On the Inner Product Predicate and a Generalization of Matching Vector Families

Balthazar Bauer[1], Jevgēnijs Vihrovs *[2], and Hoeteck Wee †[3]

[1]ENS, Paris, 45 Rue d'Ulm, 75005 Paris, France
[2]Centre for Quantum Computer Science, Faculty of Computing, University of Latvia, Raiņa 19, Riga, Latvia, LV-1586
[3]CNRS and ENS, Paris, 45 Rue d'Ulm, 75005 Paris, France

Abstract

Motivated by cryptographic applications such as predicate encryption, we consider the problem of representing an arbitrary predicate as the inner product predicate on two vectors. Concretely, fix a Boolean function P and some modulus q. We are interested in encoding x to $\vec{\mathbf{x}}$ and y to $\vec{\mathbf{y}}$ so that

$$P(x, y) = 1 \iff \langle \vec{\mathbf{x}}, \vec{\mathbf{y}} \rangle = 0 \bmod q,$$

where the vectors should be as short as possible. This problem can also be viewed as a generalization of matching vector families, which corresponds to the equality predicate. Matching vector families have been used in the constructions of Ramsey graphs, private information retrieval (PIR) protocols, and more recently, secret sharing.

Our main result is a simple lower bound that allows us to show that known encodings for many predicates considered in the cryptographic literature such as greater than and threshold are essentially optimal for prime modulus q. Using this approach, we also prove lower bounds on encodings for composite q, and then show tight upper bounds for such predicates as greater than, index and disjointness.

1 Introduction

There are many situations in cryptography where one is interested in computing some function F of a sensitive input x but the computational model is restricted so that only "simple" functions F can be directly computed. For instance, the entries of x may be encrypted so that only *affine* functions can be computed, or distributed between multiple non-interacting parties so that only *local* functions can be computed, or simply that we only know how to construct schemes for handling simple functions.

For all of these reasons, it is useful to be able to "encode" complex functions as simple functions. An extremely influential example of an "encoding" in the cryptographic literature is that of garbling

*Work done in part while interning at Centre for Quantum Technologies at NUS. Supported by the ERC Advanced Grant MQC.

†Work done in part while visiting Centre for Quantum Technologies at NUS. Supported in part by ERC Project aSCEND (H2020 639554).

schemes (or randomized encodings), which have found applications in many areas of cryptography and elsewhere (see [Yao82, FKN94, IK00, AIK06, App11, BHR12, PS03] and references therein).

In this work, we consider the problem of *inner product encoding*, namely, representing an arbitrary predicate as the inner product predicate on two vectors. Concretely, fix a Boolean function P (a predicate) and some modulus q (may be composite as well as prime). We are interested in mappings $x \mapsto \vec{\mathbf{x}}, y \mapsto \vec{\mathbf{y}}$ that map to vectors in $\mathbb{Z}_q^\ell$ such that for all x, y:

$$P(x, y) = 1 \iff \langle \vec{\mathbf{x}}, \vec{\mathbf{y}} \rangle = 0 \bmod q,$$

and ℓ is as small as possible. This notion is motivated by the study of predicate encryption in [KSW08], where q is typically very large, for instance, as large as the domains of P, and can also be viewed as a natural generalization of matching vector families to arbitrary predicates.

As an example, consider the equality predicate over $[n]$. Here, if $q = 2$, then it is not difficult to show that the vectors must have length $\Omega(n)$. On the other hand, if $q > n$, then it is sufficient to use vectors of length 2: the inner product of $(1, x)$ and $(y, -1)$ is 0 mod q iff $x = y$. More generally, for any predicate $P : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ and any prime $q \geq 2$, the "truth table" construction achieves vectors of length $\min\{|\mathcal{X}|, |\mathcal{Y}|\}$.

Interestingly, inner product predicate encoding for the equality predicate have been studied in combinatorics and complexity theory, where they are known as matching vector families. Moreover, matching vector families have found many applications, including the construction of Ramsey graphs, private information retrieval (PIR) protocols [Gro00, Yek08, Efr12, DGY11, DG15], and more recently, secret-sharing schemes [LVW17, LVW18, LV18]. Here, prior works showed that if q is a prime, then we must use vectors of length $\Omega(n^{\frac{1}{q-1}})$ [DGY11].

1.1 Our results

Our main results are nearly tight bounds for many predicates considered in the cryptographic literature such as greater than and threshold, for both prime and composite modulus q. In particular, we have the following results for prime modulus q:

- Greater than predicate for numbers in $[n]$ requires vectors of length n. This rules out the possibility of deriving the predicate encryption for range queries with $O(\sqrt{n})$ ciphertext and secret key sizes in [BW07] as a special case of inner product predicate encryption.

- Threshold for n-bit strings and threshold t requires vectors of length 2^{n-t+1}. This rules out the possibility of constructing full-fledged functional encryption schemes by carrying out FHE decryption in the lattice-based predicate encryption of Gorbunov, Vaikuntanathan and Wee [GVW15] using a pairing-based functional encryption scheme for the inner product predicate.

We then investigate encodings for composite q, specifically when q is a product of k distinct primes. In many cases, a lower bound of ℓ/k for composite q follows naturally if our method gives lower bound ℓ for prime q. For predicates such as greater than, index and disjointness, we are able to show tight lower and upper bounds for both prime and composite q. The full summary of upper and lower bounds is shown in Table 1, and the listed predicates are described in Section 3.

Finally, we also consider probabilistic inner product predicate encoding. For example, there is a probabilistic encoding of length $O((\log n)^2)$ for the greater than predicate for numbers in $[n]$, while any deterministic encoding must have length $\Omega(n)$, if q is prime.

predicate	q prime		q product of k primes	
	upper	lower	upper	lower
$\mathbf{EQ}_n$ [1]	$O(qn^{\frac{1}{q-1}})$	$\Omega(n^{\frac{1}{q-1}})$	$2^{\tilde{O}\left((\log n)^{1/k}\right)}$	$\Omega(\log n)$
$\mathbf{GT}_n$	n	n	n/k	n/k
$\mathbf{DISJ}_n$ [2], $\mathbf{INDEX}_n, \mathbf{NEQ}_n$	n	n	n/k	n/k
$\mathbf{ETHR}_n^t$ [2] [3]	$n+1$	$n/2$	$n+1$	$n/2k$
$\mathbf{MPOLY}_n^{d,q}$	n^d	n^d	n^d	n^d/k
$\mathbf{THR}_n^t$	n^{n-t+1}	2^{n-t+1}	n^{n-t+1}	$2^{n-t+1}/k$
$\mathbf{OR{-}EQ}_n^q$	2^n	2^n	2^n	$2^n/k$

Table 1: Summary of upper and lower bounds

Our lower bound technique. Our lower bound technique is remarkably simple. Suppose that q is prime and we can represent a predicate $P : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ as an inner product predicate on vectors of length r corresponding to mappings $x \mapsto \vec{\mathbf{x}}, y \mapsto \vec{\mathbf{y}}$. Following [BDL13], we consider a matrix F of dimensions $|\mathcal{X}| \times |\mathcal{Y}|$ over $\mathbb{Z}_q$ whose (x, y)'th entry is $\langle \vec{\mathbf{x}}, \vec{\mathbf{y}} \rangle \bmod q$. Then the matrix F has rank at most r, because we can write F as the product of two matrices of dimensions $|\mathcal{X}| \times r$ and $r \times |\mathcal{Y}|$. Concretely, $F = UV$ where the x'th row of U is $\vec{\mathbf{x}}^{\mathrm{T}}$ and the y'th column of V is $\vec{\mathbf{y}}$. This means that to show a lower bound on r, it suffices to show that F has large rank, e.g. by exhibiting a full rank submatrix.

