Vol.2009-AL-124 No.8
2009/5/11

Average/Worst-Case Gap of Quantum Query Complexities

Andris Ambainis,[†1] Kazuo Iwama,[†2]
Masaki Nakanishi,[†3] Harumichi Nishimura,[†4]
Rudy Raymond,[†5] Seiichiro Tani[†6,†7]
and Shigeru Yamashita [†8]

The difference between average-case and worst-case complexities is one of the central topics in theoretical computer science, and it has been extensively studied for decades. In the quantum setting, however, only a few results are known. This paper shows a super-linear tight gap between the average-case and worst-case quantum query complexities over the families of Boolean functions defined by a natural parameter. More concretely, we consider the *size of the on-set* of a Boolean function f, i.e., the number of inputs for which $f = 1$, as the parameter, and give the tight gap over the families of N-variable Boolean functions with on-set size M for $\mathrm{poly}(N) \le M \le 2^N/2$.

1. Introduction

1.1 Background

The difference between average-case and worst-case complexities is one of the central topics in theoretical computer science, and it has been extensively studied for decades. In the quantum setting, however, only a few results are known. (i) For a MAJORITY function, there is an almost quadratic gap between the average-case and worst-case query complexities over all inputs of the function[8),9)]. (ii) If we consider the average-case and worst-case behaviors of complexities over all Boolean functions (for the worst input of each function), only a linear gap is possible for query complexity[3),13),20)] and exact communication complexity[12)]. This paper shows a super-linear tight gap between the average-case and worst-case quantum query complexities over the families of Boolean functions defined by a natural parameter. More concretely, we consider the *size of the on-set* of a Boolean function f, i.e., the number of inputs for which $f = 1$, as the parameter, and give the tight gap over the families of N-variable Boolean functions with on-set size M for $\mathrm{poly}(N) \le M \le 2^N/2$.[⋆1] (Note that it is sufficient to consider the range of $M \le 2^N/2$ because of the symmetry of function values, 0 and 1.)

The research on quantum query complexity started with the Deutcsh-Jozsa algorithm[14)] and other algorithms for computing partial functions (e.g., Simon's algorithm[21)]), followed by Grover's quantum search algorithm[15)], which also computes the Boolean OR function of N variables with $O(\sqrt{N})$ queries. Since then, a sequence of results have extensively appeared in the literature, showing that similar speed-ups are possible for many other, more general Boolean functions. For example, if a Boolean function is given by a constant-depth balanced AND-OR trees (OR is by a single-depth tree), it can be computed in $O(\sqrt{N})$ queries with quantum search on bounded-error inputs[17)]. This was recently extended to any AND-OR tree with $O(N^{\frac{1}{2}+o(1)})$ queries by using the quantum walk technique[5)]. In general, however, the worst-case quantum query complexity is polynomially related to the worst-case classical query complexity for any Boolean function[9)]. In contrast, there is an exponential gap between the average-case classical and quantum query complexities of a certain Boolean function for uniform distribution of inputs, and the gap can be even larger for non-uniform distribution of inputs[8)]. As for the gap between the average-case and worst-case quantum query complexities, they are $O(N^{1/2+\epsilon})$[8)] and $\Omega(N)$[9)], respectively, over all inputs for MAJORITY functions. On the contrary, the average of complexity over all Boolean functions (for the worst input of each function) was proved to be at least $N/4 - 2\sqrt{N}\log N$[3)], which was improved to $N/4 + \Omega(\sqrt{N})$[20)], and the worst-case complexity is at most $N/2 + \sqrt{N}$[13)], respectively.

†1 Institute of Mathematics and Computer Science, University of Latvia, Latvia.
†2 School of Informatics, Kyoto University.
†3 Graduate School of Information Science, NAIST.
†4 School of Science, Osaka Prefecture University.
†5 Tokyo Research Laboratory, IBM Japan.
†6 NTT Communication Science Laboratories, NTT Corporation.
†7 JST ERATO-SORST QCI Project
†8 College of Information Science and Engineering, Ritsumeikan University.

⋆1 For $M = 2^{\Omega(N)}$, our gap is at most linear. This is consistent with the linear possible gap over all Boolean functions.

