# Graph Entropy and Quantum Sorting Problems[*]

## [Extended Abstract]

Andrew Chi-Chih Yao
Computer Science Department
Princeton University
Princeton, NJ 08544
yao@cs.princeton.edu

## ABSTRACT

Let $P = (X, <_P)$ be a partial order on a set of $n$ elements $X = \{x_1, x_2, \cdots, x_n\}$. Define the *quantum sorting problem* $\text{QSORT}_P$ as: given $n$ distinct numbers $x_1, x_2, \cdots, x_n$ consistent with $P$, sort them by a quantum decision tree using comparisons of the form "$x_i : x_j$". Let $Q_\epsilon(P)$ be the minimum number of queries used by any quantum decision tree for solving $\text{QSORT}_P$ with error less than $\epsilon$ (where $0 < \epsilon < 1/10$ is fixed). It was proved by Høyer, Neerbek and Shi (*Algorithmica* **34** (2002), 429-448) that, when $P_0$ is the empty partial order, $Q_\epsilon(P_0) \geq \Omega(n \log n)$, i.e., the classical information lower bound holds for quantum decision trees when the input permutations are unrestricted.

In this paper we show that the classical information lower bound holds, up to an additive linear term, for quantum decision trees for any partial order $P$. Precisely, we prove $Q_\epsilon(P) \geq c \log_2 e(P) - c'n$ where $c, c' > 0$ are constants and $e(P)$ is the number of linear orderings consistent with $P$. Our proof builds on an interesting connection between sorting and Körner's graph entropy that was first noted and developed by Kahn and Kim (*JCSS* **51**(1995), 390-399).

## Categories and Subject Descriptors

F.1 [**Theory of Computation**]: Computation by Abstract Devices

## General Terms

Theory

## Keywords

Graph entropy, information lower bound, partial order, quantum algorithms, sorting

[*]This research was supported in part by the National Science Foundation grant CCR-0310466

## 1. INTRODUCTION

How much can a computation be helped by the use of quantum algorithms has often been studied in the black-box model, which is also called the oracle model, or the quantum decision tree model. To determine the value of a function $f(x_1, x_2, \cdots, x_N)$, queries of the form "$x_i = ?$" are successively asked and the quantum state gets updated by this information. After a predetermined number of steps, the quantum state is measured to produce the output. This standard model has been extensively studied in recent years, and several lower bound techniques have been developed and applied. For detailed descriptions of the model, we refer the readers to recent literature on this subject (see e.g., [1, 2, 3, 4, 5, 10, 19]).

Exactly when and how much speed-up can be achieved is still unresolved for many problems in the quantum decision tree model. In this paper we focus on a specific issue which relates to the information lower bound for classical decision trees. To identify an unknown item taken from a pool of $M$ possibilities, it takes at least $\log_2 M$ tests in the classical framework if the information obtained from one test is only 1 bit. However, this is no longer true in general when quantum states can be used to collect and process information.

Two of the fundamental search problems in which information bounds play a part are the ordered table search, and the sorting problem. In both cases, there is a natural information bound in the classical decision tree model, with a matching upper bound. In the quantum setting, the ordered table search problem has been extensively studied (Farhi et al [8], Ambainis [2], Buhrman and de Wolf [6]), and it was demonstrated in these papers that the information bound $\Omega(\log n)$ is asymptotically valid for quantum decision trees. For the sorting problem, Høyer, Neerbek and Shi [10] showed that the information lower bound also remains valid for quantum decision trees, i.e., $\Omega(n \log n)$ quantum queries are needed.

In this paper, we study a class of problems known as the sorting problems for partial orders. Let $P$ be a partial order on a set of $n$ elements $\{x_1, x_2, \cdots, x_n\}$. Given $n$ input numbers consistent with $P$, Fredman [9] showed that there is a classical decision tree using $\log e(P) + 2n$ binary comparisons $x_i : x_j$ to determine the linear orderings of these numbers, where $e(P)$ is the number of linear orderings consistent with $P$. Subsequent work by Kahn and Saks [13] (see also [11, 12]) showed that $O(\log e(P))$ comparisons are sufficient. Thus, the information lower bound $\log_2 e(P)$ is asymptotically tight for classical decision trees. For quantum decision trees, it is not clear whether the information lower bound is valid (except it is known to be true when $P$ is empty as shown in [10]). This is the focus of our inquiry.

