Communication complexity by compressed sensing and Bose-Chowla theorem

Yang Liu[*] Shengyu Zhang[*]

Abstract

Communication complexity of XOR functions $f(x \oplus y)$ has attracted increasing attention in recent years, because of its connections to Fourier analysis, and its exhibition of exponential separations between classical and quantum communication complexities of total functions. However, the complexity of certain basic functions still seems elusive especially in the private-coin SMP model. In particular, an exponential gap exists between quantum upper and lower bounds for deciding whether x and y have Hamming distance at least d, despite the sequence of related efforts [GKdW04, HSZZ06, ZS09] since Yao asked it as an open question [Yao03]. In this paper we resolve this question by providing optimal randomized and quantum protocols.

We then apply the result and show efficient protocols for all symmetric XOR functions and linear threshold functions, answering an open question in [LLZ11] and another one in [MO10]. Finally, we consider matrix functions and show upper bounds for the matrix rank decision problem; the public-coin randomized SMP result matches the quantum two-way lower bound in [SW12].

Motivated from data sketching applications, we aim at efficiency of *computation* besides communication. Our protocols for the matrix rank decision problem are computationally efficient in the classical setting, and other protocols are computationally efficient if Alice and Bob have quantum computers or non-uniform classical circuits. The main techniques used in our protocols are compressed sensing, and the Bose-Chowla theorem from combinatorial number theory. To the best of our knowledge, this is the first time that these two techniques are applied to communication complexity, in which we believe that more applications could be found.

[*] Department of Computer Science and Engineering and The Institute of Theoretical Computer Science and Communications, The Chinese University of Hong Kong. Email: `{yliu, syzhang}@cse.cuhk.edu.hk`

1 Introduction

Communication complexity Communication complexity studies the minimum amount of communication needed for a computation task with input variables distributed to two or more parties. Since the seminal work [Yao79], communication complexity has attracted a great deal of attention mainly because of its connections to many other computational settings. Various modes of computation are studied, including deterministic, randomized and quantum, whose corresponding communication complexities are denoted by D, R, Q, respectively. For randomized and quantum communication protocols, it is often allowed to have a small error probability ϵ. The error bound ϵ is specified by a subscript, and usually omitted if $\epsilon = 1/3$. Different communication models are also investigated, such as two-way (Alice and Bob send messages back and forth), one-way (Alice sends one message to Bob) and SMP (Alice and Bob each send one message to a third party Referee) models. In the randomized and quantum models, Alice and Bob may share public randomness or quantum entanglement, indicated by superscripts *pub* and $*$, respectively. One also uses superscripts "1" or "$\|$" to specify that the communication model is one-way or SMP, respectively.

An important class of functions is that of XOR functions: $F(x,y) = f(x \oplus y)$ for some function f on $\{0,1\}^n$. On one hand, the functions have the obvious symmetry across all rows (and all columns) of the communication matrix $M_F = [F(x,y)]_{x,y}$, which can indeed be used to show some results. For example, the rank of the communication matrix $M_{f\circ\oplus}$ is nothing but $\|\hat{f}\|_0$, the number of nonzero Fourier coefficients. The logarithm of an approximate version of $\|\hat{f}\|_1$ serves as a lower bound of $\mathsf{Q}(f \circ \oplus)$ [LS09]. On the other hand, we still do not know how to use the structure to derive more results in communication complexity, especially in designing efficient protocols. For example, the notorious logrank conjecture, even restricted to XOR functions, still remains unsolved despite recent efforts [ZS09, MO10, KS13, TWXZ13], and communication complexity for some very basic functions such as Hamming Distance is still unknown in some models.

Data sketching The SMP model has an intimate relation to data sketching in computation over massive data sets, where a data set is partitioned and distributed to two or more parties. Each party computes a compressed "sketch" of the stored data, and then sends it to a central processing unit that uses only the sketches to complete certain tasks. This corresponds well to the SMP model in communication complexity, except that it requires not only communication efficiency but also *computational* efficiency, for all parties and the central processing unit.

Like in communication complexity, data sketching also has public-coin and private-coin variants. The private-coin model is of interest because public randomness is not always available in practical applications. Even if it is, the public randomness may be accessed by an adversary with the aim to attack the system [MNS11]. For example, sometimes the data input is given by the adversary, who knows the public randomness in advance; in this case, the adversary can assign the input for which the randomness gives the wrong answer. See [MNS11] for a more detailed review of the public-coin model and an introduction to the private-coin model with some results on specific functions (which are actually also XOR functions).

In view of the connection to data sketching, we want to design communication and time efficient protocols in SMP models, in both public-coin and private-coin variants, with an emphasis on the latter. In this paper, we will show a number of protocols that are efficient in both communication and computation. The communication costs in most of our protocols match known lower bounds for communication complexity even without the computational limit.

Next we explain our results in more details. (Also see Appendix A for a tabular summary.)

Hamming Distance and symmetric functions One function of particular interest is *Hamming Distance*, denoted $\mathsf{Ham}_{n,d}$, which decides whether the two given n-bit strings differ at less than d locations. When $d = 1$, the function becomes $\mathsf{Equality}$ function, one of the most studied functions in communication complexity in all models [Yao79, NS96, Amb96, BK97, BCWdW01].

Previous results on $\mathsf{Ham}_{n,d}$ are summarized as follows. For lower bounds, even quantum two-way protocols with shared entanglement need $\Omega(d)$-qubits of communication [HSZZ06]; for upper bounds, if public coins are available, there are randomized SMP protocols with $O(d \log d)$ bits [HSZZ06], improving upon previous results [Yao03, GKdW04]. Therefore all the communication complexities are pinned down (up to a small factor of $\log(d)$), except in the private-coin SMP model, at which we now take a closer look. For $\mathsf{R}^{\|}(\mathsf{Ham}_{n,d})$, the best known lower bound is $\Omega(\sqrt{n})$, obtained by a simple reduction to the $\mathsf{Equality}$ function whose complexity $\mathsf{R}^{\|}$ is known to be $\Theta(\sqrt{n})$. The best known upper bound for $\mathsf{R}^{\|}(\mathsf{Ham}_{n,d})$ is $O(\sqrt{n} d \log d)$, by a folklore result $\mathsf{R}^{\|}(F) = O(\sqrt{n} \cdot \mathsf{R}^{\|,pub}(F))$. For the quantum setting, the best known upper bound is $2^{\tilde{O}(d)}$, obtained by applying a general transformation $\mathsf{Q}^{\|}(F) = 2^{O(\mathsf{R}^{\|,pub}(F))} \log n$ ([Yao03]). Closing the exponential gap between lower and upper bounds for $\mathsf{Q}^{\|}(\mathsf{Ham}_{n,d})$ was an open question asked by Yao [Yao03], and [LLZ11] also asked about $\mathsf{R}^{\|}(\mathsf{Ham}_{n,d})$.

In this paper, we completely pin down both randomized and quantum communication complexities of $\mathsf{Ham}_{n,d}$ in the private-coin SMP model (up to a log factor). To our surprise, the dependence of the classical complexity on d is only *additive*.

Theorem 1. *The randomized and quantum communication complexities for $\mathsf{Ham}_{n,d}$ in the private-coin SMP model are tightly bounded as follows.*

$$\Omega(\sqrt{n} + d) \leq \mathsf{R}^{\|}(\mathsf{Ham}_{n,d}) \leq O(\sqrt{n} + d \log n), \qquad \Omega(d) \leq \mathsf{Q}^{\|}(\mathsf{Ham}_{n,d}) \leq O(d \log n).$$

In addition, the quantum upper bound can be achieved by a protocol with encoding and decoding both in poly(n) time. The classical upper bound can be achieved with encoding and decoding by a non-uniform family of circuits of poly(n) size.

Apart from the fact that Hamming distance is a fundamental concept, understanding of the communication complexity of $\mathsf{Ham}_{n,d}$ turns out to be crucial to study the class of symmetric XOR functions. Suppose that $f(z) = D(|z|)$ for some function $D : \{0, 1, \dots, n\} \to \{0, 1\}$. Define r_0 and r_1 to be the minimum integers such that $r_0, r_1 \leq n/2$ and $D(k) = D(k + 2)$ for all $k \in [r_0, n - r_1)$; set $r = \max\{r_0, r_1\}$. It was proven in [ZS09] that $\Omega(r) \leq \mathsf{Q}^*(f \circ \oplus) \leq \mathsf{R}(f \circ \oplus) \leq \tilde{O}(r)$ and $\mathsf{R}^1(f \circ \oplus) = \tilde{O}(r^2)$. The last bound was further improved in [LLZ11] which shows that even $\mathsf{R}^{\|,pub}(f \circ \oplus) = O(r \log^3 r / \log \log r)$, leaving the private-coin SMP complexity, both randomized and quantum, as an open question. In this paper, we answer it by showing (almost) tight bounds for both complexities. We also slightly improve the upper bound for $\mathsf{R}^{\|,pub}(f \circ \oplus)$ when r is large.