1.2 Our contribution

Let $\mathcal{F}_{N,M}$ be a family of N-variable Boolean functions f_N whose on-set is of size M, namely, f_N has output 1 (true) for M 0/1 assignments among the total 2^N ones. We assume $M \in \{1, 2, \ldots, 2^N/2\}$ because of the symmetry of function values, 0 and 1. Let $Q(f_N)$ be the (true) query complexity of f_N. We then investigate the asymptotic behaviors of the following three functions of N for every M:

(1) $Q_{\mathrm{worst}}(\mathcal{F}_{N,M}) \equiv \max_{f \in \mathcal{F}_{N,M}} Q(f)$.

(2) $Q_{\mathrm{best}}(\mathcal{F}_{N,M}) \equiv \min_{f \in \mathcal{F}_{N,M}} Q(f)$.

(3) $Q_{\mathrm{avg}}(\mathcal{F}_{N,M})$ is an arbitrary function such that when f_N is uniformly distributed over $\mathcal{F}_{N,M}$, $\Pr_{f_N \in \mathcal{F}_{N,M}}[Q(f_N) = \Theta(\tilde{C}_M(N))] \to 1$ as N goes to infinity if it exists, and undefined otherwise. We call $Q_{\mathrm{avg}}(\mathcal{F}_{N,M})$ the "average-case" quantum query complexities over all functions in $\mathcal{F}_{N,M}$.

Our results are as follows:

(i) For $1 \le M \le 2^{N/(\log N)^{2+\epsilon}}$ with any small positive constant ϵ,

$$Q_{\mathrm{worst}}(\mathcal{F}_{N,M}) = \Theta\left(\sqrt{N\frac{\log M}{c + \log N - \log\log M}} + \sqrt{N}\right)$$

for some positive constant c (Strictly speaking, the lower bound of $Q_{\mathrm{worst}}(\mathcal{F}_{N,M})$ is valid for broader range $1 \le M \le 2^N/2$).

(ii) For $1 \le M \le 2^{\frac{N}{2+\epsilon}}$ with small positive constant ϵ,

$$Q_{\mathrm{best}}(\mathcal{F}_{N,M}) = \Theta(\sqrt{N}).$$

(iii) For $1 \le M \le 2^N/2$,

$$Q_{\mathrm{avg}}(\mathcal{F}_{N,M}) = \Theta\left(\frac{\log M}{c + \log N - \log\log M} + \sqrt{N}\right)$$

for some positive constant c.

Notice that (i) and (iii) imply a super-linear gap between the worst-case and average-case quantum query complexities.

1.3 Technical Outlines for the Above Results (i)-(iii)

(i) For the upper bound, we use an algorithm[7] for the Oracle Identification Problem (OIP), which is defined as follows: Given an oracle x and a set S of M oracle candidates out of 2^N ones, determine which oracle in S is identical to x with the promise that x is a member of S. We set S to the on-set of f_N, run the algorithm, and finally verify with Grover search that the output of the algorithm is equal to the given N bits. To achieve the tight bound, we refine the complexity analysis of the algorithm, leading to the improvement of the query complexity of OIP. For the lower bound, we give a function with on-set size M which has the matching query complexity.

(ii) The upper bound is shown by giving a function with on-set size M whose query complexity is $O(\sqrt{N})$. The lower bound is proved by combining counting argument with $Q(f_N) = \Omega(\sqrt{s(f_N)})$[9] where $s(f_N)$ is the sensitivity of f_N.

(iii) For the upper bound, we encode the given N-bit string $x \in \{0,1\}^N$ as a quantum state $|\psi_x\rangle^{\otimes k}$ so that, for almost all Boolean functions in $\mathcal{F}_{N,M}$, $|\psi_x\rangle$ and $|\psi_y\rangle$ are almost orthogonal if $x \neq y$ for $x, y \in f_N^{-1}(1)$. We then perform state discrimination procedure[16] using $|\psi_x\rangle^{\otimes k}$ to test if x is in the on-set of f_N, and verify the result with Grover search. More concretely, let $|\psi_x\rangle = 1/(\sqrt{N})\sum_{i=1}^{N}(-1)^{x_i}|i\rangle$ for $x \in \{0,1\}^N$. We prove that for every two different states $|\psi_x\rangle$ and $|\psi_y\rangle$ where $x, y \in f_N^{-1}(1)$, it holds that $|\langle\psi_x|\psi_y\rangle| \le 2\sqrt{\log M/N}$ for almost all Boolean functions $f_N \in \mathcal{F}_{N,M}$ with $M \le 2^N/2$. The number k of the copies of $|\psi_x\rangle$ is set to $O(\sqrt{N\frac{\log M}{c+\log N - \log\log M}})$[16]. For the lower bound, we use the following facts. (1) The number of functions in $\mathcal{F}_{N,M}$ is $\binom{2^N}{M}$ (2) The number of Boolean functions computable with success probability more than 1/2 with at most $d/2$ queries is at most $T(N,d) = 2\sum_{i=0}^{D-1}\binom{2^N-1}{i}$ for $D = \sum_{i=0}^{d}\binom{N}{i}$[19][11]. We then calculate the largest d such that $T(N,d)/\binom{2^N}{M} \to 0$ for $N \to \infty$.

