Depth-Independent Lower Bounds on the Communication Complexity of Read-Once Boolean Formulas

Rahul Jain[1], Hartmut Klauck[2], and Shengyu Zhang[3]

[1] Centre for Quantum Technologies and Department of Computer Science, National University of Singapore. Email: rahul@comp.nus.edu.sg

[2] Centre for Quantum Technologies and SPMS, Nanyang Technological University. Email: hklauck@gmail.com

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

Abstract. We show lower bounds of $\Omega(\sqrt{n})$ and $\Omega(n^{1/4})$ on the randomized and quantum communication complexity, respectively, of all n-variable read-once Boolean formulas. Our results complement the recent lower bound of $\Omega(n/8^d)$ by Leonardos and Saks [LS09] and $\Omega(n/2^{O(d \log d)})$ by Jayram, Kopparty and Raghavendra [JKR09] for randomized communication complexity of read-once Boolean formulas with depth d.
We obtain our result by "embedding" either the Disjointness problem or its complement in any given read-once Boolean formula.

1 Introduction

A *read-once Boolean formula* $f : \{0,1\}^n \to \{0,1\}$ is a function which can be represented by a Boolean formula involving AND and OR such that each variable appears, possibly negated, at most once in the formula. An *alternating* AND-OR *tree* is a layered tree in which each internal node is labeled either AND or OR and the leaves are labeled by variables; each path from the root to the any leaf alternates between AND and OR labeled nodes. It is well known (see eg. [HW91]) that given a read-once Boolean formula $f : \{0,1\}^n \to \{0,1\}$ there exists a unique alternating AND-OR tree, denoted T_f, with n leaves labeled by input Boolean variables $z_1, \ldots, z_n$, such that the output at the root, when the tree is evaluated according to the labels of the internal nodes, is equal to $f(z_1 \ldots z_n)$. Given an alternating AND-OR tree T, let f_T denote the corresponding read-once Boolean formula evaluated by T.
Let $x, y \in \{0,1\}^n$ and let $x \wedge y, x \vee y$ represent the *bit-wise* $\mathsf{AND}, \mathsf{OR}$ of the strings x and y respectively. For $f : \{0,1\}^n \to \{0,1\}$, let $f^\wedge : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be given by $f^\wedge(x,y) = f(x \wedge y)$. Similarly let $f^\vee : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be given by $f^\vee(x,y) = f(x \vee y)$. Recently Leonardos and Saks [LS09], investigated the *two-party randomized communication complexity* with constant error, denoted $\mathsf{R}(\cdot)$, of $f^\wedge, f^\vee$ and

showed the following. (Please refer to [KN97] for familiarity with basic definitions in communication complexity.)

Theorem 1 ([LS09]) *Let $f : \{0,1\}^n \to \{0,1\}$ be a read-once Boolean formula such that T_f has depth d. Then*

$$\max\{\mathsf{R}(f^\wedge), \mathsf{R}(f^\vee)\} \geq \Omega(n/8^d).$$

In the theorem, the depth of a tree is the number of edges on a longest path from the root to a leaf. Independently, Jayram, Kopparty and Raghavendra [JKR09] proved randomized lower bounds of $\Omega(n/2^{O(d \log d)})$ for general read-once Boolean formulas and $\Omega(n/4^d)$ for a special class of "balanced" formulas.
It follows from results of Snir [Sni85] and Saks and Wigderson [SW86] (via a generic simulation of trees by communication protocols [BCW98]) that for the read-once Boolean formula with their canonical tree being a *complete binary* alternating AND-OR trees, the randomized communication complexity is $O(n^{0.753...})$, the best known so far. However in this situation, the results of [LS09,JKR09] do not provide any lower bound since $d = \log_2 n$ for the complete binary tree. We complement their result by giving universal lower bounds that do not depend on the depth. Below $\mathsf{Q}(\cdot)$ represents the two-party *quantum communication complexity.*

Theorem 2 *Let $f : \{0,1\}^n \to \{0,1\}$ be a read-once Boolean formula. Then,*

$$\max\{\mathsf{R}(f^\wedge), \mathsf{R}(f^\vee)\} \geq \Omega(\sqrt{n}).$$
$$\max\{\mathsf{Q}(f^\wedge), \mathsf{Q}(f^\vee)\} \geq \Omega(n^{1/4}).$$

Remark:
1. Note that the maximum in Theorem 1 and 2 is necessary since for example if f is the AND of the n input bits then it is easily seen that $\mathsf{R}(f^\wedge)$ is 1.
2. The randomized lower bound in the above theorem is easy to observe for balanced trees, as is also remarked in [LS09].
3. We obtain our result by "embedding" either the Disjointness problem or its complement in any given read-once Boolean formula. This simple idea was also used by Zhang [Zha09] and independently by Sherstov [She10] to show some relations between decision tree complexity and communication complexity.

