Electronic Colloquium on Computational Complexity, Revision 2 of Report No. 141 (2026)

Subexponential Upper Bounds for 3-Restricted Matching Vector Families

Divesh Aggarwal* Maciej Obremski†

August 11, 2026

Abstract

A Matching Vector (**MV**) family modulo a positive integer $m \geq 2$ is a pair of ordered lists $U = (u_1, \ldots, u_K)$ and $V = (v_1, \ldots, v_K)$ with $u_i, v_j \in \mathbb{Z}_m^n$ such that $\langle u_i, v_i \rangle = 0 \pmod m$ for every $i \in [K]$, while $\langle u_i, v_j \rangle \neq 0 \pmod m$ for every $i \neq j$. It is called r-restricted if the set of all inner products $\langle u_i, v_j \rangle$ has size at most r, including the diagonal value 0. Restricted matching vector families are central to the matching-vector construction of constant-query subexponential locally decodable codes: an r-restricted family of size K in $\mathbb{Z}_m^n$ gives an r-query matching-vector code with message length K and codeword length $N = m^n$.

The best previous bound tailored to the 3-restricted case was $MV(m, n, 3) \leq \exp\big(O_m(n/\log n)\big)$ for fixed m, due to Bhowmick, Dvir and Lovett [BDL14], who showed this bound under the polynomial Freiman Ruzsa conjecture, which was proved by Gowers, Green, Manners, and Tao [GGMT24]. In this work we prove the subexponential upper bound

$$MV(m, n, 3) \leq \exp\big(O_m(\sqrt{n \log n})\big)$$

for every fixed modulus m. Consequently, every 3-query matching-vector code over a fixed modulus has codeword length

$$N \geq \exp\left(\Omega_m\left(\frac{(\log K)^2}{\log \log K}\right)\right).$$

Our proof answers an explicit technical question left by the work of Aggarwal, Dutta, Li, Obremski, and Saraogi [ADL+25]: the entropy-growth method for sums of matching vectors can be pushed far beyond the initial constant-sum regime. We show that, for 3-restricted families, entropy continues to grow for sums of length L as large as $\Theta_m(\sqrt{n/\log n})$. The new ingredient is an exact-count rank bootstrap: a long-sum collision changes the exact number of one residue class seen from a suitable pivot, and exact-weight polynomials convert a disjoint collision family into a low-rank identity matrix.

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 can be recovered with probability at least $1 - \varepsilon$ by a randomized decoding procedure that reads only r positions of the codeword, even when up to δN codeword

*National University of Singapore. dcsdiva@nus.edu.sg.
†National University of Singapore. obremski.math@gmail.com.

ISSN 1433-8092

positions have been corrupted. Locally decodable codes were formally introduced by Katz and Trevisan [KT00], and had already appeared implicitly in the theory surrounding probabilistically checkable proofs [ALM+98, AS98]. They have since played a role in private information retrieval, worst-case to average-case reductions, derandomization, data structures, fault-tolerant computation, and related areas; see Yekhanin's survey [Yek12].

A central question is how large the message length K can be as a function of the block length N when the number of queries r is fixed. For $r = 2$, the Hadamard code gives $K = \log_2 N$, and this is optimal up to a constant factor by the lower bounds of Kerenidis and de Wolf [KdW04] and Goldreich, Karloff, Schulman and Trevisan [GKST06]. For $r \geq 3$, the landscape is much less settled. The strongest known general lower-bound techniques still leave a large gap from the best constructions; for example, a line of work culminating in [Woo07, Woo12, BCG20] gives polynomial lower bounds on the codeword length for constant-query LDCs, and Alrabiah, Guruswami, Kothari and Manohar proved the near-cubic lower bound $K = O(N^{1/3} \log^2 N)$ for 3 queries [AGKM23]. On the construction side, the only known constant-query subexponential LDCs are matching-vector codes, beginning with Yekhanin and Efremenko and developed further by Dvir, Gopalan and Yekhanin [Yek08, Efr09, DGY11] that achieve $K = (\log N)^{\Omega(\log \log N)}$.

1.2 Matching vector families and matching-vector codes

Matching vector families are combinatorial-algebraic objects that also arise in Ramsey-type constructions and weak representations of Boolean functions [FW81, Gro95, Alo98, Gop06]. Their most prominent role in this setting is as the algebraic core of matching-vector locally decodable codes [DGY11, Yek12].

Definition 1.1 (Matching vector family)**.** *Let $S \subseteq \mathbb{Z}_m \setminus \{0\}$. An S-matching vector family in $\mathbb{Z}_m^n$ of size K is a pair of ordered lists*

$$U = (u_1, \ldots, u_K), \qquad V = (v_1, \ldots, v_K), \qquad u_i, v_i \in \mathbb{Z}_m^n,$$

such that

$$\langle u_i, v_i \rangle = 0 \pmod{m} \quad \text{for all } i \in [K],$$

and

$$\langle u_i, v_j \rangle \in S \quad \text{for all } i \neq j.$$

An S-matching vector family is often called $(|S|+1)$-restricted. More generally, an ordered pair (U, V) is r-restricted if all inner products $\langle u_i, v_j \rangle$, including the diagonal value 0, take at most r different values. We write $MV(m, n, r)$ for the largest size of an r-restricted matching vector family in $\mathbb{Z}_m^n$, and $MV(m, n)$ when no restriction on the number of off-diagonal residues is imposed. The standard matching-vector construction transforms an r-restricted family of size K in $\mathbb{Z}_m^n$ into an r-query locally decodable code with message length K and codeword length $N = m^n$ [DGY11, Yek12]. Thus upper bounds on $MV(m, n, r)$ give lower bounds for every LDC obtained by the matching-vector route.

1.3 Upper bounds for matching vector families

A natural question for matching vector families asks how large K can be in terms of m, n, and r. Dvir, Gopalan and Yekhanin proved the essentially tight upper bound $MV(m, n) \leq O(m^{n/2})$ when m is prime, and also gave a general upper bound of $O(m^{n-1+o_m(1)})$ for arbitrary moduli

[DGY11]. Bhowmick, Dvir and Lovett developed the first general method for composite moduli, showing that

$$MV(m,n,r) \le r^{O(r\log r)} m^{n/2}$$

for restricted families [BDL14]. They also proved that the polynomial Freiman-Ruzsa theorem would imply, for fixed m, the unrestricted bound $MV(m,n) \le m^{O_m(n/\log n)}$; this implication is now unconditional by the bounded-torsion form of Marton's conjecture proved by Gowers, Green, Manners and Tao [GGMT24].