2. Preliminaries

We assume the oracle (or black-box) model in the quantum setting. In this model, an input (i.e., a problem instance) is given as an oracle. For any input $x = (x_1, \ldots, x_N) \in \{0,1\}^N$, a unitary operator O, corresponding to a single query to an oracle, maps $|i\rangle|b\rangle|w\rangle$ to $|i\rangle|b \oplus x_i\rangle|w\rangle$ for each $i \in [N] = \{1, 2, \ldots, N\}$ and $b \in \{0,1\}$, where w denotes workspace. A *quantum computation* of the oracle model is a sequence of unitary transformations $U_0 \to O \to U_1 \to O \to \cdots \to O \to U_t$, where U_j may be any unitary transformation that does not depend on the input. The above computation sequence involves t oracle calls, which is our measure of the complexity: The *quantum query complexity* $Q(P)$ of a problem P whose input is given as an N-bit string is defined to be the number of quantum

queries needed to solve P with bounded-error, i.e., with success probability at least $1/2 + c$ where c is some constant.

In this paper, our problem P is to evaluate the value (0 or 1) of a Boolean function $f(x_1, \ldots, x_N)$ over N variables, assuming that the truth table of f is known. The *on-set* of f is the set of assignments $(x_1, \ldots, x_N)$ satisfying $f(x_1, \ldots, x_N) = 1$. We denote the family of all functions whose on-set is of size M by $\mathcal{F}_{N,M}$.

A proof in this paper uses an improvement of a previous algorithm[7] for the oracle identification problem.

Definition 1 (Oracle Identification Problem (OIP)[6),7)]) Given an oracle x and a set S of M oracle candidates out of 2^N ones, determine which oracle in S is identical to x with the promise that x is a member of S.

Theorem 1 (Optimal bound of OIP[7]) OIP can be quantumly solved with a constant success probability by making $O(\sqrt{N\frac{\log M}{\log N}})$ queries to the given oracle if $\mathrm{poly}(N) \leq M \leq 2^{N^d}$ for some constant d $(0 < d < 1)$.
In this paper, the base of the logarithm is 2 when we do not explicitly write the base.

3. Worst-Case Analysis

In this section, we study the upper bound of the quantum query complexities of "all" Boolean functions in $\mathcal{F}_{N,M}$. Before showing the bound, we present two technical lemmas. The first lemma is useful for precise analysis.

Lemma 1 For $1 < z \leq 2^N$, let $d(z) = \frac{\log z}{4(\log eN - \log\log z)}$, where e is the base of the natural logarithm. Then, it holds that $d(z)$ is monotone non-decreasing, and

$$\left(\frac{eN}{d(z)}\right)^{d(z)} \leq z. \tag{1}$$

Proof The monotone non-decreasing property can be easily checked since for any $1 < z \leq z' \leq 2^N$, $d(z) \leq d(z')$. The rest of the proof follows by the following bound from recalling the definition of $d(z)$ and taking the log of the left-hand side of Eq. 1.

$$\begin{aligned}
d(z)\log\frac{eN}{d(z)} &= \frac{1}{4}\frac{\log z}{\log(eN) - \log\log z}\log\left(\frac{eN}{\frac{1}{4}\frac{\log z}{\log(eN)-\log\log z}}\right) \\
&= \frac{1}{4}\frac{\log z}{\log(eN) - \log\log z} \\
&\quad \times (\log(eN) - \log\log z + \log 4 + \log(\log eN - \log\log z)) \\
&= \frac{1}{4}\log z\left(1 + \frac{2 + \log y}{y}\right) \\
&\leq \log z,
\end{aligned}$$

for $y = \log(eN) - \log\log z$, where the last inequality is due to $\log y/y \leq 1$ for $y \geq 1$. ■

The second lemma is a well-known quantum lower bound[4] (the next formulation[2] is different from the original).

