $\mathbb{E}[\omega^{\langle X,Y\rangle}]$, can be defined as follows:

$$\mathbb{E}[\omega^{\langle X,Y\rangle}] \;=\; \mathbb{E}_{x\sim X, y\sim Y}[\omega^{\langle x,y\rangle}] \;=\; \sum_{x,y} X(x)Y(y)\omega^{\langle x,y\rangle}\,.$$

▶ **Lemma 10** (Lemma 2.14 from [7]). *Let ω be a primitive root of unity of order m. Let X and Y be two probability distributions over $\mathbb{Z}_m^n$. If $|\mathbb{E}[\omega^{\langle X,Y\rangle}]| \geq \epsilon$, then $\mathbf{H}_2(X) + \mathbf{H}_2(Y) \leq n\log m + 2\log(1/\epsilon)$.*

▶ **Lemma 11** (Lemma 3.3 from [21]). *Let ω be a primitive root of unity of order m. Let X and Y be two probability distributions over $\mathbb{Z}_m^n$, then $|\mathbb{E}[\omega^{\langle X-X,Y\rangle}]| \geq |\mathbb{E}[\omega^{\langle X,Y\rangle}]|^2$. Here $X - X$ is shorthand for taking the difference of two samples drawn independently according to the distribution X.*

▶ Remark 12. Lemma 11 was stated for distributions over $\mathbb{F}^n$ for any finite field $\mathbb{F}$ in [21], but the proof trivially extends to distributions over $\mathbb{Z}_m^n$. For completeness, we reproduce the proof from [21] in the appendix.

▶ **Lemma 13** ([21]). *Let ω be a primitive root of unity of order m. Let X and Y be two probability distributions over $\mathbb{Z}_m^n$. For any integer $x, y \geq 0$, $|\mathbb{E}[\omega^{\langle 2^x\cdot X - 2^x\cdot X, 2^y\cdot Y - 2^y\cdot Y\rangle}]| \geq |\mathbb{E}[\omega^{\langle X,Y\rangle}]|^{2^{x+y+2}}$. Here $2^x \cdot X$ represents the distribution of adding 2^x independent samples of X.*

Proof. Notice that $(X - X) - (X - X) = 2X - 2X$. Hence this follows from applying Lemma 11 $(x+1) + (y+1) = x + y + 2$ times. ◀

2.2 Matching Vector Families and Codes

Much of these definitions follow the same notation as Yekhanin's survey [25].

▶ **Definition 14** (Locally Decodable Codes). *A q-ary code $C : \Sigma^K \to \Sigma^N$ with $|\Sigma| = q$ is (r, δ, ε)-locally decodable if there exists a randomized decoding algorithm $\mathcal{A}$ such that*

1. *For all $x \in \Sigma^K$, $i \in [K]$ and all $y \in \Sigma^N$ such that $\Delta(C(x), y) \leq \delta$:*

$$\Pr_{\mathcal{A}}[\mathcal{A}^y(i) = x(i)] \geq 1 - \varepsilon$$

2. *$\mathcal{A}$ makes atmost r queries to y.*

The best-known constructions of locally decodable codes are derived from constructions of matching vector families over $\mathbb{Z}_m$ for composite m's. The following theorem demonstrates the parameters of locally decodable codes obtained from given parameters of a matching vector family.

▶ **Theorem 15** ([25]). *Let $\mathcal{U}, \mathcal{V}$ be a S-matching vector family over $\mathbb{Z}_m^n$ with $|\mathcal{U}| = |\mathcal{V}| = K$. Suppose $m \mid q - 1$ for some prime power q, then there exists a linear code $C : \mathbb{F}_q^K \to \mathbb{F}_q^{m^n}$ that is $(|S| + 1, \delta, (|S| + 1)\delta)$-locally decodable for every δ.*

In the low-query complexity regime, $r = O(1)$, the best-known matching vector families are based on Grolmusz's construction of set systems.

▶ **Theorem 16** ([25, Lemma 4.8]). *Let $m = \prod_{i=1}^{t} p_i$ be a product of distinct primes. Let w be a positive integer. Let $\{e_i\}_{i=1}^{t}$ be integers such that $p_i^{e_i} \geq w^{1/t}$ for all i. Let $d = \max_i p_i^{e_i}$ and $h \geq w$ be arbitrary. Then there exists an $\binom{h}{w}$-sized σm-bounded family of matching vectors in $\mathbb{Z}_m^n$ where $n = \binom{h}{\leq d}$ and σ is an arbitrary real larger than $\sum_{i \in [t]} \frac{1}{p_i}$.*

Then, using Theorem 15, this matching vector family construction can be transformed into a linear binary LDC with the following parameters:

▶ **Theorem 17** ([18], as stated in [25], Theorem 3.11). *For every integer $t \geq 2$ and for all $K \geq 2$, there exists an $r = 3 \cdot 2^{t-2}$-query (binary) linear locally decodable code over $\mathbb{F}_{2^t}$ encoding K-long messages to $2^{t^{(\log K)^{1/t}(\log\log K)^{1-1/t}}}$-long codewords and allowing $\delta = O(1/r)$ fraction of errors.*

In particular, when $r = 3$, there exists a code with codeword length $N = 2^{2^{\widetilde{O}(\sqrt{\log K})}}$

2.3 Upper Bounds for 2-restricted Matching Vector Families

In this section, we prove an upper bound on any $\{\alpha\}$-matching vector family over $\mathbb{Z}_m^n$. Without loss of generality, one can assume that m is a prime power p^s, because otherwise if $m = p^s q^t$, then without loss of generality $\alpha \not\equiv 0 \pmod{p^s}$, and therefore the same set of vectors form a matching vector family over $\mathbb{Z}_{p^s}^n$.

▶ **Lemma 18.** *For any prime p and positive integer s, $\mathbf{MV}(p^s, n, 2) \leq ns + 1$.*

Proof. Suppose $\alpha \not\equiv 0 \pmod{p^s}$. Assume towards contradiction that $\mathcal{U}, \mathcal{V}$ is a $\{\alpha\}$-matching vector family over $\mathbb{Z}_{p^s}^n$ with $|\mathcal{U}| = |\mathcal{V}| = K \geq ns + 2$.

Define $\boldsymbol{u}_i' := \boldsymbol{u}_{ns+2} - \boldsymbol{u}_i$ for $\boldsymbol{u}_i \in \mathcal{U}$. By definition of the $\{\alpha\}$-matching vector family, for $i, j \leq ns + 1$ we have:

$$\langle \boldsymbol{u}_i', \boldsymbol{v}_j \rangle = \begin{cases} \alpha, & i = j \\ 0, & i \neq j \end{cases} .$$

We claim that there exist integer coefficients $c_1, \ldots, c_{ns+1}$, not all zero, such that $-(p-1) \leq c_i \leq p - 1$, and

$$\sum_{i=1}^{ns+1} c_i \boldsymbol{u}_i' = 0 .$$

To see this, consider all possible sums $\sum_{i=1}^{ns+1} d_i \boldsymbol{u}_i' \in \mathbb{Z}_{p^s}^n$, for $d_i \in \{0, 1, \ldots, p-1\}$. There are p^{ns+1} such sums that can each take one of $(p^s)^n = p^{ns}$ values. By the pigeon-hole principle, there exist two distinct sums that are equal, i.e.,

$$\sum_{i=1}^{ns+1} d_i \boldsymbol{u}_i' = \sum_{i=1}^{ns+1} d_i' \boldsymbol{u}_i' ,$$

for some tuples $(d_1, \ldots, d_{ns+1}) \neq (d_1', \ldots, d_{ns+1}')$. The claim follows by taking $c_i = d_i - d_i'$ for $i \in \{1, \ldots, ns + 1\}$.