The case $r = 3$ has received special attention because it corresponds to 3-query matching-vector codes. Alrabiah, Guruswami, Kothari and Manohar's near-cubic LDC lower bound implies an $O(m^{n/3})$ upper bound for 3-restricted matching vector families [AGKM23]. Aggarwal, Dutta, Li, Obremski and Saraogi then proved the sharper fixed-modulus estimate

$$MV(m,n,3) \le m^{n/6+O_m(\log n)},$$

combining an entropy argument for moduli with at least two distinct prime factors with the polynomial prime-power bound due to Gopi [ADL+25, Gop24]. Their proof studies sums of a small number of independently sampled matching vectors and shows that the entropy of these sums grows enough to beat the $m^{n/2}$ barrier. They asked whether this entropy-growth approach could be pushed to longer sums.

1.4 Our result

We answer this question affirmatively for 3-restricted matching vector families. Our main theorem replaces the previous fixed-modulus exponential bound in n by a subexponential bound.

Theorem 1.2 (Main theorem)**.** *For every fixed integer $m \ge 2$, there is a constant C_m such that*

$$MV(m,n,3) \le \exp\bigl(C_m\sqrt{n\log n}\bigr)$$

for all $n \ge 2$.

For fixed m, this is asymptotically stronger than the general fixed-modulus bound $MV(m,n) \le \exp(O_m(n/\log n))$ obtained by combining the Bhowmick–Dvir–Lovett framework with the bounded-torsion form of Marton's conjecture [BDL14, GGMT24]. Thus the 3-restricted structure permits substantially more entropy growth than is known for unrestricted matching vector families.

Since a 3-restricted matching vector family gives a 3-query matching-vector locally decodable code of codeword length $N = m^n$, the theorem implies the following lower bound for matching-vector codes.

Corollary 1.3. *For every fixed m, every 3-query matching-vector code over modulus m with message length K and codeword length $N = m^n$ satisfies, for sufficiently large K,*

$$N \ge \exp\left(\Omega_m\left(\frac{(\log K)^2}{\log\log K}\right)\right).$$

The proof keeps the entropy-vs-support skeleton of Aggarwal et al. [ADL+25], but changes the collision analysis. Instead of proving entropy growth only for a constant sum length determined by the modulus, we control collisions for sums of length $L \asymp_m \sqrt{n/\log n}$. The key idea is to look at how often each residue appears in a collision. If one residue appears a different number of times, we can detect this difference using a low-rank test.

The next subsection gives the proof overview.

1.5 Proof overview

The proof is based on a simple entropy tension. We replace a single matching vector by a sum $U_1 + \cdots + U_L$ of many matching vectors. In the ideal collision-free case, this sum would have entropy close to $L \log K$. More generally, its entropy is $L \log K$ diminished by the frequency with which different L-tuples give the same sum. If we can show that this entropy grows almost linearly in L, then comparing it with the ambient support bound gives the desired upper bound: every such sum still lies in $\mathbb{Z}_m^n$, so its entropy can never exceed

$$\log |\mathbb{Z}_m^n| = n \log m.$$

Thus $L \log K \lesssim n \log m$, up to the collision losses. This is the basic mechanism behind the subexponential bound. In the actual argument the U_i's are sampled without replacement; this keeps the collision-counting step cleaner.

We now explain how the collision losses are controlled. If the family has only one nonzero off-diagonal residue, then after reducing modulo a prime-power divisor on which this residue remains nonzero, Lemma 2.2 gives the required polynomial bound. Hence, after Lemma 2.3, the only non-prime-power case requiring further work is the genuine two-residue case described there.

The cleanest case is $m = 6$, and in this exposition we take the two off-diagonal residues to be

$$\alpha = 2, \qquad \beta = 3\ .$$

This setting captures the difficulty of the problem.

Thus modulo 2, the values 0 and α collapse together, while β remains visible.[1] In the case $m = 6$, this indicator is simply parity:

$$\langle u_j, v_i \rangle \bmod 2 = \mathbf{1}_{\langle u_j, v_i \rangle = 3}.$$

For the entropy lower bound we use sums without replacement:

$$S_L^*(U) = \sum_{i \in A} u_i, \qquad \text{where } A \sim \binom{[K]}{L} \text{ is sampled uniformly.}$$

We need to count collisions among these sums in order to evaluate the entropy correctly. Write

$$M_L(U) = \max_x \# \left\{ A \in \binom{[K]}{L} : \sum_{i \in A} u_i = x \right\}$$

for the largest L-sum fiber, meaning the largest preimage of a single value under the map $A \mapsto \sum_{i \in A} u_i$. This is the standard additive combinatorics terminology. Since $S_L^*(U)$ is obtained by choosing A uniformly from $\binom{[K]}{L}$, we have

$$H_\infty(S_L^*(U)) = \log \binom{K}{L} - \log M_L(U).$$

Thus $M_L(U)$ measures exactly the entropy loss caused by collisions.

[1] This is a general phenomenon in the minimal-modulus case. After projecting to a suitable prime-power component, 0 and one of the two off-diagonal residues vanish together, while the other off-diagonal residue remains nonzero. Lemma 4.1 is the bounded-rank version of this observation.

The base case $L = 3$ is the fiber bound proved in [ADL+25]; in the notation above, it gives $M_3(U) \le n + 1$. We prove the recursive bound

$$M_L(U) \le L(1 + n + \cdots + n^L) M_{L-1}(U), \qquad L \ge 4. \tag{1}$$

We now build a hypergraph from a fixed fiber. Fix a value x, and let

$$\mathcal{F} = \left\{ A \in \binom{[K]}{L} : \sum_{i \in A} u_i = x \right\}$$

be the corresponding L-sum fiber. We view $\mathcal{F}$ as a hypergraph with vertex set $[K]$. Its hyperedges are the sets $A \in \mathcal{F}$, and each such edge has size L, so the hypergraph is L-uniform.

We first bound the maximum degree of this hypergraph. Fix a vertex $i \in [K]$. Every edge $A \in \mathcal{F}$ containing i gives an $(L-1)$-set $A \setminus \{i\}$ satisfying

$$\sum_{j \in A \setminus \{i\}} u_j = x - u_i.$$

Thus the number of edges containing i is at most the largest $(L-1)$-sum fiber, namely $M_{L-1}(U)$. Hence

$$\Delta(\mathcal{F}) \le M_{L-1}(U),$$

where $\Delta(\mathcal{F})$ denotes the maximum vertex degree.

