Matching-Vector Families and LDCs Over Large Modulo

Zeev Dvir[*], Guangda Hu[†]

Abstract

We prove new upper bounds on the size of families of vectors in $\mathbb{Z}_m^n$ with restricted modular inner products, when m is a large integer. More formally, if $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_t \in \mathbb{Z}_m^n$ and $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_t \in \mathbb{Z}_m^n$ satisfy $\langle \boldsymbol{u}_i, \boldsymbol{v}_i \rangle \equiv 0 \pmod{m}$ and $\langle \boldsymbol{u}_i, \boldsymbol{v}_j \rangle \not\equiv 0 \pmod{m}$ for all $i \neq j \in [t]$, we prove that $t \leq O(m^{n/2+8.47})$. This improves a recent bound of $t \leq m^{n/2+O(\log(m))}$ by [BDL13] and is the best possible up to the constant 8.47 when m is sufficiently larger than n.

The maximal size of such families, called 'Matching-Vector families', shows up in recent constructions of locally decodable error correcting codes (LDCs) and determines the rate of the code. Using our result we are able to show that these codes, called Matching-Vector codes, must have encoding length at least $K^{19/18}$ for K-bit messages, regardless of their query complexity. This improves a known super linear bound of $K2^{\Omega(\sqrt{\log K})}$ proved in [DGY11].

1 Introduction

A Matching-Vector family (MV family) in $\mathbb{Z}_m^n$ is defined as a pair of ordered lists $U = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_t)$ and $V = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_t)$ with $\boldsymbol{u}_i, \boldsymbol{v}_j \in \mathbb{Z}_m^n$, satisfying the following property: for all $i \in [t]$, $\langle \boldsymbol{u}_i, \boldsymbol{v}_i \rangle \equiv 0 \pmod{m}$ whereas for all $i \neq j \in [t]$, $\langle \boldsymbol{u}_i, \boldsymbol{v}_j \rangle \not\equiv 0 \pmod{m}$. Here $\langle \cdot, \cdot \rangle$ denotes the standard inner product. If one restricts the entries of the vectors in the family to be in the set $\{0, 1\}$ the inner products corresponds to the sizes of the intersections (modulo m) and, in this case, MV families are

---

[*]Department of Computer Science and Department of Mathematics, Princeton University, Princeton NJ. Email: `zeev.dvir@gmail.com`. Research partially supported by NSF grants CCF-0832797, CCF-1217416 and by the Sloan fellowship.

[†]Department of Computer Science Princeton University, Princeton NJ. Email: `guangdah@cs.princeton.edu`. Research partially supported by NSF grants CCF-0832797, CCF-1217416 and by the Sloan fellowship.

more commonly referred to as families of sets with restricted modular intersections. MV families were studied previously in the context of Ramsey graphs [Gro00], circuit complexity [BBR94] and, more recently, were used to construct Locally Decodable Codes (LDCs) [Yek08, Efr09, DGY11], which are error correcting codes with super-efficient decoding properties. We will elaborate more on the connection to LDCs after we state our results.

We denote by $\mathrm{MV}(m,n)$ the size of the largest MV family in $\mathbb{Z}_m^n$ (the size of the family is t in the above notation). It is an interesting (and mostly open) question to determine the value (or even order of magnitude) of $\mathrm{MV}(m,n)$ for arbitrary m and n. Upper and lower bounds on $\mathrm{MV}(m,n)$ can be roughly divided into two kinds, corresponding to the relative size of the two parameters. One typical regime is when m is small and n tends to infinity and the other is when $m >> n$ (of course there are intermediate scenarios as well).

Although our work focuses on the regime when m is much larger than n, we first describe the known results for the other regime, namely when m is a fixed constant and n tends to infinity. These regime is further divided into the case when m is prime and when m is composite. When m is a small *prime* and n tends to infinity, the value of $\mathrm{MV}(m,n)$ is known to be of the order of n^{m-1} [BF98]. When m is a small *composite*, the picture is very different and there are exponential gaps between known lower and upper bounds on $\mathrm{MV}(m,n)$. A surprising construction by Grolmuzs [Gro00] shows that $\mathrm{MV}(m,n) \geq \exp\left(c \cdot \frac{\log(n)^r}{(\log\log n)^{r-1}}\right)$ when m has r distinct prime factors (here c is an absolute constant). That is, $\mathrm{MV}(m,n)$ can be super-polynomial in n (that is $n^{\omega(1)}$) for m as small as 6 (compared with the polynomial upper bound n^{m-1} for prime m). A trivial upper bound on $\mathrm{MV}(m,n)$ is m^n since an MV family cannot contain the same vector twice. The best upper bound on $\mathrm{MV}(m,n)$ for small composite m was proved in [BDL13] and is $m^{n/2+O(\log m)}$. Assuming the Polynomial-Freiman-Ruzsa (PFR) conjecture [TV07] this can be improved to $\mathrm{MV}(m,n) \leq C_m^{n/\log(n)}$ with C_m a constant depending only on m.

Our work focuses on the regime when m is larger than n. In this setting, a construction of [YGK12] gives MV families of size $\left(\frac{m+1}{n-2}\right)^{n/2-1}$ [YGK12]. For a large *prime* m, this construction almost matches an upper bound of $O(m^{n/2})$ proved in [DGY11]. For composite m, the best upper bound on $\mathrm{MV}(m,n)$ for large m prior to this work was the same $m^{n/2+O(\log m)}$ bound from [BDL13]. Notice that, when $m > 2^n$, this bound is meaningless since it exceeds the trivial bound of m^n. In

m	upper bound for MV(m,n)
general prime	$O(m^{n/2})$ [DGY11]
small, fixed prime	$O(n^{m-1})$ [BF98]
general composite	$m^{n/2+O(\log m)}$ [BDL13]
small, fixed composite	$2^{O_m(n/\log n)}$ [BDL13] (assuming PFR)
general composite	$O(m^{n/2+8.47})$ (Theorem 1.1)

Table 1: List of upper bounds on MV(m,n)

this work we extend the proof method developed in [BDL13] to give the following bound:

Theorem 1.1. *For all integers $m > 1, n$ we have $MV(m,n) \le 100m^{n/2+8.47}$. When m is a product of distinct primes the constant 8.47 can be replaced with $4 + o(1)$.*

For small n, this bound is tight up to the constant 8.47 as the [YGK12] construction shows. When m is small, this still gives some improvement over the $m^{n/2+O(\log m)}$ bound of [BDL13] but not as dramatic (and probably far from being tight).

The main tool in our proof is Fourier analysis in the spirit of [BDL13], with which we repeatedly reduce m to one of its factor (eventually reaching the case of $m = 1$). The distribution of $\langle \boldsymbol{v_i}, \boldsymbol{u_j} \rangle$ over random $i, j \in [t]$ is far from the uniform distribution (since the probability of obtaining zero is small). This fact is used to find a large coefficient in its Fourier spectrum. This coefficient is then used to carve out a large sub family which is again an MV family, but over some proper factor of m. The proof ends when we reach the case of prime m. The difference between our proof and the one in [BDL13] is in the choice of the large coefficient (or character). We are able to show that a large character appears that has nicer number theoretic properties and so are able to analyze the loss in each step in a better way – getting rid of the $O(\log m)$ factor in the exponent.