Theorem 2. *For any symmetric function $f : \{0, 1\}^n \to \{0, 1\}$,*

$$\Omega(r) \leq \mathsf{R}^{\|,pub}(f \circ \oplus) \leq O(r \log n), \qquad \Omega(\sqrt{n} + r) \leq \mathsf{R}^{\|}(f \circ \oplus) \leq O(\sqrt{n} + r \log n),$$

$$\Omega(r) \leq \mathsf{Q}^{\|}(f \circ \oplus) \leq O(r \log n).$$

In addition, the quantum upper bound can be achieved by a protocol with encoding and decoding both in poly(n) time. The classical upper bound can be achieved with encoding and decoding by a non-uniform family of circuits of poly(n) size.

Two remarks about the theorem: First, in [Yao03], Yao also asked whether $\mathsf{Q}^{\|}(F) = O(\mathsf{R}^{\|,pub}(F)\log n)$ for all Boolean functions F. Though we could not answer the question in its full generality, the above theorem does confirm it for all symmetric XOR functions. Second, the upper bounds in the theorem actually hold even if f is not symmetric on strings with Hamming weight smaller than r_0 or larger than $n - r_1$: f can take arbitrary values on those strings, and our protocols still work with the same complexity. This naturally leads to question of what can be said about non-symmetric functions. We next extend our study to beyond the scope of symmetric functions.

Linear threshold functions The first class of non-symmetric functions that we study contains linear threshold functions (LTF), which can be viewed as an extension of the Hamming Distance function by allowing different weights on different variables. To be more precise, a linear threshold function $f : \{0,1\}^n \to \{0,1\}$ with weights $\{w_i\}$ and threshold θ is defined by $f(z) = 1$ iff $\sum_i w_i z_i \geq \theta$. Define the *margin* m to be the smallest gap between θ and the possible weighted summation: $m_b = \min_{z:f(z)=b} |\sum_i w_i z_i - \theta|$, and let $m = \min\{m_0, m_1\}$. Montanaro and Osborne studied the public-coin communication complexity of LTFs in the SMP model, and proved that $\mathsf{R}^{\|,pub}(f \circ \oplus) = O((\theta/m)^2)$. The authors asked whether the bound can be improved to $\tilde{O}((\theta/m))$. In this paper, we answer this affirmatively by showing that $\mathsf{R}^{\|,pub}(f \circ \oplus) = O((\theta/m)\log n)$. Furthermore, the same bound holds for the private-coin SMP model as well, if quantum messages are used.

Theorem 3. *For any LTF f of threshold θ and margin m, we have that*

$$\mathsf{Q}^{\|}(f \circ \oplus) = O\Big(\frac{\theta}{m}\log n\Big), \quad \mathsf{R}^{\|,pub}(f \circ \oplus) = O\Big(\frac{\theta}{m}\log n\Big) \quad \textit{and} \quad \mathsf{R}^{\|}(f \circ \oplus) = O\Big(\frac{\theta}{m}\log n + \sqrt{n}\Big).$$

In addition, the quantum upper bound can be achieved by a protocol with encoding and decoding both in poly(n) time. The classical upper bound can be achieved with encoding and decoding by a non-uniform family of circuits of poly(n) size.

Matrix functions The second class of functions that we study beyond symmetric functions are those on matrices. The input is an $n \times n$ matrix and the functions are invariant to row and column permutations, but not arbitrary permutations of all entries, thus it contains less symmetries than the aforementioned symmetric XOR functions. A typical example is the rank decision function. Define function $\mathbb{F}\text{-}\mathtt{rank}_{n,r} : \mathbb{F}^{n\times n} \to \{0,1\}$ by $\mathbb{F}\text{-}\mathtt{rank}_{n,r}(X,Y) = 1$ iff the matrix $X+Y$ has rank less than r, where the rank and the summation $X+Y$ are both over $\mathbb{F}$. When $\mathbb{F} = \mathbb{F}_2$, the function is an XOR function.

In [SW12], Sun and Wang studied the communication complexity of deciding whether $X+Y$ is full rank, where the input and computation are over $\mathbb{F}_p$. They showed a tight quantum lower bound of $\Omega(n^2 \log p)$, which also implies a quantum lower bound of $\Omega(r^2 \log p)$ for $\mathbb{F}_p\text{-}\mathtt{rank}_{n,r}$. In this paper, we show that this bound is tight, and can be achieved even by randomized SMP protocols with public coins. For private-coin SMP models, we give upper bounds as follows. Note that thanks to the explicit and efficient compressed sensing for low-rank matrices in [FS12], we can have computational efficiency in both quantum *and classical* cases.

Theorem 4. *For $f = \mathbb{F}_q\text{-}\mathtt{rank}_{n,r}$, $\mathsf{D}(f \circ +) = \Theta(n^2 \log q)$, and the public-coin randomized and quantum communication complexities in two-way, one-way, SMP models are all of the order $\Theta(r^2 \log q)$. For the SMP private-coin complexities, we have*

$$\Omega(n\sqrt{\log q} + r^2 \log q) \leq \mathsf{R}^{\|}(f \circ +) \leq O(nr \log q + n \log n),$$

$$\Omega(r^2 \log q) \leq \mathsf{Q}^{||}(f \circ +) \leq \min\{q^{O(r^2)}, O(nr \log q + n \log n)\},$$

and the protocols can be constructed explicitly with poly(n) encoding and decoding time, with communication slightly increased to $\mathsf{R}^{||} = O(nr \log(q+n))$ *and* $\mathsf{Q}^{||} = \min\{q^{O(r^2)}, O(nr \log(q+n))\}$.

Techniques The main techniques used in our protocols are compressed sensing and the Bose-Chowla theorem from combinatorial number theory. To the best of our knowledge, this is the first time that these two techniques are applied to communication complexity. It seems to us that these two techniques are particularly useful for designing protocols with both computational and communication efficiency, and we believe that they will find more applications in communication protocol designing.

Compressed sensing is usually used to recover sparse vectors and low rank matrices over $\mathbb{R}$. Our task is to recover sparse vectors and low rank matrices over finite fields. For low rank matrices over finite field, explicit and efficient schemes were discovered very recently [FS12]. For sparse recovery, unfortunately, to the best of our knowledge, there is no explicit and efficient construction known; existing schemes [BDF+11, DeV07, HAN10, Ind08] all need significantly more than $O(k \log n)$. [1]

2 Preliminaries and notation

For an n-bit string $x \in \{0,1\}^n$, we use $|x|$ to denote its Hamming weight, namely the number of 1's. For a matrix $M \in \mathbb{F}_2{}^{m \times n}$, denote $ker(M)$ the kernel, namely the subspace of $\mathbb{F}_2{}^n$ mapped to 0 by M. Denote the image space of M by $Im(M)$.

A function $F(x,y)$ on $\{0,1\}^n \times \{0,1\}^n$ is an *XOR function* if $F(x,y) = f(x \oplus y)$ for some function f on n-bit strings, where $x \oplus y$ is the bit-wise XOR of x and y. An XOR function is symmetric if f is symmetric, namely $f(z)$ depends only on the number of 1's in z, or equivalently, $f(z) = D(|z|)$ for some function $D : \{0,1,\ldots,n\} \to \{0,1\}$. Define r_0 and r_1 to be the minimum integers such that $r_0, r_1 \leq n/2$ and $D(k) = D(k+2)$ for all $k \in [r_0, n-r_1)$; set $r = \max\{r_0, r_1\}$. By definition, $D(k)$ depends only on the parity of k when $k \in [r_0, n-r_1)$. Suppose $D(k) = T(\mathsf{Parity}(k))$ for $k \in [r_0, n-r_1)$.

An important symmetric XOR function is the *Hamming Distance* function, defined as follows. $\mathsf{Ham}_{n,d}(x,y) = 1$ if $|x \oplus y| < d$ and $\mathsf{Ham}_{n,d}(x,y) = 0$ if $|x \oplus y| \geq d$.

The class of linear threshold functions (LTF) contains those $f : \{0,1\}^n \to \{0,1\}$ defined by $f(z) = 1$ if $\sum_i w_i z_i \geq \theta$ and $f(z) = 0$ if $\sum_i w_i z_i < \theta$, where $\{w_i\}$ are the weights and θ is the threshold. Define $W_0 = \max_{z:f(z)=0} \sum_i w_i z_i$ and $W_1 = \min_{z:f(z)=1} \sum_i w_i z_i$, and define $m_0 = \theta - W_0$ and $m_1 = W_1 - \theta$. The *margin* of f is $m = \max\{m_0, m_1\}$. Note that the function remains the same if $\{w_i\}$ are fixed and θ varies in $(W_0, W_1]$. Thus without loss of generality, we can assume that $\theta = (W_0 + W_1)/2$, in which case $m_0 = m_1 = m$.