Define the *quantum sorting problem* $\mathrm{QSORT}_P$ as: given $n$ distinct numbers $x_1, x_2, \cdots, x_n$ consistent with $P$, sort them by a quantum decision tree using comparisons of the form "$x_i : x_j$". Let $Q_\epsilon(P)$ be the minimum number of queries used by any quantum decision tree for solving $\mathrm{QSORT}_P$ with error less than $\epsilon$ (where $0 < \epsilon < 1/10$ is fixed).

Our main result is that the classical information lower bound holds, up to additive linear term, for quantum decision trees for any partial order $P$. Let $0 < \epsilon < 1/10$ be any fixed constant.

**Theorem 1** There exist absolute constants $c, c'$ such that $Q_\epsilon(P) \geq c \log e(P) - c'n$.

Our proof relies on the general approach used in [10], and builds on an interesting connection between sorting and Körner's graph entropy in the classical decision tree setting that was first noted and developed by Kahn and Kim [11]. The main underlying idea in our proof is that, when adapted for partial order $P$, the complexity measure used in [10] turns out to be almost the same as the entropy considered in [11].

## 2. PRELIMINARIES

### 2.1 Review of Lower Bounds in Høyer et al

We review the lower bound proof by Høyer, Neerbek and Shi [10] on quantum sorting. We begin with a general framework. Let $f : S \rightarrow \{0,1\}^m$, where $S \subseteq \{0,1\}^N$. Consider a quantum decision tree for computing $f$ with probability error bounded by $\epsilon > 0$. Let $|\xi_x^j>$ be the quantum state after $j$ oracle steps, when $x \in S$ is the oracle. Then $|\xi_x^0>$ is equal to some fixed initial state (independent of $x$). After $T$ steps at the end of computation, we must have $| < \xi_x^T | \xi_y^T > | \leq \epsilon'$ if $f(x) \neq f(y)$, where $\epsilon' = 2(\epsilon(1-\epsilon))^{1/2}$.

To prove a lower bound, a weight function $\omega : S \times S \rightarrow [0, \infty)$ is chosen. Define, for each $0 \leq j \leq T$,

$$W_j = \sum_{x,y \in S} w(x,y) < \xi_x^j | \xi_y^j > .$$

If one can manage to show that, independent of the quantum algorithm used, there is an upper bound $\delta$ to $|W_j - W_{j+1}|$, and a lower bound $M$ to $W_0 - W_T$, then one obtains a lower bound $T \geq M/\delta$ on the quantum complexity.

Consider the sorting of $n$ numbers $x_1, x_2, \cdots, x_n$ by comparisons of the form $x_i : x_j$. In this case we have a function $f : S \rightarrow \{0,1\}^m$ where $S \subseteq \{0,1\}^N$ (with $N = n(n-1)$ and $m \geq \log_2(n!)$). The set $S$ consists of all the oracles $x = (x_{i,j} | i \neq j)$ that represent a set of $n(n-1)$ bits consistent with some underlying linear orderings of the $x_i$'s. Thus, we can identify each oracle $x^\sigma \in S$ with a unique permutation $\sigma$ of $\{1, 2, \cdots, n\}$, such that $x_{i,j}^\sigma = 1$ if and only if $\sigma(i) < \sigma(j)$.

In [10], a lower bound to sorting $n$ elements was obtained by an ingenious choice of the weight function $\omega$. Let $\sigma$ be any permutation of $\{1, 2, \cdots, n\}$. For every $1 \leq k \leq n-1$ and $1 \leq d \leq n-k$, define a new permutation $\sigma^{(k,d)} = (k, k+1, \cdots, k+d) \circ \sigma$. In other words, let $(x_1, x_2, \cdots, x_n) = (\sigma(1), \sigma(2), \cdots, \sigma(n))$ be the assignment of values to $x_i$'s corresponding to $\sigma$. Then the assignment $(x'_1, x'_2, \cdots, x'_n)$ corresponding to $\sigma^{(k,d)}$ is obtained from $(x_1, x_2, \cdots, x_n)$ by replacing the entry $i$ with $i+1$ for $k \leq i < k+d$, and the entry $k+d$ with $i$. Thus, the following is true for $\tau = \sigma^{(k,d)}$:

$$\sigma^{-1}(i) = \begin{cases} \tau^{-1}(k) & \text{if } i = k+d \\ \tau^{-1}(i+1) & \text{if } k \leq i < k+d \\ \tau^{-1}(i) & \text{otherwise} \end{cases}$$