Theorem 2 (Quantum adversary method[4]) Let $\mathcal{A} \subseteq F^{-1}(0)$ and $\mathcal{B} \subseteq F^{-1}(1)$ be sets of inputs to function F. Let $R(A,B) \geq 0$ be a real-valued function, and for $A \in \mathcal{A}$, $B \in \mathcal{B}$, and location i, let

$$\theta(A,i) = \frac{\sum_{B^* \in \mathcal{B}\ :\ A(i) \neq B^*(i)} R(A,B^*)}{\sum_{B^* \in \mathcal{B}} R(A,B^*)},$$
$$\theta(B,i) = \frac{\sum_{A^* \in \mathcal{A}\ :\ A^*(i) \neq B(i)} R(A^*,B)}{\sum_{A^* \in \mathcal{A}} R(A^*,B)},$$

where $A(i)$ and $B(i)$ denotes the value of the ith variable for A and B, respectively, the denominators are all nonzero. Then the number of quantum queries needed to evaluate F with probability at least $9/10$ is $\Omega(1/\upsilon_{\mathrm{geom}})$, where

$$\upsilon_{\mathrm{geom}} = \max_{\substack{A \in \mathcal{A},\ B \in \mathcal{B},\ i\ : \\ R(A,B) > 0,\ A(i) \neq B(i)}} \sqrt{\theta(A,i)\,\theta(B,i)}.$$

We give an upper bound for the query complexities of all Boolean functions in $\mathcal{F}_{N,M}$.

Theorem 3 (Upper Bound) Let $1 \leq M \leq 2^{N/(\log N)^{2+\epsilon}}$, for some constant positive ϵ. Then, any Boolean function $f \in \mathcal{F}_M$ has quantum query complexity $O\left(\sqrt{N\frac{\log M}{\log N - \log\log M}} + \sqrt{N}\right)$.

Proof Recall that OIP is the problem that we are requested to find a hidden

oracle, with the promise that it is a member of oracle candidate set S. To use this for evaluation of the Boolean function f, let S be the on-set of f, which can be constructed from the known truth table of f. Note that $|S| = M$ since $f \in \mathcal{F}_{N,M}$. We then invoke the OIP algorithm of Theorem 1 to find the hidden oracle with $O(\sqrt{N\frac{\log M}{\log N}})$ queries, assuming the promise that the current oracle x is in S (actually, the promise does not hold if $f(x) = 0$). Let $z \in \{0,1\}^N$ be the string obtained by the OIP algorithm.

If $f(x) = 1$, the promise of the above OIP is indeed satisfied; z is equal to x with high probability.

If $f(x) = 0$, the promise does not hold; the OIP algorithm outputs some answer $z \in S$ such that $z \neq x$. To recognize this case, it suffices to check whether z is equal to x by using Grover search[15] with $O(\sqrt{N})(\in O(\sqrt{N\frac{\log M}{\log N}}))$ queries. This completes the proof for $N^k \leq M \leq 2^{N^d}$ for some constant k and any constant $0 < d < 1$. If $M < N^k$, we just pad 0-instances to make the oracle candidate set have size at least N^k, and then apply the OIP algorithm with query complexity $O(\sqrt{N})$.

For bigger M, we cannot use the OIP algorithm[7]. To expand the range of M for which the algorithm can work, we change the value of the parameters β and MAX_QUERIES in the algorithm to the following values, respectively,

$$\beta = (\log M(\log\log M)^2 \log(eN/\log M))/(2N),$$
$$\text{MAX\_QUERIES}(N,M) = 30\sigma\sqrt{N\log M\log(eN/\log M)}/\log(1/\beta).$$

Those new values can be used because we can have a better bound for the value of γ satisfying $\sum_{k=0}^{2\gamma N}\binom{N}{k} \leq |S|^{1/2}$ by virtue of Lemma 1, namely,

$$\gamma = \Theta\left(\frac{\log|S|}{\log N - \log\log|S|}\right).$$

The details of the proof are the same with those in the original algorithm[7]. ■

The following corollary is immediate.

Corollary 1 For $M \in poly(N)$, any function $f \in \mathcal{F}_{N,M}$ has quantum query complexity $O(\sqrt{N})$.

Theorem 3 is tight: We can show the following matching bound.

Theorem 4 Let $1 \leq M \leq 2^{N-1}$. Then, there is a function $f \in \mathcal{F}_M$ whose quantum query complexity is $\Omega\left(\sqrt{N\frac{\log M}{c+\log N-\log\log M}} + \sqrt{N}\right)$ for some constant $c > 0$.