2 Proofs

In this section we show the proof of Theorem 2. We start with the following definition.

Definition 1 (Embedding) *We say that a function*

$$g_1 : \{0,1\}^r \times \{0,1\}^r \to \{0,1\}$$

can be embedded *into a function* $g_2 : \{0,1\}^t \times \{0,1\}^t \to \{0,1\}$*, if there exist maps* $h_a : \{0,1\}^r \to \{0,1\}^t$ *and* $h_b : \{0,1\}^r \to \{0,1\}^t$ *such that* $\forall x, y \in \{0,1\}^r,\ g_1(x,y) = g_2(h_a(x), h_b(y))$*.*

It is easily seen that if g_1 can be embedded into g_2 then the communication complexity of g_2 is at least as large as that of g_1.
Let us define the *Disjointness* problem $\mathsf{DISJ}_n : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ as $\mathsf{DISJ}_n(x,y) = \bigwedge_{i=1,\dots,n}(x_i \vee y_i)$ (where the usual negation of the variables is left out for notational simplicity). Similarly the *Non-Disjointness* problem $\mathsf{NDISJ}_n : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is defined as $\mathsf{NDISJ}_n(x,y) = \bigvee_{i=1,\dots,n}(x_i \wedge y_i)$. We shall also use the following well-known lower bounds.

Fact 1 ([KS92,Raz92]) $\mathsf{R}(\mathsf{DISJ}_n) = \Omega(n), \mathsf{R}(\mathsf{NDISJ}_n) = \Omega(n)$*.*

Fact 2 ([Raz03]) $\mathsf{Q}(\mathsf{DISJ}_n) = \Omega(\sqrt{n}), \mathsf{Q}(\mathsf{NDISJ}_n) = \Omega(\sqrt{n})$*.*

Recall that for the given read-once Boolean formula $f : \{0,1\}^n \to \{0,1\}$ its canonical tree is denoted T_f. We have the following lemma which we prove in Section 2.1.

Lemma 3
1. *Let T_f have its last layer consisting only of* AND *gates. Let m_0 be the largest integer such that DISJ_{m_0} can be embedded into $f^\vee$ and m_1 be the largest integer such that NDISJ_{m_1} can be embedded into $f^\vee$. Then $m_0 m_1 \geq n$.*
2. *Let T_f have its last layer consisting only of* OR *gates. Let m_0 be the largest integer such that DISJ_{m_0} can be embedded into $f^\wedge$ and m_1 be the largest integer such that NDISJ_{m_1} can be embedded into $f^\wedge$. Then $m_0 m_1 \geq n$.*

With this lemma, we can prove the lower bounds on $\max\{\mathsf{R}(f^\wedge), \mathsf{R}(f^\vee)\}$ and $\max\{\mathsf{Q}(f^\wedge), \mathsf{Q}(f^\vee)\}$ as follows. For an arbitrary read-once formula f with n variables, consider the sets of leaves

$$L_{odd} = \{\text{leaves in } T_f \text{ on odd levels}\}, L_{even} = \{\text{leaves in } T_f \text{ on even levels}\}$$

At least one of the two sets has size at least $n/2$; without loss of generality, let us assume that it is L_{odd}. Depending on whether the root is AND or OR, this set consists only of AND gates or OR gates, corresponding to case 1 or 2 in Lemma 3. Then by the lemma, either $\mathsf{DISJ}_{\sqrt{n/2}}$ or $\mathsf{NDISJ}_{\sqrt{n/2}}$ can be embedded in f (by setting the leaves in L_{even} to 0's). By Fact 1 and 2, we get the lower bounds in Theorem 2.

2.1 Proof of Lemma 3

We shall prove the first statement; the second statement follows similarly. We first prove the following claim.