1.1 MV families and Locally Decodable Codes

A (q, δ, ϵ)-Locally Decodable Code, or LDC, encodes a K-symbol message x to an N-symbol codeword $C(x)$, such that every symbol x_i ($i \in [K]$) can be recovered with probability at least $1 - \epsilon$ by a randomized decoding procedure that makes only q queries to $C(x)$, even if δN locations of the codeword $C(x)$ have been corrupted. Understanding the minimum length $N = N(k)$ of an LDC with constant q is a central research question that is still far from being solved. For

$q = 1, 2$, this question is completely answered. There are no LDCs for $q = 1$ [KT00] and the best LDCs for $q = 2$ have exponential length [GKST02, KdW04]. However, for $q > 2$ there are huge gaps between lower bound and LDC constructions. The best known lower bound is $N = \tilde{\Omega}(K^{1+1/(\lceil r/2 \rceil - 1)})$ for $k > 4$ [Woo07] and $N = \Omega(K^2)$ for $k = 3, 4$ [KdW04, Woo10], while the best construction has super-polynomial length. Constructions of LDCs have been studied extensively for more than a decade. Until recently, all constructions of LDC with constant q had exponential encoding length. In a breakthrough work of Yekhanin [Yek08] and following improvements [Efr09, Rag07, KY09, IS10, CFL$^+$10, DGY11, BET10], a new family of LDCs based on Matching Vector families was introduced. These codes, called Matching-Vector codes, rely on constructions of MV families and can have sub-exponential length for q as small as 3 [Efr09]. Using Grolmuzs construction as a building block, one obtains an encoding length of roughly

$$N \sim \exp \exp \left((\log K)^{O(\log \log q / \log q)} (\log \log K) \right).$$

The size of the MV family used in the code construction is critical. In its simplest form, an MV code using an MV family of size t in $\mathbb{Z}_m^n$ will send $K = t$ bits of message into $N = m^n$ bits of encoding and will require $q = m$ queries to decode. Several improvements are possible for reducing the number of queries below m but these are case-based and hard to generalize for arbitrary m.

Our improved bound on the size of MV families allows us to prove an unconditional lower bound on the encoding length of MV codes, *regardless of the query complexity.*

Theorem 1.2. *For any MV-code with message length K and codeword length N we have $N > K^{\frac{19}{18}}$. This bound is regardless of the number of queries.*

This theorem improves on a bound of $N > K 2^{\Omega(\sqrt{\log K})}$ proved in [DGY11].

1.2 Organization

We begin in Section 2 with a number of preliminary lemmas and notations that will be used throughout the proof. In Section 3 we prove our main technical lemma which is the heart of our proof. The lemma is used iteratively in the proof of our main theorem which is given in Section 4.

The proof of the stronger bound for the case when m is a product of distinct primes is given in the appendix.

2 Preliminaries

2.1 Fourier lemma

We consider a probability distribution μ over $\mathbb{Z}_m$. Let $\omega_m = e^{\frac{2\pi}{m}i}$ be an order m primitive root of unity. It is not difficult to see $\mathbb{E}_{x\sim\mu}[\omega_m^{jx}] = 0$ for all $j \in \{1, 2, \ldots, m-1\}$ if μ is the uniform distribution. We will show that $\mathbb{E}_{x\sim\mu}[\omega_m^{jx}]$ is bounded away from zero for some $j \in \{1, 2, \ldots, m-1\}$ if μ is far from being uniform.

In [BDL13], it was shown that

$$\max_{1 \leq j \leq m-1} \left| \mathbb{E}_{x\sim\mu}[\omega_m^{jx}] \right| = \Omega(\frac{1}{m^{1.5}})$$

if the statistical distance between μ and the uniform distribution is big, i.e. $\frac{1}{2}\sum_{x\in\mathbb{Z}_m} |\mu(x) - \frac{1}{m}| = \Omega(\frac{1}{m})$. In the following lemma, we prove a better lower bound that depends only on $s = \text{order}(\omega_m^j)$ under a stronger condition $|\mu(0) - \frac{1}{m}| = \Omega(\frac{1}{m})$.

Consider μ as a function from $\mathbb{Z}_m$ to $\mathbb{C}$. For $0 \leq j \leq m-1$, the Fourier coefficient $\hat{\mu}(j)$ is

$$\hat{\mu}(j) = \frac{1}{m}\sum_{x\in\mathbb{Z}_m} \mu(x)\omega_m^{-jx} = \frac{1}{m}\mathbb{E}_{x\sim\mu}[\omega_m^{-jx}].$$

One can see that $\hat{\mu}(0) = \frac{1}{m}$. The set of functions $\{\omega_m^{jx} \mid 0 \leq j \leq m-1\}$ is an orthogonal basis for all functions from $\mathbb{Z}_m$ to $\mathbb{C}$, and the function $\mu(x)$ can be written as

$$\mu(x) = \sum_{j=0}^{m-1} \hat{\mu}(j)\omega_m^{jx}. \tag{1}$$

Lemma 2.1. *Let $\mu : \mathbb{Z}_m \mapsto [0, 1]$ be a probability distribution over $\mathbb{Z}_m$ (i.e. $\sum_{x\in\mathbb{Z}_m} \mu(x) = 1$). If $\mu(0) \leq \frac{1}{100m}$, there must exist $j \in \{1, 2, \ldots, m-1\}$ such that $\left| \mathbb{E}_{x\sim\mu}[\omega_m^{jx}] \right| \geq \frac{1}{sf(s)}$, where $s = \frac{m}{\gcd(j,m)}$ is the order of ω_m^j for $\omega_m = e^{\frac{2\pi}{m}i}$, and $f : \mathbb{Z}^+ \mapsto \mathbb{R}$ is any function satisfying $\sum_{s=2}^{\infty} \frac{1}{f(s)} \leq 0.99$.*

5

Proof. By setting $x = 0$ in (1), we have

$$\mu(0) = \sum_{j=0}^{m-1} \hat{\mu}(j)\omega_m^{j \cdot 0} = \sum_{j=0}^{m-1} \hat{\mu}(j) = \frac{1}{m} + \frac{1}{m} \sum_{j=1}^{m-1} \mathbb{E}_{x \sim \mu} [\omega_m^{-jx}].$$

Therefore

$$\sum_{j=1}^{m-1} \left| \mathbb{E}_{x \sim \mu} [\omega_m^{jx}] \right| \geq \left| \sum_{j=1}^{m-1} \mathbb{E}_{x \sim \mu} [\omega_m^{jx}] \right| = \left| \sum_{j=1}^{m-1} \mathbb{E}_{x \sim \mu} [\omega_m^{-jx}] \right| = m \cdot \left| \mu(0) - \frac{1}{m} \right| \geq 0.99. \tag{2}$$

For every $d \mid m$ $(1 \leq d \leq m-1)$, define $T_d = \{j \mid \gcd(j, m) = d, 1 \leq j \leq m-1\}$. For all $j \in T_d$, the order of ω_m^j is $s_d = \frac{m}{d}$ $(2 \leq s_d \leq m)$. We also see $T_d = \{k \cdot d \mid 1 \leq k < s_d, \gcd(k, s_d) = 1\}$, hence $|T_d| = \varphi(s_d) < s_d$.