We have that at least one of $c_1, \ldots, c_{ns+1}$ is non-zero. Without loss of generality, let $c_{ns+1} \neq 0$. Notice that $c_{ns+1} \in \{-(p-1), \ldots, -1\} \cup \{1, \ldots, p-1\}$, and hence has a multiplicative inverse modulo p^s. Let $c'_i = c_{ns+1}^{-1} c_i \pmod{p^s}$ for $i \in [ns+1]$. Thus, we have that

$$\boldsymbol{u}'_{ns+1} \;=\; -\sum_{i=1}^{ns} c'_i \boldsymbol{u}'_i \ .$$

Now take the inner product on both sides with $\boldsymbol{v}_{ns+1}$. Note that, $\langle \boldsymbol{u}'_{ns+1}, \boldsymbol{v}_{ns+1} \rangle = \alpha$, while

$$\left\langle \sum_{i=1}^{ns} c'_i \boldsymbol{u}'_i, \boldsymbol{v}_{ns+1} \right\rangle \;=\; \sum_{i=1}^{ns} c'_i \langle \boldsymbol{u}'_i, \boldsymbol{v}_{ns+1} \rangle \;=\; 0 \ .$$

Therefore, we obtain that $\alpha = 0$ which is a contradiction. ◀

▶ Remark 19. By the above lemma, we get $\mathbf{MV}(m, n, 2) \leq n \log m + 1$, since trivially, $s \leq \log m / \log p \leq \log m$.

2.4 Upper Bounds for Prime Powers

In this section, we provide a polynomial upper bound for matching vector families over $\mathbb{Z}_m$ where m is a prime power. The sketch was personally communicated to us by Sivakanth Gopi. To prove this, we need the following lemma from [14]. We state it without proof, but it follows from Lucas' Theorem.

▶ **Lemma 20.** *Let p be a prime and $s \in \mathbb{Z}^+$. There exists a polynomial $f(x_1, \ldots, x_n)$ of degree $p^s - 1$ such that for $x \in \{0,1\}^n$, $f(x) \neq 0 \pmod{p}$ iff $\sum_{i=1}^n x_i$ is divisible by p^s. Specifically, $f(x) = \prod_{j=1}^{s-1} \left(1 - \binom{\sum_{i=1}^n x_i}{p^j}^{p-1}\right)$.*

Here $\binom{\sum_{i=1}^n x_i}{k} = \sum_{i_1 < \ldots < i_k} x_{i_1} \ldots x_{i_k}$ is always a valid integer polynomial for $k \leq n$. Note that in [14], the lemma is technically stated with $1 - f$ as opposed to f. We are now ready to show the upper bound. Also, note that f takes only 2 values 1 and $0 \pmod{p}$, depending on the divisibility property.

▶ **Theorem 21** (Upper bound for prime-power [15]). *Let $m = p^s$ for some prime p and $s \in \mathbb{Z}^+$. Then, for any matching vector family $\mathcal{U}, \mathcal{V} \subseteq \mathbb{Z}_m^n$ of size $t = |\mathcal{U}| = |\mathcal{V}|$ we have that $t \leq (2emn)^{m-1}$.*

Proof. We will first convert the matching vector family $\mathcal{U}, \mathcal{V}$ into a matching vector family consisting of vectors with *only zero or one* coordinates, we stress that the matching vector family is still considered over $\mathbb{Z}_m$ and $0, 1$ are interpreted as elements of $\mathbb{Z}_m$, not $\mathbb{F}_2$. Thereafter, we will argue by showing an upper bound on the rank of a specially constructed matrix, via the polynomial defined in Lemma 20.

Converting into a binary matching vector family. We will transform vectors in $\mathcal{U}, \mathcal{V}$ into vectors in $\{0,1\}^{n'} \subset \mathbb{Z}_m^{n'}$ for $n' = nm^2$, such that the pairwise inner-product $\langle \boldsymbol{u}, \boldsymbol{v} \rangle$ for $\boldsymbol{u} \in \mathcal{U}, \boldsymbol{v} \in \mathcal{V}$ is preserved over $\mathbb{Z}_m$. For any $\boldsymbol{u} \in \mathcal{U}$ (resp. $\boldsymbol{v} \in \mathcal{V}$), map each coordinate $\boldsymbol{u}_i$ (resp. $\boldsymbol{v}_i$) into the $m \times m$ matrix[1] $\boldsymbol{u}'_i$ (resp. $\boldsymbol{v}'_i$) with all 1's in the first $\boldsymbol{u}_i$ rows (resp. first $\boldsymbol{v}_i$ columns). It is easy to see that $\boldsymbol{u}_i \cdot \boldsymbol{v}_i = \langle \boldsymbol{u}'_i, \boldsymbol{v}'_i \rangle$ and

[1] Matrix is interpreted as a vector simply by reading entries of the matrix row by row.

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \sum_{i \in [n]} \boldsymbol{u}_i \cdot \boldsymbol{v}_i = \sum_{i \in [n]} \langle \boldsymbol{u}'_i, \boldsymbol{v}'_i \rangle = \langle \boldsymbol{u}', \boldsymbol{v}' \rangle$$

As a result, we can turn any matching vector family $\mathcal{U}, \mathcal{V} \subseteq \mathbb{Z}_m^n$ into one with only binary vectors in $\mathbb{Z}_m^{mn^2}$. For the rest of the argument, denote $\mathcal{U}, \mathcal{V} \subseteq \mathbb{Z}_m^{n'}$ to be the binary vectors obtained like above from the original $\mathcal{U}, \mathcal{V}$ where $n' = nm^2$.

Constructing a useful matrix. Suppose $\mathcal{U} = \{\boldsymbol{u}^{(1)}, \dots, \boldsymbol{u}^{(t)}\}, \mathcal{V} = \{\boldsymbol{v}^{(1)}, \dots, \boldsymbol{v}^{(t)}\}$ where all $\boldsymbol{u}^{(i)}, \boldsymbol{v}^{(j)} \in \{0,1\}^{n'} \subset \mathbb{Z}_m^{n'}$. Recall $m = p^s$, let $f(x)$ be such a polynomial as defined in Lemma 20. By the same lemma, $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \sum_{i=1}^{n'} \boldsymbol{u}_i \boldsymbol{v}_i = 0 \pmod{m}$ iff $f(\boldsymbol{u}_1\boldsymbol{v}_1, \dots, \boldsymbol{u}_{n'}\boldsymbol{v}_{n'}) \neq 0 \pmod{p}$. Now, we define a matrix $M \in \mathbb{Z}_p^{t \times t}$ whose rows are indexed by vectors in $\mathcal{U}$ and whose columns are indexed by vectors in $\mathcal{V}$ such that:

$$M_{i,j} = M(\boldsymbol{u}^{(i)}, \boldsymbol{v}^{(j)}) \ := \ f(\boldsymbol{u}_1^{(i)}\boldsymbol{v}_1^{(j)}, \dots, \boldsymbol{u}_{n'}^{(i)}\boldsymbol{v}_{n'}^{(j)}) \ .$$

Note that $M(\boldsymbol{u}^{(i)}, \boldsymbol{v}^{(j)}) \neq 0$ iff $i = j$. Hence, M is full rank diagonal matrix (infact, it is the identity matrix!).