We will use the following communication complexity results.

Theorem 5 (Babai-Kimmel, [BK97]). $\mathsf{R}^{||}(f) = \Omega\Big(\sqrt{\mathsf{D}^{||}(f)}\Big)$.

Lemma 6 (Forklore). $\mathsf{R}^{||}(f) = O\big(\sqrt{n} \cdot \mathsf{R}^{||,pub}(f)\big)$.

[1] Actually, [FS12] also gives a general (and efficient) transform from sparse recovery to low rank recovery, so if there were efficient sparse recovery schemes, their technical construction of low rank discovery would have been a simple corollary.

Theorem 7 (Yao, [Yao03]). $\mathsf{Q}^{\|}(f) = 2^{O(\mathsf{R}^{\|,pub}(f))} \log n$.

For the last theorem, while the original paper only stated the result for total functions f, it is not hard to verify that the transformation (from a public-coin randomized protocol to a private-coin quantum protocol) works for partial functions as well, and the transformation is explicitly constructed. Similarly, Lemma 6 also applies to partial functions.

3 Communication and computation efficient protocols for the Hamming Distance problem

In this section, we shall show two efficient protocols for the $\mathsf{Ham}_{n,d}$ problem. Assume that $d = o(n)$; otherwise the lower bound already gives $\mathsf{Q}^*(\mathsf{Ham}_{n,d}) = \Theta(n)$. Both protocols need to break the problem into two promise problems

$$\mathsf{Ham}_{n,d|2d}(x,y) = \begin{cases} 1 & \text{if } |x \oplus y| < d \\ 0 & \text{if } |x \oplus y| > 2d \end{cases}, \quad \mathsf{Ham}_{n,d,2d}(x,y) = \begin{cases} 1 & \text{if } |x \oplus y| < d \\ 0 & \text{if } d \leq |x \oplus y| \leq 2d \end{cases}.$$

In an SMP protocol, the players run protocols for these two promise problems. If the answer to $\mathsf{Ham}_{n,d|2d}(x,y)$ is 0, then Referee outputs 0. Otherwise Referee outputs the result of the protocol for $\mathsf{Ham}_{n,d,2d}(x,y)$. A case-by-case analysis (dividing possible inputs into cases of $|x \oplus y| \leq d$, $d < |x \oplus y| \leq 2d$, $|x \oplus y| \geq 2d$) gives the correctness of the protocol.

It is shown in [HSZZ06] that $\mathsf{R}^{\|,pub}(\mathsf{Ham}_{n,d|2d}) = O(1)$, and the general transformations from randomized public-coin SMP protocol to randomized and quantum private-coin protocols (Lemma 6 and Theorem 7, respecitvely) can be applied on it to get randomized and quantum private-coin protocols, thereby showing

$$\mathsf{R}^{\|}(\mathsf{Ham}_{n,d|2d}) = O(\sqrt{n}), \quad \mathsf{Q}^{\|}(\mathsf{Ham}_{n,d|2d}) = O(\log n).$$

Therefore, both upper bounds in Theorem 1 are established as long as we can prove that

$$\mathsf{R}^{\|}(\mathsf{Ham}_{n,d,2d}) = O(d \log n).$$

We will give two protocols for $\mathsf{R}^{\|}(\mathsf{Ham}_{n,d,2d})$ in the next two subsections, both actually accomplishing a more difficult mission of computing the *entire* string $x \oplus y$ (under the condition that $|x \oplus y| \leq 2d$) rather than just computing its Hamming weight. In addition, both protocols are actually *deterministic*. Thus for the original problem $\mathsf{Ham}_{n,d}$, all the randomized or quantum parts in the corresponding optimal protocols are actually in distinguishing $|x \oplus y| \leq d$ and $|x \oplus y| > 2d$.

3.1 Protocol for $\mathsf{D}^{\|}(\mathsf{Ham}_{n,d,2d})$ using compressed sensing

We first give a protocol that is simple, but computationally inefficient. It has m strings $r_1, ..., r_m$, each of n bits, to be specified later. The protocol is as follows.

1. Alice: Send $E(x) = (\langle x, r_1 \rangle, ..., \langle x, r_m \rangle)$.
2. Bob: Send $E(y) = (\langle y, r_1 \rangle, ..., \langle y, r_m \rangle)$.
3. Referee: Decode $(x \oplus y)$ from $E(x \oplus y) = E(x) \oplus E(y)$.

Now we analyze the protocol, during the process of which we will also determine $r_1, ..., r_m$. For Referee to be able to recover $x \oplus y$ from $E(x \oplus y)$, one needs that all strings $z \in \{0,1\}^n$ with Hamming weight at most $k = 2d$ have different codewords $E(z)$. Thus it is enough to show that all nonzero strings $z \in \{0,1\}^n$ with Hamming weight at most $2k$ have nonzero $E(z)$. We will show the existence of such r_i's by a probabilistic argument. Consider choosing r_i uniformly at random, then for any nonzero $z \in \{0,1\}^n$, the probability of $E(z) = 0$ is exactly 2^{-m}. Now a union bound gives

$$\mathbf{Pr}[\exists z \text{ with } |z| \le 2k \text{ s.t. } E(z) = 0] \le \sum_{i=1}^{2k} \binom{n}{i} 2^{-m}$$

which is strictly smaller than 1, if $m > ck \log n$ for some constant c (recall that $k = 2d = o(n)$). So there exists a fixed choice of $(r_1, ..., r_m)$ for Alice and Bob to use in the deterministic protocol.

Though the protocol achieves the optimal communication complexity, it is not explicit and not computationally efficient: The r_i's are not explicitly given, and Referee needs to check a whole list of $\sum_{i=1}^{2k} \binom{n}{i}$ codewords of all possible low-weight strings, to decode $x \oplus y$. If $d = \omega(1)$, then this decoding cannot be done in $poly(n)$ time.

3.2 Protocol for $\mathsf{D}^{\|}(\mathsf{Ham}_{n,d,2d})$ using Bose-Chowla theorem

In this section, we will give a protocol which achieves the optimal communication complexity and is computationally efficient. We will need the following theorem in combinatorial number theory.

Theorem 8 (Bose and Chowla, [BC62])**.** *For any prime p and integer $n > 0$, we can find $q = p^n$ strictly positive integers $d_1, d_2, \dots, d_q$, each less than q^k, such that the sums*

$$d_{i_1} + d_{i_2} + \cdots + d_{i_k} \mod (q^k - 1)$$

for all possible $1 \le i_1 \le i_2 \cdots \le i_k \le q$ are distinct.

Now we explain how to use this theorem to give a protocol $\mathcal{P}_k$ that can compute $x \oplus y$ using $k \log n$ bits, when given a promise that $|x \oplus y| \le k$. This would then give a deterministic protocol for $\mathsf{Ham}_{n,d,2d}(x, y)$ using only $O(d \log n)$ bits. The protocol is as follows. Fix an integer $q \in [n, 2n)$ s.t. q is a power of 2.

1. Alice: Send $|x|$, and $H_x = \sum_{i:x_i=1} d_i \mod (q^k - 1)$ to Referee.
2. Bob: Send $|y|$, and $H_y = \sum_{i:y_i=1} d_i \mod (q^k - 1)$ to Referee.
3. Referee: Use the messages from Alice and Bob to exactly recover $x \oplus y$.

The last step is not fully specified, but the decoding will be clear from the proof of the following theorem (the proof is deferred to Appendix B).

Theorem 9. *The protocol $\mathcal{P}_k$ always computes $x \oplus y$ correctly with $O(k \log n)$ bits of communication. The computation on all parties can be implemented in polynomial time on a quantum computer, and by a non-uniform family of classical circuits of size $poly(n)$.*

4 Applications to symmetric XOR functions and LTFs

In this section, we apply the efficient protocols for the $\mathsf{Ham}_{n,d}$ problem in the last section to solve more general classes of functions. The first class contains all symmetric XOR functions, addressed in Section 4.1. The second class contains linear threshold functions, addressed in Section 4.2.