Define the weight function

$$\begin{aligned} \omega(\sigma, \tau) &= \frac{1}{d} \text{ if } \tau = \sigma^{(k,d)} \text{ for some } k \text{ and } d; \\ \omega(\sigma, \tau) &= 0 \text{ otherwise.} \end{aligned}$$

With this choice, they were able to show the bounds below, for any quantum decision tree.

**Lemma 1**[10] For each $0 \leq j < T$, $|W_j - W_{j+1}| \leq 2\pi n!$.

**Lemma 2**[10] $W_0 = n!(nH_n - n)$, $W_T \leq \epsilon' W_0$, where $H_n = \sum_{1 \leq i \leq n} 1/j$.

It follows from Lemmas 1 and 2 that $T \geq \Omega(n \log n)$ for the sorting problem (with $n$ unrestricted input numbers).

### 2.2 Review of Polytopes and Graph Entropy

We first review some concepts and results from Stanley [21]. Let $P = (X, <_P)$ be a partial order on a set $X = \{x_1, x_2, \cdots, x_n\}$. A point $y = (y_1, y_2, \cdots, y_n) \in R^n$ is said to be *consistent* with $P$, if $y_i \leq y_j$ whenever $x_i \leq_P x_j$. For any $y$ consistent with $P$, and for each $1 \leq i \leq n$, let $d_i(y) = y_i$ if $x_i$ is a minimal element in the partial order $P$; otherwise, let $d_i(y)$ be the minimum of $y_i - y_j$ for any $j$ satisfying $x_j <_P x_i$.

The *order polytope* $\mathcal{O}(P)$ is the set of all the points $y = (y_1, y_2, \cdots, y_n) \in [0,1]^n$ consistent with $P$. The *chain polytope* $\mathcal{C}(P)$ is the set of points $z = (z_1, z_2, \cdots, z_n) \in [0,1]^n$ satisfying $z_{i_1} + z_{i_2} + \cdots + z_{i_k} \leq 1$ if $x_{i_1} <_P x_{i_2} <_P \cdots <_P x_{i_k}$ is a chain in $P$. Define a *transfer map* $\phi : \mathcal{O}(P) \rightarrow \mathcal{C}(P)$ by the formula $\phi(y) = (d_1(y), d_2(y), \cdots, d_n(y))$.

Let $\Delta(P)$ be the set of all permutations of $\{1, 2, \cdots, n\}$, and recall $e(P) = |\Delta(P)|$. For each $\sigma \in \Delta(P)$, let $\mathcal{O}_\sigma(P)$ be the subset of points in $\mathcal{O}(P)$ consistent with the permutation $\sigma$.

**Lemma 3** [21] (a) For any $\sigma \in \Delta(P)$, $\phi$ is a linear, measure-preserving bijection when its domain is restricted to $\mathcal{O}_\sigma(P)$. (b) $\phi$ is a continuous, piecewise-linear, measure-preserving bijection from $\mathcal{O}(P)$ onto $\mathcal{C}(P)$.

We now review a special *entropy* in connection with partial orders, which is based on the concept *graph entropy* first introduced by Körner [14]. Graph entropy has an extensive literature, including applications to complexity theory (e.g. Körner [15], Newman, Ragde and Wigderson [16], Radhakhrishnan [17]). We refer the readers to Csiszár et al [7], or Simonyi's survey [20] for additional information. We restrict our discussions here to those needed for this paper.