On the other hand, by Lucas' Theorem, f has degree at most $p^s - 1 = m - 1$ and n' variables. Notice that there are at most $\binom{n'}{m-1} + \binom{n'}{m-2} + \dots + \binom{n'}{0} \leq \binom{n'+m-1}{m-1}$ possible monomials up to degree $m-1$. For any polynomial f of degree $m-1$ we can consider the vector $\widetilde{\boldsymbol{f}} \in \mathbb{Z}_m^{\binom{n'+m-1}{m-1}}$, a vector of coefficients of each of $\binom{n'+m-1}{m-1}$ monomials. One can also take any input $x_1, \dots, x_{n'}$ and consider a vector $\widetilde{\boldsymbol{x}} \in \mathbb{Z}_m^{\binom{n'+m-1}{m-1}}$ consisting of evaluations of each of the monomials on $x_1, \dots, x_{n'}$. Notice that then $f(x_1, \dots, x_{n'}) = \langle \widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{f}} \rangle$. Similarly, for each $\boldsymbol{u} \in \mathcal{U}, \boldsymbol{v} \in \mathcal{V}$ we can find $\widetilde{\boldsymbol{u}}, \widetilde{\boldsymbol{v}} \in \mathbb{Z}_m^{\binom{n'+m-1}{m-1}}$, such that $f(\boldsymbol{u}_1\boldsymbol{v}_1, \dots, \boldsymbol{u}_{n'}\boldsymbol{v}_{n'}) = \langle \widetilde{\boldsymbol{u}}, \widetilde{\boldsymbol{v}} \rangle$, for example $\widetilde{\boldsymbol{u}}$ consists of evaluations of monomials on input $u_1, \dots, u_{n'}$ while $\widetilde{\boldsymbol{v}}$ combines evaluations of monomials on $v_1, \dots, v_{n'}$ with coefficients of polynomial f. Thus, an obvious counting argument gives the desired upper bound:

$$t \ = \ \mathrm{rank}(M) \ \leq \ \binom{n' + m - 1}{n'} \ = \ \binom{nm^2 + m - 1}{m - 1} \ \leq \ (2enm)^{m-1} \ . \qquad ◀$$

Note that the above proof fails immediately for composites. For one, we don't have a version of the Lucas' Theorem. Moreover, there is no unified notion of rank that can be utilized, for example by reducing it to working over a prime field.

▶ Remark 22. For any constant $m = p^s$ for some prime p and positive integer s, Theorem 21 implies $\mathbf{MV}(m, n, r) \leq (2emn)^{m-1} = \mathrm{poly}(n)$ for any $0 < r \leq m$.

2.5 Some properties of 3-restricted matching vector families

We will consider matching vector families in $\mathbb{Z}_m$ where m is a composite positive integer. All algebraic operations in this section, unless otherwise stated, are modulo $\mathbb{Z}_m$.

Let $\alpha, \beta \in \mathbb{Z}_m \setminus \{0\}$. Consider a matching vector family $(\mathcal{U}, \mathcal{V})$ such that $\mathcal{U} = \{u_1, \dots, u_K\}, V = \{v_1, \dots, v_K\} \subseteq \mathbb{Z}_m^n$ that satisfy the following properties:

- For all $i \in [K]$, $\langle \boldsymbol{u}_i, \boldsymbol{v}_i \rangle = 0$.
- For all $i, j \in [K]$ with $i \neq j$, we have that $\langle \boldsymbol{u}_i, \boldsymbol{v}_j \rangle \in \{\alpha, \beta\}$.

We call $(\mathcal{U}, \mathcal{V})$ an $\{\alpha, \beta\}$-matching vector family of size K.

▶ Remark 23. For the rest of the paper until the end of Section 3.2, we will assume that $(\mathcal{U},\mathcal{V})$ is not a matching vector family modulo m' for any $m'|m$. We can make this assumption without loss of generality. This is because we will prove an upper bound on K that increases as a function of m, and we get a better bound if there exists such an m'.

In our argument, it will be useful to work with composites that are exactly the product of two prime powers. This is clearly the first non-trivial case since we have a polynomial upper bound (in n) for the case of prime powers. In fact, for 3-restricted matching vector families, [7] demonstrated that this is the only non-trivial case. We include the proof below for completeness.

▶ **Lemma 24** ([7]). *Let $(\mathcal{U},\mathcal{V})$ be a $(q+1)$-restricted matching vector family in $\mathbb{Z}_m^n$. Then m has at most q prime factors.*

Proof. Suppose $m = \prod_i p_i^{r_i}$ for some primes p_i's and corresponding positive integers r_i's. Suppose the possible non-zero values of inner products are $\alpha_1,\ldots,\alpha_q \pmod m$. For each α_j, there must exist i such that $\alpha_j \neq 0 \pmod{p_i^{r_i}}$. We can take m' to be the product of atmost q prime powers that satisfy the above condition for the respective α_j's. Thus the $(q+1)$-restricted matching vector family in $\mathbb{Z}_m^n$ induces (via the modulo m' operation) a matching vector family in $\mathbb{Z}_{m'}^n$ of the same size, and with the same number of restrictions. ◀

From the previous lemma, we can assume $m = p^s q^t$ for any $\{\alpha,\beta\}$-matching vector family $(\mathcal{U},\mathcal{V})$ where p, q are some primes and s, t are some positive integers. In the next lemma, we can show some more restrictions on the values taken by α and β.

▶ **Lemma 25.** *Let $m = p^s q^t$ be a composite positive integer such that for any m' which is a divisor of m, $(\mathcal{U},\mathcal{V})$ is not a matching vector family modulo m'. Then we have $\alpha = 0 \pmod{p^s q^{t-1}}$, $\alpha \neq 0 \pmod{q^t}$, $\beta = 0 \pmod{p^{s-1} q^t}$, $\beta \neq 0 \pmod{p^s}$.*

Proof. By the same argument as Lemma 24, we can assume without loss of generality that $\alpha \neq 0 \pmod{q^t}$ and $\beta \neq 0 \pmod{p^s}$. Now, suppose for a contradiction $\beta \neq 0 \pmod{p^{s-1}q^t}$. Clearly, $\alpha \neq 0 \pmod{p^{s-1}q^t}$ Then $(\mathcal{U},\mathcal{V})$ is a $\{\alpha,\beta\}$-matching vector family modulo $m' = p^{s-1}q^t \mid m$ which is a contradiction. Similarly, we can show $\alpha = 0 \pmod{p^s q^{t-1}}$. ◀

3 Upper Bound for 3-restricted Matching Vector Families

Let $\mathcal{U},\mathcal{V}$ be a matching vector family of size K in $\mathbb{Z}_m^n$ with non-zero residues α and β. By Lemma 24, we have that $m = p^s q^t$ for primes p and q and integers $s, t \geq 1$ with $p < q$. Further, using Lemma 25, we have that $\alpha = 0 \pmod{p^s q^{t-1}}$, $\alpha \neq 0 \pmod{q^t}$, $\beta = 0 \pmod{p^{s-1}q^t}$, $\beta \neq 0 \pmod{p^s}$. Define, a parameter τ (which is a small constant) as follows:

$$\tau := \frac{1}{2pq}.$$

Let $\mathcal{S}_{q,K} = \{A \subset [K] \mid |A| = q\}$ be the set of all subsets of $[K]$ of size q. Let $\omega = e^{\frac{2\pi i}{m}}$ be a primitive root of unity of order m.