If the lemma was not true, we have

$$\sum_{j=1}^{m-1} \left| \mathbb{E}_{x \sim \mu} [\omega_m^{jx}] \right| = \sum_{\substack{d \mid m \\ d < m}} \left(\sum_{j \in T_d} \left| \mathbb{E}_{x \sim \mu} [\omega_m^{jx}] \right| \right) < \sum_{\substack{d \mid m \\ d < m}} \left(s_d \cdot \frac{1}{s_d f(s_d)} \right) < \sum_{s=2}^{\infty} \frac{1}{f(s)} \leq 0.99.$$

This violates inequality (2). Thus the lemma is proved. $\qquad \square$

2.2 Notations and Facts about MV Families

We use $\langle \cdot, \cdot \rangle$ to denote the inner product over $\mathbb{Z}$ between two vectors. In all calculations, we identify $\mathbb{Z}_m$ as $\{0, 1, \ldots, m-1\}$ and treat the numbers as on $\mathbb{Z}$. Conventionally, we consider $a \bmod 1$ to be 0 for any integer a.

Notation 2.2. *Let r be a positive integer. For an integer v, define $v^{(r)} \in \{0, 1, \ldots, r-1\}$ to be v modulo r. For a vector $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)$, define $\boldsymbol{v}^{(r)} = (v_1^{(r)}, v_2^{(r)}, \ldots, v_n^{(r)})$. For a list of vectors $V = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_t)$, define $V^{(r)} = (\boldsymbol{v}_1^{(r)}, \boldsymbol{v}_2^{(r)}, \ldots, \boldsymbol{v}_t^{(r)})$.*

Notation 2.3. *Let r be a positive integer. For an integer v, define $v^{[r]} \in \mathbb{Z}$ to be $(v - v^{(r)})/r$. For a vector $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)$, define $\boldsymbol{v}^{[r]} = (v_1^{[r]}, v_2^{[r]}, \ldots, v_n^{[r]})$. Thus $\boldsymbol{v} = r\boldsymbol{v}^{[r]} + \boldsymbol{v}^{(r)}$ for any vector $\boldsymbol{v}$. For a list of vectors $V = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_t)$, define $V^{[r]} = (\boldsymbol{v}_1^{[r]}, \boldsymbol{v}_2^{[r]}, \ldots, \boldsymbol{v}_t^{[r]})$.*

Definition 2.4. *Let $U = (\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_t)$ and $V = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_t)$ be two lists of vectors in $\mathbb{Z}_m^n$. (U, V) is a* matching vector family *if $\langle \boldsymbol{u}_i, \boldsymbol{v}_i \rangle \equiv 0 \pmod{m}$ for all $i \in [t]$ and $\langle \boldsymbol{u}_i, \boldsymbol{v}_j \rangle \not\equiv 0 \pmod{m}$ for all $i \neq j \in [t]$. The number t is the size of the MV family and is denoted by $|(U, V)|$.*

Claim 2.5. *For an MV family (U, V) where $U = (\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_t)$, $V = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_t)$ and $i \neq j \in [t]$, we have $\boldsymbol{u}_i \neq \boldsymbol{u}_j$ and $\boldsymbol{v}_i \neq \boldsymbol{v}_j$.*

Proof. Assume $\boldsymbol{u}_i = \boldsymbol{u}_j$ for $i \neq j$, we have $\langle \boldsymbol{u}_i, \boldsymbol{v}_j \rangle = \langle \boldsymbol{u}_i, \boldsymbol{v}_i \rangle \equiv 0 \pmod{m}$. This violates the definition of MV family. $\square$

Notation 2.6. *Let U, V, U', V' be 4 lists of vectors in $\mathbb{Z}_m^n$, and say $U = (\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_t)$, $V = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_t)$. We write $(U', V') \subseteq (U, V)$ if there exists a set $T \subseteq [t]$ such that $U' = (\boldsymbol{u}_i : i \in T)$ and $V' = (\boldsymbol{v}_i : i \in T)$. Observe that if (U, V) is an MV family, so is (U', V').*

Definition 2.7. *(r_1, r_2, r_3) is a* partition *of m if $r_1, r_2, r_3 \in \mathbb{Z}^+$ and $r_1 r_2 r_3 = m$. (r_1, r_2, r_3 are not assumed to be coprime.)*

Definition 2.8. *For an MV family (U, V) where $U = (\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_t)$ and $V = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_t)$, we say (U, V)* respects *(r_1, r_2, r_3), where (r_1, r_2, r_3) is a partition of m, if the following conditions are satisfied:*

1. *$\exists \boldsymbol{u}_0 \in \mathbb{Z}_{r_1}^n$ such that $\boldsymbol{u}_i^{(r_1)} = \boldsymbol{u}_0$ for all $i \in [t]$,*

2. *$\exists \boldsymbol{v}_0 \in \mathbb{Z}_{r_2}^n$ such that $\boldsymbol{v}_i^{(r_2)} = \boldsymbol{v}_0$ for all $i \in [t]$,*

3. *$\langle \boldsymbol{u}_i^{[r_1]}, \boldsymbol{v}_0 \rangle$ modulo r_2 is the same for all $i \in [t]$,*

4. *$\langle \boldsymbol{u}_0, \boldsymbol{v}_i^{[r_2]} \rangle$ modulo r_1 is the same for all $i \in [t]$.*

Claim 2.9. *If an MV family (U, V) respects (r_1, r_2, r_3), then $\langle \boldsymbol{u}_i, \boldsymbol{v}_j \rangle \equiv 0 \pmod{r_1 r_2}$ for all $\boldsymbol{u}_i \in U, \boldsymbol{v}_j \in V$.*

Proof. Let $\boldsymbol{u}_0 = \boldsymbol{u}_i^{(r_1)}$ and $\boldsymbol{v}_0 = \boldsymbol{v}_j^{(r_2)}$. They are fixed for all $\boldsymbol{u}_i \in U$ and $\boldsymbol{v}_j \in V$. We have

$$
\begin{aligned}
\langle \boldsymbol{u}_i, \boldsymbol{v}_j \rangle &= \langle r_1 \boldsymbol{u}_i^{[r_1]} + \boldsymbol{u}_0, r_2 \boldsymbol{v}_j^{[r_2]} + \boldsymbol{v}_0 \rangle \\
&= r_1 r_2 \langle \boldsymbol{u}_i^{[r_1]}, \boldsymbol{v}_j^{[r_2]} \rangle + r_1 \langle \boldsymbol{u}_i^{[r_1]}, \boldsymbol{v}_0 \rangle + r_2 \langle \boldsymbol{u}_0, \boldsymbol{v}_j^{[r_2]} \rangle + \langle \boldsymbol{u}_0, \boldsymbol{v}_0 \rangle.
\end{aligned}
$$