For any $z = (z_1, z_2, \cdots, z_n)$ with $z_i > 0$, define

$$\psi(z) = \log_2 n - \frac{1}{n} \sum_{1 \leq i \leq n} \log_2 \frac{1}{z_i}. \qquad (1)$$

In connection with sorting problems (in the classical setting), Kahn and Kim [11] define the following entropy notion associated with $P$:

$$H(\overline{P}) = \max\{\psi(z) \,|\, z \in \mathcal{C}(P)\}. \qquad (2)$$

This can be described alternatively as the graph entropy of the comparability graph of $P$. It is known [7] that the maximum is finite and achieved at a unique point $z$ in the polytope.

**Lemma 5** [11] $\log_2 e(P) \leq nH(\overline{P}) \leq O(\log_2 e(P))$.

## 3. PROOF OF THEOREM 1

We adopt the general approach in [10] described in the previous section. Let the set of oracles $S$ consist of those labelled by permutations in $\Delta(P)$. Define the weight function $\omega$ in the same way, except of course $\omega$ is defined only on $S \times S$. Consider any quantum decision tree $B$ for QSORT$_P$ with error bounded by $\epsilon$. Define $\epsilon', W_j$ as in the previous section. Let $A_P = \sum_{\sigma,\tau \in \Delta(P)} \omega(\sigma, \tau)$.

**Lemma 6** For each $0 \le j < T$, $|W_j - W_{j+1}| \le 2\pi e(P)$.

**Lemma 7** $W_0 - W_T \ge (1 - \epsilon')A_P$.

The proof of Lemma 6 is exactly the same as the proof of Lemma 1 as given in [10]. Lemma 7 follows easily from the requirement on the initial and final quantum states produced by the quantum decision tree $B$.

**Proposition 1** There exist absolute constants $\lambda, \lambda' > 0$ such that

$$\frac{1}{e(P)} A_P \ge \lambda \log e(P) - \lambda' n.$$

It follows immediately from Lemmas 6, 7 and Proposition 1 that

$$\begin{aligned} T &\ge \frac{(1-\epsilon')A_P}{2\pi e(P)} \\ &\ge \frac{1-\epsilon'}{2\pi}(\lambda \log e(P) - \lambda' n), \end{aligned}$$

which gives Theorem 1.

We prove Proposition 1 in two steps.

**Proposition 2**

$$\frac{1}{e(P)} A_P \ge \Omega(n \mathbf{E}_{z \in \mathcal{C}(P)}(\psi(z))).$$

**Proposition 3** There exists an absolute constant $\mu > 0$ such that

$$\mathbf{E}_{z \in \mathcal{C}(P)}(\psi(z)) \ge H(\overline{P}) - \mu.$$

Clearly, Proposition 1 follows from Lemma 5, Propositions 2 and 3. We now prove Proposition 2.

Note that each permutation $\sigma \in \Delta(P)$ gives rise naturally to a point $(\sigma(1), \sigma(2), \cdots, \sigma(n))$ in $R^n$, and we shall use the notation $d_i(\sigma)$ with this understanding.

**Lemma 8** For any $\sigma \in \Delta(P)$,

$$\log_2 d_i(\sigma) \ge \log_2(n+1) + \mathbf{E}_{y \in \mathcal{O}_\sigma(P)}(\log_2 d_i(y)).$$

**Proof of Lemma 8** First consider the case when $x_i$ is not a minimal element under partial order $P$. Let $j$ be such that $x_j <_P x_i$, $d_i(\sigma) = \sigma(i) - \sigma(j)$. Then the expected value of $d_i(y)$ for a random $y \in \mathcal{O}_\sigma(P)$ is the expected difference between the $\sigma(i)$-th and the $\sigma(j)$-th smallest elements among $n$ randomly chosen real numbers in the interval $[0, 1]$. By standard results from order statistics, we have

$$\mathbf{E}_{y \in \mathcal{O}_\sigma(P)}(d_i(y)) = \frac{1}{n+1} d_i(\sigma). \qquad (3)$$