The other parameter we need is the matching number $\nu(\mathcal{F})$, i.e. the largest number of pairwise-disjoint edges in $\mathcal{F}$. Here is why the degree and matching number together control the whole hypergraph. Take a maximal matching $\mathcal{M}_0$. Its size is at most $\nu(\mathcal{F})$, and the union of its edges contains $L\nu(\mathcal{F})$ vertices. By maximality, every edge of $\mathcal{F}$ must intersect this union; otherwise we could add that edge to the matching. Now count all edges of $\mathcal{F}$ by choosing one vertex of intersection with this union. Each vertex lies in at most $\Delta(\mathcal{F})$ edges, so

$$|\mathcal{F}| \le L\nu(\mathcal{F})\Delta(\mathcal{F}). \tag{2}$$

Together with $\Delta(\mathcal{F}) \le M_{L-1}(U)$, this shows that it remains to prove

$$\nu(\mathcal{F}) \le 1 + n + \cdots + n^L.$$

Let $\mathcal{M}$ be a matching in $\mathcal{F}$. For every $A \in \mathcal{M}$ choose a pivot $i(A) \in A$ and let

$$r_A = \#\{j \in A : \langle u_j, v_{i(A)} \rangle = 3\}.$$

If $B \ne A$ is another hyperedge in the matching, then both hyperedges belong to $\mathcal{F}$, so they have the same sum:

$$\sum_{j \in A} u_j = \sum_{j \in B} u_j$$

This equality forces the number of 3-edges from B into $v_{i(A)}$ to be different from r_A. To see this, write

$$r_B = \#\{j \in B : \langle u_j, v_{i(A)} \rangle = 3\}.$$

Taking inner product with $v_{i(A)}$ gives

$$\langle \sum_{j \in A} u_j, v_{i(A)} \rangle = \langle \sum_{j \in B} u_j, v_{i(A)} \rangle \pmod 6.$$

On the A-side, the pivot $i(A)$ contributes the diagonal value 0, and the other $L-1$ terms contribute either 2 or 3. Thus the A-side is

$$2(L-1-r_A)+3r_A = 2L-2+r_A \pmod 6.$$

On the B-side all L terms are off-diagonal, so the contribution is

$$2(L-r_B)+3r_B = 2L+r_B \pmod 6.$$

The two contributions must be equal modulo 6, hence

$$r_B \equiv r_A - 2 \pmod 6.$$

In particular $r_B \neq r_A$. Now, for each hyperedge $C \in \mathcal{M}$, look at the Boolean vector

$$z^{A,C} = \big(\langle u_j, v_{i(A)}\rangle \bmod 2\big)_{j\in C}.$$

Its entries record exactly which elements of C have inner product 3 with the pivot $v_{i(A)}$. For $C = A$, this vector has exactly r_A ones. For every other $C = B$, the calculation above shows that it has a different number of ones. The following rank argument is inspired by the matrix/polynomial method used in the prime-power matching-vector bound of [ADL+25, Theorem 2.16].

Over $\mathbb{F}_2$, for each $r \in \{0,\dots,L\}$, the test that a Boolean vector has exactly r ones can be represented by a polynomial P_r of degree at most L. Indeed, any Boolean function on L bits has a multilinear polynomial representation of degree at most L. We use these polynomials to build a matrix that compares every hyperedge C against the reference count r_A determined by the row hyperedge A. Formally, define a matrix R, with rows and columns indexed by $\mathcal{M}$, by

$$R_{A,C} = P_{r_A}(z^{A,C}).$$

The discussion above says that $R_{A,A} = 1$, while $R_{A,C} = 0$ for $C \neq A$. Thus R is the identity matrix on $\mathcal{M}$, whose rank is $|\mathcal{M}|$.

It remains to explain why the same matrix also has small rank. We may take the polynomial P_r to be symmetric in its L Boolean inputs. Thus its degree-d part is a scalar multiple of the elementary symmetric polynomial

$$e_d(z) = \sum_{\substack{S\subseteq C\\ |S|=d}} \prod_{j\in S} z_j.$$

In our matrix, $z_j = \langle u_j, v_{i(A)}\rangle \bmod 2$, so a typical degree-$d$ term has the form

$$\prod_{j\in S} \langle u_j, v_{i(A)}\rangle.$$

Although there are many possible subsets S, all terms in a fixed row use the same right-hand vector $v_{i(A)}$. Therefore they fold into one tensor inner product:

$$\prod_{j\in S} \langle u_j, v_{i(A)}\rangle = \langle \bigotimes_{j\in S} u_j, v_{i(A)}^{\otimes d}\rangle.$$

After summing over all d-subsets $S \subseteq C$, the degree-d part of $R_{A,C}$ has the form

$$\langle W_{C,d}, v_{i(A)}^{\otimes d}\rangle,$$

where $W_{C,d}$ is a vector in the d-fold tensor space depending only on the column hyperedge C. We also set $W_{C,0} = 1$. Choose P_{r_A} in its symmetric multilinear form, and let $\alpha_{A,d}$ denote the coefficient of the degree-d part in row A. This coefficient depends only on A, through the value r_A, since P_{r_A} is the symmetric polynomial testing whether the number of ones is r_A.

Therefore each entry of R can be written as

$$R_{A,C} = \sum_{d=0}^{L} \langle W_{C,d}, \alpha_{A,d} v_{i(A)}^{\otimes d} \rangle.$$

Equivalently, if we concatenate all degrees, then

$$R_{A,C} = \langle W_{C,0} \oplus W_{C,1} \oplus \cdots \oplus W_{C,L}, \alpha_{A,0} \oplus \alpha_{A,1} v_{i(A)} \oplus \cdots \oplus \alpha_{A,L} v_{i(A)}^{\otimes L} \rangle. \tag{3}$$

Now we can split the matrix R into two pieces. Let the shared coordinate space be

$$\mathbb{F}_2 \oplus \mathbb{F}_2^n \oplus (\mathbb{F}_2^n)^{\otimes 2} \oplus \cdots \oplus (\mathbb{F}_2^n)^{\otimes L}.$$