As an example, consider the greater than predicate on $[n]$ for any prime modulus q. Then, the matrix F is an $n \times n$ upper triangular matrix where all the entries on and above the diagonal are non-zero. This matrix has rank n, hence any correct construction must have dimension at least n. Note that the above lower bound argument breaks down when q is composite. In fact, if $q = 2^n$, there is an encoding for greater than with dimension 1: take $x \mapsto 2^x, y \mapsto 2^{n-y}$. Correctness follows from the fact that $2^x \cdot 2^{n-y} = 0 \bmod 2^n \Leftrightarrow x \geq y$, and the construction extends also to the setting where q is a product of n distinct primes.

In order to extend our lower bounds to composite q that is the product of k distinct primes, we observe that if $F \bmod q$ contains a triangular submatrix of dimensions $\ell \times \ell$, then there exists some prime factor p of q such that $F \bmod p$ contains a triangular submatrix of dimensions $\ell/k \times \ell/k$; this follows from looking at the CRT decomposition of q and a pigeonhole argument. This simple observation allows us to translate many of our lower bounds to the composite modulus setting, which we prove to be essentially optimal via new upper bounds.

For instance, for the "greater than" predicate, we obtain a tight bound of n/k when q is a product of k distinct primes; this is sharp contrast to standard matching vector families (i.e., the equality predicate), where we have constructions of length $2^{\tilde{O}((\log n)^{1/k})}$ when q is a product of k distinct primes. For the upper bound, we begin with a construction of length 1 for $k = n$ and then derive the more general construction by treating the inputs as vectors of length n and then dividing that into n/k blocks each of length k.

Finally, we extend our results to the randomized setting. Here, we use a similar argument to show that the minimum size of a probabilistic inner product encoding is upper bounded by the

[1] Bounds from previous works, see Section 4.1 for references.
[2] For sufficiently large q.
[3] Assuming $t \leq n-2$, see Section 4.6.

probabilistic rank introduced by Alman and Williams [AW17].

Organization. The paper is organized as follows. We describe our lower bound method in Section 2. The notation and predicates used throughout the rest of the paper are defined in Section 3. In Section 4 we describe lower and upper bounds for these predicates. Finally, we consider probabilistic encodings in Section 5.

2 Main Theorem

In this section we describe our lower bound technique. Let $P : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a predicate, and $q \geq 2$ be the integer modulus. We say that a matrix $F : \mathcal{X} \times \mathcal{Y}$ *represents* P modulo q if for all $x \in \mathcal{X}, y \in \mathcal{Y}$, we have $F_{x,y} = 0 \bmod q$ iff $P(x,y) = 1$.

An *inner product encoding* of P of length ℓ is a pair of mappings from $\mathcal{X}, \mathcal{Y}$ to Z_q^ℓ that map x, y to $\vec{\mathbf{x}}, \vec{\mathbf{y}}$ in a way that the matrix $F : \mathcal{X} \times \mathcal{Y}$ defined by $F_{x,y} = \langle \vec{\mathbf{x}}, \vec{\mathbf{y}} \rangle \bmod q = (\sum_{i=1}^{\ell} \vec{\mathbf{x}}_i \cdot \vec{\mathbf{y}}_i) \bmod q$ represents P. Denote the length of the shortest reduction from P to inner product modulo q by $\mathrm{DI}(P,q)$ (Deterministic Inner product). Then we have the following simple and effective lower bound method.

Theorem 1. *For any predicate P and any prime $q \geq 2$, we have*

$$\mathrm{DI}(P,q) = \min_F \operatorname{rank}(F),$$

where F is any matrix that represents P modulo q.

Proof. We show that if P can be represented by a matrix F modulo q, then the necessary and sufficient length of the encoding from P to F is exactly $\operatorname{rank}(F)$. The decomposition rank definition states that the rank of an $m \times n$ matrix F is the smallest integer r such that F can be factored as $F = UV$, where U is an $m \times r$ matrix and V is a $r \times n$ matrix. Let $U_{x,*}$ be the row vector of U that corresponds to $x \in \mathcal{X}$ and $V_{*,y}$ be the column vector of V that corresponds to $y \in \mathcal{Y}$. Then the pair of mappings $x \mapsto U_{x,*}^{\mathrm{T}}$ and $y \mapsto V_{*,y}$ is a correct encoding of P, which is also the shortest possible for F. □

Therefore, to show a lower bound on the length of an encoding for P, it is sufficient to exhibit a set of rows R and a set of columns C such that for any matrix F that represents P, the submatrix $F[C,R]$ is a full rank submatrix. Typically we find a large full rank upper triangular submatrix and apply Theorem 1. Other times, we prove a lower bound for some predicate Q, and then prove that the same lower bound holds for P by showing a predicate reduction from Q to P (see Section 3 for details).

For composite q, we have the following lower bound:

Theorem 2. *Let $q = p_1 \cdots p_k$ be a product of k distinct primes. Let P be a predicate such that every matrix F that represents P modulo q is a triangular $n \times n$ matrix such that all numbers on the main diagonal are non-zero modulo q. Then*

$$\mathrm{DI}(P,q) \geq n/k.$$

Proof. Let F represent P modulo q. Let $F^{(i)} = F \bmod p_i$ (all entries taken modulo p_i). Since all entries on the main diagonal of F are non-zero, there exists $i \in [k]$ such that there at least n/k non-zero entries on the main diagonal of $F^{(i)}$ by pigeonhole principle. As $F^{(i)}$ is also a triangular matrix, the rank of $F^{(i)}$ modulo p_i is at least n/k. By Theorem 1, the length of any encoding from P to $F^{(i)}$ modulo p_i must be at least n/k, hence also $\mathrm{DI}(P, q) \geq n/k$. □

3 Definitions and Predicates

In this section, first we describe some of the notation used throughout the paper. Then we define the predicates examined in the paper, and define the predicate reduction.

Notation. We denote the set of all subsets of $[n]$ by $2^{[n]}$. For a set $S \subseteq [n]$, define the characteristic vector $\chi(S) \in \{0,1\}^n$ by

$$\chi(S)_i = \begin{cases} 1, & \text{if } i \in S, \\ 0, & \text{otherwise.} \end{cases}$$

Conversely, for a vector $x \in \{0,1\}^n$, let $\chi^{-1}(x)$ be the characteristic set of x.