4.1 Symmetric XOR functions

In this section we prove Theorem 2. Recall that when $|x \oplus y| \in [r_0, n-r_1)$, $f(x \oplus y) = T(\mathsf{Parity}(x \oplus y))$ for some function T which depends only on the parity of number of 1's in $x \oplus y$. The protocol $\mathcal{P}(f, r)$ is described as follows.

```
1. Run the best protocol for $\mathsf{Ham}_{n,r_0}$ on input $(x, y)$, and one for $\mathsf{Ham}_{n,r_1+1}$ on input $(\bar{x}, y)$.
2. Run the deterministic protocol $\mathcal{P}_{r_0}$ on input $(x, y)$, and $\mathcal{P}_{r_1}$ on input $(\bar{x}, y)$.
3. Alice and Bob also send $\mathsf{Parity}(x)$ and $\mathsf{Parity}(y)$, respectively.
4. if the outcomes of Step 1 imply that $|x \oplus y| < r_0$ or $|x \oplus y| = n - |\bar{x} \oplus y| \geq n - r_1$
5.     Apply $f$ on $x \oplus y$ computed in Step 2.
6. else
7.     Output $T(\mathsf{Parity}(x) \oplus \mathsf{Parity}(y))$.
```

The correctness of the protocol follows from that of individual protocols it uses, and we omit details here. The communication cost also depends on the inner protocols. More specifically, for public-coin randomized protocol, the three communication steps take $O(d \log d)$, $O(d \log n)$ and $O(1)$ bits, respectively, thus the overall cost is $O(d \log n)$. For private-coin randomized protocol, the three communication steps take $O(\sqrt{n} + d \log n)$, $O(d \log n)$ and $O(1)$ bits, respectively, thus the overall cost is $O(\sqrt{n} + d \log n)$. For private-coin quantum protocol, the three communication steps take $O(d \log n)$, $O(d \log n)$ and $O(1)$ bits, respectively, thus the overall cost is $O(d \log n)$. This proves the upper bounds in Theorem 2.

4.2 Linear threshold functions

Recall that a linear threshold function $f : \{0,1\}^n \to \{0,1\}$ with weights $\{w_i\}$ and threshold θ is defined by $f(z) = 1$ if $\sum_i w_i z_i \geq \theta$ and $f(z) = 0$ otherwise. Also recall from Section 2 that θ can be assumed to be the middle point of the closest pair of inputs with different function values, and the margin is half of their distance.

We will first show that all the weights w_i can be assumed to be at least the margin m. For a string $z \in \{0,1\}^n$ and a subset $S = \{i_1, \ldots, i_s\} \subseteq [n]$ with $i_1 < \cdots < i_s$, the restriction of z on S is $z_S = z_{i_1} \ldots z_{i_s}$. (The proof is deferred to Appendix B.)

Lemma 10. *For any LTF f of n variables with weights $\{w_i : i \in [n]\}$, threshold θ and margin m, we can select a subset $S \subseteq [n]$ s.t. the LTF f' on variables z_S with the same weight $\{w_i : i \in S\}$ and threshold θ, has the new margin $m' \geq m$ and weights $w_i \geq m$ for all $i \in S$, and f' is consistent with f in the strong sense that $f'(z_S) = f(z)$, for all z.*

Next we will use a fact in [MO10] about the random projections in $\mathbb{F}_2$.

Lemma 11 (Montanaro and Osborne, [MO10])**.** *For any $z \in \{0,1\}^n$, if we draw a random variable $r \in \{0,1\}^n$ by picking the i-th bit to be 1 with probability $p_i = (1 - (1 - \frac{1}{\theta})^{w_i})/2$, then $\mathbf{E}[\langle r, z \rangle] = (1 - (1 - \frac{1}{\theta})^{\sum_i w_i z_i})/2$, where the inner product is over $\mathbb{F}_2$.*

Now we are ready to prove Theorem 3.

Proof. By Lemma 10, we get a function f' and a set S with all weights $w_i \geq m$ for $i \in S$. The protocol is designed for f', which also gives the correct answer to f. Similar to the one for symmetric XOR functions, the protocol breaks the inputs into two cases by two thresholds θ and 2θ. Define $f'_{\theta|2\theta}(z_S) = 1$ if $w(z_S) < \theta$ and $f'_{\theta|2\theta}(z_S) = 0$ if $w(z_S) > 2\theta$. First, it is not hard to see that $\mathsf{R}^{\|,pub}(f'_{\theta|2\theta}) = O(1)$. Actually, if $f'_{\theta|2\theta}(z_s) = 0$, then $w(z_S) > 2\theta$ and thus $\mathbf{Pr}[\langle r, z_S\rangle] > (1-(1-1/\theta)^{2\theta})/2$, and if $f'_{\theta|2\theta}(z_S) = 1$, then $\mathbf{Pr}[\langle r, z_S\rangle] < (1-(1-1/\theta)^{\theta})/2$. Since there is a constant gap between the two bounds, we can use constant samples to estimate the value $\mathbf{Pr}[\langle r, z_S\rangle]$ to distinguish these two cases, as long as public coins are available to get the r's. By Lemma 6 and Theorem 7, we know that $\mathsf{R}^{\|}(f'_{\theta|2\theta})) = O(\sqrt{n})$ and $\mathsf{Q}^{\|}(f'_{\theta|2\theta})) = O(\log n)$.

Now we assume that $w(z_S) \leq 2\theta$. Then

$$\sum_{i\in S} z_i \leq \frac{\sum_{i\in S} w_i z_i}{\min_{i\in S} w_i} \leq \frac{\sum_{i\in S} w_i z_i}{m} \leq \frac{2\theta}{m},$$

where in the second inequality we used Lemma 10. Thus we can use the randomized private-coin protocol for Hamming Distance, with promise that $z_S = x_S \oplus y_S$ has Hamming weight at most $2\theta/m$, to completely pin down z_S, which also gives us $f'(z_S) = f(z)$. This part can be done by deterministic communication of $O(\frac{\theta}{m}\log|S|) = O(\frac{\theta}{m}\log n)$ bits. Putting the two parts of communication cost together, we get the desired bounds. The computational efficiency follows from that of the protocol for $\mathsf{D}^{\|}(\mathsf{Ham}_{n,d,2d})$ using the Bose-Chowla theorem. □

5 Communication complexity for functions on matrices

In this section we study the rank decision function. Recall that $\mathbb{F}_q$-$\mathtt{rank}_{n,d}$ is the function defined by $\mathbb{F}$-$\mathtt{rank}_{n,r}(X,Y) = 1$ if $\mathtt{rank}(X+Y) < r$ and $\mathbb{F}$-$\mathtt{rank}_{n,r}(X,Y) = 0$ otherwise. In Appendix B.3, we show tight lower bound for deterministic communication complexity.

5.1 Public-coin randomized communication complexity

Theorem 12. $\mathsf{R}^{\|,pub}(\mathbb{F}_q\text{-}\mathtt{rank}_{n,r}) = O(r^2 \log q)$.

Proof. The protocol $\mathcal{P}(r)$ is given as follows.

1. Alice and Bob use public coins to sample k random matrices $L_1, \cdots, L_k \in {\mathbb{F}_q}^{r\times n}$ and k random matrices $R_1, \cdots, R_k \in {\mathbb{F}_q}^{n\times r}$ uniformly and independently, where $k = O(1)$ is specified later.
2. Alice sends $L_i X R_i, i = 1, \cdots, k$ to Referee.
3. Bob sends $L_i Y R_i, i = 1, \cdots, k$ to Referee.
4. Referee checks the values $\mathtt{rank}(L_i X R_i + L_i Y R_i), i = 1, \cdots, k$. If all these values are less than r, output "1"; otherwise output "0".

Correctness Let $Z = X + Y$. If $\mathtt{rank}(Z) < r$, then $\mathtt{rank}(L_i X R_i + L_i Y R_i) = \mathtt{rank}(L_i Z R_i) < r$ for all $i = 1, \cdots, k$ for sure. We claim that if $\mathtt{rank}(Z) \geq r$, then all matrices $L_i Z R_i$ have rank at least r with a constant probability. Indeed, for each i, $\mathtt{rank}(Z R_i) = r$ if and only if the r columns of R_i are linearly independent and all outside $ker(Z)$. Since $\mathtt{rank}(Z) \geq r$, we have that $\mathtt{rank}(ker(Z)) \leq n - r$. Thus the probability that the r columns of R_i are linearly independent and all outside $ker(Z)$ is

$$\frac{q^n - q^{n-r}}{q^n}\frac{q^n - q^{n-r+1}}{q^n}\cdots\frac{q^n - q^{n-1}}{q^n} > \prod_{i=1}^{\infty}(1-q^{-i}) \overset{\text{def}}{=} c(q), \tag{1}$$

where $c(q) > 1/4$ is a constant for all q. Using the same argument for $L_i(ZR_i)$, we have that conditioned on $\mathtt{rank}(ZR_i) = r$, the probability that $\mathtt{rank}(L_iZR_i) = r$ is also greater than $c(q)$. Thus

$$\Pr(\mathtt{rank}(L_iZR_i) = r) \geq c(q)^2 > 1/16. \tag{2}$$

By setting $k = 160$, we have that with probability $1 - e^{-10}$, at least one of $\mathtt{rank}(L_iXR_i + L_iYR_i)$ is r. This completes the proof for the correctness.