Define M_v to be the matrix whose row indexed by $A \in \mathcal{M}$ is

$$\alpha_{A,0} \oplus \alpha_{A,1} v_{i(A)} \oplus \alpha_{A,2} v_{i(A)}^{\otimes 2} \oplus \cdots \oplus \alpha_{A,L} v_{i(A)}^{\otimes L},$$

and define M_W to be the matrix whose column indexed by $C \in \mathcal{M}$ is

$$W_{C,0} \oplus W_{C,1} \oplus W_{C,2} \oplus \cdots \oplus W_{C,L},$$

written as a column vector in the same direct sum. By (3), the (A, C)-entry of $M_v M_W$ is exactly $R_{A,C}$. Hence

$$R = M_v M_W.$$

The shared dimension of this product is at most

$$1 + n + \cdots + n^L.$$

Therefore $\operatorname{rank}(R) \le 1 + n + \cdots + n^L$. Since R is the identity matrix on $\mathcal{M}$, we get

$$|\mathcal{M}| \le 1 + n + \cdots + n^L.$$

Since $\mathcal{M}$ was an arbitrary matching in $\mathcal{F}$, this bounds $\nu(\mathcal{F})$. Plugging this into the hypergraph counting bound (2), together with $\Delta(\mathcal{F}) \le M_{L-1}(U)$, gives the recurrence (1). Iterating the recurrence from the base case $M_3(U) \le n + 1$, we get

$$M_L(U) \le (n+1) \prod_{\ell=4}^{L} \ell (1 + n + \cdots + n^{\ell}).$$

Since $1 + n + \cdots + n^{\ell} \le (\ell + 1) n^{\ell}$, taking logarithms gives

$$\log M_L(U) \le O\left(\sum_{\ell=4}^{L} \ell \log n + \sum_{\ell=4}^{L} \log \ell \right) = O(L^2 \log n + L \log L).$$

Thus

$$M_L(U) \le \exp(O(L^2 \log n + L \log L)).$$

Consequently

$$H_\infty(S_L^*(U)) \geq L \log K - O(L^2 \log n + L \log L).$$

The upper bound on the entropy is now immediate: $S_L^*(U)$ is supported on $\mathbb{Z}_6^n$, so $H_\infty(S_L^*(U)) \leq n \log 6$. Combining this with the lower bound above gives

$$L \log K - O(L^2 \log n + L \log L) \leq n \log 6.$$

Rearranging,

$$L \log K \leq n \log 6 + O(L^2 \log n + L \log L).$$

Dividing by L, we get

$$\log K \leq O\left(\frac{n}{L} + L \log n + \log L\right).$$

The first term decreases with L, while the second increases with L. We balance them by choosing L on the order of $\sqrt{n/\log n}$. Then

$$\frac{n}{L} = O(\sqrt{n \log n}), \qquad L \log n = O(\sqrt{n \log n}),$$

and $\log L$ is smaller. Hence

$$\log K \leq O(\sqrt{n \log n}),$$

which proves Theorem 1.2 for $m = 6$.

For general fixed m, the formal proof splits into a few cases. If m is a prime power, then the known polynomial bound already finishes the proof. Otherwise, the 3-restricted problem reduces to the case where m is a product of two prime powers, say $p^a q^b$; this is Lemma 2.3. In that case, after relabeling the two nonzero residues, one residue is zero modulo p^a while the other is not; symmetrically, the second residue is zero modulo q^b while the first is not. Thus the mod-2 indicator used for $m = 6$ is replaced by another bounded-degree function for one of the two residues, built using Lucas' theorem and digit polynomials modulo p. Once this is in place, the hypergraph and rank argument above goes through in the same way.

2 Preliminaries

2.1 Entropy

For a finitely supported random variable X, its min-entropy is

$$H_\infty(X) = -\log \max_x \mathbb{P}[X = x].$$

We will use only the elementary support bound

$$H_\infty(X) \leq \log |\operatorname{supp}(X)|.$$

In particular, if X takes values in $\mathbb{Z}_m^n$, then

$$H_\infty(X) \leq n \log m.$$

2.2 Prime-power and 2-restricted bounds

We shall use two standard rank bounds from the prior work. The first is the polynomial bound for prime-power moduli.[2]

Lemma 2.1 (Prime-power bound [ADL+25, Theorem 2.16])**.** *Let $m = p^a$ be a prime power. Then every matching vector family over $\mathbb{Z}_m^n$ has size at most*

$$(2emn)^{m-1}.$$

In particular, for fixed m one has $MV(m, n, 3) \le n^{O_m(1)}$.

The second is the elementary bound for 2-restricted families over a prime power.

Lemma 2.2 (2-restricted prime-power bound [ADL+25])**.** *Let $\gamma \neq 0 \pmod{p^a}$. If $(u_i, v_i)_{i=1}^T$ is a matching vector family over $\mathbb{Z}_{p^a}^n$ with all off-diagonal inner products equal to γ, then*

$$T \le an + 1.$$

2.3 Minimal-modulus reduction

The following standard reduction is due to Bhowmick, Dvir and Lovett and was used explicitly by Aggarwal et al.

Lemma 2.3 (Minimal modulus for 3-restricted families [ADL+25, Lemmas 2.19 and 2.20])**.** *Let (U, V) be a 3-restricted matching vector family modulo m, with off-diagonal residues α, β. Suppose m is minimal among divisors of the original modulus for which the same lists remain a matching vector family. Then either m is a prime power, or*

$$m = p^a q^b$$

for two distinct primes $p < q$ and positive integers a, b. In the latter case, after possibly interchanging α and β,

$$\alpha \equiv 0 \pmod{p^a q^{b-1}}, \qquad \alpha \not\equiv 0 \pmod{q^b},$$

and

$$\beta \equiv 0 \pmod{p^{a-1} q^b}, \qquad \beta \not\equiv 0 \pmod{p^a}.$$

Thus, after passing to a minimal divisor, every non-prime-power case that is not already covered by the prime-power or 2-restricted bounds has exactly the two-residue structure described above.

3 Proof setup and induction scheme

This section sets up the non-prime-power case after the minimal-modulus reduction. We work with a modulus

$$m = p^a q^b, \qquad p < q,$$

as in Lemma 2.3.

[2] This lemma is due to Gopi [Gop24].

The proof compares two entropy estimates for the same random sum. The upper bound is the trivial support bound

$$H_\infty(S_L^*(U)) \le n \log m,$$

because $S_L^*(U)$ takes values in $\mathbb{Z}_m^n$. The lower bound is the main work: we prove that $S_L^*(U)$ has high min-entropy because large collisions among L-sums are rare. The gap between these two estimates gives the final upper bound on the family size.