For analysis, we define a few random variables as follows: Let $\boldsymbol{U}_i$ be uniform and independent vectors drawn from $\mathcal{U}$. Let ℓ be the *smallest* integer such that $q \leq 2^\ell$. We have:

1. $\boldsymbol{U} := \boldsymbol{U}_i$.
2. $\boldsymbol{U}^{\S} := \sum_{i=1}^{2^\ell} \boldsymbol{U}_i - \sum_{i=2^\ell+1}^{2^{\ell+1}} \boldsymbol{U}_i$. That is, sample *with replacement* $2 \cdot 2^\ell$ independent vectors from $\mathcal{U}$, add up the first 2^ℓ vectors and subtract away the remaining 2^ℓ vectors.

3. $\boldsymbol{U}^{\dagger} := \sum_{i=1}^{q} \boldsymbol{U}_i$. That is, sample *with replacement* q independent vectors from $\mathcal{U}$ and add them up.
4. $\boldsymbol{U}^{\ddagger} := \sum_{i=q+1}^{2^\ell} \boldsymbol{U}_i - \sum_{i=2^\ell+1}^{2^{\ell+1}} \boldsymbol{U}_i$. Here we define $\boldsymbol{U}^{\ddagger}$ such that $\boldsymbol{U}^{\S} = \boldsymbol{U}^{\dagger} + \boldsymbol{U}^{\ddagger}$.
5. $\boldsymbol{U}^{\star}$ is distributed as follows: pick a subset $A \sim \mathcal{S}_{q,K}$ uniformly at random and $\boldsymbol{U}^{\star} := \sum_{i\in A} \boldsymbol{u}_i$. In other words, sample *without replacement* q vectors from $\mathcal{U}$ and add them up. This is in contrast to $\boldsymbol{U}^{\dagger}$ which is sampling *with replacement.*

Random variables for $\mathcal{V}$ are defined similarly.

Sampling without replacement helps in lower bounding the min-entropy of $\boldsymbol{U}^{\star}$. On the other hand, we upper bound the min-entropy of $\boldsymbol{U}^{\dagger}$ where sampling with replacement (and hence full independence) is useful. And we are able to relate these two quantities as these two distributions are really close to each other.

The introduction of $\boldsymbol{U}^{\S}$ (and hence $\boldsymbol{U}^{\ddagger} = \boldsymbol{U}^{\S} - \boldsymbol{U}^{\dagger}$) serves as a tool for lower bounding the bias of $\boldsymbol{U}^{\dagger}$ via Lemma 13 because we are restricted to sums containing powers of 2 elements.

3.1 Upper Bound on Min-Entropy

In the following lemma, we prove that since $\mathcal{U}, \mathcal{V}$ constitute a 3-restriced **MV** family over $\mathbb{Z}_m^n$, then $\langle \boldsymbol{U}, \boldsymbol{V}\rangle$ has to be *far* from uniform. This is captured via Fourier analysis.

▶ **Lemma 26** (Large bias for **MV** family). *Let $K \geq 10pq$. Then, we have that $\left|\mathbb{E}[\omega^{\langle \boldsymbol{U},\boldsymbol{V}\rangle}]\right| \geq \tau$.*

Proof. This proof relies on the fact that the random variable $\omega^{\langle \boldsymbol{U},\boldsymbol{V}\rangle}$ is essentially supported on two values, $\{\omega^\alpha, \omega^\beta\}$, since the mass on ω^0 (corresponding to $\langle \boldsymbol{U}, \boldsymbol{V}\rangle = 0$) is very small. To complete the argument, we show that $\alpha - \beta$ is an integer different from $m/2$, and thus the mass on ω^α doesn't cancel that on ω^β. The detailed argument is given below.

Recall that $\langle \boldsymbol{U}, \boldsymbol{V}\rangle \in \{0, \alpha, \beta\}$ and $\Pr[\langle \boldsymbol{U}, \boldsymbol{V}\rangle = 0] = \frac{1}{K}$. Let $\lambda := \Pr[\langle \boldsymbol{U}, \boldsymbol{V}\rangle = \alpha]$, $\lambda \in \left[0, 1 - \frac{1}{K}\right]$. Then $\Pr[\langle \boldsymbol{U}, \boldsymbol{V}\rangle = \beta] = 1 - \lambda - \frac{1}{K}$. Since $\alpha = 0 \pmod{p^s q^{t-1}}$ and $\beta = 0 \pmod{p^{s-1} q^t}$, let $\frac{\alpha}{m} = \frac{c}{q}$ and $\frac{\beta}{m} = \frac{d}{p}$, for some $0 < c < q$ and $0 < d < p$. By definition of the bias, we have

$$\begin{aligned}
\left|\mathbb{E}[\omega^{\langle \boldsymbol{U},\boldsymbol{V}\rangle}]\right| &= \left|\sum_{u,v\in\mathbb{Z}_m^n} \boldsymbol{U}(u)\boldsymbol{V}(v)\omega^{\langle u,v\rangle}\right| \\
&\geq \min_{\lambda\in[0,1-\frac{1}{K}]} \left|\frac{1}{K} + \lambda\omega^\alpha + (1 - 1/K - \lambda)\omega^\beta\right| \\
&\geq \min_{\lambda\in[0,1-\frac{1}{K}]} \left|\lambda\omega^\alpha + (1 - 1/K - \lambda)\omega^\beta\right| - \frac{1}{K} \\
&= \min_{\lambda\in[0,1-\frac{1}{K}]} \left|\lambda + (1 - 1/K - \lambda)\omega^{\beta-\alpha}\right| - \frac{1}{K} \\
&= \min_{\lambda\in[0,1-\frac{1}{K}]} \sqrt{\lambda^2 + (1 - 1/K - \lambda)^2 + 2\lambda(1 - 1/K - \lambda)\cos\left(2\pi\left(\frac{c}{q} - \frac{d}{p}\right)\right)} - \frac{1}{K} \\
&\geq \min_{\lambda\in[0,1-\frac{1}{K}]} \sqrt{\lambda^2 + (1 - 1/K - \lambda)^2 + 2\lambda(1 - 1/K - \lambda)\cos\left(\frac{(pq-1)\pi}{pq}\right)} - \frac{1}{K} \\
&\geq \left(\frac{K-1}{K}\right) \cdot \sin\left(\frac{\pi}{2pq}\right) - \frac{1}{K} \\
&\geq \tau\,.
\end{aligned}$$

The third equality follows from the identity that for constants c_i, and θ,

$$|c_1 + c_2 e^{i\theta}|^2 = |(c_1 + c_2\cos\theta) + ic_2\sin\theta|^2 = c_1^2 + c_2^2 + 2c_1c_2\cos\theta \ .$$

The third last inequality follows from the observation (which we discuss below) that $\left|2\pi\frac{cp-dq}{pq} - \pi\right| \geq \frac{\pi}{pq}$, and that cos θ for $\theta \in (0, 2\pi)$ is minimized when $|\theta - \pi|$ is minimized. If pq is odd, the above is trivially true since after multiplying left-hand-side by $\frac{pq}{\pi}$ we obtain $|2(cp - dq) - pq|$ and this is a difference of odd and even number and thus the absolute value has to be at least 1. Now, suppose $p = 2$, via similar argument we have

$$\left|\frac{2c - dq}{q} - 1\right| \geq \frac{1}{q} > \frac{1}{2q} \ ,$$

since $c \neq 0 \pmod{q}$ and q is an odd prime.