It is easy to verify that the above formula is valid in the other case (when $x_i$ is a minimal element under $P$). Now using the convexity of logarithm and Equation (3), we obtain

$$\begin{aligned} \mathbf{E}_{y \in \mathcal{O}_\sigma(P)}(\log_2 d_i(y)) &\le \log_2(\mathbf{E}_{y \in \mathcal{O}_\sigma(P)}(d_i(y)) \\ &= \log_2(\frac{1}{n+1} d_i(\sigma)). \end{aligned}$$

This proves Lemma 8. *Q.E.D.*

It follows from Lemma 8 that

$$\begin{aligned} &\frac{1}{e(P)} \sum_{\sigma \in \Delta(P)} \log_2 d_i(\sigma) \\ &\ge \log_2 n + \frac{1}{e(P)} \sum_{\sigma \in \Delta(P)} \mathbf{E}_{y \in \mathcal{O}_\sigma(P)}(\log_2 d_i(y)) \\ &= \log_2 n + \mathbf{E}_{y \in \mathcal{O}(P)}(\log_2 d_i(y)). \qquad (4) \end{aligned}$$

From the definition of $A_P$ and $\omega$, we have

$$\begin{aligned} A_P &\ge \sum_{\sigma \in \Delta(P)} \sum_{1 \le i \le n} (1 + \frac{1}{2} + \cdots + \frac{1}{d_i(\sigma) - 1}) \\ &\ge \sum_{\sigma \in \Delta(P)} \sum_{1 \le i \le n} \Omega(\log_2 d_i(\sigma)). \qquad (5) \end{aligned}$$

It follows from (4) and (5) that

$$\begin{aligned} \frac{1}{e(P)} A_P &\ge \Omega(n \log_2 n + \sum_{1 \le i \le n} \mathbf{E}_{y \in \mathcal{O}(P)}(\log_2 d_i(y))) \\ &= \Omega(n \log_2 n + \mathbf{E}_{y \in \mathcal{O}(P)}(\sum_{1 \le i \le n} \log_2 d_i(y))). \end{aligned}$$

By Lemma 3, $\phi_P$ is a 1-1 onto measure-preserving mapping from $\mathcal{O}(P)$ to $\mathcal{C}(P)$. This leads to

$$\begin{aligned} \frac{1}{e(P)} A_P &\ge \Omega(n \log_2 n + \mathbf{E}_{z=(z_1,\cdots,z_n) \in \mathcal{C}(P)}(\sum_{1 \le i \le n} \log_2 z_i)) \\ &= \Omega(n \mathbf{E}_{z \in \mathcal{C}(P)}(\psi(z))). \end{aligned}$$

This proves Proposition 2.

To finish the proof of Proposition 1, we now prove Proposition 3. First we derive two lemmas. Let $Q_n$ be the set of all $y = (y_1, \cdots, y_n) \in R^n$ such that $y_i \ge 0$ and $\sum_{1 \le i \le n} y_i \le 1$.

**Lemma 9** Let $\mu \ge 100$. Take a random $y = (y_1, y_2, \cdots, y_n)$ uniformly chosen from $Q_n$. Then,

$$\Pr\{\sum_{1 \le i \le n} \log_2 \frac{1}{y_i} \ge n \log_2 n + \mu n\} \le e^{-\mu n/4}.$$

**Proof of Lemma 9** The proof is technical, and will be delayed to the Appendix.

Let $F_n \subseteq R^n$ be the set of all $w = (w_1, w_2, \cdots, w_n)$ such that $w_i \ge 0$ for all $i$ and $\sum_{1 \le i \le n} w_i \le n$. For any $w = (w_1, w_2, \cdots, w_n) \in F_n$, let $Y$ denote the function on $F_n$ defined by $Y(w) = \frac{1}{n} \sum_{1 \le i \le n} \log_2 w_i$.

**Lemma 10** Let $\mu = 200$, and let $D \subseteq F_n$ be a polytope of volume no less than 1. Then $\mathbf{E}_{w \in D}(Y) \ge -\mu$, for all sufficiently large $n$.