Proof If $1 \leq M \leq N^2$, the upper bound $O(\sqrt{N})$ given in Theorem 3 matches the lower bound given in Theorem 5 in the next section; the theorem follows.

Suppose that $N^2 \leq M \leq 2^{N-1}$. Consider a Boolean function f such that for a k satisfying

$$D = \sum_{i=0}^{k}\binom{N}{i} \leq M \quad \text{and} \quad \sum_{i=0}^{k+1}\binom{N}{i} > M,$$

$f(x) = 1$ for all x with $\text{Ham}(x) \leq k$ and $M - D$ elements x with $\text{Ham}(x) = k+2$, and $f(x) = 0$ for all other elements. Here $\text{Ham}(x)$ denotes the Hamming weight of x. The quantum query complexity of such function is $\Omega(\sqrt{Nk})$ by Theorem 2: That is, let $\mathcal{A} \subseteq f^{-1}(0)$ and $\mathcal{B} \subseteq f^{-1}(1)$ be defined, respectively, as the set of x's with $\text{Ham}(x) = k+1$ and k. For any $A \in \mathcal{A}$ and $B \in \mathcal{B}$, let us define the relation R in Theorem 2 such that $R(A,B) = 1$ if A and B differ in exactly one position. Then, it can be shown that $\theta(A,i) = 1/(k+1)$, and $\theta(B,i) = 1/(N-k)$, and hence the lower bound. To complete the proof, it suffices to show that $k = \Omega(d(M/N))$ for the function $d(\cdot)$ defined in Lemma 1. By simple algebra, it holds that (by assuming $d(M/N)$ is an integer for simplicity):

$$\sum_{i=0}^{d(M/N)}\binom{N}{i} \leq N\binom{N}{d(M/N)} \leq N\left(\frac{eN}{d(M/N)}\right)^{d(M/N)} \leq M,$$

where the last inequality is due to Lemma 1. This implies that $k = \Omega(d(M/N)) = \Omega(\log M/(\log eN - \log\log N))$. ■

4. Best-Case Analysis

This section gives a tight lower bound for the query complexities of all Boolean functions in $\mathcal{F}_{N,M}$. In particular, this lower bound with Corollary 1 implies that if M is in $poly(N)$, then any function $f \in \mathcal{F}_{N,M}$ has essentially the same complexity up to a constant factor as the OR function.

Theorem 5 (Lower Bound) If $1 \leq M \leq 2^{\frac{N}{2+\epsilon}}$ for any positive constant ϵ, any $f \in \mathcal{F}_{N,M}$ has quantum query complexity $\Omega(\sqrt{N})$.

Proof We use the sensitivity argument. Recall that the sensitivity $s_x(f)$ of a Boolean function f on $x \in \{0,1\}^N$ is the number of variables x_i such that $f(x) \neq f(x^i)$, where x^i is the string obtained from x by flipping the value of x_i. The sensitivity $s(f)$ of f is the maximum of $s_x(f)$ over all x. The results of Beals et al.[9] implies $Q(f) = \Omega(\sqrt{s(f)})$. By the definition of $s(f)$ and the result by Beals et al., we can see that $Q(f) = \Omega(\sqrt{|Z|})$, where Z is the set of 0-*points*, elements whose values of f is 0, "around" an arbitrarily chosen element in the on-set (1-*point*). Here, "around" means the Hamming distance is 1. Therefore, if there is a 1-point around which there are $\Omega(N)$ 0-points, $Q(f) = \Omega(\sqrt{N})$.

To prove by contradiction, we assume that, around every 1-point, there are $o(N)$ 0-points, i.e., there are $(N - o(N))$ 1-points. Suppose that $(0, 0, \ldots, 0)$ is a 1-point (otherwise, we can give a similar argument using some 1-point). Set $S_0 = \{(0, 0, \ldots, 0)\}$. Define S_k inductively to be the set of all 1 points around all points in S_{k-1}, whose Hamming weight is k. By assumption, the number of 1-points around every point in S_{k-1} is $N - o(N) = N(1 - \alpha)$ for any small $\alpha = o(1)$. For each point x in S_{k-1}, there exist at most $(k-1)$ 1-points around x in S_{k-2}. Thus, for each point x in S_{k-1}, there exist at least $(N(1-\alpha) - (k-1))$ 1-points around x in S_k. Similarly, for each point x in S_k, there exist at most k 1-points around x in S_{k-1}. Thus, $|S_k| \geq |S_{k-1}|(N(1-\alpha) - (k-1))/k$. From this inductive inequality and $|S_0| = 1$, we have $|S_k| \geq (N(1-\alpha))(N(1-\alpha) - 1)(N(1-\alpha) - 2) \cdots (N(1-\alpha) - (k-1))/k!$. The number of inputs x such that $f(x) = 1$ and the Hamming weight of x is at most k is $T(k) = |S_0| + \cdots + |S_k|$. We will show $T(k) > M$ for some $k \leq N/2$, a contradiction, as follows. $T(k) > |S_k| \geq (N(1-\alpha))(N(1-\alpha)-1) \cdots (N(1-\alpha) - (k-1))/k! > \left(\frac{N(1-\alpha)}{k}\right)^k$. For $k = \frac{N}{2+\epsilon}$, we obtain $T(k) > 2^{\frac{N}{2+\epsilon}} \geq M$. ■