The first term is 0 modulo $r_1 r_2$. The second term is fixed modulo $r_1 r_2$ because $\langle \boldsymbol{u}_i^{[r_1]}, \boldsymbol{v}_0 \rangle$ is fixed modulo r_2. Similarly, the third term is also a constant modulo $r_1 r_2$. Therefore $\langle \boldsymbol{u}_i, \boldsymbol{v}_j \rangle$ modulo $r_1 r_2$ is the same for all $\boldsymbol{u}_i \in U$ and $\boldsymbol{v}_j \in V$. Note that when $i = j$, $\langle \boldsymbol{u}_i, \boldsymbol{v}_j \rangle \equiv 0 \pmod{r_1 r_2}$ since (U, V) is an MV family. Therefore $\langle \boldsymbol{u}_i, \boldsymbol{v}_j \rangle \equiv 0 \pmod{r_1 r_2}$ for all $\boldsymbol{u}_i \in U, \boldsymbol{v}_j \in V$. $\qquad\square$

Claim 2.10. *Every MV family (U, V) respects $(1, 1, m)$.*

Proof. Let $\boldsymbol{u}_0$ and $\boldsymbol{v}_0$ be the zero vector. All the conditions are satisfied. $\qquad\square$

Claim 2.11. *If an MV family (U, V) respects $(r_1, r_2, 1)$, then it must has size 1.*

Proof. Since $r_1 r_2 = m$, by Claim 2.9 we have $\langle \boldsymbol{u}_i, \boldsymbol{v}_j \rangle \equiv 0 \pmod{m}$ for all $\boldsymbol{u}_i \in U, \boldsymbol{v}_j \in V$. By the definition of MV family, the size of (U, V) must be 1. $\qquad\square$

3   Proof of the Main Lemma

Consider an MV family (U, V), where $U = (\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_t)$ and $V = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_t)$. We pick $\boldsymbol{u} \in U$ and $\boldsymbol{v} \in V$ uniformly at random and consider the distribution of $\langle \boldsymbol{u}, \boldsymbol{v} \rangle^{(m)}$. The inner product is 0 with probability $1/t$. Thus the distribution is far from uniform when $t >> m$. We will take advantage of this fact and prove our key lemma. For an MV family (U, V) respecting (r_1, r_2, r_3), we can find a large subfamily and reduce r_3 to some smaller number.

Let $f : \mathbb{Z}^+ \mapsto \mathbb{R}$ be a function satisfying $\sum_{s=2}^{\infty} \frac{1}{f(s)} \leq 0.99$. We will specify $f(s)$ in later proofs.

Lemma 3.1. *If an MV family (U, V) respects (r_1, r_2, r_3) for some $r_3 \geq 2$ and $|(U, V)| = t \geq 100m$, then there exists $s \mid r_3$ with $s \geq 2$ and an MV family $(U', V') \subseteq (U, V)$ with $|(U', V')| \geq t/(s^{n/2+4} f(s)^2)$ that respects either $(r_1 s, r_2, r_3/s)$ or $(r_1, r_2 s, r_3/s)$.*

Proof. We prove the lemma in 4 steps.

Step 1: Finding a nice character with a large bias.

By Claim 2.9, $\frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{r_1 r_2}$ is an integer for all $\boldsymbol{u} \in U, \boldsymbol{v} \in V$. We can also see $\frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{r_1 r_2} \equiv 0 \pmod{r_3}$ iff $\langle \boldsymbol{u}, \boldsymbol{v} \rangle \equiv 0 \pmod{m}$. Consider the distribution of $\left(\frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{r_1 r_2} \right)^{(r_3)} \in \mathbb{Z}_{r_3}$, where $\boldsymbol{u}$ and $\boldsymbol{v}$ are uniformly drawn from U and V respectively. We have

$$\mathbf{Pr}\left[\left(\frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{r_1 r_2} \right)^{(r_3)} = 0 \right] = \mathbf{Pr}\left[\langle \boldsymbol{u}, \boldsymbol{v} \rangle \equiv 0 \pmod{m} \right] = \frac{1}{t} \leq \frac{1}{100m} \leq \frac{1}{100 r_3}.$$

Applying Lemma 2.1 on $\mathbb{Z}_{r_3}$, there exists a $j \in \{1, 2, \ldots, r_3 - 1\}$ such that

$$\left| \mathop{\mathbb{E}}_{\substack{\boldsymbol{u} \sim U \\ \boldsymbol{v} \sim V}} \left[\omega_{r_3}^{j \frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{r_1 r_2}} \right] \right| \geq \frac{1}{s f(s)}, \tag{3}$$

where $\omega_{r_3} = e^{\frac{2\pi i}{r_3}}$ and $s = \frac{r_3}{\gcd(j, r_3)}$ is the order of $\omega_{r_3}^j$. Note that we have dropped the modulo r_3 operation because $(\omega_{r_3}^j)^{r_3} = 1$. It follows that

$$\mathop{\mathbb{E}}_{\substack{\boldsymbol{u}, \widetilde{\boldsymbol{u}} \sim U \\ \boldsymbol{v} \sim V}} \left[\omega_{r_3}^{j \frac{\langle \boldsymbol{u} - \widetilde{\boldsymbol{u}}, \boldsymbol{v} \rangle}{r_1 r_2}} \right] = \mathop{\mathbb{E}}_{\boldsymbol{v} \sim V} \left| \mathop{\mathbb{E}}_{\boldsymbol{u} \sim U} \left[\omega_{r_3}^{j \frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{r_1 r_2}} \right] \right|^2 \geq \left| \mathop{\mathbb{E}}_{\substack{\boldsymbol{u} \sim U \\ \boldsymbol{v} \sim V}} \left[\omega_{r_3}^{j \frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{r_1 r_2}} \right] \right|^2 \geq \frac{1}{s^2 f(s)^2}.$$

Therefore there exists a fixed $\widetilde{\boldsymbol{u}} \in U$ such that

$$\left| \mathop{\mathbb{E}}_{\substack{\boldsymbol{u} \sim U \\ \boldsymbol{v} \sim V}} \left[\omega_{r_3}^{j \frac{\langle \boldsymbol{u} - \widetilde{\boldsymbol{u}}, \boldsymbol{v} \rangle}{r_1 r_2}} \right] \right| = \left| \mathop{\mathbb{E}}_{\substack{\boldsymbol{u} \sim U \\ \boldsymbol{v} \sim V}} \left[\omega_{r_3}^{j \langle \frac{\boldsymbol{u} - \widetilde{\boldsymbol{u}}}{r_1}, \boldsymbol{v} \rangle / r_2} \right] \right| \geq \frac{1}{s^2 f(s)^2}.$$

Since $\boldsymbol{u}^{(r_1)} = \widetilde{\boldsymbol{u}}^{(r_1)}$, we have $\boldsymbol{u} - \widetilde{\boldsymbol{u}} = r_1(\boldsymbol{u}^{[r_1]} - \widetilde{\boldsymbol{u}}^{[r_1]})$. The above inequality can be written as

$$\left| \mathop{\mathbb{E}}_{\substack{\boldsymbol{u} \sim U \\ \boldsymbol{v} \sim V}} \left[\omega_{r_3}^{j \langle \boldsymbol{u}^{[r_1]} - \widetilde{\boldsymbol{u}}^{[r_1]}, \boldsymbol{v} \rangle / r_2} \right] \right| \geq \frac{1}{s^2 f(s)^2}. \tag{4}$$

Step 2: Partitioning into buckets.

