Electronic Colloquium on Computational Complexity, Report No. 11 (2011)

Tight bounds on the randomized communication complexity of symmetric XOR functions in one-way and SMP models*

Ming Lam Leung† Yang Li‡ Shengyu Zhang§

Abstract

We study the communication complexity of symmetric XOR functions, namely functions $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ that can be formulated as $f(x,y) = D(|x \oplus y|)$ for some predicate $D : \{0,1,...,n\} \to \{0,1\}$, where $|x \oplus y|$ is the Hamming weight of the bitwise XOR of x and y. We give a public-coin randomized protocol in the Simultaneous Message Passing (SMP) model, with the communication cost matching the known lower bound for the *quantum* and *two-way* model up to a logarithm factor. As a corollary, this closes a quadratic gap between quantum lower bound and randomized upper bound for the one-way model, answering an open question raised in Shi and Zhang [SZ09].

1 Introduction

Communication complexity quantifies the minimum amount of communication needed for two (or sometimes more) parties to jointly compute some function f. Since introduced by Yao [Yao79], it has attracted significant attention in the last three decades, not only for its elegant mathematical structure but also for its numerous applications in other computational models [KN97, LS07].

The two parties involved in the computation, usually called Alice and Bob, can communicate in different manners, and here we consider the three well-studied models, namely the two-way model, the one-way model and the simultaneous message passing (SMP) model. In the two-way model, Alice and Bob are allowed to communicate interactively in both directions, while in the one-way model, Alice can send message to Bob and Bob does not give feedback to Alice. An even weaker communication model is the SMP model, where Alice and Bob are prohibited to exchange information directly, but instead they each send a message to a third party Referee, who then announces a result. A randomized protocol is called private-coin if Alice and Bob each flip their own and private random coins. If they share the same random coins, then the protocol is called public-coin. The private coin model differs from the public coin model by at most an additive factor of $O(\log n)$ in the two-way and one-way models [New91].

We use $R^{priv}(f)$ to denote the communication complexity of a best private-coin randomized protocol that computes f with error at most $1/3$ in the two-way protocol. Similarly, we use the $R^{||,priv}(f)$ to denote the communication complexity in the private-coin SMP model, and $R^{1,priv}(f)$ for the private-coin one-way model. Changing the superscript "priv" to "pub" gives the notation for the communication complexities in the public-coin models. If we allow Alice and Bob to use quantum protocols, then $Q(f), Q^1(f), Q^{||}(f)$ represent the quantum communication

*This work is supported by Hong Kong General Research Fund No. 419309 and No. 418710.

†Department of Computer Science and Engineering, The Chinese University of Hong Kong. Email: mlleung@cse.cuhk.edu.hk

‡Email: danielliy@gmail.com

§Department of Computer Science and Engineering, The Chinese University of Hong Kong. Email: syzhang@cse.cuhk.edu.hk

ISSN 1433-8092

complexity in two-way model, one-way model and SMP model, separately. In the quantum case the communication complexity is evaluated in terms of the number of qubits in the communication. If Alice and Bob share prior entanglement, then we use a star in the superscript to denote the communication complexity.

Arguably the most fundamental issue in communication complexity is to determine the largest gap between the quantum and classical complexity. In particular, there is no super-constant separation between quantum and classical complexities in the one-way model; actually, it could well be the truth that they are the same up to a constant factor for all total Boolean functions.

One way to understand the question is to study special classes of functions. An important class of Boolean functions is that of XOR functions, namely those in the form of $f(x \oplus y)$ where $x \oplus y$ is the bitwise XOR of x and y. Some well studied functions such as the Equality function and the Hamming Distance function are special cases of XOR functions. XOR functions belong to a larger class of "composed functions"; see [LZ10] for some recent studies.