Cost The communication cost of the protocol is $2k = O(1)$ times the length of a message L_iXR_i, which has r^2 entries each from $\mathbb{F}_q$. Thus the total communication cost is $O(r^2 \log q)$. The computational cost is clearly $poly(n)$. □

5.2 Private-coin SMP

We first give a randomized private-coin SMP protocol which is simpler to understand but not explicit or efficient.

Theorem 13. $\mathsf{R}^{||}(\mathbb{F}_q\text{-}\mathtt{rank}_{n,r}) = O(rn\log q + n\log n)$.

Proof. Let $m = 4rn$. Fix any linear error correction code $E_2 : {\mathbb{F}_q}^{n\times n} \to 100n^2$ with constant error tolerance. The protocol is as follows, where the matrices $R_i \in {\mathbb{F}_q}^{n\times n}$ are to be determined later.

1. Alice: Send $E_1(X) = (\langle X, R_1\rangle, ..., \langle X, R_m\rangle)$, a random subset $I \subseteq [100n^2]$ of size $|I| = 100n$, and $E_2(X)|_I$.
2. Bob: Send $E_1(Y) = (\langle Y, R_1\rangle, ..., \langle Y, R_m\rangle)$, a random subset $J \subseteq [100n^2]$ of size $|J| = 100n$, and $E_2(Y)|_J$.
3. Referee:

```
(a) Find a $Z$ with $\mathtt{rank}(Z) < r$ satisfying $E_1(Z) = E_1(X) + E_1(Y)$.
(b) **if** such $Z$ does not exist, **then** output "$\mathtt{rank}(X+Y) \geq r$".
(c) **else if** $I \cap J = \emptyset$ **then** output "Fail"
(d) **else if** $E_2(X)|_{I\cap J} + E_2(Y)|_{I\cap J} = E_2(Z)|_{I\cap J}$
(e)     output "$\mathtt{rank}(X+Y) < r$"
(f) **else** output "$\mathtt{rank}(X+Y) \geq r$".
```

The analysis follows the same line as that of the compressed sensing protocol for $\mathsf{Ham}_{n,k}$, except that the number of $n \times n$ matrices over $\mathbb{F}_q$ of rank at most $2r$ is at most $2rq^{4rn-\binom{2r}{2}}$, the bound given in the proof of Theorem 18. Thus by taking $m = 4rn$, there exists $r_1, \ldots, r_m$ s.t. $E_1(X) \neq 0$ for all matrices $X \in {\mathbb{F}_q}^{n\times n}$ with rank at most $2r$.

For the correctness of the whole protocol, note that conditioned on $\mathtt{rank}(X+Y) < r$, we have $Z = X + Y$, and the Equality test also passes. If $\mathtt{rank}(X+Y) \geq r$, then since the output Z has $\mathtt{rank}(Z) < r$, we know that $Z \neq X+Y$, and thus the Equality test fail with high probability. Indeed, by Birthday Paradox, $|I \cap J| > 0$ with high probability, and the locations in $I \cap J$ are uniformly at random. Thus by the property of error correction code of E_2, $E_2(X)|_{I\cap J} + E_2(Y)|_{I\cap J} \neq E_2(Z)|_{I\cap J}$ with high probability.

The communication cost: The compressed sensing part takes $O(rn\log(q))$, and the Equality testing part takes $O(n\log n + n\log(q))$, so overall the complexity is $O(rn\log(q) + n\log n)$. □

The protocol is computationally inefficient. Fortunately, for low-rank matrix recovery, there is computationally efficient compressed sensing available.

Construction 1 : Let $n \geq r \geq 1$. Let $\mathbb{K}$ be an extension of $\mathbb{F}_q$ such that $g \in \mathbb{K}$ is of order $\geq n$. Let $D_{k,l} \in \mathbb{K}^{n\times n}$ be the matrix defined by $(D_{k,l})_{i,j} = g^{lj}$ if $i + j = k$, and $(D_{k,l})_{i,j} = 0$ otherwise. Define $D_r = \{D_{k,l}\}_{0\leq k\leq 2n-2,\ 0\leq l<r}$, and $D'_r = \{D_{k,l}\}_{0\leq k\leq 2n-2,\ 0\leq l<\min\{r,k+1,2n-(k+1)\}}$.

Theorem 14 (Forbes-Shpilka, [FS12]). *Let $1 \leq r \leq n/2$, then D'_{2r} (from Construction 1) of size $O(nr)$ can be computed in poly(n) time, and there is an algorithm which can recover every $X \in \mathbb{K}^{n\times n}$ with $\texttt{rank}(X) \leq r$ exactly from $\{\langle X, D_i\rangle : D_i \in D'_{2r}\}$ in poly(n) time.*

Further, the above result can be extended to any field $\mathbb{F}_q$.

Theorem 15 (Forbes-Shpilka, [FS12]). *Let $1 \leq r \leq n$. Over any field $\mathbb{F}_q$, there is an explicit set of parameters $M_i \in \mathbb{F}^{n\times n}, i = 1, \ldots, m = O(nr\max\{\log_q n, 1\})$, which can be constructed in poly(n) time, and there is an algorithm which can exactly recover every matrix $X \in \mathbb{F}^{n\times n}$ with $\texttt{rank}(X) \leq r$ from $(\langle X, M_1\rangle, \ldots, \langle X, M_m\rangle)$ in poly(n) time.*

We can replace the compressed sensing part by the new measurements provided by the above theorem, thus achieving the computational efficiency.

The quantum upper bounds in Theorem 4 can be obtained by combining the randomized upper bounds, and the $q^{O(r^2)}$ upper bound obtained by a simple application of Theorem 7 and Theorem 12. This completes the proof of Theorem 4.

6 Adversarial model and concluding remarks

Last, we consider the adversarial model as in [MNS11], and obtain efficient protocols for $\mathsf{Ham}_{n,d}$ with bounded soundness error, perfect completeness and recovery, sketch size $\tilde{O}(\sqrt{n} + d)$, update time $\tilde{O}(1)$ and communication complexity $\tilde{O}(d)$. The sketch size can be reduced to $\tilde{O}(d)$ if using quantum protocols. We also obtain protocols for $\texttt{rank}_{n,r}$ with similar parameters except for sketch size and communication complexity, which change to $\tilde{O}(nr)$. Due to space limit, the description of the model and the results in it are deferred to Appendix C.

Techniques. This paper uses compressed sensing or Bose-Chowla theorem to solve three open questions in a *simple and unified* way. We hope that this is preferred to the hypothetical situation that three complicated and different methods are used to solve them.

Why care $\texttt{rank}_{n,r}$? In the investigation of the $\mathsf{Ham}_{n,d}$ problem, which had an exponential gap between lower and upper bounds before the present paper, the effort to close the gap led to the discovery of the application of compressed sensing to communication protocols designing. The most prominent open question left in this paper is the randomized and quantum private-coin SMP complexity for the $\texttt{rank}_{n,r}$ problem. Further study this function is called for, not only because it is a natural extension of the singularity property studied in [SW12], but also for the following reason. It seems that quantum fingerprint, the seemingly only known technique to design quantum private-coin SMP protocols [BCWdW01, Yao03], is not enough to exponentially improve the quantum upper bound. Efforts to close the gaps for $\texttt{rank}_{n,r}$ may hopefully stimulate new techniques for both randomized and quantum protocols designing.

References

[Amb96] Andris Ambainis. Communication complexity in a 3-computer model. *Algorithmica*, 16(3):298–301, 1996.

[BC62] Raj Chandra Bose and Sarvadaman Chowla. Theorems in the additive theory of numbers. *Commentarii Mathematici Helvetici*, 37(1):141–147, 1962.

[BCWdW01] Harry Buhrman, Richard Cleve, John Watrous, and Ronald de Wolf. Quantum fingerprinting. *Physical Review Letters*, 87(16), 2001.

[BDF$^+$11] Jean Bourgain, Stephen J. Dilworth, Kevin Ford, Sergei Konyagin, and Denka Kutzarova. Explicit constructions of rip matrices and related problems. *Duke Mathematical Journal*, 159(1):145–185, 2011.

[BK97] László Babai and Peter G. Kimmel. Randomized simultaneous messages: Solution of a problem of Yao in communication complexity. In *IEEE Conference on Computational Complexity*, pages 239–246, 1997.

[DeV07] Ronald A. DeVore. Deterministic constructions of compressed sensing matrices. *Journal of Complexity*, 23(4-6):918–925, 2007.

[FS12] Michael A. Forbes and Amir Shpilka. On identity testing of tensors, low-rank recovery and compressed sensing. In *Proceedings of the 44th Annual ACM Symposium on Theory of Computing*, pages 163–172, 2012.

[GKdW04] Dmitry Gavinsky, Julia Kempe, and Ronald de Wolf. Quantum communication cannot simulate a public coin. *arXiv:quant-ph/0411051*, 2004.