For simplicity, denote the characteristic vector of $\{i\}$ by $\mathbf{e}_i$ (the length is usually inferred from the context). The characteristic vectors of $\varnothing$ and $[n]$ are denoted by 0^n and 1^n. We denote the identity matrix of dimension n by I_n, and all ones matrix by J_n.

For a truth expression T, we define $[T]$ to be 1 if T is true, and 0 if T is false. For example, $[x = y] = 1$ iff $x = y$.

For a number $x \in [2^n]$, let $\mathrm{bin}(x) \in \{0,1\}^n$ be the binary representation of $x - 1$.

Predicates. We consider the predicates listed below.

- Equality: $\mathcal{X} = \mathcal{Y} = [n]$ and $\mathbf{EQ}_n(x, y) = [x = y]$.
- Greater than: $\mathcal{X} = \mathcal{Y} = [n]$ and $\mathbf{GT}_n(x, y) = [x > y]$.
- Inequality: $\mathcal{X} = \mathcal{Y} = [n]$ and $\mathbf{NEQ}_n(x, y) = [x \neq y]$.
- Index: $\mathcal{X} = \{0,1\}^n, \mathcal{Y} = [n]$ and $\mathbf{INDEX}_n(x, i) = [x_i = 0]$. Here, x_i denotes the i'th coordinate of x. Note that we can also interpret x as the characteristic vector of a subset of $[n]$. Because in our model $0 \bmod q$ corresponds to "true", we have defined the index to be true if the bit value in the corresponding position is 0.
- Disjointness: $\mathcal{X} = \mathcal{Y} = 2^{[n]}$ and $\mathbf{DISJ}_n(S, T) = [S \cap T = \varnothing]$.
- Exact threshold: $\mathcal{X} = \mathcal{Y} = 2^{[n]}$ and $\mathbf{ETHR}_n^t(S, T) = [|S \cap T| = t]$, where $t \in [n]$ is the threshold parameter.
- Threshold: $\mathcal{X} = \mathcal{Y} = 2^{[n]}$ and $\mathbf{THR}_n^t(S, T) = [|S \cap T| \geq t]$, where $t \in [n]$ is the threshold parameter.
- Multilinear polynomials: $\mathcal{X} = \mathbb{Z}_q^n$, $\mathcal{Y} \subseteq \{p \mid p \in \mathbb{Z}_q[x_1, \ldots, x_n], \deg(p) \leq d\}$, the latter is the set of all multilinear polynomials of degree at most d. Then $\mathbf{MPOLY}_n^{d,q}(x, p) = [p(x_1, \ldots, x_n) = 0 \bmod q]$.
- Disjunction of equality tests: $\mathcal{X} = \mathcal{Y} = \mathbb{Z}_q^n$ and $\mathbf{OR{-}EQ}_n^q(x, y) = [\bigvee_{i=1}^n x_i = y_i]$.

Reductions We say that a predicate $P_1 : \mathcal{X}_1 \times \mathcal{Y}_1 \to \{0,1\}$ can be *reduced* to a predicate $P_2 : \mathcal{X}_2 \times \mathcal{Y}_2 \to \{0,1\}$ if there exist two mappings $f : \mathcal{X}_1 \to \mathcal{X}_2$ and $g : \mathcal{Y}_1 \to \mathcal{Y}_2$ such that $P_2(f(x), g(y)) = P_1(x, y)$ for all $x \in \mathcal{X}_1, y \in \mathcal{Y}_1$ (or mappings $f : \mathcal{X}_1 \to \mathcal{Y}_2$ and $g : \mathcal{Y}_1 \to \mathcal{X}_2$). In that case we write $P_2 \Rightarrow P_1$.

For example, consider the following reductions:

- $\mathbf{DISJ}_n \Rightarrow \mathbf{INDEX}_n \Rightarrow \mathbf{NEQ}_n$.

  The reduction $\mathbf{DISJ}_n \Rightarrow \mathbf{INDEX}_n$ holds since $\mathbf{INDEX}_n(x, i) = \mathbf{DISJ}_n(\chi^{-1}(x), \{i\})$. On the other hand, $\mathbf{INDEX}_n \Rightarrow \mathbf{NEQ}_n$, as $\mathbf{NEQ}_n(i, j) = \mathbf{INDEX}_n(\mathbf{e}_i, j)$.

- $\mathbf{INDEX}_n \Rightarrow \mathbf{GT}_n$.

  As $\mathbf{GT}_n(x, y) = \mathbf{INDEX}_n(\chi([y]), x)$, the reduction follows.

- Let $P : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be any predicate. Then $\mathbf{INDEX}_{\min\{|\mathcal{X}|,|\mathcal{Y}|\}} \Rightarrow P$.

  Let T be the $\mathcal{X} \times \mathcal{Y}$ truth table of P defined by $T_{x,y} = P(x, y)$. Then we have $P(x, y) = \mathbf{INDEX}_{|\mathcal{X}|}(T_x, y)$ and $\mathbf{INDEX}_{|\mathcal{X}|} \Rightarrow P$. Similarly, we also have $\mathbf{INDEX}_{|\mathcal{Y}|} \Rightarrow P$.

Effectively, then an inner product encoding for P_2 implies an encoding for P_1 and a lower bound for P_1 implies a lower bound for P_2. This makes it easier to prove upper and lower bounds. For example, as later we prove that $\mathrm{DI}(\mathbf{INDEX}_n, q) = n$ for prime q (see Section 4.2), the last reduction implies that $\mathrm{DI}(P, q) \le \min\{|\mathcal{X}|, |\mathcal{Y}|\}$ for all predicates P.

If q is a product of k distinct primes, then $\mathrm{DI}(P, q) \le \min\{|\mathcal{X}|, |\mathcal{Y}|\}/k$ for the same reason. Therefore, for any predicate, if $k = \min\{|\mathcal{X}|, |\mathcal{Y}|\}$, there is an encoding of $\mathcal{X}$ and $\mathcal{Y}$ simply to numbers modulo q.

4 Deterministic Encodings

In this section, we apply our technique to provide lower bounds on deterministic inner product encodings for many well-known predicates. For each of them, first we discuss the encodings and then proceed to prove lower bounds.

4.1 Equality

An encoding for $\mathbf{EQ}_n$ over q is a matching family of vectors modulo q [DGY11]. The maximum size of a matching family of vectors of length ℓ modulo q is denoted by $\mathrm{MV}(q, \ell)$ and has been studied extensively. Lower and upper bounds on $\mathrm{MV}(q, \ell)$ give upper and lower bounds on $\mathrm{DI}(\mathbf{EQ}_n, q)$, respectively (in the relevant literature, usually q and ℓ are denoted by m and n, respectively).