We partition the set U into buckets according to $\boldsymbol{u}^{[r_1]} - \widetilde{\boldsymbol{u}}^{[r_1]}$ modulo s: $U = \bigcup_{\boldsymbol{w} \in \mathbb{Z}_s^n} B(\boldsymbol{w}, U)$, where

$$\widetilde{B}(\boldsymbol{w}, U) = \left\{ \boldsymbol{u} \in U \ \middle| \ \left(\boldsymbol{u}^{[r_1]} - \widetilde{\boldsymbol{u}}^{[r_1]} \right)^{(s)} = \boldsymbol{w} \right\}.$$

9

We also partition V into buckets $B(\boldsymbol{w}, V) = \{\boldsymbol{v} \in V \mid (\boldsymbol{v}^{[r_2]})^{(s)} = \boldsymbol{w}\}$ for all $\boldsymbol{w} \in \mathbb{Z}_s^n$. Define $p_{\boldsymbol{w}} = |\widetilde{B}(\boldsymbol{w}, U)|/t$ to be the density of $\widetilde{B}(\boldsymbol{w}, U)$ and $q_{\boldsymbol{w}} = |B(\boldsymbol{w}, V)|/t$ be the density of $B(\boldsymbol{w}, V)$.

Picking $\boldsymbol{u}$ uniformly from U can be equivalently considered as two steps: 1. For each bucket $\widetilde{B}(\boldsymbol{w}, U)$, pick a representative $\boldsymbol{u}_{\boldsymbol{w}} \in \widetilde{B}(\boldsymbol{w}, U)$ uniformly; 2. Pick one bucket according to the probability distribution $p_{\boldsymbol{w}}$, and output the representative. For inequality (4), we split the procedure of picking $\boldsymbol{u} \sim U$ into these two steps.

$$
\begin{aligned}
\frac{1}{s^2 f(s)^2} \;\leq\;& \left| \mathop{\mathbb{E}}_{\substack{\boldsymbol{u} \sim U \\ \boldsymbol{v} \sim V}} \left[\omega_{r_3}^{j\langle \boldsymbol{u}^{[r_1]} - \widetilde{\boldsymbol{u}}^{[r_1]}, \boldsymbol{v}\rangle/r_2} \right] \right| \\
=\;& \left| \mathop{\mathbb{E}}_{\substack{\text{for each } \boldsymbol{w}, \\ \boldsymbol{u}_{\boldsymbol{w}} \sim \widetilde{B}(\boldsymbol{w}, U)}} \mathop{\mathbb{E}}_{\boldsymbol{w} \sim p_{\boldsymbol{w}}} \mathop{\mathbb{E}}_{\boldsymbol{v} \sim V} \left[\omega_{r_3}^{j\langle \boldsymbol{u}_{\boldsymbol{w}}^{[r_1]} - \widetilde{\boldsymbol{u}}^{[r_1]}, \boldsymbol{v}\rangle/r_2} \right] \right| \\
\leq\;& \mathop{\mathbb{E}}_{\substack{\text{for each } \boldsymbol{w}, \\ \boldsymbol{u}_{\boldsymbol{w}} \sim \widetilde{B}(\boldsymbol{w}, U)}} \left| \mathop{\mathbb{E}}_{\boldsymbol{w} \sim p_{\boldsymbol{w}}} \mathop{\mathbb{E}}_{\boldsymbol{v} \sim V} \left[\omega_{r_3}^{j\langle \boldsymbol{u}_{\boldsymbol{w}}^{[r_1]} - \widetilde{\boldsymbol{u}}^{[r_1]}, \boldsymbol{v}\rangle/r_2} \right] \right|.
\end{aligned}
$$

There exists a fixed list of representatives from each bucket $(\boldsymbol{u}_{\boldsymbol{w}} \in B(\boldsymbol{w}, U) : \boldsymbol{w} \in \mathbb{Z}_s^n)$ such that

$$
\frac{1}{s^2 f(s)^2} \leq \left| \mathop{\mathbb{E}}_{\substack{\boldsymbol{w} \sim p_{\boldsymbol{w}} \\ \boldsymbol{v} \sim V}} \left[\omega_{r_3}^{j\langle \boldsymbol{u}_{\boldsymbol{w}}^{[r_1]} - \widetilde{\boldsymbol{u}}^{[r_1]}, \boldsymbol{v}\rangle/r_2} \right] \right|. \tag{5}
$$

For every $\boldsymbol{w} \in \mathbb{Z}_s^n$ and $\boldsymbol{u} \in B(\boldsymbol{w}, U)$, we use $\boldsymbol{u}'$ to denote the vector $(\boldsymbol{u}^{[r_1]} - \widetilde{\boldsymbol{u}}^{[r_1]})^{[s]}$. Thus

$$
\boldsymbol{u}_{\boldsymbol{w}}^{[r_1]} - \widetilde{\boldsymbol{u}}^{[r_1]} = s\boldsymbol{u}'_{\boldsymbol{w}} + (\boldsymbol{u}_{\boldsymbol{w}}^{[r_1]} - \widetilde{\boldsymbol{u}}^{[r_1]})^{(s)} = s\boldsymbol{u}'_{\boldsymbol{w}} + \boldsymbol{w}.
$$

Hence inequality (5) can be written as

$$
\frac{1}{s^2 f(s)^2} \leq \left| \mathop{\mathbb{E}}_{\substack{\boldsymbol{w} \sim p_{\boldsymbol{w}} \\ \boldsymbol{v} \sim V}} \left[\omega_{r_3}^{j\langle s\boldsymbol{u}'_{\boldsymbol{w}} + \boldsymbol{w}, \boldsymbol{v}\rangle/r_2} \right] \right|. \tag{6}
$$

Step 3: Finding a large bucket.