The above lower bound is tight, since it is easy to construct a Boolean function for any $M \leq 2^{\frac{N}{2+\epsilon}}$ such that its query complexity is $O(\sqrt{N})$. Thus we have shown that, for $1 \leq M \leq 2^{N/(\log N)^{2+\epsilon}}$, there are Boolean functions, f_1 and f_2, which are "easiest" and "hardest" in class $\mathcal{F}_{N,M}$, such that $Q(f_1) = \Theta(\sqrt{N})$ and $Q(f_2) = \Theta\left(\sqrt{N\frac{\log M}{c + \log N - \log\log M}} + \sqrt{N}\right)$.

5. Average-Case Analysis

This section considers the "average" behavior of functions in $\mathcal{F}_{N,M}$, that is, upper and lower bounds for the quantum query complexities of almost all functions in $\mathcal{F}_{N,M}$. To prove the upper bound, we need the following lemma that bounds the inner product of any two quantum states created by a single query to f.

Lemma 2 Let $|\psi_x\rangle = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (-1)^{x_i} |i\rangle$ for $x \in \{0,1\}^N$. For every two different states $|\psi_x\rangle$ and $|\psi_y\rangle$ where $x, y \in f^{-1}(1)$, it holds that $|\langle\psi_x|\psi_y\rangle| \leq 2\sqrt{\frac{\log M}{N}}$ for almost all Boolean functions $f \in \mathcal{F}_{N,M}$ with $M \leq 2^{N-1}$.

Proof Since $|\langle\psi_x|\psi_y\rangle| \leq 1$ obviously holds for every two quantum states, we will only show the lemma when $M \leq 2^{N/4}$. Notice that by the definition,

$$\begin{aligned}\langle\psi_x|\psi_y\rangle &= \frac{1}{N}\sum_{i=1}^{N}(-1)^{x_i \oplus y_i} \\ &= \frac{1}{N}\sum_{i=1}^{N}(1 - 2(x_i \oplus y_i)) \\ &= \frac{1}{N}(N - 2\mathrm{Ham}(x,y)),\end{aligned}$$

where $\mathrm{Ham}(x,y)$ is the Hamming distance of x and y.

We can prove the following claim: If f is uniformly distributed over $\mathcal{F}_{N,M}$, then for every $x, y \in f^{-1}(1)$ and $M \leq 2^{N/4}$, $\mathrm{Ham}(x,y) \geq N\left(\frac{1}{2} - \sqrt{\frac{(2+\epsilon)}{\log e}\frac{\log M}{2N}}\right)$ holds for any $\epsilon > 0$ with probability $1 - o(1)$. The lemma then follows from the claim by choosing ϵ appropriately. ■

Now, we are ready to show an upper bound for the average behavior of quantum query complexities of functions in $\mathcal{F}_{N,M}$.

Theorem 6 (Upper Bound) For some constant $c > 0$ and $1 \leq M \leq 2^{N-1}$, the quantum query complexities of almost all Boolean functions in $\mathcal{F}_{N,M}$ is $O(\frac{\log M}{c + \log N - \log\log M} + \sqrt{N})$.

Proof We use Lemma 2 in conjunction with the *quantum state discrimination*

result[16]: Assume that a set of M quantum states $\{|\psi_x\rangle\}$ satisfies $|\langle\psi_x|\psi_y\rangle|^2 \leq F < 1$ for any $x \neq y$. If $k = O(\log M/\log(1/F))$ copies of each state $|\psi_x\rangle$ are given, we can distinguish which one of M quantum states their copies are, i.e., know the index x with probability at least 2/3.