While the general XOR function seems hard to study, recently Shi and Zhang [SZ09] considered symmetric XOR functions, *i.e.* $f(x \oplus y) = D(|x \oplus y|)$ for some $D : \{0, 1, ..., n\} \to \{0, 1\}$. 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 $k \in [r_0, n - r_1)$ and set $r = \max\{r_0, r_1\}$. Shi and Zhang proved that the quantum lower bound for symmetric XOR functions in the two-way model is $\Omega(r)$, and on the other hand, they also gave randomized protocol in communication of $\tilde{O}(r)$ in the two-way model and $\tilde{O}(r^2)$ in the one-way model. Pinning down the quantum and randomized communication complexity of symmetric XOR functions in the one-way model was raised as an open problem.

In this work, we close the quadratic gap by proving a randomized upper bound of $\tilde{O}(r)$, which holds even for the SMP model. Namely,

Theorem 1 *For any symmetric XOR function f,*

$$R^{\|,pub}(f) = O(r\log^3 r/\log\log r) \tag{1}$$

Combining this upper bound with Shi and Zhang's quantum lower bound in the two-way model, we have the following.

Corollary 2 *The randomized and quantum communication complexities of symmetric XOR functions are $\tilde{\Theta}(r)$, in the two-way, the one-way and the public-coin SMP models.*

A good question for further exploration is the private-coin SMP model.

2 Preliminaries

In this part we review some known results on the randomized and quantum communication complexity of the Hamming Distance function and the Equality function.

Let $\mathsf{Ham}_n^{(d)}$ be the boolean function such that $\mathsf{Ham}_n^{(d)}(x, y) = 1$ if and only if the two n-bit strings x and y have Hamming distance at most d. Yao [Yao03] showed a randomized upper bound of $O(d^2)$ in the public-coin SMP model, lated improved by Gavinsky, Kempe and de Wolf [GKdW04] to $O(d \log n)$ and further by Huang, Shi, Zhang and Zhu [HSZZ06] to $O(d \log d)$. Let $\mathcal{HD}_{d,\epsilon}$ denote the $O(d \log d \log(1/\epsilon))$-cost randomized protocol by repeating the [HSZZ06] protocol for $O(\log(1/\epsilon))$ times so that the error probability is below ϵ.

The parity function $\mathsf{Parity}(x)$ is defined as $\mathsf{Parity}(x) = 1$ if and only if $|x|$ is odd.

A function $f : \{0, 1\}^n \times \{0, 1\}^n \to \{0, 1\}$ is a symmetric XOR function if $f(x, y) = S(x \oplus y)$ for some symmetric function S. That is, $f(x, y) = D(|x \oplus y|)$ where $D : \{0, 1, \ldots, n\} \to \{0, 1\}$. Let $\tilde{D}(k) = D(n-k)$ and $\tilde{S}(x, y) = \tilde{D}(|x \oplus y|)$. 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, $D(k)$ only depends on the parity of k when $k \in [r_0, n - r_1]$. Suppose $D(k) = T(\mathsf{Parity}(k))$ for $k \in [r_0, n - r_1]$ (for some function T).

All the logarithms in this paper are based 2.

3 A public-coin protocol in the SMP model

This section gives the protocol in Theorem 1. We will first give a subprocedure $\mathcal{P}_k$ which computes the function in the special case of $|x \oplus y| \leq k$. It is then used as a building block for the general protocol $\mathcal{P}$.

In the protocols we will use random partitions. A *random k-partition* of $[n]$ is a random function p mapping $[n]$ to $[k]$, *i.e.* mapping each element in $[n]$ to $[k]$ uniformly at random and independently. We call the set $\{i \in [n] : p(i) = j\}$ the *block* $B(j)$. A simple fact about the random partition is the following.

Lemma 3 *For any string $z \in \{0,1\}^n$ with at most k 1's, a random k-partition has*

$$\mathbf{Pr}[\textit{All } k \textit{ blocks have less than } c \textit{ 1's}] \geq 1 - O(1/k^2). \tag{2}$$

where $c = 4 \log k / \log \log k$.