For the entropy lower bound, we track collisions among sums without replacement. For an integer $L \ge 1$ and a set $A \in \binom{[K]}{L}$, write

$$\Sigma_L(A) = \sum_{i \in A} u_i.$$

The largest L-sum fiber is

$$M_L(U) = \max_x \#\{A \in \binom{[K]}{L} : \Sigma_L(A) = x\}.$$

Thus $M_L(U)$ is the largest number of L-subsets that produce the same sum. Now choose an L-subset $A \subseteq [K]$ uniformly at random and set

$$S_L^*(U) = \Sigma_L(A) = \sum_{i \in A} u_i.$$

Then the largest atom of $S_L^*(U)$ has probability at most $M_L(U)/\binom{K}{L}$, so bounding fibers is exactly what gives entropy growth.

The induction proves that these fibers stay small. The base case is a short-sum statement inherited from the previous 3-query argument [ADL+25]. As in the overview, the inductive step fixes one L-sum fiber $\mathcal{F}$, views it as an L-uniform hypergraph on $[K]$, and bounds its size through

$$|\mathcal{F}| \le L\, \nu(\mathcal{F})\, \Delta(\mathcal{F}),$$

where $\nu(\mathcal{F})$ is the matching number and $\Delta(\mathcal{F})$ is the maximum vertex degree. The degree is controlled by $M_{L-1}(U)$. The main organizing step is the matching-number bound. As in the overview, disjoint hyperedges in one fiber are separated by the exact number of β-residues seen from a pivot. The residue-indicator section below turns this exact-count test into a low-rank matrix, and combining this with the hypergraph count gives the recurrence that drives the whole proof.

4 A residue indicator for composite moduli

The next step is to replace the simple mod-2 test used in the $m = 6$ example. There, among the possible inner-product values $0, \alpha, \beta$, the residue β is the one that is nonzero modulo 2, while 0 and α vanish modulo 2. Thus the indicator

$$\Delta(x, y) := \langle x, y \rangle \bmod 2$$

indicates exactly the β-entries. Moreover,

$$\langle x, y \rangle \bmod 2 = \sum_{t=1}^{n} x_t y_t \pmod 2,$$

so this indicator is linear in the coordinate products $x_t y_t$ (equivalently, bilinear in x and y). In particular, it can be written as an inner product

$$\Delta(x, y) = \langle \Phi(x), \Psi(y) \rangle_{\mathbb{F}_2}$$

where $\Phi(x)$ and $\Psi(y)$ are the coordinatewise reductions of x and y modulo 2. The general indicator below is the analogue of this statement for $m = p^a q^b$: it should still identify exactly the β-residue, and it should still have a short inner-product representation.

For a general modulus $m = p^a q^b$, the indicator is only slightly more subtle. We use Lucas' theorem to identify the relevant base-p digit of the inner product. To obtain the short inner-product representation, we first analyze the resulting polynomial over $\mathbb{Q}$, and only then reduce modulo p. This avoids doing the division step directly inside $\mathbb{F}_p$, where it may not be valid.

Lemma 4.1 (Residue indicator)**.** *Let $m = p^a q^b$ and let α, β satisfy Lemma 2.3. There is a prime field $\mathbb{F}_p$ and a function*

$$\Delta : \mathbb{Z}_m^n \times \mathbb{Z}_m^n \to \mathbb{F}_p$$

such that, whenever $\langle x, y \rangle \in \{0, \alpha, \beta\}$,

$$\Delta(x, y) = \begin{cases} 1, & \langle x, y \rangle = \beta, \\ 0, & \langle x, y \rangle \in \{0, \alpha\}. \end{cases}$$

Moreover, Δ has a short inner-product representation: for some integer $R \le n^{d_m}$, with $d_m = O_m(1)$, there are maps

$$\Phi, \Psi : \mathbb{Z}_m^n \to \mathbb{F}_p^R$$

such that

$$\Delta(x, y) = \langle \Phi(x), \Psi(y) \rangle_{\mathbb{F}_p}.$$

Consequently, products of indicator values also have short inner-product representations. Here $x^{(\ell)}, y^{(\ell)} \in \mathbb{Z}_m^n$ denote the ℓ-th pair of inputs to the indicator. For every $h \ge 1$,

$$\prod_{\ell=1}^{h} \Delta(x^{(\ell)}, y^{(\ell)}) = \left\langle \bigotimes_{\ell=1}^{h} \Phi(x^{(\ell)}), \bigotimes_{\ell=1}^{h} \Psi(y^{(\ell)}) \right\rangle_{\mathbb{F}_p}$$

in a space of dimension $R^h \le n^{O_m(h)}$. In particular, every matrix whose (i, j) entry is

$$\prod_{\ell=1}^{h} \Delta(x_i^{(\ell)}, y_j^{(\ell)})$$

has rank at most $n^{O_m(h)}$ over $\mathbb{F}_p$.

Proof. We split the proof into two parts. First we construct the indicator. Then we show that it has a short inner-product representation.

Constructing the indicator. Project all inner products modulo p^a. By Lemma 2.3, both 0 and α vanish modulo p^a, while

$$\beta \equiv p^{a-1} u \pmod{p^a}$$

for some unit $u \in \mathbb{F}_p^\times$. Therefore it is enough to distinguish the residue $p^{a-1}u$ from the residue 0 modulo p^a.

Set

$$k := p^{a-1}.$$

For an integer T, Lucas' theorem says that

$$\binom{T}{k} \pmod{p}$$

is the $(a-1)$-st base-p digit of T. Hence

$$\binom{T}{k} \equiv 0 \pmod{p} \qquad \text{if } T \equiv 0 \pmod{p^a},$$

whereas

$$\binom{T}{k} \equiv u \pmod{p} \qquad \text{if } T \equiv p^{a-1}u \pmod{p^a}.$$

Consider the expression $\binom{\langle x,y\rangle}{k}$. In this expression, we evaluate $\langle x, y\rangle$ using the canonical representatives of the coordinates in $\{0, \ldots, m-1\}$; by Lucas' theorem, the resulting value modulo p depends only on $\langle x, y\rangle \bmod p^a$. We define

$$\Delta(x, y) := u^{-1}\binom{\langle x, y\rangle}{k} \pmod{p}. \tag{6}$$

This gives exactly the required values on the three residues $0, \alpha, \beta$.

Finding a short inner-product representation. We now prove that Δ can be written as an inner product in dimension $n^{O_m(1)}$.

First ignore the reduction modulo p. Represent each coordinate of $x, y \in \mathbb{Z}_m^n$ by an integer in $\{0, \ldots, m-1\}$, and define the integer-valued matrix

$$D_{x,y} := \binom{\langle x, y\rangle}{k}.$$

This matrix has rows indexed by $x \in \mathbb{Z}_m^n$ and columns indexed by $y \in \mathbb{Z}_m^n$. We claim that D has rank at most $n^{O_m(1)}$ over $\mathbb{Q}$.