For prime q, a tight $\mathrm{DI}(\mathbf{EQ}_n, q) = \Theta(qn^{\frac{1}{q-1}})$ bound is known [DGY11]. If q is a product of k primes, we have a $2^{\tilde{O}\left((\log n)^{1/k}\right)}$ upper bound from [Gro00]. For any composite q, we also have an $\Omega(\log n)$ lower bound from [DH13].

Here, first we show two simple upper bounds for $q = 2$ and $q \ge n$. Then we reprove the optimal lower bound for $q = 2$ using our rank lower bound.

Upper bounds. For $q = 2$, we construct an encoding of length n. Let $\vec{\mathbf{x}} = \mathbf{e}_x$ and $\vec{\mathbf{y}} = 1^n - \mathbf{e}_y$. Then $\langle \vec{\mathbf{x}}, \vec{\mathbf{y}} \rangle = \langle \mathbf{e}_x, 1^n \rangle - \langle \mathbf{e}_x, \mathbf{e}_y \rangle = 1 - [x = y]$, thus it is a correct inner product encoding and $\mathrm{DI}(\mathbf{EQ}_n, 2) \leq n$.

Let q be any integer such that $q \geq n$. Let $\vec{\mathbf{x}} = (1, x)$ and $\vec{\mathbf{y}} = (y, -1)$. Then $\langle \vec{\mathbf{x}}, \vec{\mathbf{y}} \rangle = y - x$, so it is 0 iff $x = y$. Therefore, $\mathrm{DI}(\mathbf{EQ}_n, q) \leq 2$.

Lower bound. We show a matching lower bound for case $q = 2$. There is a unique matrix F over $\mathbb{Z}_2$ that represents $\mathbf{EQ}_n$, namely $F_{x,y} = 0 \bmod q \Leftrightarrow x = y$. Express $F = J_n - I_n$. By sub-additivity of rank, we have $\mathrm{rank}(F) \geq \mathrm{rank}(I_n) - \mathrm{rank}(J_n) = n - 1$. Hence, by Theorem 1, any inner product encoding of $\mathbf{EQ}_n$ modulo 2 requires vectors of length at least $n - 1$, that is, $\mathrm{DI}(\mathbf{EQ}_n, 2) \geq n - 1$.

4.2 Index

We prove that $\mathrm{DI}(\mathbf{INDEX}_n, q) = \lceil n/k \rceil$, for every q that is a product of k distinct primes.

For some q, the upper bound follows from $\mathbf{DISJ}_n \Rightarrow \mathbf{INDEX}_n$ (see Section 4.5). However, there is a much simpler encoding, which we present below. Moreover, this upper bound holds for every q that is the product of k distinct primes.

Upper bound. We begin with the warm-up for the special case $k = n$. Here, consider

$$\vec{\mathbf{x}} = \prod_{i=1}^{n} p_i^{1-x_i}, \qquad \vec{\mathbf{y}} = q/p_y.$$

Then $\langle \vec{\mathbf{x}}, \vec{\mathbf{y}} \rangle = 0 \bmod q$ iff $x_y = 0$.

Next, we consider general k, n. Since $\mathbf{INDEX}_{\lceil n/k \rceil \cdot k} \Rightarrow \mathbf{INDEX}_n$, it is enough to construct an encoding for the case $k \mid n$. The data is the string $x \in \{0,1\}^n$, and the index is given by $y \in [n]$. Encode x as an $n/k \times k$ binary matrix $X_{i,j} = x_{(i-1)\cdot k + j}$, and y as an $n/k \times k$ binary matrix $Y_{i,j} = [y = (i-1) \cdot k + j]$.

Now we construct the encoding.

$$\vec{\mathbf{x}}_i = \prod_{j=1}^{k} p_j^{X_{i,j}}, \qquad \vec{\mathbf{y}}_i = \begin{cases} q/p_j, & \text{if } Y_{i,j} = 1, \\ 0, & \text{otherwise.} \end{cases}$$

Now we analyze the correctness of the protocol. Let i, j be such that $Y_{i,j} = 1$. Then $\langle \vec{\mathbf{x}}, \vec{\mathbf{y}} \rangle = \prod_{l=1}^{k} p_l^{X_{i,l}} \cdot (q/p_j)$.

- If $X_{i,j} = 1$, then $\langle \vec{\mathbf{x}}, \vec{\mathbf{y}} \rangle = 0 \bmod q$.
- If $X_{i,j} = 0$, then $p_j \nmid \langle \vec{\mathbf{x}}, \vec{\mathbf{y}} \rangle$, hence $\langle \vec{\mathbf{x}}, \vec{\mathbf{y}} \rangle \neq 0 \bmod q$.

Lower bound. The lower bound follows from $\mathbf{INDEX}_n \Rightarrow \mathbf{NEQ}_n$ (see Section 4.3).

4.3 Inequality

We show that $\mathrm{DI}(\mathbf{NEQ}_n, q) = \lceil n/k \rceil$, for every q that is the product of k distinct primes.

Upper bound. The upper bound follows from $\mathbf{INDEX}_n \Rightarrow \mathbf{NEQ}_n$ (see Section 4.2).

Lower bound. Any matrix that represents $\mathbf{NEQ}_n$ is a diagonal matrix with non-zero entries on the main diagonal. By Theorem 2, it follows that $\mathrm{DI}(\mathbf{NEQ}_n, q) \geq n/k$.

4.4 Greater Than

We show that $\mathrm{DI}(\mathbf{GT}_n, q) = \lceil n/k \rceil$, for every q that is the product of k distinct primes.

Upper bound. The upper bound follows from $\mathbf{INDEX}_n \Rightarrow \mathbf{GT}_n$ (see Section 4.2).

If q is prime, the encoding simplifies to $\vec{\mathbf{x}} = \mathbf{e}_x$ and $\vec{\mathbf{y}} = \sum_{i=1}^{y} \mathbf{e}_i$. If $k = n$, a different simple encoding is $\vec{\mathbf{x}} = \prod_{i=1}^{x-1} p_i$ and $\vec{\mathbf{y}} = \prod_{i=y+1}^{n} p_i$.

Lower bound. Let F be any matrix that represents $\mathbf{GT}_n$ modulo q. Then all entries below the main diagonal are 0, while all entries on and above the main diagonal are non-zero, hence F is a triangular matrix. By Theorem 2, we conclude that $\mathrm{DI}(\mathbf{GT}_n, q) \geq n/k$.

4.5 Disjointness

We prove that $\mathrm{DI}(\mathbf{DISJ}_n, q) = \lceil n/k \rceil$ for an appropriate choice of q that depends on n, and that $\mathrm{DI}(\mathbf{DISJ}_n, q) \geq n/k$ if q is any product of k distinct primes.