**Proof of Lemma 10**

We assume that $n$ is sufficiently large for all asymptotic inequalities (such as $n^2 e^{-n} < e^{-n/6}$) valid. Let $F_n^-$ be the set of all $w \in F_n$ satisfying $Y(w) \le -\mu/2$. Let $Q_n^-$ be the set of all $y \in Q_n$ satisfying $\sum_{1 \le k \le n} \log_2 \frac{1}{y_k} \ge n \log_2 n + \mu n/2$. Note that $F_n$ is $Q_n$ scaled up by a factor of $n$ on all sides, and $F_n^-$ is $Q_n^-$ scaled up by a factor of $n$. The probability of a random $w \in F_n$ falling into $F_n^-$ is exactly equal to the probability for a random $y \in Q_n$ falling into $Q_n^-$. By Lemma 9, we conclude that

$$\begin{aligned} Vol(F_n^-) &\le e^{-\mu n/8} Vol(F_n) \\ &= e^{-\mu n/8} \frac{n^n}{n!} \\ &\le e^{-\mu n/9}. \qquad (6) \end{aligned}$$

Let $J_n$ be the set of all $w = (w_1, w_2, \cdots, w_n) \in F_n$ with $0 \leq w_1 \leq 2^{-\mu n}$.

With the help of equation (6) we obtain

$$\begin{aligned} & \int_{F_n^-} |\log_2 w_1| dw \\ \leq & \int_{J_n} |\log_2 w_1| dw + \int_{F_n^- - J_n} |\log_2 w_1| dw \\ \leq & \frac{n^{n-1}}{(n-1)!} \int_0^{2^{-\mu n}} |\log_2 w_1| dw_1 \\ & \qquad + Vol(F_n^-) |\log_2 (2^{-\mu n})| \\ \leq & e^n (\mu n + 1) 2^{-\mu n} + e^{-\mu n/9} \mu n \\ \leq & e^{-\mu n/10}. \end{aligned}$$

This implies

$$\begin{aligned} \int_{F_n^-} |Y(w)| dw & \leq \frac{1}{n} \sum_{1 \leq i \leq n} \int_{F_n^-} |\log_2 w_i| dw \\ & = \int_{F_n^-} |\log_2 w_1| dw \\ & \leq e^{-\mu n/10}. \end{aligned} \tag{7}$$

Using (7) we obtain

$$\begin{aligned} \int_D Y(w) dw & = \int_{D - F_n^-} Y(w) dw + \int_{D \cap F_n^-} Y(w) dw \\ & \geq -Vol(D - F_n^-) \mu / 2 - \int_{F_n^-} |Y(w)| dw \\ & \geq -Vol(D) \mu / 2 - e^{-\mu n/10}. \end{aligned}$$

As $Vol(D) \geq 1$, this leads to

$$\begin{aligned} \mathbf{E}_{w \in D}(Y) & = \frac{\int_D Y(w) dw}{Vol(D)} \\ & \geq -\frac{\mu}{2} - \frac{e^{-\mu n/10}}{Vol(D)} \\ & \geq -\mu. \end{aligned}$$

This completes the proof of Lemma 10. *Q.E.D.*

We now use Lemma 10 to prove Proposition 3. By the definition of $H(\overline{P})$, Proposition 3 can be written as

$$\mathbf{E}_{z \in \mathcal{C}(P)}(\psi(z)) \geq \psi(a) - \mu,$$

where $a$ is the unique point $z \in \mathcal{C}(P)$ for $\psi(z)$ to achieve its maximum. As seen in [7, 11], $C(P)$ is contained in the positive quadrant (all $z_i \geq 0$) of $R^n$ bounded by $\sum_{1 \leq i \leq n} \frac{z_i}{a_i} \leq n$.

Make the change of variables $z_i = a_i w_i$ for $1 \leq i \leq n$. Then $z \rightarrow w$ gives a one-to-one linear mapping of $C(P)$ in the $z_i$ space onto a polytope $C'(P)$ in the $w_i$ space, with the transformation $dz = (\prod_{1 \leq i \leq n} a_i) dw$ between the infinitesimal volume elements. Note that $C'(P)$ (which includes all points in $[0,1]^n$) is a polytope of volume $\geq 1$ in the positive quadrant of $R^n$ bounded by $\sum_{1 \leq i \leq n} w_i \leq n$, i.e., $C'(P) \subseteq F_n$. By Lemma 10, we have

$$\mathbf{E}_{w \in C'(P)}(Y) \geq -\mu. \tag{8}$$

Now note that

$$\begin{aligned} \psi(z) & = \log_2 n + \frac{1}{n} \sum_{1 \leq i \leq n} \log_2 (a_i w_i) \\ & = \log_2 n + \frac{1}{n} \sum_{1 \leq i \leq n} \log_2 (a_i) + \frac{1}{n} \sum_{1 \leq i \leq n} \log_2 (w_i) \\ & = \psi(a) + Y(w). \end{aligned}$$