Indeed, over $\mathbb{Q}$, the binomial coefficient is a degree-k polynomial:

$$\binom{z}{k} = \frac{z(z-1)\cdots(z-k+1)}{k!} = \frac{1}{k!}\sum_{r=0}^{k} s(k, r)z^r, \tag{4}$$

where $s(k, r)$ are the signed Stirling numbers of the first kind. Substituting $z = \langle x, y\rangle$, it is enough to bound the rank of the matrix with entries $\langle x, y\rangle^r$, for each $0 \le r \le k$. For fixed r, use the tensor identity

$$\langle x, y\rangle^r = \langle x^{\otimes r}, y^{\otimes r}\rangle. \tag{7}$$

The vectors $x^{\otimes r}$ and $y^{\otimes r}$ live in a space of dimension n^r, with the usual convention that the 0-fold tensor space is one-dimensional. Therefore the matrix $(\langle x, y\rangle^r)_{x,y}$ has rank at most n^r over $\mathbb{Q}$. Summing over $0 \le r \le k$, and using $k = p^{a-1} = O_m(1)$, gives

$$\operatorname{rank}_{\mathbb{Q}}(D) \le \sum_{r=0}^{k} n^r = n^{O_m(1)}. \tag{8}$$

Now reduce the entries of D modulo p. Let $\overline{D}$ be the resulting matrix over $\mathbb{F}_p$. Its rank over $\mathbb{F}_p$ cannot be larger than the rank of D over $\mathbb{Q}$. To see this, any minor of $\overline{D}$ that is nonzero over $\mathbb{F}_p$ comes from an integer determinant of D that is not divisible by p, and hence is nonzero over $\mathbb{Q}$. Thus every nonzero minor over $\mathbb{F}_p$ is also a nonzero minor over $\mathbb{Q}$.

Combining this observation with (8), we get

$$\operatorname{rank}_{\mathbb{F}_p}(\overline{D}) \le n^{O_m(1)}.$$

Multiplying by the nonzero scalar $u^{-1} \in \mathbb{F}_p$ does not change the rank, so the matrix of $\Delta(x, y)$ over $\mathbb{F}_p$ also has rank at most $n^{O_m(1)}$. A rank factorization of this matrix over $\mathbb{F}_p$ gives maps

$$\Phi, \Psi : \mathbb{Z}_m^n \to \mathbb{F}_p^R, \qquad R \le n^{d_m},$$

such that

$$\Delta(x, y) = \langle \Phi(x), \Psi(y) \rangle_{\mathbb{F}_p}.$$

The tensor-product identity in the statement gives the corresponding representation for products of indicator values. Since that representation has dimension $R^h \le n^{O_m(h)}$, the associated matrices have rank at most $n^{O_m(h)}$ over $\mathbb{F}_p$. □

5 Induction - fiber size bound

As in the overview, the fiber-growth argument proceeds by induction on the sum length L. Throughout this section, the modulus is a minimal non-prime-power modulus

$$m = p^a q^b, \qquad p < q,$$

as in Lemma 2.3. Thus the two nonzero inner-product residues α, β satisfy

$$\alpha \equiv 0 \pmod{p^a q^{b-1}}, \qquad \alpha \not\equiv 0 \pmod{q^b},$$

and

$$\beta \equiv 0 \pmod{p^{a-1} q^b}, \qquad \beta \not\equiv 0 \pmod{p^a}.$$

For an integer $L \ge 1$ and a set $A \in \binom{[K]}{L}$, write

$$\Sigma_L(A) = \sum_{i \in A} u_i.$$

The largest L-sum fiber is

$$M_L(U) = \max_x \#\{A \in \binom{[K]}{L} : \Sigma_L(A) = x\}.$$

The earlier 3-query proof gives the following base case. The mechanism is that two different sets in the same short-sum fiber are forced to behave like a 2-restricted matching vector family.

Lemma 5.1 (Base fiber bound [ADL+25])**.** *With notation as above,*

$$M_q(U) \le bn + 1.$$

Proof. This is the short-sum base case from Aggarwal et al. [ADL+25]. The key point is that two distinct sets in the same q-sum fiber must be disjoint, and their cross inner products can have only one value α modulo q^b. Choosing one representative from each set in the fiber would therefore produce a 2-restricted matching vector family over $\mathbb{Z}_{q^b}^n$, with diagonal value 0 and off-diagonal value α. Lemma 2.2 rules this out once the fiber is too large, and gives $|\mathcal{F}| \le bn + 1$ for every q-sum fiber $\mathcal{F}$, which is exactly $M_q(U) \le bn + 1$. □

We now prove the fiber-growth recurrence described in the setup section.

Theorem 5.2 (Long-sum fiber bound)**.** *Let (U, V) be a 3-restricted matching vector family modulo a minimal non-prime-power divisor $m = p^a q^b$, with $p < q$, as in Lemma 2.3. Then for all $L \ge q$,*

$$M_L(U) \le \exp\big(O_m(L^2 \log n + L \log L)\big).$$

Proof. The base case $L = q$ is Lemma 5.1. We prove a recurrence for $L > q$.

Fix an L-sum fiber

$$\mathcal{F} = \{A \in \binom{[K]}{L} : \Sigma_L(A) = x\}.$$

View $\mathcal{F}$ as an L-uniform hypergraph on $[K]$. Let $\Delta(\mathcal{F})$ be its maximum vertex degree and $\nu(\mathcal{F})$ its matching number. If $\mathcal{M}_0$ is a maximal matching, then every edge of $\mathcal{F}$ intersects the union of the edges in $\mathcal{M}_0$; otherwise that edge could be added to the matching. This union contains exactly $L|\mathcal{M}_0| \le L\nu(\mathcal{F})$ vertices, and each vertex lies in at most $\Delta(\mathcal{F})$ edges. Hence

$$|\mathcal{F}| \le L\,\nu(\mathcal{F})\,\Delta(\mathcal{F}).$$

If a set $A \in \mathcal{F}$ contains a fixed vertex i, then $A \setminus \{i\}$ lies in an $(L-1)$-sum fiber. Therefore

$$\Delta(\mathcal{F}) \le M_{L-1}(U).$$

It remains to bound $\nu(\mathcal{F})$.

Let $\mathcal{M} \subseteq \mathcal{F}$ be a pairwise-disjoint subfamily. For each $A \in \mathcal{M}$, choose a special pivot $i(A) \in A$ (any fixed point will do). Let

$$r_A = \#\{j \in A : \langle u_j, v_{i(A)}\rangle = \beta\}.$$

Since the diagonal term is 0, this equals the number of β-edges from $A \setminus \{i(A)\}$ into $v_{i(A)}$.