Upper bound. We start with a simple encoding for $k = n$ that works for any product of n distinct primes q. Recall that the sets S and T are the input to disjointness. Let

$$\vec{\mathbf{x}} = \prod_{i=1}^{n} p_i^{1-\chi(S)_i}, \qquad \vec{\mathbf{y}} = \prod_{i=1}^{n} p_i^{1-\chi(T)_i}.$$

Then $\langle \vec{\mathbf{x}}, \vec{\mathbf{y}} \rangle = \prod_{i=1}^{n} p_i^{2-\chi(S)_i-\chi(T)_i}$ is $0 \bmod q$ iff S and T are disjoint. If $k < n$, then for any p_i it is possible that although some of the products $\vec{\mathbf{x}}_i \cdot \vec{\mathbf{y}}_i$ are not divisible by p_i, their sum might be divisible by p_i, hence the encoding doesn't work for any q.

For the general case, the following variation of Dirichlet's theorem will be useful for us.

Theorem 3 (Dirichlet)**.** *For any integer $q \geq 2$, there are infinitely many primes p such that $p = 1 \bmod q$.*

Let $q = p_1 \cdots p_k$ be a product of k distinct primes $p_1, \ldots, p_k$ to be defined later. We construct an encoding of length n/k for the case $k \mid n$. Encode $S \subseteq [n]$ as an $n/k \times k$ binary matrix $X_{i,j} = \chi(S)_{(i-1)\cdot k+j}$. Similarly encode T as Y. Let

$$\vec{\mathbf{x}}_i = \prod_{j=1}^{k} p_j^{1-X_{i,j}}, \qquad \vec{\mathbf{y}}_i = \prod_{j=1}^{k} p_j^{1-Y_{i,j}}.$$

Now we find the appropriate primes $p_1, \ldots, p_k$ for the general case. We construct them and prove the correctness by induction on k.

Base case. If $k = 1$, then q is a prime itself. Pick any prime q such that $q > n$. We have $\langle \vec{\mathbf{x}}, \vec{\mathbf{y}} \rangle = \sum_{i=1}^{n} q^{2-\chi(S)_i-\chi(T)_i}$. If x and y are disjoint, then $q \mid \langle \vec{\mathbf{x}}, \vec{\mathbf{y}} \rangle$. Suppose that S and T are not disjoint. Let $b = |S \cap T|$. Then $\langle \vec{\mathbf{x}}, \vec{\mathbf{y}} \rangle = (\sum_{i \in S \cap T} 1) \bmod q = b \bmod q$. As $b \leq n$, we have $b < q$, therefore $\langle \vec{\mathbf{x}}, \vec{\mathbf{y}} \rangle \neq 0 \bmod q$.

Inductive step. Assume that there exists a correct encoding for some q such that it is a product of $k-1$ distinct primes $p_1, \dots, p_{k-1}$. Let p_k be a prime such that $p_k > (n/k) \cdot (p_1 \cdots p_{k-1})^2$ and $p_k = 1 \bmod (p_1 \cdots p_{k-1})$ (such exist by Theorem 3).

Suppose that $\langle \vec{\mathbf{x}}, \vec{\mathbf{y}} \rangle = p_k \cdot a + b$, where $b \in \{0, \dots, p_k - 1\}$. Examine the sets $S^{(k)} = \{i \in [n/k] \mid ik \in S\}$ and $T^{(k)} = \{i \in [n/k] \mid ik \in T\}$.

- Suppose that $S^{(k)}$ and $T^{(k)}$ are not disjoint. Then the set $I = S^{(k)} \cap T^{(k)}$ is non-empty. If $i \notin I$, then at least one of $X_{i,k}$ and $Y_{i,k}$ is 0, thus $\vec{\mathbf{x}}_i \cdot \vec{\mathbf{y}}_i = \prod_{j=1}^{k} p_j^{2-X_{i,j}-Y_{i,j}}$ is divisible by p_k. Thus, we have that $b = \sum_{i \in I} \prod_{j=1}^{k-1} p_j^{2-X_{i,j}-Y_{i,j}} < (n/k) \cdot (p_1 \cdots p_{k-1})^2 < p_k$. Therefore, $\langle \vec{\mathbf{x}}, \vec{\mathbf{y}} \rangle = b \bmod p_k \neq 0 \bmod p_k$.
- Suppose that $S^{(k)}$ and $T^{(k)}$ are disjoint. Then for all $i \in [n/k]$, we have that $p_k \mid \vec{\mathbf{x}}_i \vec{\mathbf{y}}_i$. Therefore, $p_k \mid \langle \vec{\mathbf{x}}, \vec{\mathbf{y}} \rangle$.

  Moreover, since $p_k = 1 \bmod (p_1 \cdots p_{k-1})$, we have that $\vec{\mathbf{x}}_i \bmod (p_1 \cdots p_{k-1}) = \prod_{j=1}^{k-1} p_j^{1-X_{i,j}}$ and $\vec{\mathbf{y}}_i \bmod (p_1 \cdots p_{k-1}) = \prod_{j=1}^{k-1} p_j^{1-Y_{i,j}}$. Therefore, $\langle \vec{\mathbf{x}}, \vec{\mathbf{y}} \rangle \bmod (p_1 \cdots p_{k-1})$ is equal to 0 iff the sets $S \setminus S^{(k)}$ and $T \setminus T^{(k)}$ are disjoint by the inductive hypothesis.

Lower bound. The lower bound follows from $\mathbf{DISJ}_n \Rightarrow \mathbf{INDEX}_n$ (see Section 4.2).

4.6 Exact Threshold

Upper bound. The following encoding modulo $q \geq n$ of length $n+1$ is due to Katz, Sahai and Waters [KSW08]. For all $1 \leq i \leq n$, let $\vec{\mathbf{x}}_i = \chi(S)_i$, and let $\vec{\mathbf{x}}_{n+1} = 1$. For all $1 \leq i \leq n$, let $\vec{\mathbf{y}}_i = \chi(T)_i$, and let $\vec{\mathbf{y}}_{n+1} = -t$. Then $\langle \vec{\mathbf{x}}, \vec{\mathbf{y}} \rangle$ is equal to 0 iff $|S \cap T| = t$. Therefore, $\mathrm{DI}(\mathbf{ETHR}_n^t, q) \leq n+1$.

Surprisingly, if $t \geq n-1$, there exist constant size encodings.

- If $t = n$, there is an encoding of length 2. The encoding is as follows: $\vec{\mathbf{x}} = (1, [S = [n]])$ and $\vec{\mathbf{y}} = (1, -[T = [n]])$. Then we have $\langle \vec{\mathbf{x}}, \vec{\mathbf{y}} \rangle = 1 - [S = [n]] \cdot [T = [n]]$, which is 0 iff $S = T = [n]$.
- If $t = n-1$, there is an encoding of length 3. The encoding for S and T is as follows:

$$\vec{\mathbf{x}} = \begin{cases} (1,0,0), & \text{if } |S| = n, \\ (0,i,1), & \text{if } |S| = [n] \setminus \{i\}, \\ (1,-1,1), & \text{otherwise.} \end{cases} \qquad \vec{\mathbf{y}} = \begin{cases} (1,0,0), & \text{if } |T| = n, \\ (0,1,-i), & \text{if } |T| = [n] \setminus \{i\}. \\ (1,1,1), & \text{otherwise.} \end{cases}$$

  It is easy to check by hand that $\langle \vec{\mathbf{x}}, \vec{\mathbf{y}} \rangle = 0$ iff $|S \cap T| = n-1$. Note that we require $q \geq n+2$.