Proof Consider the complement event. There are k possible blocks to violate the condition, $\binom{k}{c}$ choices for the c 1's (out of k 1's) put in the "bad" block, and for each of these 1's, the probability of it mapped to the block is $1/k$. Thus the union bound gives

$$\mathbf{Pr}[\text{There exists a block with } c \text{ 1's}] \leq k \cdot \binom{k}{c} \cdot \frac{1}{k^c} \leq \left(\frac{ek}{c}\right)^c \cdot \frac{1}{k^{c-1}} = \left(\frac{e}{c}\right)^c \cdot k \tag{3}$$

It is easily verified that the chosen c makes this bound $O(1/k^2)$. □

Now the protocol P_k is as in **Box** $\mathcal{P}_k$. Recall that $\mathcal{HD}_{d,\epsilon}$ is the $O(d \log d \log(1/\epsilon))$ randomized protocol with error probability below ϵ.

Box $\mathcal{P}_k$:

A public-coin randomized protocol $\mathcal{P}_k$ for functions $f(x,y) = D(|x \oplus y|)$, with promise $|x \oplus y| \leq k$, in the SMP model

Input: $x \in \{0,1\}^n$ to Alice and $y \in \{0,1\}^n$ to Bob, with promise $|x \oplus y| \leq k$
Output: One bit $\bar{f}$ by Referee satisfying $\bar{f} = f(x,y)$ with probability at least 0.9.

Protocol:
Alice and Bob:

1. Use public coins to generate a common random k-partition $[n] = \uplus_{i=1}^{k} B(i)$.
2. **for** $i = 1$ **to** k
   for $j = 0$ **to** $c = 4 \log k / \log \log k$
   run (Alice and Bob's part of) the protocol $\mathcal{HD}_{j,\epsilon}$ on input $(x_{B(i)}, y_{B(i)})$ with $\epsilon = 1/(10k \log c)$, sending a pair of messages $(m_{a,i,j}(x_{B(i)}), m_{b,i,j}(y_{B(i)}))$.

Referee:

1. **for** $i = 1$ **to** k
   (a) On receiving $\{(m_{a,i,j}(x_{B(i)}), m_{b,i,j}(y_{B(i)})) : j = 1, \ldots, c\}$, run (Referee's part of) the protocol $\mathcal{HD}_{j,\epsilon}$ which outputs h_{ij}.
   (b) Use binary search in $(h_{i1}, \ldots, h_{ic})$ to find the Hamming distance h_i of $(x_{B(i)}, y_{B(i)})$.
2. Output $D(\sum_{i=1}^{k} h_i)$.

Lemma 4 *If $|x \oplus y| \leq k$, then* Referee *outputs $D(|x \oplus y|)$ with probability at least 0.9. The cost of protocol $\mathcal{P}_k$ is $O(k \log^3 k / \log\log k)$.*

Proof First, by Lemma 3, each block contains at most c different indices i *s.t.* $x_i \neq y_i$. Namely, the Hamming distance of $x_{B(i)}$ and $y_{B(i)}$ is at most c. Thus running the protocols $\mathcal{HD}_{j,\epsilon}$ for $j = 0, ..., c$ would give information to find the Hamming distance h_i of $(x_{B(i)}, y_{B(i)})$. In each block $B(i)$, h_i is correctly computed as long as each of the $\lceil \log c \rceil$ values h_{ij} on the (correct) path of the binary search is correct. Thus a union bound gives the overall error probability upper bounded by $k(\log c)\epsilon = 1/10$. The cost of the protocol is $O(k \cdot c \cdot c \log c \log(1/\epsilon)) = O(k \log^3 k / \log\log k)$. □

With the protocol $\mathcal{P}_k$ in hand, we now construct the general protocol as in **Box $\mathcal{P}$.**

Box $\mathcal{P}$

A public-coin randomized protocol $\mathcal{P}$ for functions $f(x, y) = S(x \oplus y)$ in the SMP model

Input: $x \in \{0,1\}^n$ to Alice and $y \in \{0,1\}^n$ to Bob
Output: One bit b which equals to $f(x, y)$ with probability at least 2/3.