Claim 1 *For an n-leaf ($n > 2$) alternating* AND-OR *tree T such that all its internal nodes just above the leaves have exactly two children (denoted the x-child and the y-child), let $s(T)$ denote the number of such nodes directly above the leaves. Let $m_0(T)$ be the largest integer such that* DISJ_{m_0} *can be embedded into f_T and $m_1(T)$ be the largest integer such that* NDISJ_{m_1} *can be embedded into f_T. Then $m_0(T)m_1(T) \geq s(T)$.*

Proof. The proof is by induction on depth d of T. When $n > 2$, the condition of the tree makes $d > 1$, so the base case is $d = 2$.

Base Case $d = 2$: In this case T consists either of the root labeled AND with $s(T)$ (fan-in 2) children labeled ORs or it consists of the root labeled OR with $s(T)$ (fan-in 2) children labeled ANDs. We consider the former case and the latter follows similarly. In the former case f_T is clearly $\mathsf{DISJ}_{s(T)}$ and hence $m_0(T) = s(T)$. Also $m_1(T) \geq 1$ as follows. Let us choose the first two children v_1, v_2 of the root. Further choose the x child of v_1 and the y child of v_2 which are kept free and the values of all other input variables are set to 0. It is easily seen that the function (of input bits x, y) now evaluated is NDISJ_1. Hence $m_0(T)m_1(T) \geq s(T)$.

Induction Step $d > 2$: Assume the root is labeled AND (the case when the root is labeled OR follows similarly). Let the root have r children $v_1, \ldots, v_r$ which are labeled OR and let the corresponding subtrees be $T_1, \ldots, T_r$ rooted at $v_1, \ldots, v_r$ respectively. Without loss of generality let the first r' (with $0 \leq r' \leq r$) of these trees be of depth 1 in which case the corresponding $s(\cdot) = 0$. It is easily seen that

$$s(T) = r' + \left(\sum_{i=r'+1}^{r} s(T_i) \right) .$$

For $i > r'$, we have from the induction hypothesis that $m_1(T_i)m_0(T_i) \geq s(T_i)$.
It is clear that $m_0(T) \geq \sum_{i=1}^{r} m_0(T_i)$, since we can simply combine the Disjointness instances of the subtrees. Also we have

$$m_1(T) \geq \max\{m_1(T_{r'+1}), \ldots, m_1(T_r), 1\},$$

because we can either take any one of the subtree instances (and set all other inputs to 0), or at the very least can pick a pair of x, y leaves (as in the base case above) and fix the remaining variables appropriately to realize a single AND gate which amounts to embedding NDISJ_1. Now,

$$m_0(T)m_1(T) \geq \left(\sum_{i=1}^{r} m_0(T_i)\right) \cdot (\max\{m_1(T_1), \ldots, m_1(T_r), 1\})$$
$$\geq r' + \left(\sum_{i=r'+1}^{r} m_0(T_i)m_1(T_i)\right)$$
$$\geq r' + \left(\sum_{i=r'+1}^{r} s(T_i)\right) = s(T) \ .$$

Now we prove Lemma 3: Let us view $f^{\vee} : \{0,1\}^{2n} \to \{0,1\}$ as a read-once Boolean formula, with input (x,y) of $f^{\vee}$ corresponding to the x- and y- children of the internal nodes just above the leaves. Note that in this case $T_{f^{\vee}}$ satisfies the conditions of the above claim and $s(T_{f^{\vee}}) = n$. Hence the proof of the first statement in Lemma 3 finishes.