By inequality (6),

$$
\begin{aligned}
\left(\frac{1}{s^2 f(s)^2}\right)^2
&\leq \left|\sum_{\boldsymbol{w}\in\mathbb{Z}_s^n}\sum_{\boldsymbol{v}\in V} p_{\boldsymbol{w}} \cdot \frac{1}{t} \cdot \omega_{r_3}^{j\langle s\boldsymbol{u}_{\boldsymbol{w}}'+\boldsymbol{w},\boldsymbol{v}\rangle/r_2}\right|^2 \\
&\leq \left(\sum_{\boldsymbol{w}\in\mathbb{Z}_s^n} p_{\boldsymbol{w}}^2\right) \cdot \left(\sum_{\boldsymbol{w}\in\mathbb{Z}_s^n}\left|\sum_{\boldsymbol{v}\in V} \frac{1}{t}\cdot\omega_{r_3}^{j\langle s\boldsymbol{u}_{\boldsymbol{w}}'+\boldsymbol{w},\boldsymbol{v}\rangle/r_2}\right|^2\right) \\
&= \left(\sum_{\boldsymbol{w}\in\mathbb{Z}_s^n} p_{\boldsymbol{w}}^2\right) \cdot \left(\sum_{\boldsymbol{w}\in\mathbb{Z}_s^n}\sum_{\boldsymbol{v},\widetilde{\boldsymbol{v}}\in V} \frac{1}{t^2}\cdot\omega_{r_3}^{j\langle s\boldsymbol{u}_{\boldsymbol{w}}'+\boldsymbol{w},\boldsymbol{v}-\widetilde{\boldsymbol{v}}\rangle/r_2}\right) \\
&= \left(\sum_{\boldsymbol{w}\in\mathbb{Z}_s^n} p_{\boldsymbol{w}}^2\right) \cdot \left(\sum_{\boldsymbol{v},\widetilde{\boldsymbol{v}}\in V} \frac{1}{t^2}\cdot\sum_{\boldsymbol{w}\in\mathbb{Z}_s^n}\omega_{r_3}^{j\langle s\boldsymbol{u}_{\boldsymbol{w}}'+\boldsymbol{w},\boldsymbol{v}^{[r_2]}-\widetilde{\boldsymbol{v}}^{[r_2]}\rangle}\right) \\
&= \left(\sum_{\boldsymbol{w}\in\mathbb{Z}_s^n} p_{\boldsymbol{w}}^2\right) \cdot \left(\sum_{\boldsymbol{v},\widetilde{\boldsymbol{v}}\in V} \frac{1}{t^2}\cdot\sum_{\boldsymbol{w}\in\mathbb{Z}_s^n}\omega_{r_3}^{j\langle \boldsymbol{w},\boldsymbol{v}^{[r_2]}-\widetilde{\boldsymbol{v}}^{[r_2]}\rangle}\right) \\
&= \left(\sum_{\boldsymbol{w}\in\mathbb{Z}_s^n} p_{\boldsymbol{w}}^2\right) \cdot \left(\sum_{\boldsymbol{v},\widetilde{\boldsymbol{v}}\in V} \frac{1}{t^2}\cdot s^n\cdot \mathbb{1}_{\boldsymbol{v}^{[r_2]}\neq\widetilde{\boldsymbol{v}}^{[r_2]}}\right) \\
&= \left(\sum_{\boldsymbol{w}\in\mathbb{Z}_s^n} p_{\boldsymbol{w}}^2\right) \cdot \left(\sum_{\boldsymbol{w}\in\mathbb{Z}_s^n} q_{\boldsymbol{w}}^2\right)\cdot s^n.
\end{aligned}
\tag{7}
$$

In the last step we used the fact $\boldsymbol{v}^{[r_2]}\neq\widetilde{\boldsymbol{v}}^{[r_2]}$ for two $\boldsymbol{v},\widetilde{\boldsymbol{v}}\in V$. This can be seen by contrapositive. If $\boldsymbol{v}^{[r_2]}=\widetilde{\boldsymbol{v}}^{[r_2]}$, we have $\boldsymbol{v}=\widetilde{\boldsymbol{v}}$ since $\boldsymbol{v}^{(r_2)}=\widetilde{\boldsymbol{v}}^{(r_2)}$. This contradicts Claim 2.5.

By (7), we see either $\sum p_{\boldsymbol{w}}^2 \geq 1/(s^{n/2+2}f(s)^2)$ or $\sum q_{\boldsymbol{w}}^2 \geq 1/(s^{n/2+2}f(s)^2)$. Without loss of generality, assume $\sum p_{\boldsymbol{w}}^2 \geq 1/(s^{n/2+2}f(s)^2)$. By

$$
\max\{p_{\boldsymbol{w}}\} = \max\{p_{\boldsymbol{w}}\}\cdot\sum p_{\boldsymbol{w}} \geq \sum p_{\boldsymbol{w}}^2,
$$

there exists a bucket $\widetilde{B}(\boldsymbol{w}_0, U)$ with size at least $t/(s^{n/2+2}f(s)^2)$. Let $\widetilde{U}$ be that bucket, and $\widetilde{V}$ be the subset of V with the same indices. Then $(\widetilde{U},\widetilde{V})\subseteq (U,V)$ is an MV family of size at least $t/(s^{n/2+2}f(s)^2)$. Next, we will find a subfamily $(U',V')\subseteq(\widetilde{U},\widetilde{V})$ that respects $(r_1 s, r_2, r_3/s)$.

Step 4: Analyzing the elements in the large bucket.

Let $\boldsymbol{u}_0$ and $\boldsymbol{v}_0$ denote $\boldsymbol{u}^{(r_1)}$ and $\boldsymbol{v}^{(r_2)}$ respectively for $\boldsymbol{u} \in U, \boldsymbol{v} \in V$. For every $\boldsymbol{u} \in \widetilde{U}$, we know $(\boldsymbol{u}^{[r_1]} - \widetilde{\boldsymbol{u}}^{[r_1]})^{(s)}$ equals the same vector $\boldsymbol{w}_0$ by the definition of the bucket. Therefore $\boldsymbol{u}^{[r_1]} - \widetilde{\boldsymbol{u}}^{[r_1]} = s\boldsymbol{u}' + \boldsymbol{w}_0$ and

$$\boldsymbol{u} = r_1\boldsymbol{u}^{[r_1]} + \boldsymbol{u}_0 = r_1(\widetilde{\boldsymbol{u}}^{[r_1]} + s\boldsymbol{u}' + \boldsymbol{w}_0) + \boldsymbol{u}_0 = r_1 s\boldsymbol{u}' + \left(r_1\widetilde{\boldsymbol{u}}^{[r_1]} + r_1\boldsymbol{w}_0 + \boldsymbol{u}_0\right). \tag{8}$$

We can see $\boldsymbol{u}^{(r_1 s)} = r_1\widetilde{\boldsymbol{u}}^{[r_1]} + r_1\boldsymbol{w}_0 + \boldsymbol{u}_0$ is the same for all $\boldsymbol{u} \in \widetilde{U}$. Also $\boldsymbol{v}^{(r_2)} = \boldsymbol{v}_0$ is the same for all $\boldsymbol{v} \in \widetilde{V}$. These two conditions are still satisfied for any subfamily of $(\widetilde{U}, \widetilde{V})$. It suffices to find $(U', V') \subseteq (\widetilde{U}, \widetilde{V})$ such that

- $\langle \boldsymbol{u}^{[r_1 s]}, \boldsymbol{v}_0 \rangle$ modulo r_2 is the same for all $\boldsymbol{u} \in U'$. By (8) we have $\boldsymbol{u}^{[r_1 s]} = \boldsymbol{u}'$, so we need $\langle \boldsymbol{u}', \boldsymbol{v}_0 \rangle$ modulo r_2 to be the same for all $\boldsymbol{u} \in U'$.

- $\langle r_1\widetilde{\boldsymbol{u}}^{[r_1]} + r_1\boldsymbol{w} + \boldsymbol{u}_0, \boldsymbol{v}^{[r_2]} \rangle$ modulo $r_1 s$ is the same for all $\boldsymbol{v} \in V'$.