[HAN10] Jarvis Haupt, Lorne Applebaum, and Robert Nowak. On the restricted isometry of deterministically subsampled fourier matrices. In *Proceedings of the 44th Annual Conference on Information Sciences and Systems*, pages 1–6, 2010.

[HSZZ06] Wei Huang, Yaoyun Shi, Shengyu Zhang, and Yufan Zhu. The communication complexity of the Hamming Distance problem. *Information Processing Letters*, 99(4):149–153, 2006.

[Ind08] Piotr Indyk. Explicit constructions for compressed sensing of sparse signals. In *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 30–33, 2008.

[KS13] Raghav Kulkarni and Miklos Santha. Query complexity of matroids. In *Proceedings of the 8th International Conference on Algorithms and Complexity*, 2013.

[Lan93] Georg Landsberg. Uber eine anzahlbestimmung und eine damit zusammenhängende reihe. *Journal Fur Die Reine Und Angewandte Mathematik*, 111:87–88, 1893.

[LLZ11] Ming Lam Leung, Yang Li, and Shengyu Zhang. Tight bounds on the communication complexity of symmetric XOR functions in one-way and SMP models. In *Proceedings of the 8th Annual Conference on Theory and Applications of Models of Computation*, pages 403–408, 2011.

[LS09] Troy Lee and Adi Shraibman. Lower bounds on communication complexity. *Foundations and Trends in Theoretical Computer Science*, 3(4):263–398, 2009.

[MNS11] Ilya Mironov, Moni Naor, and Gil Segev. Sketching in adversarial environments. *SIAM Journal on Computing*, 40(6):1845–1870, 2011.

[MO10] Ashley Montanaro and Tobias Osborne. On the communication complexity of XOR functions. *arXiv:*, 0909.3392v2, 2010.

[NS96] Ilan Newman and Mario Szegedy. Public vs. private coin flips in one round communication games. In *Proceedings of the Twenty-Eighth Annual ACM Symposium on the Theory of Computing*, pages 561–570, 1996.

[Sho97] Peter Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 26:1484–1509, 1997.

[SW12] Xiaoming Sun and Chengu Wang. Randomized communication complexity for linear algebra problems over finite fields. In *Proceedings of the 29th International Symposium on Theoretical Aspects of Computer Science*, pages 477–488, 2012.

[TWXZ13] Hing Yin Tsang, Chung Hoi Wong, Ning Xie, and Shengyu Zhang. Fourier sparsity, spectral norm, and the log-rank conjecture. In *Proceedings of the 54th Annual IEEE Symposium Foundations of Computer Science*, 2013.

[Yao79] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing. In *Proceedings of the Eleventh Annual ACM Symposium on Theory of Computing (STOC)*, pages 209–213, 1979.

[Yao03] Andrew Chi-Chih Yao. On the power of quantum fingerprinting. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, pages 77–81, 2003.

[ZS09] Zhiqiang Zhang and Yaoyun Shi. Communication complexities of symmetric XOR functions. *Quantum Information & Computation*, 9(3):255–263, 2009.

A Summary of results in a table

Our results on the communication complexity part are summarized in Table 1.

B Proofs

B.1 Proof of Theorem 9

Proof. Let $A := \{i : x_i = 0, y_i = 1\}$, $B := \{j : x_j = 1, y_j = 0\}$, then we have

$$|y| - |x| = |A| - |B|, \qquad |A| + |B| \leq k, \tag{3}$$

and

$$H_y - H_x = \sum_{i \in A} d_i - \sum_{j \in B} d_j \mod (2^k - 1). \tag{4}$$

The correctness of the protocol is guaranteed if we can prove the following claim.

function	previous bounds	our results
$\mathsf{Ham}_{n,d}$	$\Omega(\sqrt{n}+d) \le \mathsf{R}^{\|} \le \tilde{O}(\sqrt{n}d)$ $\Omega(d) \le \mathsf{Q}^{\|} \le 2^{\tilde{O}(d)}$	$\mathsf{R}^{\|} = \tilde{\Theta}(\sqrt{n}+d)$, $\mathsf{Q}^{\|} = \tilde{\Theta}(d)$
symmetric XOR	$\Omega(\sqrt{n}+r) \le \mathsf{R}^{\|} \le \tilde{O}(\sqrt{n}r)$ $\Omega(r) \le \mathsf{Q}^{\|} \le 2^{\tilde{O}(r)}$	$\mathsf{R}^{\|} = \tilde{\Theta}(\sqrt{n}+r)$, $\mathsf{Q}^{\|} = \tilde{\Theta}(r)$
LTF	$\mathsf{R}^{\|,pub} = O((\theta/m)^2))$	$\mathsf{Q}^{\|},\ \mathsf{R}^{\|,pub} = \tilde{O}(\theta/m)$ $\mathsf{R}^{\|} = \tilde{O}(\theta/m+\sqrt{n})$
$\mathtt{rank}_{n,r}$ over $\mathbb{F}_q$	$\mathsf{Q} = \Omega(r^2 \log q)$	$\mathsf{Q}, \mathsf{Q}^1, \mathsf{R}, \mathsf{R}^1, \mathsf{Q}^{\|,pub}, \mathsf{R}^{\|,pub} = \Theta(r^2 \log q)$, $\tilde{\Omega}(n+r^2) \le \mathsf{R}^{\|} \le \tilde{O}(nr)$ $\tilde{\Omega}(r^2) \le \mathsf{Q}^{\|} \le \min\{q^{O(r^2)}, \tilde{O}(nr)\}$

Table 1: Summary of our results on communication complexity

Claim 1. *A and B are uniquely determined by $|x|, |y|, H_x, H_y$.*

Suppose there are two pairs (A, B), and (A', B') both satisfy Eq.(3) and (4), then we have

$$|A| - |B| = |A'| - |B'| = |y| - |x|, \quad |A| + |B| \le k, \quad |A'| + |B'| \le k,$$

and

$$\sum_{i \in A} d_i - \sum_{j \in B} d_j = \sum_{i \in A'} d_i - \sum_{j \in B'} d_j \mod (2^k - 1).$$

Rearranging the (in)equalities, we have

$$|A| + |B'| = |A'| + |B| \le k,$$

and

$$\sum_{i \in A} d_i + \sum_{j \in B'} d_j = \sum_{i \in A'} d_i + \sum_{j \in B} d_j \mod (2^k - 1).$$

These imply that the two multisets $A + B'$ and $A' + B$ are equal, where we define the sum of two sets S and T to be the multiset which collect all elements of S and T, with multiplicity reserved. Then the above fact can be written as $A + B' = A' + B$. Since

$$A \cap B = \emptyset, \quad \text{and } A' \cap B' = \emptyset,$$

we would have $A = A'$ and $B = B'$, contradicting the assumption. Therefore A and B are uniquely determined by $|x|, |y|, H_x, H_y$, based on which Referee can output $x \oplus y$.

Communication cost The message from Alice contains $|x|$, which has $\log n$ bits, and H_x, which has at most $\log(q^k - 1) = O(k \log n)$ bits. Overall the message is $O(k \log n)$ bits, and so is Bob's message.

Computational cost Though the statement of Bose-Chowla theorem is about the existence of d_i's, its proof is actually constructive. And it is not hard to see from its proof that there is a *poly*(n)-time randomized algorithm to generate the integers $d_1, \ldots, d_q$, provided that one has a computational oracle to solve the DiscreteLog problem. (On the other hand, recovering $(i_1, ..., i_k)$

from $d_{i_1}+d_{i_2}+\cdots+d_{i_k} \mod (q^k-1)$ is actually easy; it only needs multiplications of field elements.) With a quantum computer, DiscreteLog can be solved in polynomial time [Sho97]. One can also hardwire the d_i's into a non-uniform family of circuits of $poly(n)$ size. □

B.2 Proof of Lemma 10

Proof. We will change f to f' by removing a sequence of variables one by one. Suppose that there is a weight $w_i < m$ in f. We will show that the variable z_i can be removed without affecting the function. More rigorously, we will show that the LTF function f_1 with the same weights and threshold as f, but restricted to $S = [n] - \{i\}$, would be consistent with f, yet the margin does not decrease. For each z, define $w(z) = \sum_i w_i z_i$, and denote by $z^{(i)}$ the strings obtained from z by flipping the i-th bit, and by z' the string obtained from z by removing z_i. Consider the following two cases.