3 Concluding Remarks

1. The randomized communication complexity varies between $\Theta(n)$ for the Tribes_n function (a read-once Boolean formula whose canonical tree has depth 2) [JKS03] and $O(n^{0.753\ldots})$ for functions corresponding to completely balanced AND-OR trees (which have depth $\log n$). It will probably be hard to prove a generic lower bound much larger than $\sqrt{n}$ for all read-once Boolean formulas $f : \{0,1\}^n \to \{0,1\}$, since the best known lower bound on the randomized query complexity of every read-once Boolean formula is $\Omega(n^{.51})$ [HW91] and communication complexity lower bounds immediately imply slightly weaker query complexity lower bounds (via the generic simulation of trees by communication protocols [BCW98]).
2. Ambainis et al. [ACR+07] show how to evaluate any alternating AND-OR tree T with n leaves by a quantum query algorithm with slightly more than $\sqrt{n}$ queries; this also gives the same upper bound for the communication complexity of $\max\{\mathsf{Q}(f_T^{\wedge}), \mathsf{Q}(f_T^{\vee})\}$. On the other hand, it is easily seen that the *parity* of n bits can be computed by a formula of size $O(n^2)$ involving $\mathsf{AND}, \mathsf{OR}$. Therefore it is easy to show that the function *Inner Product modulo* 2 i.e. the function $\mathsf{IP}_m : \{0,1\}^m \times \{0,1\}^m \to \{0,1\}$ given by $\mathsf{IP}_m(x,y) = \sum_{i=1}^m x_i y_i \bmod 2$, with $m = \sqrt{n}$ can be reduced to the evaluation of an alternating AND-OR tree of size $O(n)$ and logarithmic depth. Since it is known that $\mathsf{Q}(\mathsf{IP}_{\sqrt{n}}) = \Omega(\sqrt{n})$ [CvDNT99], we get an example of an alternating AND-OR tree T with n leaves and $\log n$ depth such that $\mathsf{Q}(f_T^{\wedge}) = \Omega(\sqrt{n})$. Since the same lower bound also holds for shallow trees such as OR, hence $\Theta(\sqrt{n})$ might turn out to be the correct bound on $\max\{\mathsf{Q}(f_T^{\wedge}), \mathsf{Q}(f_T^{\vee})\}$ for all alternating AND-OR trees T with n leaves regardless of the depth.

Acknowledgments: Research of Rahul Jain and Hartmut Klauck is supported by the internal grants of the Centre for Quantum Technologies (CQT), which is funded by the Singapore Ministry of Education and the Singapore National Research Foundation. Research of Shengyu Zhang is partially supported by the Hong Kong grant RGC-419309, and partially by CQT for his research visit.

References

[ACR+07] A. Ambainis, A. Childs, B. Reichardt, R. Spalek, and S. Zhang. Any AND-OR formula of size n can be evaluated in time $n^{1/2+o(1)}$ on a quantum computer. In *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science*, pages 363–372, 2007.

[BCW98] H. Buhrman, R. Cleve, and A. Wigderson. Quantum vs. classical communication and computation. In *Proceedings of the Thirtieth Annual ACM Symposium on the Theory of Computing (STOC)*, pages 63–68, 1998.

[CvDNT99] R. Cleve, W. van Dam, M. Nielsen, and A. Tapp. Quantum entanglement and the communication complexity of the inner product function. *Lecture Notes in Computer Science*, 1509:61–74, 1999.

[HW91] R. Heiman and A. Wigderson. Randomized vs. deterministic decision tree complexity for read-once boolean functions. In *Proceedings of the 6th Structures in Complexity Theory Conference*, pages 172–179, 1991.

[JKR09] T.S. Jayram, S. Kopparty, and P. Raghavendra. On the communication complexity of read-once ac^0 formula. In *Proceedings of the 24th Annual IEEE Conference on Computational Complexity*, pages 329–340, 2009.

[JKS03] T. S. Jayram, R. Kumar, and D. Sivakumar. Two applications of information complexity. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing (STOC)*, pages 673–682, 2003.

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

[KS92] B. Kalyanasundaram and G. Schnitger. The probabilistic communication complexity of set intersection. *SIAM Journal on Discrete Mathematics*, 5(4):545–557, 1992. Earlier version in Structures'87.

[LS09] N. Leonardos and M. Saks. Lower bounds on the randomized communication complexity of read-once functions. In *Proceedings of the 24th Annual IEEE Conference on Computational Complexity*, pages 341–350, 2009.

[Raz92] A. Razborov. On the distributional complexity of disjointness. *Theoretical Computer Science*, 106(2):385–390, 1992.

[Raz03] A. Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya of the Russian Academy of Science, mathematics*, 67(1):159–176, 2003. quant-ph/0204025.

[She10] A. Sherstov. On quantum-classical equivalence for composed communication problems. *Quantum Information and Computation*, 10(5-6):435–455, 2010.

[Sni85] M. Snir. Lower bounds for probabilistic linear decision trees. *Theoretical Computer Science*, 38:69–82, 1985.

[SW86] M. Saks and A. Wigderson. Probabilistic boolean decision trees and the complexity of evaluating game trees. In *Proceedings of the 27th Annual IEEE Symposium on Foundations of Computer Science*, pages 29–38, 1986.

[Zha09] S. Zhang. On the tightness of the Buhrman-Cleve-Wigderson simulation. In *Proceedings of the 20th International Symposium on Algorithms and Computation*, pages 434–440, 2009.