Notice that when $x \in f^{-1}(1)$, we can easily see that the quantum state discrimination succeeds to output x with only $k = O(\log M/(c+\log N - \log\log M))$ queries. Namely, we create k copies of quantum state $|\psi_x\rangle$ defined in Lemma 2, each of which only requires one query, and know which $x \in f^{1}(1)$ with high probability by checking it with the Grover search by spending additional $O(\sqrt{N})$ queries.

On the other hand, when $x \in f^{-1}(0)$, the quantum state discrimination might return any $x' \in f^{-1}(1)$ (or, falsely distinguished the state). However, $x \neq x'$ can also be tested by applying the Grover search. This completes the proof. ■

We can show the optimality of Theorem 6 as follows.

Theorem 7 (Lower Bound) For some constant $c > 0$ and $1 \leq M \leq 2^{N-1}$, almost all Boolean functions in $\mathcal{F}_{N,M}$ have quantum query complexity $\Omega(\frac{\log M}{c+\log N-\log\log M} + \sqrt{N})$.

Proof We shall prove the theorem for $2^{\sqrt{N}} < M \leq 2^{N-1}$ since $\Omega(\sqrt{N})$ holds for $M \leq 2^{\sqrt{N}}$ and assuming $M \leq 2^{N-1}$ does not lose the generality. Let us denote the number of queries d by the monotone non-decreasing function $d(z)$ in Lemma 1.

First, notice that the number of functions in $\mathcal{F}_{N,M}$ is $\binom{2^N}{M}$, which is at least $\left(\frac{2^N}{M}\right)^M = 2^{M'}$, for $M' = M(N - \log M)$. Secondly, notice that the number of Boolean functions computable with success probability more than 1/2 with at most $d/2$ queries is at most $T(N,d) = 2\sum_{i=0}^{D-1}\binom{2^N-1}{i}$ for $D = \sum_{i=0}^{d}\binom{N}{i}$. This bound is derived from the following two properties of a *sign-representing polynomial* p, a real-valued polynomial with properties that $p(x)$ is positive whenever $f(x) = 0$ and $p(x)$ is negative whenever $f(x) = 1$: (i) The *unbounded-error* query complexity of a Boolean function f, where the success probability is only guaranteed to be more than 1/2, is exactly half of the minimum degree of its sign-representing polynomial[19][11]. (ii) The number of Boolean functions whose minimum degrees of sign-representing polynomials are at most d is $T(N,d)$[1].

We shall complete the proof of the theorem by the following three claims. The first and second claims show that, for $z = \frac{M'}{(N+1)^2}$, the value of $T(N,d(z))$ (or, the number of functions recognizable with queries at most $d(z)/2$) is very small compared to $2^{M'}$, i.e., $T(N,d(z))/|\mathcal{F}_{N,M}| = o(1)$. The third claim proves the theorem's statement on the number of queries, i.e., $d\left(\frac{M'}{(N+1)^2}\right)/2 = \Omega(\log M/(\log(eN) - \log\log M))$.

Claim 1 For $z = \frac{M'}{(N+1)^2}$, it holds that $T(N,d(z)) = o(1)2^{ND}$.

Claim 2 For $z = \frac{M'}{(N+1)^2}$, it holds that $ND \leq M'$.

The theorem follows since, by Claims 1 and 2, $T(N,d)/|\mathcal{F}_{N,M}| = o(1)2^{DN-M'} = o(1)$, for the number of queries $d\left(\frac{M'}{(N+1)^2}\right)$ whose lower bound is proven by the following claim.

Claim 3

$$d\left(\frac{M'}{(N+1)^2}\right) = \Omega\left(\frac{\log M}{c+\log N-\log\log M}\right).$$

Below are the proofs of the claims.

Proof [Claim 1] By definition of $T(N,d)$, we have

$$\begin{aligned} T(N,d) &= 2\sum_{i=0}^{D-1}\binom{2^N-1}{i} \leq 2D\binom{2^N-1}{D-1} \leq 2D\left(\frac{e(2^N-1)}{D-1}\right)^D \\ &= 2^{1+\log D+D\log e-D\log(D-1)+ND} \\ &= o(1)2^{ND}, \end{aligned}$$

where the last equality is due to $2^{1+\log D+D\log e-D\log(D-1)} \leq 2^{2D-D\log(D-1)} = o(1)$ as $N \to \infty$. ■