Protocol:

1. Run the protocol $\mathcal{HD}_{r_0,1/10}$ and the protocol $\mathcal{HD}_{n-r_1,1/10}$ on $(\bar{x}, y)$.
2. Run the protocol $\mathcal{P}_{r_0}$ for function S on (x, y) and the protocol $\mathcal{P}_{r_1}$ for function $\tilde{S}$ on $(\bar{x}, y)$.
3. Alice: send $\mathsf{Parity}(x)$
4. Bob: send $\mathsf{Parity}(y)$.
5. Referee:
   (a) If $\mathcal{HD}_{r_0,1/10}$ on (x, y) outputs 1, then output $\mathcal{P}_{r_0}$ on (x, y) and halt.
   (b) If $\mathcal{HD}_{n-r_1,1/10}$ on $(\bar{x}, y)$ outputs 1, then output $\mathcal{P}_{r_1}$ on $(\bar{x}, y)$ and halt.
   (c) Output $T(\mathsf{Parity}(x) \oplus \mathsf{Parity}(y))$.

Theorem 5 *The protocol $\mathcal{P}$ outputs the correct value with probability at least 2/3, and the complexity cost is $O(r \log^3 r / \log\log r)$.*

Proof *Correctness*: If $|x \oplus y| \leq r_0$, then with probability at least 0.9, the protocol $\mathcal{HD}_{r_0,1/10}(x, y)$ outputs 1, thus Referee outputs $\mathcal{P}_{r_0}$ on (x, y), which equals to $f(x, y)$ with probability at least 0.9 by the correctness of the protocol $\mathcal{P}_{r_0}$. Thus the overall success probability is at least $0.81 > 2/3$.

If $|x \oplus y| \geq n - r_1$, then $|\bar{x} \oplus y| \leq r_1$ and with probability at least 0.9, the protocol $\mathcal{HD}_{r_1,1/10}(\bar{x}, y)$ outputs 1, thus Referee outputs $\mathcal{P}_{r_1}(\tilde{S}, \bar{x}, y)$, which equals to

$$\tilde{S}(\bar{x} \oplus y) = \tilde{D}(n - |x \oplus y|) = D(|x \oplus y|) \tag{4}$$

with probability at least 0.9 by the correctness of the protocol $\mathcal{P}_{r_1}$. Thus the overall success probability is at least $0.81 > 2/3$.

If $r_0 < |x \oplus y| < n - r_1$, then the protocol proceeds to the very last step with probability at least $1 - 0.1 - 0.1 = 0.8$. And once this happens, then Referee outputs the correct value with certainty, since $f(x, y) = T(\mathsf{Parity}(x \oplus y)) = T(\mathsf{Parity}(x) \oplus \mathsf{Parity}(y))$.

Complexity: The cost is twice of the cost of the protocol P_r, plus twice of the cost of the protocol $\mathcal{HD}_{r,1/10}$, plus 2, which in total is $O(r \log^3 r / \log\log r)$. □

References

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

[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.

[KN97] Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, Cambridge, UK, 1997.

[LS07] Troy Lee and Adi Shraibman. *Lower bounds on communication complexity*, volume 263399 of *4*. Foundations & Trends in Theoretical Computer Science, 2007.

[LZ10] Troy Lee and Shengyu Zhang. Composition theorems in communication complexity. In *Proceedings of the 37th International Colloquium on Automata, Languages and Programming (ICALP)*, pages 475–489, 2010.

[New91] Ilan Newman. Private vs. common random bits in communication complexity. *Information Processing Letters*, 39(2):67–71, 1991.

[SZ09] Yaoyun Shi and Zhiqiang Zhang. Communication complexities of XOR functions. *Quantum Information and Computation*, 9(3&4):255–263, 2009.

[Yao79] Andrew Yao. Some complexity questions related to distributive computing. In *Proceedings of the eleventh annual ACM symposium on Theory of computing*, pages 209–213, 1979.

[Yao03] Andrew Yao. On the power of quantum fingerprinting. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 77–81, 2003.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de