Lower bound. We show that for $1 \leq t \leq n-2$, we have $\mathrm{DI}(P, q) \geq \max\{n-t+2, t+2\}/k \geq (n/2+2)/k$.

(a) First we prove that if $t \geq 1$, the length of any encoding must be at least $(n-t+2)/k$. We show that by using two reductions.

    Firstly, we have $\mathbf{ETHR}_n^t \Rightarrow \mathbf{ETHR}_{n-t+1}^1$, because we can map $S \mapsto S \cup \{n-t+2, \dots, n\}$. Secondly, we prove that $\mathbf{ETHR}_m^1 \Rightarrow \mathbf{GT}_{m+1}$. Consider the following mappings:

$$f = \begin{cases} 1 \mapsto \varnothing, \\ i \mapsto [i-1], \end{cases} \qquad g = \begin{cases} j \mapsto \{j\}, \\ m+1 \mapsto \varnothing. \end{cases} \tag{1}$$

Consider a pair of numbers $x, y \in [m+1]$. If $x = 1$, then $\mathbf{GT}_{m+1}(x, y) = 0$ and also $\mathbf{ETHR}^1_m(f(x), g(y)) = \mathbf{ETHR}^1_m(\varnothing, g(y)) = 0$. If $y = m+1$, then $\mathbf{GT}_{m+1}(x, y) = 0$ and $\mathbf{ETHR}^1_m(f(x), g(y)) = \mathbf{ETHR}^1_m(f(x), \varnothing) = 0$. Otherwise, $\mathbf{ETHR}^1_m(f(x), g(y)) = \mathbf{ETHR}^1_m([x-1], \{y\}) = \mathbf{GT}_{m+1}(x, y)$. Hence the reduction is correct.

Therefore, we conclude that

$$\mathrm{DI}(\mathbf{ETHR}^t_n, q) \geq \mathrm{DI}(\mathbf{ETHR}^1_{n-t+1}, q) \geq \mathrm{DI}(\mathbf{GT}_{n-t+2}, q) \geq (n-t+2)/k$$

by the lower bound on greater than of Section 4.4.

(b) Now we prove that if $t \leq n-2$, the length of any encoding is at least $(t+2)/k$. Again, we exhibit two reductions.

Firstly, $\mathbf{ETHR}^t_n \Rightarrow \mathbf{ETHR}^t_{t+2}$ simply mapping any set to itself. Secondly, $\mathbf{ETHR}^{m-2}_m \Rightarrow \mathbf{NEQ}_m$. This is because we can map $x \mapsto [m] \setminus \{x\}$ for any $x \in [m]$. Then the size of the intersection $|([m] \setminus \{x\}) \cap ([m] \setminus \{y\})|$ is equal to $m-2$ if $x \neq y$, and $m-1$, if $x = y$.

Therefore, it follows that

$$\mathrm{DI}(\mathbf{ETHR}^t_n, q) \geq \mathrm{DI}(\mathbf{ETHR}^t_{t+2}, q) \geq \mathrm{DI}(\mathbf{NEQ}_{t+2}, q) \geq (t+2)/k$$

by the lower bound on inequality of Section 4.3.

Therefore, for any $1 \leq t \leq n-2$, any encoding must have length at least $\max\{n-t+2, t+2\}/k$ and we have that $\mathrm{DI}(\mathbf{ETHR}^t_n, q) = \Omega(n)$.

4.7 Multilinear Polynomials

First we show a known encoding that gives $\mathrm{DI}(\mathbf{MPOLY}^d_n, q) \leq \binom{n}{\leq d} = O(n^d)$. Then we show a lower bound of $\mathrm{DI}(\mathbf{MPOLY}^d_n, q) \geq \binom{n}{d}/k = \Omega(n^d/k)$. For prime q, we show an optimal lower bound $\mathrm{DI}(\mathbf{MPOLY}^d_n, q) \geq \binom{n}{\leq d}$.

Upper bound. The following is a simple construction by [KSW08]. For $S \subseteq [n]$, let $X_S = \prod_{i \in S} x_i$ and let $p = \sum_{S \subseteq [n], |S| \leq d} a_S X_S$ be a multilinear polynomial of degree at most d. For each subset $S \subseteq [n]$ such that $|S| \leq d$, let $\vec{\mathbf{x}}_S = X_S$ and $\vec{\mathbf{y}}_S = a_S$; then $\langle \vec{\mathbf{x}}, \vec{\mathbf{y}} \rangle$ is precisely equal to $p(x)$. Since a multilinear polynomial of degree at most d on n variables has at most $\binom{n}{\leq d} = \sum_{i=0}^{d} \binom{n}{i} \leq (n+1)^d$ monomials, it follows that $\mathrm{DI}(\mathbf{MPOLY}^d_n, q) = O(n^d)$.

Lower bound. We show a reduction $\mathbf{MPOLY}^d_n \Rightarrow \mathbf{NEQ}_{\binom{n}{d}}$. Let S be the bijection from the numbers in $\left[\binom{n}{d}\right]$ to subsets of $[n]$ of size d. For a pair of inputs $x, y \in \left[\binom{n}{d}\right]$, consider mappings $x \mapsto \chi(S(x))$ and $y \mapsto X_{S(y)}$. Since $\mathbf{MPOLY}^d_n(\chi(S(x)), X_{S(y)}) = 0$ iff $x \neq y$, it is a correct reduction. Thus, $\mathrm{DI}(\mathbf{MPOLY}^d_n, q) \geq \binom{n}{d}/k = \Omega(n^d/k)$ by the lower bound from Section 4.3.

Note that if q is prime, we can get a tight lower bound of $\binom{n}{\leq k}$. Since any two distinct polynomials disagree on some inputs, each polynomial must be mapped to a different vector. Therefore, the number of possible vectors must be at least the number of possible polynomials, $|\mathbb{Z}_q^n| \geq |\mathcal{Y}|$. The total number of possible monomials of degree at most d is $\binom{n}{\leq d}$. Each monomial can have any coefficient in $\mathbb{Z}_q$. It implies that $q^m \geq q^{\binom{n}{\leq d}}$ and $m \geq \binom{n}{\leq d}$.

4.8 Threshold

First we show an upper bound of $\mathrm{DI}(\mathbf{THR}_n^t, q) = O(n^{n-t+1})$ for $q \geq n$, and then a lower bound of $\mathrm{DI}(\mathbf{THR}_n^t, q) \geq 2^{n-t+1}/k$.