Case 1: $f(z) = 1$. If $z_i = 0$, then removing z_i does not change $w(z)$, namely $w(z') = w(z)$, thus $f_1(z)$ remains 1 since f_1 has the same threshold. Now assume that $z_i = 1$. We know that $w(z) \geq \theta + m$ from the definition of m, but actually, we can say more by claiming that $w(z) \geq \theta + m + w_i$. Suppose that this is not the case, then on one hand, flipping z_i from 1 to 0 would make

$$w(z^{(i)}) = w(z) - w_i < \theta + m + w_i - w_i = \theta + m. \tag{5}$$

On the other hand, since $w(z) \geq \theta + m$ and $w_i < m$, we have $w(z^{(i)}) = w(z) - w_i \geq \theta + m - w_i > \theta$, and thus $f(z^{(i)})$ is still 1 by definition of f. Now the weight bound in Eq.(5) violates the definition of m. Therefore we know that $w(z) \geq \theta + m + w_i$, and thus $w(z') = w(z) - w_i \geq \theta + m$. This implies that $f_1(z) = 1$, and the margin of f_1 incurred by these inputs z is at least m.

Case 2: $f(z) = 0$. Similarly, we can consider two subcases depending on the value of z_i. The involved case is $z_i = 0$, where we can show that $w(z) \leq \theta - m - w_i$. Thus removing z_i causes $w(z') \leq \theta - m$, thus $f_1(z)$ remains 0.

Combining the two cases, we also see that the new margin is at least m. So removing z_i does not change f or decrease m. Continue this procedure until all $w_i \geq m$. The remaining indices i form the set S. □

B.3 Deterministic communication complexity of matrix rank decision problem

We will first prove a lower bound for the deterministic communication complexity of $\mathbb{F}_q\text{-}\mathtt{rank}_{n,d}$.

Lemma 16. *For any function $f : I^n \times I^n \to \{0,1\}$, if each row and each column of the communication matrix M_f has at most K (and at least one) 1's, then $D(f) \geq n \log_2 |I| - 2\log_2 K$. If each row and each column of the communication matrix M_f has exactly K 1's, then $D(f) \geq n \log_2 |I| - \log_2 K$.*

Proof. Since each row and column has at most K 1's, any 1-rectangle has size at most K^2. Since any c-bit deterministic communication protocol partitions the $|I|^{2n}$ inputs into at most 2^c monochromatic rectangles, we have

$$D(f) \geq \log_2 \frac{|f^{-1}(1)|}{K^2} \geq \log_2 \frac{|I|^n}{K^2} = n \log_2 |I| - 2\log_2 K.$$

If furthermore each row/column has exactly K 1's, then the analysis can be slightly tightened as follows.

$$D(f) \geq \log_2 \frac{|f^{-1}(1)|}{K^2} = \log_2 \frac{|I|^n K}{K^2} = n \log_2 |I| - \log_2 K.$$

□

We will also need a fact about counting the number of matrices of rank r over $\mathbb{F}_q$.

Lemma 17 (Landsberg, [Lan93])**.** *The number of $n \times n$ matrices of rank r over $\mathbb{F}_q$ is*

$$q^{\binom{r}{2}} \cdot \frac{(q^{n-r+1}-1)^2(q^{n-r+2}-1)^2 \cdots (q^n-1)^2}{(q-1)(q^2-1)\cdots(q^r-1)}.$$

Theorem 18. $\mathsf{D}(\mathbb{F}_q\text{-}\mathtt{rank}_{n,r}) = \Omega(n^2 \log q)$.

Proof. Assume that $r < n/4$, because otherwise the lower bound of $\Omega(r^2 \log q)$ in [SW12] already gives the theorem. Note that for XOR functions, each row or column of the communication matrix has exactly K 1's, where K is the number of matrices over $\mathbb{F}_q$ with rank less than r. By Lemma 17, we have

$$\begin{aligned} K &= \sum_{k=0}^{r-1} q^{\binom{k}{2}} \cdot \frac{(q^{n-k+1}-1)^2(q^{n-k+2}-1)^2 \cdots (q^n-1)^2}{(q-1)(q^2-1)\cdots(q^k-1)} \\ &\leq rq^{\binom{r}{2}} \cdot \frac{(q^{n-r+1}-1)^2(q^{n-r+2}-1)^2 \cdots (q^n-1)^2}{(q-1)(q^2-1)\cdots(q^r-1)} \\ &\leq rq^{\binom{r}{2}} \cdot \frac{q^{2rn-\binom{r}{2}}}{q^{\binom{r}{2}}} \\ &= rq^{2rn-\binom{r}{2}}. \end{aligned}$$

Applying Lemma 16, we obtain the lower bound

$$\mathsf{D}(\mathbb{F}_q\text{-}\mathtt{rank}_{n,r}) \geq n^2 \log_2 q - \log_2 K \geq (\log_2 q)\Big(n^2 - 2rn + \binom{r}{2}\Big) - \log_2 r = \Omega(n^2 \log q),$$

when $r < n/4$. □

Corollary 19. $\mathsf{R}^{||}(\mathbb{F}_q\text{-}\mathtt{rank}_{n,r}) = \Omega(n\sqrt{\log q})$.

Proof. By Theorem 5 and 18, we have

$$\mathsf{R}^{||}(\mathbb{F}_q\text{-}\mathtt{rank}_{n,r}) = \Omega\Big(\sqrt{\mathsf{D}^{||}(\mathbb{F}_q\text{-}\mathtt{rank}_{n,r})}\Big) = \Omega\Big(\sqrt{\mathsf{D}(\mathbb{F}_q\text{-}\mathtt{rank}_{n,r})}\Big) = \Omega(n\sqrt{\log q}).$$

□

C The adversarial sketch model

In this section we strengthen our results to fit the adversarial sketch model, introduced in [MNS11]. Similar to the common sketch model mentioned in Section 1, the parties still generate a sketch of their inputs. But motivated from cryptographic reasons, the parties do not have secure public randomness when generating sketches. So it can be considered as private-coin model in this regard. Formally, there are two phases: *sketch phase* and *interaction phase*. In the first phase, sketch phase, an input $x \in \{0,1\}^n$ is given by a sequence of $\texttt{insert}(i)$ and $\texttt{delete}(i)$ operations on an initially all-0 string. Each $\texttt{insert}(i)$ operation updates the i-th bit of x to be 1 and each $\texttt{delete}(i)$ operation updates the i-th bit of x to be 0.[2] Upon receiving an operation $\texttt{insert}\ (i)$, each party can operate her sketch in any way, but the input is given in one pass, and it is not available after the sketch phase. In the second phase, interaction phase, the two parties can communicate and they aim to compute the function value $f(x,y)$. The complexity measures are sketch size (the maximum size of the sketch in the entire sketch phase), update time (the time to update the sketch for one insert/delete operation in the sketch phase) and communication complexity (in the interaction phase).

The model can be extended to the quantum setting by allowing the sketches and the communication to be quantum.

The Hamming Distance problem in the adversarial sketch model In the paper [MNS11], the authors studied the Equality function of deciding whether the two n-bit input strings are equal, with the promise that both inputs' Hamming weights are at most K. To state their result, let us formally define the concept of sparse recovery. We call a set V of vectors in $\{0,1\}^n$ an *r-sparse-recovery set* if every vector x with Hamming weight at most r can be uniquely determined by $S_V = \{(u, \langle x, u\rangle) : u \in V\}$. An r-sparse-recovery algorithm for V is an algorithm that exactly reconstructs x from the S_V. The paper [MNS11] proved the following results.

Lemma 20 ([MNS11])**.** *There exists an explicit construction of an r-sparse-recovery set of size $r \cdot \mathrm{polylog}(n)$, and it has a corresponding deterministic r-sparse-recovery algorithm with running time $O(r \cdot \mathrm{polylog}(n))$. Besides, if the input is given in the form of a sequence of* insert *and* delete *operations, the update time for processing each* insert *or* delete *operation is $O(\mathrm{polylog}(n))$.*

The above result is achieved by the follow scheme. An r-sparse vector $x \in \{0,1\}^n$is encoded as $C(x) = \{C_0(x), C_1(x), \ldots, C_{\log_2 r+1}(x)\}$, where each C_i has $O(r \log n)$ entries, and each entry is a vector in ${\mathbb{F}_q}^l$ for some $l = O(\mathrm{polylog}(n))$ and prime $q : n < q \leq 2n$. It has the property that on each insert/delete operation, we only need to add or subtract a certain vector in ${\mathbb{F}_q}^l$ to at most $O(\mathrm{polylog}(n))$ entries in $C(x)$. In addition, for any two strings $x \neq y, |x| \leq r, |y| \leq r$, there exists an index i, $0 \leq i \leq \log_2 r + 1$, such that $C_i(x)$ differs from $C_i(y)$ on at least 3/4-fraction of their entries. These two properties are achieved by bounded-neighbor dispersers.

Based on the above construction, [MNS11] obtained the following result for Equality.