Take distinct $A, B \in \mathcal{M}$. Since they are disjoint and have the same L-sum, pairing

$$\sum_{j\in A} u_j = \sum_{j\in B} u_j$$

with $v_{i(A)}$ and moving the A-side to the B-side gives

$$\sum_{j\in B}\langle u_j, v_{i(A)}\rangle - \sum_{j\in A\setminus\{i(A)\}} \langle u_j, v_{i(A)}\rangle \equiv 0 \pmod m.$$

The pivot term $j = i(A)$ drops out because $\langle u_{i(A)}, v_{i(A)}\rangle = 0$. For $\gamma \in \{\alpha, \beta\}$, write

$$\#_\gamma(C \to i(A)) = \#\{j \in C : \langle u_j, v_{i(A)}\rangle = \gamma\}.$$

Then the last congruence can be written as

$$a_\alpha \alpha + a_\beta \beta \equiv 0 \pmod m,$$

where

$$a_\alpha = \#_\alpha(B \to i(A)) - \#_\alpha(A \setminus \{i(A)\} \to i(A)),$$

and similarly for a_β with β. The two count differences satisfy

$$a_\alpha + a_\beta = |B| - |A \setminus \{i(A)\}| = 1,$$

because B contributes L off-diagonal terms and $A \setminus \{i(A)\}$ contributes $L-1$. Reducing modulo q^b, the β-term vanishes, while $\alpha = q^{b-1}u_\alpha$ with $u_\alpha \not\equiv 0 \pmod q$. Hence

$$a_\alpha q^{b-1} u_\alpha \equiv 0 \pmod{q^b},$$

and therefore $q \mid a_\alpha$, while α is divisible by q^{b-1} but not by q^b). If $a_\beta = 0$, then $a_\alpha = 1$, contradicting $q \mid a_\alpha$. Hence $a_\beta \neq 0$. Thus

$$\#_\beta(B \to i(A)) \neq \#_\beta(A \setminus \{i(A)\} \to i(A)) = r_A.$$

So the exact β-count into the pivot $v_{i(A)}$ distinguishes A from all other disjoint hyperedges B.

Let $\Delta(x, y)$ be the indicator from Lemma 4.1. As in the overview, we now use a low-degree polynomial that checks an exact count. For every $r \in \{0, \dots, L\}$, there is a polynomial

$$P_r(z_1, \dots, z_L) \in \mathbb{F}_p[z_1, \dots, z_L]$$

of degree at most L that equals 1 on Boolean inputs of Hamming weight r and equals 0 on Boolean inputs of all other Hamming weights. To see this, let

$$e_d(z_1, \dots, z_L) = \sum_{\substack{S \subseteq [L] \\ |S| = d}} \prod_{s \in S} z_s$$

be the elementary symmetric polynomial of degree d. On a Boolean vector of weight w, it has value $e_d(z) = \binom{w}{d}$ in $\mathbb{F}_p$. The binomial basis $w \mapsto \binom{w}{d}$, $0 \le d \le L$, is triangular in the values $w = 0, \dots, L$ with diagonal 1, so we can choose coefficients $c_{r,d} \in \mathbb{F}_p$ such that

$$P_r(z) = \sum_{d=0}^{L} c_{r,d} e_d(z)$$

is the exact-weight-r polynomial.

Define a matrix R over $\mathbb{F}_p$, with rows and columns indexed by $\mathcal{M}$, by

$$R_{A,B} = P_{r_A}\big((\Delta(u_j, v_{i(A)}))_{j \in B}\big).$$

For $B = A$, the Hamming weight is r_A, so $R_{A,A} = 1$. For $B \neq A$, the Hamming weight is not r_A, so $R_{A,B} = 0$. Hence R is the identity matrix indexed by $\mathcal{M}$.

It remains to bound the rank of R. Write

$$P_{r_A} = \sum_{\ell=0}^{L} c_{A,\ell} e_\ell$$

in the elementary-symmetric basis. Fix a degree ℓ. Expanding e_ℓ gives

$$e_\ell\big((\Delta(u_j, v_{i(A)}))_{j \in B}\big) = \sum_{\substack{S \subseteq B \\ |S| = \ell}} \prod_{j \in S} \Delta(u_j, v_{i(A)}).$$

Each product on the right is exactly the kind of product handled by Lemma 4.1: the tensor-product representation of Δ gives an inner-product representation of dimension $n^{O_m(\ell)}$. More concretely, the row side depends on the pivot $v_{i(A)}$, while the column side for a subset $S \subseteq B$ depends on the vectors $(u_j)_{j\in S}$. After summing over all ℓ-subsets $S \subseteq B$, we simply replace that column vector by the sum of the corresponding subset-vectors. Thus the whole degree-ℓ matrix still has rank at most $n^{O_m(\ell)}$. Multiplying row A by the scalar $c_{A,\ell}$ does not increase this rank, even though the scalar may depend on A. Therefore the degree-ℓ contribution has rank at most $n^{O_m(\ell)}$, and summing over all degrees gives

$$\operatorname{rank}(R) \le \sum_{\ell=0}^{L} n^{O_m(\ell)} \le n^{O_m(L)}.$$

Since R is the identity matrix indexed by the matching $\mathcal{M}$, we conclude that

$$|\mathcal{M}| = \operatorname{rank}(R) \le n^{O_m(L)}.$$

Thus

$$\nu(\mathcal{F}) \le n^{O_m(L)}$$

and consequently

$$M_L(U) \le L\, n^{O_m(L)} M_{L-1}(U).$$

As in the overview, iterating this recurrence from the base case $M_q(U) \le bn + 1$ gives

$$\log M_L(U) \le O_m(L^2 \log n + L \log L).$$

□

Corollary 5.3 (Entropy growth)**.** *For all $L \ge q$,*

$$H_\infty(S_L^*(U)) \ge L \log K - O_m(L^2 \log n + L \log L).$$

Proof. The largest atom of $S_L^*(U)$ has probability at most $M_L(U)/\binom{K}{L}$. Since $\binom{K}{L} \ge (K/L)^L$, the claim follows from Theorem 5.2. □

6 Final entropy calculation

Proof of Theorem 1.2. Let (U, V) be a 3-restricted matching vector family modulo m of size K. Pass to a minimal divisor $m_0 \mid m$ for which the same lists remain a matching vector family. Since m is fixed, replacing m by m_0 only changes constants, and $n \log m_0 \le n \log m$.