Upper Bound. The idea is to encode the threshold into multilinear polynomial evaluation. Let $x = \chi(S)$ and $y = \chi(T)$. Examine the following polynomial:

$$p_y(x) = \left(\sum_{i=1}^{n} x_i y_i - t\right) \cdot \left(\sum_{i=1}^{n} x_i y_i - (t+1)\right) \cdot \ldots \cdot \left(\sum_{i=1}^{n} x_i y_i - n\right).$$

Firstly, $\sum_{i=1}^{n} x_i y_i = |S \cap T|$, thus $p_y(x) = 0$ iff $|S \cap T| \geq t$. Secondly, the degree of each factor is 1, hence $\deg(p_y) = n-t+1$. Note that the polynomial p_y is still multilinear, since all the variables are 0 or 1. Therefore, we have a reduction $\mathbf{MPOLY}_n^{n-t+1} \Rightarrow \mathbf{THR}_n^t$. The upper bound from Section 4.7 implies that $\mathrm{DI}(\mathbf{THR}_n^t, q) \leq \mathrm{DI}(\mathbf{MPOLY}_n^{n-t+1}, q) \leq \binom{n}{\leq n-t+1} = O(n^{n-t+1})$.

Lower Bound. First of all, we have $\mathbf{THR}_n^t \Rightarrow \mathbf{THR}_{n-t+1}^1$, as we can map a set $S \subseteq [n-t+1]$ to $S \cup \{n-t+2, \ldots, n\}$. Next we prove that $\mathrm{DI}(\mathbf{THR}_m^1, q) \geq 2^m/k$.

Let F be any matrix representing $\mathbf{THR}_m^1$. We show that F is a triangular matrix with all entries on the main diagonal being non-zero. Then the claim follows by Theorem 2.

Order the rows of F by the increasing order of the size of the sets they correspond to. Then order the columns of F in such a way that the sets corresponding to the i-th row and the i-th column are the complements of each other.

As the complements don't overlap, the numbers on the main diagonal of F are non-zero. Now examine any entry on the i-th row and j-th column such that $i \geq j$. Let S correspond to the set of the i-th row and T correspond to the set of the j-th column. Since the columns are ordered by the decreasing size of the sets, we have that $|S| \geq m - |T|$, or equivalently $|S| + |T| \geq m$.

If $|S| + |T| > m$, then the sets must overlap and the value of $F_{i,j}$ is 0. If $|S| + |T| = m$, then the only way S and T do not overlap is if T is the complement of S. In any case all the numbers below the main diagonal are 0, and non-zero on the main diagonal. See Figure 1 for an example.

$$\begin{array}{c} \\ 000 \\ 001 \\ 010 \\ 100 \\ 011 \\ 101 \\ 110 \\ 111 \end{array}
\begin{array}{c}
\begin{array}{cccccccc} 111 & 110 & 101 & 011 & 100 & 010 & 001 & 000 \end{array} \\
\begin{bmatrix}
* & * & * & * & * & * & * & * \\
0 & * & 0 & 0 & * & * & 0 & * \\
0 & 0 & * & 0 & * & 0 & * & * \\
0 & 0 & 0 & * & 0 & * & * & * \\
0 & 0 & 0 & 0 & * & 0 & 0 & * \\
0 & 0 & 0 & 0 & 0 & * & 0 & * \\
0 & 0 & 0 & 0 & 0 & 0 & * & * \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & *
\end{bmatrix}
\end{array}$$

Figure 1: An example of F for $m = 3$. Stars represent non-zero elements.

4.9 Disjunctions of Equality Tests

We show that for prime q, we have $\mathrm{DI}(\mathbf{OR{-}EQ}_n^q, q) \leq 2^n$ and if q is a product of k distinct primes, then $\mathrm{DI}(\mathbf{OR{-}EQ}_n^q, q) \geq 2^n/k$.

Upper bound. We prove that $\mathbf{MPOLY}_n^{n,q} \Rightarrow \mathbf{OR{-}EQ}_n^q$. Examine a multilinear polynomial

$$p_y(x) = \prod_{i=1}^{n}(x_i - y_i).$$

Clearly, $p_y(x) = 0 \bmod q$ iff at least one equality holds. Therefore, if we map $x \mapsto x$ and $y \mapsto p_y$, then we have a correct reduction to multilinear polynomial evaluation. By the upper bound from Section 4.7, we have $\mathrm{DI}(\mathbf{OR{-}EQ}_n^q, q) \leq \mathrm{DI}(\mathbf{MPOLY}_n^{n,q}, q) \leq \sum_{i=0}^{n}\binom{n}{i} = 2^n$.

Lower bound. We prove that $\mathbf{OR{-}EQ}_n^q \Rightarrow \mathbf{NEQ}_{2^n}$. For the input $x, y \in [2^n]$ to $\mathbf{NEQ}_n$, map $x \mapsto \mathrm{bin}(x)$ and $y \mapsto \mathrm{bin}(y) \oplus 1^n$. As $x \neq y$ iff there exists an i such that $\mathrm{bin}(x)_i \neq \mathrm{bin}(y)_i$, we have that $x \neq y$ iff $\mathbf{OR{-}EQ}_n^q(\mathrm{bin}(x), \mathrm{bin}(y) \oplus 1^n) = 1$. The lower bound follows by Section 4.3.

5 Randomized Constructions

We can formulate the problem in the randomized setting as follows. Let $P : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a predicate. Consider all pairs of mappings $\mathcal{U} = \{(x \mapsto \vec{\mathbf{x}}, y \mapsto \vec{\mathbf{y}}) \mid \vec{\mathbf{x}}, \vec{\mathbf{y}} \in \mathbb{Z}_q^\ell \text{ for some } \ell\}$. These also include mappings that are incorrect inner product encodings of P. Let μ be a probability distribution over $\mathcal{U}$. Then μ is a *probabilistic inner product encoding* modulo q with error ϵ, if

$$\Pr[P(x,y) \neq [\langle \vec{\mathbf{x}}, \vec{\mathbf{y}} \rangle = 0 \bmod q] \mid (x \mapsto \vec{\mathbf{x}}, y \mapsto \vec{\mathbf{y}}) \sim \mu] \leq \epsilon.$$

We consider the length of the longest encoding under μ to be the length of μ and denote it by $\mathrm{RI}^\mu(P, q)$ (Randomized Inner product). Then define

$$\mathrm{RI}_\epsilon(P, q) = \min_\mu \mathrm{RI}^\mu(P, q),$$

where μ ranges over all probabilistic inner product encodings of P modulo q with error ϵ.

Next we reproduce the definition of the probabilistic rank (over $\mathbb{Z}_q$) by Alman and Williams [AW17]:

Definition 1 (Probabilistic Matrix)**.** For $n, m \in \mathbb{N}$, define a *probabilistic matrix* over $\mathbb{Z}_q$ to be a distribution of matrices $\mathcal{M} \subset \mathbb{Z}_q^{n \times m}$. A probabilistic matrix $\mathcal{M}$ *computes* a matrix $A \in \mathbb{Z}_q^{n \times m}$ with error $\epsilon > 0$ if for every entry $(i,j) \in [n] \times [m]$,

$$\Pr_{M \sim \mathcal{M}}[A_{i,j} \neq M_{i,j}] \leq \epsilon.$$

Definition 2 (Probabilistic Rank)**.** Let q be prime. Then a probabilistic matrix $\mathcal{M}$ has *rank* r if the maximum rank of an M in support of $\mathcal{M}$ is r. Define the ϵ*-probabilistic rank* of a matrix $A \in \mathbb{Z}_q^{n \times m}$ to be the minimum rank of a probabilistic matrix computing M with error ϵ. Denote it by $\mathrm{rank}_\epsilon(A)$.