The second last inequality follows from taking $\lambda = \frac{K-1}{2K}$ which minimizes the expression, and using $1 - \cos\theta = 2\sin^2(\theta/2)$, giving a value of the square root to be

$$\left(2\left(\frac{K-1}{2K}\right)^2 - 2\cdot\left(\frac{K-1}{2K}\right)^2\cos\left(\frac{\pi}{pq}\right)\right)^{1/2} = \left(\frac{K-1}{K}\right)\sin\left(\frac{\pi}{2pq}\right) \ .$$

Finally, the last inequality follows from $\sin x \geq \frac{x}{2}$ for $x \in [0, 1]$ and that $K > 10pq$. ◀

Next, we will show that $\boldsymbol{U}^\star$ has low min-entropy.

▶ **Lemma 27.** $\left|\mathbb{E}[\omega^{\langle \boldsymbol{U}^\dagger + \boldsymbol{U}^\ddagger, \boldsymbol{V}^\dagger + \boldsymbol{V}^\ddagger\rangle}]\right| \geq \tau^{16q^2}$.

Proof. Recall that ℓ is the *smallest* integer such that $q \leq 2^\ell$. Thus, $2^{\ell-1} < q$, i.e., $2^\ell < 2q$. By Lemma 13, we can control the bias of $\langle \boldsymbol{U}^\S, \boldsymbol{V}^\S\rangle$:

$$\begin{aligned}
\left|\mathbb{E}[\omega^{\langle \boldsymbol{U}^\dagger + \boldsymbol{U}^\ddagger, \boldsymbol{V}^\dagger + \boldsymbol{V}^\ddagger\rangle}]\right| &= \left|\mathbb{E}[\omega^{\langle \boldsymbol{U}^\S, \boldsymbol{V}^\S\rangle}]\right| \\
&\geq \left|\mathbb{E}[\omega^{\langle \boldsymbol{U}, \boldsymbol{V}\rangle}]\right|^{2^{2(\ell+1)}} \\
&\geq \left|\mathbb{E}[\omega^{\langle \boldsymbol{U}, \boldsymbol{V}\rangle}]\right|^{4\cdot(2q)^2} \\
&= \tau^{16q^2} \ .
\end{aligned}$$

◀

Define the event E_U (resp. E_V) where each $\boldsymbol{U}_i$ (resp. $\boldsymbol{V}_i$) is unique when $\boldsymbol{U}^\dagger$ (resp. $\boldsymbol{V}^\dagger$) is sampled. Notice that $\boldsymbol{U}^\dagger|E_U$ is distributed exactly as $\boldsymbol{U}^\star$. Also, $\Pr[\neg E_U] = \Pr[\neg E_V] \leq \frac{q^2}{K}$ and so, by the union bound $\Pr[\neg(E_U \wedge E_V)] \leq \frac{2q^2}{K}$.

▶ **Lemma 28.** $\left|\mathbb{E}[\omega^{\langle \boldsymbol{U}^\star + \boldsymbol{U}^\ddagger, \boldsymbol{V}^\star + \boldsymbol{V}^\ddagger\rangle}]\right| \geq \tau^{16q^2} - \frac{2q^2}{K}$.

Proof. The proof uses the above observation that $\boldsymbol{U}^\dagger|E_U$ (resp. $\boldsymbol{V}^\dagger|E_V$) is distributed exactly as $\boldsymbol{U}^\star$ (resp. $\boldsymbol{V}^\star$), and that both E_U and E_V hold simultaneously with probability at least $1 - \frac{2q^2}{K}$. In particular,

$$\begin{aligned}\left|\mathbb{E}[\omega^{\langle \boldsymbol{U}^\dagger+\boldsymbol{U}^\ddagger,\boldsymbol{V}^\dagger+\boldsymbol{V}^\ddagger\rangle}]\right| &= \Big|\mathbb{E}[\omega^{\langle \boldsymbol{U}^\dagger+\boldsymbol{U}^\ddagger,\boldsymbol{V}^\dagger+\boldsymbol{V}^\ddagger\rangle} \mid E_U \wedge E_V]\cdot \Pr[E_U\wedge E_V] \\ &\qquad + \mathbb{E}[\omega^{\langle \boldsymbol{U}^\dagger+\boldsymbol{U}^\ddagger,\boldsymbol{V}^\dagger+\boldsymbol{V}^\ddagger\rangle} \mid \neg(E_U \wedge E_V)]\Pr[\neg(E_U\wedge E_V)]\Big| \\ &\le \left|\mathbb{E}[\omega^{\langle \boldsymbol{U}^\dagger+\boldsymbol{U}^\ddagger,\boldsymbol{V}^\dagger+\boldsymbol{V}^\ddagger\rangle} \mid E_U \wedge E_V]\right| \Pr[E_U\wedge E_V] \\ &\qquad + \left|\mathbb{E}[\omega^{\langle \boldsymbol{U}^\dagger+\boldsymbol{U}^\ddagger,\boldsymbol{V}^\dagger+\boldsymbol{V}^\ddagger\rangle} \mid \neg(E_U \wedge E_V)]\right|\Pr[\neg(E_U\wedge E_V)] \\ &= \left|\mathbb{E}[\omega^{\langle \boldsymbol{U}^\star+\boldsymbol{U}^\ddagger,\boldsymbol{V}^\star+\boldsymbol{V}^\ddagger\rangle}]\right| \Pr[E_U\wedge E_V] \\ &\qquad + \left|\mathbb{E}[\omega^{\langle \boldsymbol{U}^\dagger+\boldsymbol{U}^\ddagger,\boldsymbol{V}^\dagger+\boldsymbol{V}^\ddagger\rangle} \mid \neg(E_U \wedge E_V)]\right|\Pr[\neg(E_U\wedge E_V)] \\ &\le \left|\mathbb{E}[\omega^{\langle \boldsymbol{U}^\star+\boldsymbol{U}^\ddagger,\boldsymbol{V}^\star+\boldsymbol{V}^\ddagger\rangle}]\right| + \frac{2q^2}{K}\ .\end{aligned}$$

It follows that $\left|\mathbb{E}[\omega^{\langle \boldsymbol{U}^\star+\boldsymbol{U}^\ddagger,\boldsymbol{V}^\star+\boldsymbol{V}^\ddagger\rangle}]\right| \ge \tau^{16q^2} - \frac{2q^2}{K}$. ◀

▶ **Lemma 29.** $\mathbf{H}_\infty(\boldsymbol{U}^\star)+\mathbf{H}_\infty(\boldsymbol{V}^\star) \le n\log m - 2\log(\tau^{16q^2}-\frac{2q^2}{K})$.

Proof. By Lemma 28 and Lemma 10, we have that

$$\mathbf{H}_2(\boldsymbol{U}^\star+\boldsymbol{U}^\ddagger)+\mathbf{H}_2(\boldsymbol{V}^\star+\boldsymbol{V}^\ddagger) \le n\log m - 2\log(\tau^{16q^2}-\frac{2q^2}{K})\ .$$

Finally, notice that $\boldsymbol{U}^\star$ is independent of $\boldsymbol{U}^\ddagger$ and $\boldsymbol{V}^\star$ is independent of $\boldsymbol{V}^\ddagger$. By Lemma 9, we have:

$$\mathbf{H}_\infty(\boldsymbol{U}^\star)+\mathbf{H}_\infty(\boldsymbol{V}^\star) \le \mathbf{H}_\infty(\boldsymbol{U}^\star+\boldsymbol{U}^\ddagger)+\mathbf{H}_\infty(\boldsymbol{V}^\star+\boldsymbol{V}^\ddagger) \le n\log m - 2\log\left(\tau^{16q^2}-\frac{2q^2}{K}\right)\ ,$$

where $\mathbf{H}_\infty(X) \le \mathbf{H}_2(X)$ for any random variable X. ◀

3.2 Lower Bound on Min-Entropy

In this subsection, we will talk about sums of vectors from $\mathcal{U},\mathcal{V}$. For each $A \subseteq [K]$, we can naturally associate a sum of vectors $\sum_{i\in A}\boldsymbol{u}_i$ from $\mathcal{U}$ (and similarly for $\mathcal{V}$). Note that we only care about *unique* sums where each vector from $\mathcal{U}$ (or $\mathcal{V}$) appears only once. In this view, a *collision* corresponds to pair of sets $A, B, A \ne B$ such that $\sum_{i\in A}\boldsymbol{u}_i = \sum_{j\in B}\boldsymbol{u}_j$ (or similarly for vectors from $\mathcal{V}$). In the following we give a lower bound on the min-entropy of $\boldsymbol{U}^\star$. By symmetry, the statement holds for $\boldsymbol{V}^\star$ as well.