Theorem 21 ([MNS11])**.** *For* Equality*, given the promise that both inputs x and y have Hamming weights at most K, there exists an explicit δ-error protocol with the sketch size $O(\sqrt{K}\mathrm{polylog}(n)\log(1/\delta))$, update time $O(\mathrm{polylog}(n))$, and communication cost $O(\mathrm{polylog}(n))$.*

[2] Here we assume that the given sequence of insert and delete operations are consistent in the sense that it does not insert an element i when the element is currently in the set, or delete an element not in the set. We can also extend to the multi-set case as in [MNS11] where each element appears at most m times, but for simplicity, here we just discuss the case of the input being a set $S \subseteq [n]$, which is identified with the indicator string.

With the help of the construction, we can design efficient protocols for $\mathsf{Ham}_{n,d}$ in the adversarial sketch model.

Theorem 22. *In the adversarial sketch model, for every n, $0 \le d \le n$ and $0 < \delta < 1$, there exists an explicit δ-error protocol for $\mathsf{Ham}_{n,d}$, with the following properties:*

1. *Perfect completeness and recovery: If $|x \oplus y| < d$, the protocol always outputs* Yes *and $x \oplus y$.*
2. *Soundness: If $|x \oplus y| \ge d$, the protocol outputs* No *with probability at least $1 - \delta$.*
3. *Sketch size: $O((\sqrt{n \log n} + d \cdot \mathrm{polylog}(n)) \log(1/\delta))$.*
4. *Worst-case update time: $O(\mathrm{polylog}(n))$.*
5. *Communication complexity: $O(d \cdot \mathrm{polylog}(n) \log(1/\delta))$.*

If we use quantum protocols, then the sketch size can be decreased to $O(d \cdot \mathrm{polylog}(n) \log(1/\delta))$.

Proof. All the factors $\log(1/\delta)$ are for error reduction and we omit the standard analysis and we will omit it in what follows. We maintain two sketches, the first one for sparse recovery and the second one for verification. More specifically, we use the first sketch to make a sparse recovery for the string $x \oplus y$, assuming that $|x \oplus y| < d$. We adopt the construction in Lemma 20 with $r = d$ to achieve the sketch size $d \cdot \mathrm{polylog}(n)$ and the update time $O(\mathrm{polylog}(n))$. The second sketch is for verifying that the recovered string is indeed equal to $x \oplus y$, for which we adopt the result in Theorem 21 with $K = n$, with the sketch size $O(\sqrt{d \cdot \mathrm{polylog}(n)})$ and the update time $O(\mathrm{polylog}(n))$. In the interaction phase, Alice sends the first sketch to Bob, who then recovers a string z which is equal to $x \oplus y$ if $|x \oplus y| < d$. They then run the Equality part of the protocol in Theorem 21. The communication cost for the first part is $O(d \cdot \mathrm{polylog} n)$, the size of $C(x)$, and the communication cost for the second part is $O(\mathrm{polylog}(n))$. Thus the overall cost is $O(d \cdot \mathrm{polylog} n)$.

The proof for the quantum protocol is similar, except that we maintain a quantum sketch for the second part by a quantum fingerprint. More specifically, for a vector x, we just maintain fingerprints

$$|u_i(x)\rangle = \frac{1}{|C_i(x)|^{1/2}} \sum_{1 \le j \le |C_i(x)|} |j, C_i(x)[j]\rangle$$

for each codewords $C_i(x), i = 0, \ldots, \log_2 r + 1$ as the sketch. On each `insert`/`delete` operation, since we only need to add or subtract a certain vector in ${\mathbb{F}_q}^l$ to $O(\mathrm{polylog}(r))$ entries in each codeword, we can implement it by $\mathrm{polylog}(n)$ controlled operations, $|j\rangle\langle j| \otimes U_j$, where each U_j is an addition or subtraction, on the fingerprint. This takes $O(\mathrm{polylog}(n))$ time. For the communication complexity, in the sparse recovery part, Alice sends the whole sketch, which is of size $O(r \cdot \mathrm{polylog}(n) \log(1/\delta))$. In the equality checking part, we just run a swap-test for each pair of fingerprints $|u_i(x)\rangle$ and $|u_i(y)\rangle$. Note that if $x \ne y$, then there is an i *s.t.* a constant fraction of $C_i(x)[j]$ and $C_i(y)[j]$ are different. Since the binary representations of $C_i(x)[j]$ has only $\log\log(n)$ bits, so $\langle C_i(x)[j], C_i(y)[j]\rangle \ge 1 - 1/\log\log(n)$. Thus using $\log\log(n)$ independent copies of $|u_i(x)\rangle$ can achieve a constant error probability as in the classical case. Thus the total communication complexity $O(\mathrm{polylog}(n))$. Thus the whole communication complexity is $O(d \cdot \mathrm{polylog}(n))$. □

The Rank Decision problem in the adversarial sketch model Next we consider rank decision problem in the adversarial sketch model, in which each input matrix X is given by a sequences of $\texttt{insert}(i,j)/\texttt{delete}(i,j)$ operations. Throughout this part, we will fix the field to be $\mathbb{F}_q$. For simplicity of statements of the following results, let us assume that $q = poly(n)$.

We call a set M of matrices in $\{0,1\}^{n\times n}$ an *r-low-rank-recovery set* if every matrix X with rank at most r can be uniquely determined by $T_M = \{(A, \langle X, A\rangle) : A \in M\}$. An r-low-rank-recovery algorithm for M is an algorithm that exactly reconstructs X from T_M. We will need a technique to transform an sparse-recovery set to a low-rank-recovery set, given by a recent paper [FS12]. For any $n \times n$ matrix M, we denote $M^{(k)}$ the k-th anti-diagonal of M, which is the tuple of the entries $(M_{i,j} : i + j = k)$. Sometimes we treat it as a vector.

Theorem 23 ([FS12])**.** *Let $n \geq r \geq 1$. For each $0 \leq k \leq 2n-2$, suppose that R_k is a set of $n \times n$ matrices such that*

1. *for any $R \in R_k$ and $k' \neq k$, $R^{(k')} = 0$, and*
2. *$\{R^{(k)} : R \in R_k\}$, when viewed as a collection of vectors, form a $(2\min(r, k+1, 2n-(k+1)))$-sparse-recovery set.*

Then $R = \bigcup_k R_k$ is an r-low-rank-recovery set.

If, for each k, the set $\{R^{(k)}\}_{R\in R_k}$ has an $(2\min(r, k+1, 2n-(k+1)))$-sparse-recovery algorithm SR_k running in time t_k, then there is a deterministic algorithm performing an r-low-rank-recovery for R in time $O(rn^2 + \sum_{k=2}^{2n}(t_k + n|R_k|))$.

Now we can state our result about matrix rank decision in the adversarial sketch model.

Theorem 24. *In the adversarial sketch model, for every n, $0 \leq r \leq n$ and $0 < \delta < 1$, there exists an explicit protocol for $\texttt{rank}_{n,r}$ with the following properties:*

1. *Perfect completeness and recovery: If $\texttt{rank}(X+Y) < r$, the protocol always outputs* Yes *and $X+Y$.*
2. *Soundness: If $\texttt{rank}(X+Y) \geq r$, the protocol outputs* No *with probability at least $1-\delta$.*
3. *Sketch size: $O(nr \cdot \mathrm{polylog}(n)\log(1/\delta))$.*
4. *Worst-case update time: $O(\mathrm{polylog}(n))$.*
5. *Communication complexity: $O(nr \cdot \mathrm{polylog}(n)\log(1/\delta))$.*

Proof. The protocol contains two parts. The first part reconstructs the matrix $Z = X+Y$ assuming that $\texttt{rank}(X+Y) < r$. Denote the reconstruction result by Z', the second part is to test whether $X+Y = Z'$. We need to maintain sketches for both parts.

For the first part, we explicitly construct $(2\min(r, k+1, 2n-(k+1)))$-sparse-recovery set v_k for $k = 0, \cdots, 2n-2$ using Lemma 20. We then apply Theorem 23 on these sets v_k to form an r-low-rank-recovery set $R = \bigcup_k R_k$.

On each $\texttt{insert}/\texttt{delete}$ operation, suppose the operation is on the (i,j)-th entry in the matrix, the property of the construction of Lemma 20 guarantees that we only need to update $O(\mathrm{polylog}(n))$ entries. The sketch for this part is of size $O(nr \cdot \mathrm{polylog}(n)\log(1/\delta))$.

For the second part, it is the Equality problem on two n^2-bit strings. Applying Theorem 21 with $K = n^2$, we only need to maintain a sketch of size $O(n\sqrt{\text{polylog}(n)\log(1/\delta)})$, with update time $O(\text{polylog}(n))$.

In the interaction phase, Alice just sends her whole sketch to Bob, who then makes all the calculations locally and sends the result back to Alice. The total communication cost is $O(nr \cdot \text{polylog}(n)\log(1/\delta))$.

The completeness and soundness come from those of the protocols in the two parts. This completes the proof. □