As we can see, the probabilistic choice of a distribution μ corresponds to a matrix M sampled from $\mathcal{M}$. By a similar reasoning as in the proof of Theorem 1, we have the following theorem:

Theorem 4. *For any predicate P, prime $q \geq 2$ and error ϵ, we have*

$$\mathrm{RI}_\epsilon(P, q) \leq \min_F \mathrm{rank}_\epsilon(F),$$

where F is any matrix that represents P modulo q.

Proof. Let F be any matrix that represents P modulo q. Suppose that $\mathcal{M}$ is a probabilistic matrix that computes F. Then any M in support of $\mathcal{M}$ defines an encoding of length $\mathrm{rank}(M)$ by the decomposition rank. Therefore, there is a probability distribution over the encodings such that the maximum length is $\mathrm{rank}_\epsilon(F)$. □

For some predicates, the probabilistic rank can be much smaller than the deterministic rank. Let $T(P)$ be a truth table of a predicate P (defined by $T(P)_{x,y} = P(x, y)$). The same authors prove that $\mathrm{rank}_\epsilon(T(\mathbf{EQ}_n)) = O(1/\epsilon)$ and $\mathrm{rank}_\epsilon(T(\mathbf{LEQ}_n)) = O((\log n)^2/\epsilon)$ (see Lemmas D.1 and D.2 in [AW17]). Since the matrix $T(P)$ represents the predicate $\neg P$ (in our setting), these results imply that for any prime q:

1. $\mathrm{RI}_\epsilon(\mathbf{NEQ}_n, q) = O(1/\epsilon)$,
2. $\mathrm{RI}_\epsilon(\mathbf{GT}_n, q) = O((\log n)^2/\epsilon)$.

We conclude by showing that these results immediately imply a constant length probabilistic encoding for $\mathbf{EQ}_n$ modulo any prime:

Corollary 5. *For any prime q, we have $\mathrm{RI}_\epsilon(\mathbf{EQ}_n, q) = O(1/\epsilon)$.*

Proof. Let $\mathcal{M}$ be a probabilistic matrix that computes $T(\mathbf{EQ}_n)$ with error ϵ. The matrix $F(\mathbf{EQ}_n) = J_n - T(\mathbf{EQ}_n)$ represents $\mathbf{EQ}_n$. Therefore, the probabilistic matrix $J_n - \mathcal{M}$ computes $F(\mathbf{EQ}_n)$ with error ϵ. Since $\mathrm{rank}(F(\mathbf{EQ}_n)) \leq 1 + \mathrm{rank}(T(\mathbf{EQ}_n))$, we have that $\mathrm{RI}(\mathbf{EQ}_n, q) = O(1/\epsilon)$. □

6 Acknowledgments

We thank Srijita Kundu, Swagato Sanyal and Alexander Belov for helpful discussions, and Krišjānis Prūsis for suggestions on the presentation. We greatly thank Miklos Santha for hospitality during our stay at CQT, and Andris Ambainis for support. We are grateful to the anonymous reviewers for suggestions and useful feedback, and for pointing out the precise lower bound for prime q in Section 4.7.

References

[AIK06] Benny Applebaum, Yuval Ishai, and Eyal Kushilevitz. Cryptography in NC^0. *SIAM J. Comput.*, 36(4):845–888, 2006.

[App11] Benny Applebaum. Randomly encoding functions: A new cryptographic paradigm - (invited talk). In *ICITS*, pages 25–31, 2011.

[AW17] Josh Alman and Ryan Williams. Probabilistic rank and matrix rigidity. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2017, pages 641–652, New York, NY, USA, 2017. ACM.

[BDL13] Abhishek Bhowmick, Zeev Dvir, and Shachar Lovett. New bounds for matching vector families. In *STOC*, pages 823–832, 2013.

[BHR12] Mihir Bellare, Viet Tung Hoang, and Phillip Rogaway. Foundations of garbled circuits. In *ACM CCS*, 2012. Also Cryptology ePrint Archive, Report 2012/265.

[BW07] Dan Boneh and Brent Waters. Conjunctive, subset, and range queries on encrypted data. In *TCC*, pages 535–554, 2007.

[DG15] Zeev Dvir and Sivakanth Gopi. 2-server PIR with sub-polynomial communication. In *STOC*, pages 577–584, 2015.

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

[DH13] Zeev Dvir and Guangda Hu. Matching-vector families and LDCs over large modulo. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 513–526, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[Efr12] Klim Efremenko. 3-query locally decodable codes of subexponential length. volume 41, pages 1694–1703, 2012.

[FKN94] Uriel Feige, Joe Kilian, and Moni Naor. A minimal model for secure computation. In *STOC*, pages 554–563, 1994.

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

[GVW15] Sergey Gorbunov, Vinod Vaikuntanathan, and Hoeteck Wee. Predicate encryption for circuits from LWE. In *CRYPTO (2)*, pages 503–523, 2015. Also, Cryptology ePrint Archive, Report 2015/029.

[IK00] Yuval Ishai and Eyal Kushilevitz. Randomizing polynomials: A new representation with applications to round-efficient secure computation. In *FOCS*, pages 294–304, 2000.

[KSW08] Jonathan Katz, Amit Sahai, and Brent Waters. Predicate encryption supporting disjunctions, polynomial equations, and inner products. In *EUROCRYPT*, pages 146–162, 2008.

[LV18] Tianren Liu and Vinod Vaikuntanathan. Breaking the circuit-size barrier in secret sharing. STOC 2018. Cryptology ePrint Archive, Report 2018/333, 2018.

[LVW17] Tianren Liu, Vinod Vaikuntanathan, and Hoeteck Wee. Conditional disclosure of secrets via non-linear reconstruction. In *Advances in Cryptology - CRYPTO 2017 - 37th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 20-24, 2017, Proceedings, Part I*, pages 758–790, 2017.

[LVW18] Tianren Liu, Vinod Vaikuntanathan, and Hoeteck Wee. Towards breaking the exponential barrier for general secret sharing. In *Advances in Cryptology - EUROCRYPT 2018 - 37th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tel Aviv, Israel, April 29 - May 3, 2018 Proceedings, Part I*, pages 567–596, 2018.

[PS03] Manoj Prabhakaran and Amit Sahai. *Secure Multi-Party Computation.* IOS Press, 2003.

[Yao82] Andrew Chi-Chih Yao. Theory and applications of trapdoor functions. In *FOCS*, pages 80–91, 1982.

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