We start with a lemma that discusses the behaviour of a collision.

▶ **Lemma 30.** *Let $m = p^s q^t$ where $q > p$ and $\mathcal{U},\mathcal{V}$ be a $\{0,\alpha,\beta\}$-restricted matching vector family of size K such that $\alpha = 0 \pmod{p^s q^{t-1}}$, $\alpha \ne 0 \pmod{q^t}$ and $\beta = 0 \pmod{p^{s-1}q^t}$, $\beta \ne 0 \pmod{p^s}$. Let $A, B \in \mathcal{S}_{q,K}, A \ne B$ be such that $\sum_{i\in A}\boldsymbol{u}_i = \sum_{i\in B}\boldsymbol{u}_i$ where each $\boldsymbol{u}_i \in \mathcal{U}$. Then:*

(i) *For each $i \in A, j \in B$, $\boldsymbol{u}_i \ne \boldsymbol{u}_j$ ($A \cap B = \emptyset$).*

(ii) *For each $i \in A, j \in B$, $\langle \boldsymbol{u}_j, \boldsymbol{v}_i\rangle = \alpha \pmod{q^t}$.*

Proof.

Proof of Part (i). For any $i \in A$, we can write $\sum_{j\in B}\langle \boldsymbol{u}_j, \boldsymbol{v}_i\rangle - \sum_{j\in A}\langle \boldsymbol{u}_j, \boldsymbol{v}_i\rangle$ as a linear combination of α's and β's such that

$$a_i\alpha + b_i\beta = \sum_{j\in B}\langle \boldsymbol{u}_j, \boldsymbol{v}_i\rangle - \sum_{j\in A}\langle \boldsymbol{u}_j, \boldsymbol{v}_i\rangle = \langle \sum_{j\in B}\boldsymbol{u}_j - \sum_{j\in A}\boldsymbol{u}_j, \boldsymbol{v}_i\rangle = 0 \pmod{m} \tag{1}$$

for some integers a_i, b_i.

As $\langle \boldsymbol{u}_i, \boldsymbol{v}_i\rangle = 0 \pmod{m}$, there are q terms in the first sum and $q-1$ terms in the second sum; this is where we use the fact that $\boldsymbol{u}_i$ were picked without replacement. Therefore, we have the obvious inequalities: $-(q-1) \le a_i, b_i \le q$. Since, $\beta = 0 \pmod{q^t}$, and $a_i\alpha + b_i\beta = 0 \pmod{m}$, it follows that $a_i\alpha + b_i\beta = 0 \pmod{q^t} \implies a_i\alpha = 0 \pmod{q^t}$. Since α is divisible by q^{t-1}, and not q^t, and a_i is between $-(q-1)$ and q, it must happen that either $a_i = 0$ or $a_i = q$.

Suppose for a contradiction that $i \in A \cap B$. Let $k \in A\backslash B$ (non-empty since $A \ne B$). Then,

$$a_k\alpha + b_k\beta = \sum_{j\in B}\langle \boldsymbol{u}_j, \boldsymbol{v}_k\rangle - \sum_{j\in A}\langle \boldsymbol{u}_j, \boldsymbol{v}_k\rangle = \sum_{j\in B\backslash A}\langle \boldsymbol{u}_j, \boldsymbol{v}_k\rangle - \sum_{j\in A\backslash(B\cup\{k\})}\langle \boldsymbol{u}_j, \boldsymbol{v}_k\rangle \ .$$

By our choice of $\boldsymbol{u}_k$, fact that all inner products on the right-hand side above are non-zero, and since $|A| = |B| = q$, we know that:

1. $|B\backslash A| = |A\backslash(B\cup\{k\})| + 1$, and therefore $a_k + b_k = |B\backslash A| - |A\backslash(B\cup\{k\})| = 1$.
2. $-(q-1) \le -|A\backslash(B\cup\{k\})| \le b_k \le |B\backslash A| \le q-1$

If $a_k = 0$, then $b_k = 1$ implies that $a_k\alpha + b_k\beta = \beta \ne 0 \pmod{m}$, a contradiction to Equation (1). Also, notice that $a_k \ne q$ since for the same reason as above $a_k \le |B\backslash A| \le q-1$. This completes the proof of Part (i).

Proof of Part (ii). Now, consider any $i \in A$, clearly $i \notin B$ from our proof above. From a similar argument as above, we know $a_i + b_i = 1$. And further $a_i = q$ and $b_i = 1 - a_i = -(q-1)$. However the only way we can have $a_i = q$ is if each term in the sum $\sum_{j\in B}\langle \boldsymbol{u}_j, \boldsymbol{v}_i\rangle$ is α which completes the proof. Note that we can trivially reduce from mod m to mod q^t, since $\langle \boldsymbol{u}_j, \boldsymbol{v}_i\rangle = \alpha \ne 0 \pmod{q^t}$. ◀

▶ **Theorem 31.** *Let $m = p^s q^t$ where $q > p$ and let $\boldsymbol{U}^\star$ be defined as in the beginning of this section. Then $\mathbf{H}_\infty(\boldsymbol{U}^\star) \ge q\log\left(\frac{K}{q}\right) - \log(nt+1)$.*

Proof. We will upper bound the probability $\Pr_{\boldsymbol{U}^\star}[\boldsymbol{U}^\star = \boldsymbol{u}]$, where $\boldsymbol{u} \in \mathbb{Z}_m^n$ is an arbitrary vector. Let

$$C_{\boldsymbol{u}} := \left\{ A \in \mathcal{S}_{q,K} \mid \sum_{i\in A}\boldsymbol{u}_i = \boldsymbol{u} \right\} ,$$

be the set of sums of length q from $\mathcal{S}_{q,K}$ such that its sum is equal to $\boldsymbol{u}$. Note that each term in the sum is unique by how we sample $\boldsymbol{U}^\star$.

Now, consider any pair of collisions $A, B \in C_{\boldsymbol{u}}$. From Lemma 30, we know that for each $i \in A, j \in B$, $i \ne j$ and $\langle \boldsymbol{u}_j, \boldsymbol{v}_i\rangle = \alpha$, and all the sets in $C_{\boldsymbol{u}}$ are disjoint. In particular, if we pick exactly one vector from each sum in $C_{\boldsymbol{u}}$ (and their corresponding matching vectors in $\mathcal{V}$), the set of vectors form a $\{\alpha\}$-matching vector family in $\mathbb{Z}_{q^t}^n$. Formally, $\mathcal{U}' := \{\boldsymbol{u}_i \mid i = \min(A), A \in C_{\boldsymbol{u}}\} \subseteq \mathcal{U}$, and $\mathcal{V}' \subseteq \mathcal{V}$ with the corresponding matching vectors forms a $\{\alpha\}$-matching vector family in $\mathbb{Z}_{q^t}^n$ of size $|C_{\boldsymbol{u}}|$.