If m_0 is a prime power, Lemma 2.1 gives $K \le n^{O_m(1)}$, which is stronger than the desired bound. Otherwise write

$$m_0 = p^a q^b, \qquad p < q,$$

so the family satisfies the hypotheses of Theorem 5.2 over the minimal modulus m_0.

Choose

$$L = \left\lfloor c_m \sqrt{\frac{n}{\log n}} \right\rfloor$$

with $c_m > 0$ fixed. For all sufficiently large n, depending only on m, this choice satisfies $L \ge q$; the remaining values of n are absorbed into the constant in the theorem. If $K < L$, then

$$\log K \le \log L = O_m(\sqrt{n \log n}),$$

so we are done. Thus assume $K \geq L$, so $S_L^*(U)$ is well defined. By Corollary 5.3,

$$H_\infty(S_L^*(U)) \geq L \log K - O_m(L^2 \log n + L \log L).$$

On the other hand, $S_L^*(U)$ is supported on $\mathbb{Z}_{m_0}^n$, so

$$H_\infty(S_L^*(U)) \leq n \log m_0 \leq n \log m.$$

Combining the two inequalities gives

$$L \log K \leq n \log m + O_m(L^2 \log n + L \log L).$$

Thus

$$\log K \leq \frac{n \log m}{L} + O_m(L \log n + \log L).$$

With the above choice of L,

$$\frac{n}{L} = O_m(\sqrt{n \log n}), \qquad L \log n = O_m(\sqrt{n \log n}).$$

Thus

$$\log K \leq O_m(\sqrt{n \log n}),$$

as required. □

7 Concluding remarks and open questions

The theorem shows that the entropy-growth approach of [ADL+25] can be pushed from the initial short-sum regime to sums of length $L \asymp_m \sqrt{n/\log n}$ for 3-restricted matching vector families. The main new ingredient is the bound on long-sum fibers. Instead of requiring a collision to force all cross-edges to have one residue, the argument only needs this exact count, measured from a suitable pivot, to change. This produces an identity matrix on every large matching inside a fiber, and the residue indicator turns that identity matrix into a low-rank object.

For larger query number, the first genuinely new obstruction already appears for $m = 6$ and off-diagonal set $\{2, 3, 5\}$. The relation

$$2 + 3 - 5 = 0 \pmod 6$$

can preserve the two CRT count profiles and therefore evades the simple exact-count test used here. Thus a collision in a long-sum fiber need not produce the kind of count separator that drives the 3-restricted induction, and even the appropriate base case is no longer clear from the present method. Understanding how such relations can be packed inside a fiber seems to be the next barrier toward extending the approach to 4-restricted and more general constant-query matching vector families.

8 Use of AI

The authors used ChatGPT 5.5 extensively during proof exploration and manuscript preparation. ChatGPT proposed parts of the successful proof strategy, including the graph/hypergraph matching interpretation used in the long-sum collision argument. The authors supplied the problem, prior entropy-growth framework, and long-sum/tensoring intuitions, and selected, refined, and independently verified the final proof. Codex was additionally used for paragraph-level editing. The authors take full responsibility for the paper.

References

[ADL+25] D. Aggarwal, P. Dutta, Z. Li, M. Obremski, and S. Saraogi. Improved lower bounds for 3-query matching vector codes. In *16th Innovations in Theoretical Computer Science Conference (ITCS 2025)*, volume 325 of *LIPIcs*, 2025. Article 61.

[AGKM23] O. Alrabiah, V. Guruswami, P. K. Kothari, and P. 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 (STOC 2023)*, pages 1438–1448, 2023.

[ALM+98] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, 1998.

[Alo98] N. Alon. The Shannon capacity of a union. *Combinatorica*, 18(3):301–310, 1998.

[AS98] S. Arora and S. Safra. Probabilistic checking of proofs: A new characterization of NP. *Journal of the ACM*, 45(1):70–122, 1998.

[BCG20] A. Bhattacharyya, L. S. Chandran, and S. Ghoshal. Combinatorial lower bounds for 3-query LDCs. In *11th Innovations in Theoretical Computer Science Conference (ITCS 2020)*, volume 151 of *LIPIcs*, 2020. Article 85.

[BDL14] A. Bhowmick, Z. Dvir, and S. Lovett. New bounds for matching vector families. *SIAM Journal on Computing*, 43(5):1654–1683, 2014.

[DGY11] Z. Dvir, P. Gopalan, and S. Yekhanin. Matching vector codes. *SIAM Journal on Computing*, 40(4):1154–1178, 2011.

[Efr09] K. Efremenko. 3-query locally decodable codes of subexponential length. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing (STOC 2009)*, pages 39–44, 2009.

[FW81] P. Frankl and R. M. Wilson. Intersection theorems with geometric consequences. *Combinatorica*, 1(4):357–368, 1981.

[GGMT24] W. T. Gowers, B. Green, F. Manners, and T. Tao. Marton's conjecture in abelian groups with bounded torsion, 2024. arXiv:2404.02244.

[GKST06] O. Goldreich, H. Karloff, L. J. Schulman, and L. Trevisan. Lower bounds for linear locally decodable codes and private information retrieval. *Computational Complexity*, 15:263–296, 2006.

[Gop06] P. Gopalan. Constructing Ramsey graphs from Boolean function representations. In *21st Annual IEEE Conference on Computational Complexity (CCC 2006)*, pages 120–128, 2006.

[Gop24] S. Gopi. Upper bound for matching vector families for prime powers, 2024. Personal communication.

[Gro95] V. Grolmusz. On the weak mod m representation of Boolean functions. *Chicago Journal of Theoretical Computer Science*, 1995.

[KdW04] I. Kerenidis and R. de Wolf. Exponential lower bound for 2-query locally decodable codes via a quantum argument. *Journal of Computer and System Sciences*, 69(3):395–420, 2004.

[KT00] J. Katz and L. Trevisan. On the efficiency of local decoding procedures for error-correcting codes. In *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing (STOC 2000)*, pages 80–86, 2000.

[Woo07] D. P. Woodruff. New lower bounds for general locally decodable codes, 2007. Electronic Colloquium on Computational Complexity.

[Woo12] D. 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.

[Yek08] S. Yekhanin. Towards 3-query locally decodable codes of subexponential length. *Journal of the ACM*, 55(1):1–16, 2008.

[Yek12] S. Yekhanin. Locally decodable codes. *Foundations and Trends in Theoretical Computer Science*, 6(3):139–255, 2012.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il