Proof [Claim 2] By approximating the sum of binomials, we have, for $z = \frac{M'}{(N+1)^2} \leq \frac{M'}{N(N+1)}$,

$$D = \sum_{i=0}^{d(z)}\binom{N}{i} \leq (d(z)+1)\left(\frac{eN}{d(z)}\right)^{d(z)} \leq (N+1)z \leq \frac{M'}{N},$$

where the second last inequality is due to Lemma 1 and $d(z) \leq N$. ■

Proof [Claim 3] Recall that $d(z)$ is a monotone non-decreasing function, and therefore, because $M' = M(N - \log M) \geq M$, we have,

$$d\left(\frac{M'}{(N+1)^2}\right) \geq d\left(\frac{M}{(N+1)^2}\right) = \frac{1}{4}\frac{\log\left(\frac{M}{(N+1)^2}\right)}{\log(eN) - \log\log\left(\frac{M}{(N+1)^2}\right)}$$
$$= \Omega\left(\frac{\log M}{c + \log N - \log\log M}\right).$$

■

This completes the proof of Theorem 7. ■

References

1) M.Anthony. Classification by polynomial surfaces. *Discrete Applied Mathematics* 61:91–103, 1995.
2) S.Aaronson: Lower bounds for local search by quantum arguments. *SIAM J. Comput.* 35(4):804-824, 2006.
3) A. Ambainis. A note on quantum black-box complexity of almost all Boolean functions. *Inf. Process. Lett.* 71(1):5–7, 1999.
4) A. Ambainis. Quantum lower bounds by quantum arguments, *J. Comput. Sys. Sci.* 64:750–767, 2002.
5) A.Ambainis, A.M.Childs, B.W.Reichardt, R.Špalek, 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 *Proc. 48th FOCS*, pages 363–372, 2007.
6) A. Ambainis, K. Iwama, A. Kawachi, H. Masuda, R. H. Putra, and S. Yamashita. Quantum identification of Boolean oracles. In *Proc. 21st STACS, Lecture Notes in Comput. Sci.* 2996:105–116, 2004.
7) A. Ambainis, K. Iwama, A. Kawachi, R. Raymond, and S. Yamashita. Improved algorithms for quantum identification of Boolean oracles. *Theor. Comput. Sci.* 378(1):41–53, 2007.
8) A.Ambainis and R.de Wolf. Average-Case Quantum Query Complexity. In *Proc. 17th STACS, Lecture Notes in Comput. Sci.* 1770:133–144, 2000.
9) R.Beals, H.Buhrman, R.Cleve, M.Mosca, and R.deWolf. Quantum lower bounds by polynomials. *J. ACM* 48(4):778–797, 2001.
10) H.Buhrman, R.Cleve, R.deWolf, and C.Zalka. Bounds for small-error and zero-error quantum algorithms. In *Proc. 40th FOCS*, pages 358–368, 1999.
11) H.Buhrman, N.Vereschagin, and R.de Wolf. On computation and communication with small bias. In *Proc.22nd CCC*, pages 24–32, 2007.
12) H.Buhrman and R.de Wolf. Communication Complexity Lower Bounds by Polynomials. In *Proc. 16th CCC*, pp.120-130, 2001.
13) W.van Dam. Quantum oracle interrogation: getting all information for almost half the price. In *Proc. 39th FOCS*, pages 362–367, 1998.
14) D.Deutsch and R.Jozsa. Rapid Solution of Problems by Quantum Computation. *Proceedings of the Royal Society A*, 435:563–574, 1991.
15) L.K.Grover. A fast quantum mechanical algorithm for database search. In *Proc. 28th STOC*, pages 212–219, 1996.
16) A.W.Harrow and A.Winter. How many copies are needed for state discrimination? quant-ph/0606131v1, 2006.
17) P.Høyer, M.Mosca, and R.de Wolf. Quantum search on bounded-error inputs. In *Proc. 30th ICALP, Lecture Notes in Comput. Sci.* 2719: 291–299, 2003.
18) P.Høyer, R.Špalek. Lower bounds on quantum query complexity. *Bulletin of the EATCS* 87: 78-103, 2005.
19) A.Montanaro, H.Nishimura, and R.Raymond. Unbounded-error quantum query complexity. In *Proc. 19th ISAAC, Lecture Notes in Comput. Sci.*, 5369: 919-930, 2008.
20) R. O'Donnel and R. A. Servedio. Extremal properties of polynomial threshold functions. *J. Comput. Sys. Sci.*, 74:298–312, 2008.
21) D. R. Simon. On the Power of Quantum Computation. *SIAM J. on Comput.*, 26(5):1474-1483, 1997.