By Lemma 18, we have $|C_{\boldsymbol{u}}| \leq nt + 1$. Therefore, we have

$$\mathbf{H}_\infty(\boldsymbol{U}^\star) \;\geq\; \log\left(\frac{\binom{K}{q}}{nt+1}\right) \;\geq\; q\log\left(\frac{K}{q}\right) - \log(nt+1)\;. \qquad ◀$$

▶ Remark 32. As pointed out by a helpful anonymous reviewer, we can extend Lemma 30 to work for $A, B \in \mathcal{S}_{q',K}$ where $q' > q$ in certain cases. Suppose we again write $\sum_{j\in B}\langle \boldsymbol{u}_j, \boldsymbol{v}_i\rangle - \sum_{j\in A}\langle \boldsymbol{u}_j, \boldsymbol{v}_i\rangle$ as a linear combination of $a_{i,q'}\alpha + b_{i,q'}\beta = 0 \pmod{m}$. The proof only requires two properties from q':

1. For the first part, we require that for every $q'' < q'$ the corresponding equation $a_{i,q''}\alpha + b_{i,q''}\beta = 0 \pmod{m}$ has only trivial solutions ($a_{i,q''} = b_{i,q''} = 0$).
2. For the second part, we require that $a_{i,q'} = q'$.

For $m = 6$, $q' = 3$ but by a calculation it can be shown that for $m = 15$, we can have $q' = 6 > 5 = q$. This implies that we can improve the conclusion of Theorem 31 to $\mathbf{H}_\infty(\boldsymbol{U}^\star) \geq q'\log\left(\frac{K}{q'}\right) - \log(nt+1)$ which eventually gives us a better bound for Theorem 33. However, we omit this for clarity.

3.3 Main Theorem

We are now ready to prove our main theorem.

▶ **Theorem 33.** *Let $n, m \geq 1$ be positive integers with $n > 10m^3$. Suppose $\mathcal{U}, \mathcal{V}$ is a 3-restricted* **MV** *family of size K over $\mathbb{Z}_m^n$. Then $K \leq (2emn)^{m-1}$ if m is a prime power, and otherwise*

$$K \;\leq\; m^{\frac{n+2\log(n+1)}{2q} + 2m^2}\;,$$

where q is the second-smallest prime dividing m.

Proof. Let m' be the smallest factor of m such that $\mathcal{U}, \mathcal{V}$ is a 3-restricted **MV** family of size K over $\mathbb{Z}_{m'}^n$. If m' is a prime power, then by Theorem 21,

$$K \;\leq\; (2em'n)^{m'-1} \;\leq\; (2emn)^{m-1}\;.$$

Suppose m is not a prime power, then $K \;\leq\; (2emn)^{m-1} \;\leq\; m^{\frac{n+2\log(n+1)}{2q} + 2m^2}$; the second inequality follows, since $n > 10m^3$.

Now we consider the case when m' is not a prime power. Then, by Lemma 24, we have that $m' = p_1^s \cdot p_2^t$ for primes p_1 and p_2 and integers $s, t \geq 1$ with $p_1 < p_2$. Further, using Lemma 25, we have that $\alpha = 0 \pmod{p_1^s p_2^{t-1}}$, $\alpha \neq 0 \pmod{p_2^t}$, $\beta = 0 \pmod{p_1^{s-1}p_2^t}$, $\beta \neq 0 \pmod{p_1^s}$.

Let q be the *second-smallest* prime dividing m, and we assume that $K > m^{\frac{n+2\log(n+1)}{2q} + 2m^2}$, since otherwise we are already done. Also, note that by assumption, $p_2 \geq q$. Since $n \geq 10m^3$, and $2 \leq p_1p_2 \leq m$, this implies that

$$K \;\geq\; m^{7m^2} \;\geq\; m^2 \cdot (2m)^{4m^2} \;\geq\; 4p_2^2 \cdot (2p_1p_2)^{16p_2^2}\;.$$

By Lemma 29, we have without loss of generality (between $\boldsymbol{U}^\star$ and $\boldsymbol{V}^\star$) that

$$\begin{aligned}\mathbf{H}_\infty(\boldsymbol{U}^\star) &\leq \frac{n\log m'}{2} - \log\left(\left(\frac{1}{2p_1p_2}\right)^{16p_2^2} - \frac{2p_2^2}{K}\right)\\ &\leq \frac{n\log m}{2} - \log\left(\frac{1}{2}\cdot\left(\frac{1}{2p_1p_2}\right)^{16p_2^2}\right)\\ &\leq \frac{n\log m}{2} + 16p_2^2\cdot\log(2p_1p_2) + 1\\ &\leq \frac{n\log m}{2} + 4m^2\log m + 4m^2 + 1\ .\end{aligned}$$

Here we used that $2p_2 \leq p_1p_2 \leq m$. On the other hand, by Theorem 31, we know that

$$\mathbf{H}_\infty(\boldsymbol{U}^\star) \geq p_2\log\left(\frac{K}{p_2}\right) - \log(nt+1)\ .$$

Putting them together we have:

$$p_2\log\left(\frac{K}{p_2}\right) - \log(nt+1) \leq \frac{n\log m}{2} + 4m^2\log m + 4m^2 + 1\ .$$

This gives us that

$$\begin{aligned}K &\leq p_2\cdot m^{\frac{n}{2p_2}}\cdot m^{\frac{4m^2}{p_2}}\cdot 2^{\frac{4m^2+1}{p_2}}\cdot 2^{\frac{\log(nt+1)}{p_2}}\\ &\leq m^{\frac{n}{2p_2}}\cdot m^{\frac{4m^2}{p_2}+1}\cdot m^{\frac{1.9m^2}{p_2}}\cdot 2^{\frac{\log(nt+1)}{p_2}}\\ &\leq m^{\frac{n}{2p_2}}\cdot m^{\frac{6m^2}{p_2}}\cdot 2^{\frac{\log(nt+1)}{p_2}}\\ &\leq m^{\frac{n}{2p_2}}\cdot m^{2m^2}\cdot 2^{\frac{\log(nt+1)}{p_2}}\\ &\leq m^{\frac{n}{2q}}\cdot m^{2m^2}\cdot 2^{\frac{\log(nt+1)}{p_2}}\\ &= m^{\frac{n}{2q}}\cdot m^{2m^2}\cdot m^{\frac{\log(nt+1)}{p_2\cdot\log m}};\\ &\leq m^{\frac{n}{2q}}\cdot m^{2m^2}\cdot m^{\frac{\log(nt+1)}{p_2\cdot t\cdot\log p_2}};\\ &\leq m^{\frac{n}{2q}}\cdot m^{2m^2}\cdot m^{\frac{\log(nt+1)}{p_2\cdot t}}\\ &\leq m^{\frac{n}{2q}}\cdot m^{2m^2}\cdot m^{\frac{\log(nt+1)^{1/t}}{p_2}}\\ &\leq m^{\frac{n}{2q}}\cdot m^{2m^2}\cdot m^{\frac{\log(n+1)}{p_2}}\\ &\leq m^{\frac{n+2\log(n+1)}{2q}}\cdot m^{2m^2}\ .\end{aligned}$$

In the second inequality, we used $p_2 \leq m$, and also, $2^{4m^2+1} < m^{1.9m^2}$, which is true since $m \geq 6$. In the third inequality, we used the fact that $\frac{6m^2}{p_2} \geq \frac{5.9m^2}{p_2} + 1$, which is equivalent to showing $\frac{m^2}{10p_2} \geq 1$, which is true since $\frac{m^2}{10p_2} = \left(\frac{m}{2p_2}\right)\cdot\left(\frac{m}{5}\right) > 1$, because $m \geq 2p_2 \geq 2q \geq 6$. In the seventh inequality, we used that $m \geq p_2^t$. Finally, the second to last inequality follows from Bernoulli's inequality: $(nt+1) \leq (n+1)^t$.