Since $\langle \boldsymbol{u}^{[r_1]}, \boldsymbol{v}_0 \rangle = \langle s\boldsymbol{u}' + \widetilde{\boldsymbol{u}}^{[r_1]} + \boldsymbol{w}, \boldsymbol{v}_0 \rangle$ modulo r_2 is the same for all $\boldsymbol{u} \in U$ by (U, V) respecting (r_1, r_2, r_3), we can see that $s\langle \boldsymbol{u}', \boldsymbol{v}_0 \rangle$ modulo r_2 is the same for all $\boldsymbol{u} \in U$. Hence there are $\gcd(s, r_2)$ possible values for $\langle \boldsymbol{u}', \boldsymbol{v}_0 \rangle$ modulo r_2. We pick the most frequent value c_1 and keep only the vectors with $\langle \boldsymbol{u}', \boldsymbol{v}_0 \rangle \equiv c_1 \pmod{r_2}$ in $\widetilde{U}$ and the corresponding vectors in $\widetilde{V}$.

Since $\langle \boldsymbol{u}_0, \boldsymbol{v}^{[r_2]} \rangle$ modulo r_1 is the same for all $\boldsymbol{v} \in V$ by (U, V) respecting (r_1, r_2, r_3), we can see that there are s possible values for $\langle \boldsymbol{u}_0, \boldsymbol{v}^{[r_2]} \rangle$ modulo sr_1. We pick the most frequent value c_2 and keep only the vectors with $\langle \boldsymbol{u}_0, \boldsymbol{v}^{[r_2]} \rangle \equiv c_2 \pmod{sr_1}$ in $\widetilde{U}$ and the corresponding vectors in $\widetilde{V}$.

After the above two steps, the MV family has size at least

$$\frac{|(\widetilde{U}, \widetilde{V})|}{\gcd(s, r_2) \cdot s} \geq \frac{|(\widetilde{U}, \widetilde{V})|}{s^2} \geq \frac{t}{s^{n/2+4} f(s)^2}. \tag{9}$$

And this is the required (U', V'). $\qquad\square$

4   Proof of Theorems 1.1 and 1.2

We now prove Theorem 1.1 by repeatedly applying Lemma 3.1.

Proof of Theorem 1.1. By Claim 2.10, (U, V) is good with respect to $(1, 1, m)$. Initially we set $r_1 = 1$, $r_2 = 1$ and $r_3 = m$. By Lemma 3.1, we there is a subfamily that respects (r_1', r_2', r_3'), where $r_1' r_2' r_3' = m$ and $r_3' < m$. We repeatedly apply Lemma 3.1. Each round r_3 is reduced by some factor. We can continue this procedure until either $r_3 = 1$ or the size of the MV family becomes less than $100m$. For the case $r_3 = 1$, the size of the MV family is also less than $100m$ by Claim 2.11. Say there are k rounds, and in each round we divide r_3 by $s_1, s_2, \ldots, s_k$ respectively. We have $s_1 s_2 \cdots s_k \leq m$ and in the ith round ($i \in [k]$), the size of the MV family is decreased by a factor at most $s_i^{n/2+4} f(s_i)^2$. Therefore the original size is upper bounded by

$$|(U, V)| \leq 100m \cdot \prod_{i=1}^{k} s_i^{n/2+4} f(s_i)^2 \leq 100m \cdot m^{n/2+4} \cdot \prod_{i=1}^{k} f(s_i)^2 = 100m^{n/2+5} \prod_{i=1}^{k} f(s_i)^2.$$

Pick $f(s) = s^{1.735}$, we can verify $\sum_{s=2}^{\infty} \frac{1}{f(s)} \leq 0.99$. Therefore $|(U, V)| \leq 100m^{n/2+5}(m^{1.735})^2 = 100m^{n/2+8.47}$. $\qquad\square$

Combining with the lower bound $m^{n-1+o_m(1)}$ proved in [DGY11], we can give a universal lower bound for the length of the MV code in [DGY11]. This is a restatement of Theorem 1.2 stated in the introduction.

Corollary 4.1. *Any MV code (as constructed in [DGY11]) has encoding length at least $N > K^{\frac{19}{18}}$, where K is the message length regardless of the query complexity.*

Proof. Given an MV family in $\mathbb{Z}_m^n$ with size t, we can encode a message of length $K = t$ into a codeword of length $N = m^n$.

If $n \geq 19$, by Theorem 1.1 we have $K \leq m^{n/2+8.47}$. Hence $K \leq m^{(1/2+8.47/19)n} < m^{\frac{18}{19}n} = N^{\frac{18}{19}}$ and $N > K^{\frac{19}{18}}$.

If $n \leq 18$, it was shown in [DGY11] that $K \leq m^{n-1+o_m(1)}$. Hence $K < m^{n-\frac{18}{19}} \leq m^{n-\frac{n}{19}} =$

$m^{\frac{18}{19}n} = N^{\frac{18}{19}}$ and $N > K^{\frac{19}{18}}$. Note that here we assumed m is sufficient large. This is reasonable because we are considering encoding an arbitrarily long message and K is sufficiently large. $\qquad\square$

5 The case of distinct prime factors

If m is a product of distinct primes, the bound can be improved to $m^{n/2+4+o_m(1)}$. The proof follows the same outline as general composite m.

Theorem 5.1. *Let m be a product of distinct primes. For every MV family (U, V) in $\mathbb{Z}_m^n$, $|(U, V)| \leq 100m^{n/2+4+o_m(1)}$, where $o_m(1)$ goes to 0 as m grows.*

Proof. The proof is similar to Theorem 1.1. We only sketch the changes here.

First, we improve the size of (U', V') found in Lemma 3.1 to $t/(s^{n/2+2}f(s)^2)$. Since m is a product of distinct primes, r_1 and r_2 must be coprime to s, where s is the number in inequality (3). Let τ_1 and τ_2 be integers that $\tau_1 r_1 \equiv 1 \pmod{s}$ and $\tau_2 r_2 \equiv 1 \pmod{s}$, we have

$$\omega_{r_3}^{j\frac{\langle u,v\rangle}{r_1 r_2}} = \omega_{r_3}^{j\langle u,v\rangle\tau_1\tau_2}.$$

We partition U and V into buckets according to $\boldsymbol{u}$ modulo s and $\boldsymbol{v}$ modulo s: $U = \bigcup\limits_{\boldsymbol{w}\in\mathbb{Z}_s^n} B(\boldsymbol{w}, U)$ and $V = \bigcup\limits_{\boldsymbol{w}\in\mathbb{Z}_s^n} B(\boldsymbol{w}, V)$, where

$$B(\boldsymbol{w}, U) = \{\boldsymbol{u}\in U \mid \boldsymbol{u}^{(s)} = \boldsymbol{w}\}$$

and

$$B(\boldsymbol{w}, V) = \{\boldsymbol{v}\in V \mid \boldsymbol{v}^{(s)} = \boldsymbol{w}\}.$$