This is a contradiction to our assumption that $K > m^{\frac{n+2\log(n+1)}{2q}+2m^2}$. This finishes the proof. ◀

▶ **Corollary 34.** *For an arbitrary positive integer m, consider a 3-query matching vector codes $C : \Sigma^K \to \Sigma^N$ constructed from 3-restricted matching vector families using the construction in Theorem 15, then $N \geq \Omega(K^{6-o_K(1)})$.*

The proof follows straightforwardly from the Theorem 15 combined with Theorem 33.

References

1 Noga Alon. The shannon capacity of a union. *Combinatorica*, 18(3):301–310, 1998. doi:10.1007/PL00009824.

2 Omar Alrabiah, Venkatesan Guruswami, Pravesh K Kothari, and Peter Manohar. A near-cubic lower bound for 3-query locally decodable codes from semirandom csp refutation. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 1438–1448, 2023. doi:10.1145/3564246.3585143.

3 Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM (JACM)*, 45(3):501–555, 1998. doi:10.1145/278298.278306.

4 Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: A new characterization of np. *Journal of the ACM (JACM)*, 45(1):70–122, 1998. doi:10.1145/273865.273901.

5 László Babai and Péter Frankl. Linear algebra methods in combinatorics. *to appear*, 2020.

6 Arnab Bhattacharyya, L. Sunil Chandran, and Suprovat Ghoshal. Combinatorial Lower Bounds for 3-Query LDCs. In 11^{th} *Innovations in Theoretical Computer Science Conference (ITCS 2020)*, volume 151, pages 85:1–85:8, 2020. doi:10.4230/LIPIcs.ITCS.2020.85.

7 Abhishek Bhowmick, Zeev Dvir, and Shachar Lovett. New bounds for matching vector families. *SIAM Journal on Computing*, 43(5):1654–1683, 2014. doi:10.1137/130932296.

8 Yeow Meng Chee, San Ling, Huaxiong Wang, and Liang Feng Zhang. Upper bounds on matching families in $\mathbb{Z}_{pq}^n$. *IEEE transactions on information theory*, 59(8):5131–5139, 2013.

9 Zeev Dvir, Parikshit Gopalan, and Sergey Yekhanin. Matching vector codes. *SIAM Journal on Computing*, 40(4):1154–1178, 2011. doi:10.1137/100804322.

10 Klim Efremenko. 3-query locally decodable codes of subexponential length. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 39–44, 2009. doi:10.1145/1536414.1536422.

11 P. Frankl and R. M. Wilson. Intersection theorems with geometric consequences. *Combinatorica*, 1(4):357–368, December 1981. doi:10.1007/bf02579457.

12 Oded Goldreich, Howard Karloff, Leonard J Schulman, and Luca Trevisan. Lower bounds for linear locally decodable codes and private information retrieval. *computational complexity*, 15:263–296, 2006. doi:10.1007/S00037-006-0216-3.

13 P. Gopalan. Constructing ramsey graphs from boolean function representations. In *21st Annual IEEE Conference on Computational Complexity (CCC'06)*, pages 14 pp.–128, 2006. doi:10.1109/CCC.2006.14.

14 Sivakanth Gopi. Lecture notes: Polynomial representing or mod m. https://homes.cs.washington.edu/~anuprao/pubs/codingtheory/lecture12.pdf, 2019.

15 Sivakanth Gopi. Upper bound for matching vector families for prime-powers. *Personal Communication*, 2024.

16 W. T. Gowers, Ben Green, Freddie Manners, and Terence Tao. Marton's conjecture in abelian groups with bounded torsion, 2024. arXiv:2404.02244.

17 Vince Grolmusz. On the weak mod m representation of boolean functions. *Chicago Journal of Theoretical Computer Science*, 1995.

18 Toshiya Itoh and Yasuhiro Suzuki. Improved constructions for query-efficient locally decodable codes of subexponential length. *IEICE Transactions on Information and Systems*, 93(2):263–270, 2010. doi:10.1587/TRANSINF.E93.D.263.

19 Jonathan Katz and Luca Trevisan. On the efficiency of local decoding procedures for error-correcting codes. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 80–86, 2000. doi:10.1145/335305.335315.

20 Iordanis Kerenidis and Ronald de Wolf. Exponential lower bound for 2-query locally decodable codes via a quantum argument. *Journal of Computer and System Sciences*, 3(69):395–420, 2004. doi:10.1016/J.JCSS.2004.04.007.

21 Anup Rao. An exposition of bourgain's 2-source extractor. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 14. Citeseer, 2007.

22 David P Woodruff. New lower bounds for general locally decodable codes. In *Electronic Colloquium on Computational Complexity (ECCC)*, 2007.

23 David P Woodruff. A quadratic lower bound for three-query linear locally decodable codes over any field. *Journal of Computer Science and Technology*, 27(4):678–686, 2012. doi:10.1007/S11390-012-1254-8.

24 Sergey Yekhanin. Towards 3-query locally decodable codes of subexponential length. *Journal of the ACM (JACM)*, 55(1):1–16, 2008. doi:10.1145/1326554.1326555.

25 Sergey Yekhanin. Locally decodable codes. *Foundations and Trends® in Theoretical Computer Science*, 6(3):139–255, 2012. doi:10.1561/0400000030.

A Proof of Lemma 11

We include the proof of completeness.

Proof.

$$\begin{aligned}|\mathbb{E}[\omega^{\langle X,Y\rangle}]| &= \left|\sum_{y\in\mathbb{Z}_m^n} Y(y)\sum_{x\in\mathbb{Z}_m^n} X(x)\omega^{\langle x,y\rangle}\right| \\ &\leq \sum_{y\in\mathbb{Z}_m^n} Y(y)\left|\sum_{x\in\mathbb{Z}_m^n} X(x)\omega^{\langle x,y\rangle}\right| .\end{aligned}$$

On the other hand, since the square function is convex, by Jensen's inequality, one have

$$\begin{aligned}|\mathbb{E}[\omega^{\langle X,Y\rangle}]|^2 &\leq \sum_{y\in\mathbb{Z}_m^n} Y(y)\left|\sum_{x\in\mathbb{Z}_m^n} X(x)\omega^{\langle x,y\rangle}\right|^2 \\ &= \left|\sum_{y\in\mathbb{Z}_m^n} Y(y)\sum_{x_1,x_2\in\mathbb{Z}_m^n} X(x_1)X(x_2)\omega^{\langle x_1,y\rangle}\omega^{-\langle x_2,y\rangle}\right| \\ &= \left|\sum_{y\in\mathbb{Z}_m^n} Y(y)\sum_{x_1,x_2\in\mathbb{Z}_m^n} X(x_1)X(x_2)\omega^{\langle x_1-x_2,y\rangle}\right| \\ &= \left|\mathbb{E}[\omega^{\langle X_1-X_2,Y\rangle}]\right| .\end{aligned}$$ ◀