We still use $p_{\boldsymbol{w}}$ to denote $|B(\boldsymbol{w}, U)|/t$ and $q_{\boldsymbol{w}}$ to denote $|B(\boldsymbol{w}, V)|/t$. By inequality (3),

$$
\begin{aligned}
\left(\frac{1}{sf(s)}\right)^2 &\leq \left| \underset{\substack{\boldsymbol{u} \sim U \\ \boldsymbol{v} \sim V}}{\mathbb{E}} \left[\omega_{r_3}^{j\langle \boldsymbol{u}, \boldsymbol{v}\rangle \tau_1 \tau_2} \right] \right|^2 \\
&= \left| \sum_{\boldsymbol{w} \in \mathbb{Z}_s^n} \sum_{\boldsymbol{v} \in V} p_{\boldsymbol{w}} \cdot \frac{1}{t} \cdot \omega_{r_3}^{j\langle \boldsymbol{w}, \boldsymbol{v}\rangle \tau_1 \tau_2} \right|^2 \\
&\leq \left(\sum_{\boldsymbol{w} \in \mathbb{Z}_s^n} p_{\boldsymbol{w}}^2 \right) \cdot \left(\sum_{\boldsymbol{w} \in \mathbb{Z}_s^n} \left| \sum_{\boldsymbol{v} \in V} \frac{1}{t} \cdot \omega_{r_3}^{j\langle \boldsymbol{w}, \boldsymbol{v}\rangle \tau_1 \tau_2} \right|^2 \right) \\
&= \left(\sum_{\boldsymbol{w} \in \mathbb{Z}_s^n} p_{\boldsymbol{w}}^2 \right) \cdot \left(\sum_{\boldsymbol{v}, \widetilde{\boldsymbol{v}} \in V} \frac{1}{t^2} \cdot \sum_{\boldsymbol{w} \in \mathbb{Z}_s^n} \omega_{r_3}^{j\langle \boldsymbol{w}, \boldsymbol{v}-\widetilde{\boldsymbol{v}}\rangle \tau_1 \tau_2} \right) \\
&= \left(\sum_{\boldsymbol{w} \in \mathbb{Z}_s^n} p_{\boldsymbol{w}}^2 \right) \cdot \left(\sum_{\boldsymbol{w} \in \mathbb{Z}_s^n} q_{\boldsymbol{w}}^2 \right) \cdot s^n.
\end{aligned}
$$

We can see either $\sum p_{\boldsymbol{w}}^2 \geq 1/(s^{n/2+1} f(s))$ or $\sum q_{\boldsymbol{w}}^2 \geq 1/(s^{n/2+1} f(s))$. Without loss of generality, assume $\sum p_{\boldsymbol{w}}^2 \geq 1/(s^{n/2+1} f(s))$. By

$$
\max\{p_{\boldsymbol{w}}\} = \max\{p_{\boldsymbol{w}}\} \cdot \sum p_{\boldsymbol{w}} \geq \sum p_{\boldsymbol{w}}^2,
$$

there exists a bucket with size $|B(\boldsymbol{w}, U)| \geq t/(s^{n/2+1} f(s))$. Let $\widetilde{U}$ be that bucket, and $\widetilde{V}$ be the subset of V with the same indices. Then $(\widetilde{U}, \widetilde{V}) \subseteq (U, V)$ is an MV family of size at least $t/(s^{n/2+1} f(s))$.

Next we find $(U', V') \subseteq (\widetilde{U}, \widetilde{V})$ using the same method as in Lemma 3.1. By inequality (9),

$$
|(U', V')| \geq \frac{|(\widetilde{U}, \widetilde{V})|}{\gcd(s, r_2) \cdot s} = \frac{|(\widetilde{U}, \widetilde{V})|}{s} \geq \frac{t}{s^{n/2+2} f(s)}.
$$

At last, we use the proof of Theorem 1.1 except $f(s) = \frac{1}{3s \ln^2 s}$. One can verify $\sum_{s=2}^{\infty} \frac{1}{f(s)} < 0.99$. Let $s_1, s_2, \ldots, s_k$ be the numbers divided from r_3 in each round, by the proof of Theorem 1.1,

$$
|(U, V)| \leq 100m \prod_{i=1}^{k} s_i^{n/2+2} f(s_i) \leq 100m^{n/2+3} \prod_{i=1}^{k} (3s_i \ln^2 s_i) \leq 100m^{n/2+4} \prod_{i=1}^{k} (3 \ln^2 s_i).
$$

15

For a sufficiently large integer s, we have $3 \ln^2 s < s^\epsilon$, where ϵ is an arbitrary fixed small number. When $m \to \infty$, all $s_1, s_2, \ldots, s_k$ except a constant number of them must be that large. Take $\epsilon \to 0$, we have $|(U, V)| \le m^{n/2 + 4 + o_m(1)}$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

References

[BBR94]   David A. Mix Barrington, Richard Beigel, and Steven Rudich. Representing boolean functions as polynomials modulo composite numbers. In *Computational Complexity*, pages 455–461, 1994.

[BDL13]   Abhishek Bhowmick, Zeev Dvir, and Shachar Lovett. New bounds on matching vector families. In *45th ACM Symposium on Theory of Computing (STOC)*, 2013.

[BET10]   Avraham Ben-Aroya, Klim Efremenko, and Amnon Ta-Shma. Local list decoding with a constant number of queries. In *51st IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 715–722, 2010.

[BF98]   László Babai and Peter Frankl. *Linear algebra methods in combinatorics*. 1998.

[CFL+10] Yeow Meng Chee, Tao Feng, San Ling, Huaxiong Wang, and Liang Feng Zhang. Query-efficient locally decodable codes of subexponential length. *Electronic Colloquium on Computational Complexity (ECCC)*, TR10-173, 2010.

[DGY11]   Zeev Dvir, Parikshit Gopalan, and Sergey Yekhanin. Matching vector codes. *SIAM J. Comput.*, 40(4):1154–1178, 2011.

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

[GKST02] Oded Goldreich, Howard Karloff, Leonard J. Schulman, and Luca Trevisan. Lower bounds for linear locally decodable codes and private information retrieval. In *17th IEEE Computational Complexity Conference (CCC)*, pages 175–183, 2002.

[Gro00]   Vince Grolmusz. Superpolynomial size set-systems with restricted intersections mod 6 and explicit ramsey graphs. *Combinatorica*, 20(1):71–86, 2000.

[IS10]   Toshiya Itoh and Yasuhiro Suzuki. Improved constructions for query-efficient locally decodable codes of subexponential length. *IEICE Transactions on Information and Systems*, E93-D(2):263–270, 2010.

[KdW04]   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*, 69(3):395–420, 2004.

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

[KY09]   Kiran S. Kedlaya and Sergey Yekhanin. Locally decodable codes from nice subsets of finite fields and prime factors of Mersenne numbers. *SIAM J. Comput.*, 38(5):1952–1969, 2009.

[Rag07]   Prasad Raghavendra. A note on Yekhanin's locally decodable codes. *Electronic Colloquium on Computational Complexity (ECCC)*, TR07-016, 2007.

[TV07]   Terrence Tao and Van H. Vu. *Additive Combinatorics*. 2007.

[Woo07]   David P. Woodruff. New lower bounds for general locally decodable codes. *Electronic Colloquium on Computational Complexity (ECCC)*, TR07-006, 2007.

[Woo10]   David P. Woodruff. A quadratic lower bound for three-query linear locally decodable codes over any field. In *Proceedings of the 13th international conference on Approximation, and 14 the International conference on Randomization, and combinatorial optimization: algorithms and techniques*, APPROX/RANDOM'10, pages 766–779, Berlin, Heidelberg, 2010. Springer-Verlag.

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

[YGK12]   Chen Yuan, Qian Guo, and Haibin Kan. A novel elementary construction of matching vectors. *Information Processing Letters*, 112(12):494–496, 2012.