Thus,

$$\mathbf{E}_{z \in \mathcal{C}(P)}(\psi(z)) = \psi(a) + \mathbf{E}_{w \in C'(P)}(Y). \tag{9}$$

It follows from (8) and (9) that

$$\mathbf{E}_{z \in \mathcal{C}(P)}(\psi(z)) \geq \psi(a) - \mu.$$

This proves Proposition 3.

## 4. DISCUSSIONS

We suggest several open problems for future investigations. Firstly, can one strengthen Theorem 1 to $Q_\epsilon(P) \geq \Omega(\log e(P))$? We conjecture that in fact the quantity $A_P$ in Lemma 7 provides such a lower bound, even though the estimation techniques used in the present paper are not strong enough to prove it.

Secondly, to what extent do various classical sorting lower bounds remain valid? For example, suppose we allow any ternary polynomial tests $p(x_1, x_2, \cdots, x_n) : 0$, what is the number of quantum queries needed to sort $n$ numbers?

Lastly, there are other information lower bounds whose validity in quantum case is non-obvious. For example, see Shi [18]. It would be very interesting to study systematically all the classical information lower bounds for decision trees.

## 5. REFERENCES

[1] S. Aaronson. Quantum lower bounds for the collision problem. In *Proc. 34th Annual ACM Symposium on Theory of Computing*, pages 635–642. ACM, 2002.

[2] A. Ambainis. A better lower bound for quantum algorithms searching an ordered list. In *Proc. 40th Annual IEEE Symposium on Foundations of Computer Science*, pages 352–357. IEEE, 1999.

[3] A. Ambainis. Quantum lower bounds by quantum arguments. In *Proc. 32nd Annual ACM Symposium on Theory of Computing*, pages 636–643. ACM, 2000.

[4] R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. de Wolf. Quantum lower bounds by polynomials. In *Proc. 39th Annual IEEE Symposium on Foundations of Computer Science*, pages 352–361. IEEE, 1998.

[5] C. Bennett, E. Bernstein, G. Brassard, and U. Vazirani. Strengths and weaknesses of quantum computation. *SIAM J. on Computing*, 26:1510–1523, 1997.

[6] H. Buhrman and R. de Wolf. A lower bound for quantum search of an ordered list. *Information Processing Letters*, 70:205–209, 1999.

[7] C. Csiszár, J. Körner, L. Lovász, K. Marton, and G. Simonyi. Entropy splitting for antiblocking corners and perfect graphs. *Combinatorica*, 10:27–40, 1990.

[8] E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser. A limit on the speed of quantum computation for insertion into an ordered list. *arXiv.org e-Print archive*, quant-ph/9812057, 1998.

[9] M. Fredman. How good is the information bound in sorting. *Theoretical Computer Science*, 1:355–361, 1976.

[10] P. Høyer, J. Neerbek, and Y. Shi. Quantum complexities of ordered searching, sorting, and element distinctness. *Algorithmica*, 34:429–448, 2002.

[11] J. Kahn and J. Kim. Entropy and sorting. *Journal of Computer and System Sciences*, 51:390–399, 1995.

[12] J. Kahn and N. Linial. Balancing poset extensions via Brunn-Minkowski. *Combinatorica*, 11:363–368, 1991.

[13] J. Kahn and M. Saks. Balancing poset extensions. *Order*, 1:113–126, 1984.

[14] J. Körner. Coding of an information source having ambiguous alphabet and the entropy of graphs. In *Transactions of 6th Prague Conference on Information Theory, etc.*, pages 411–425. Academia, Prague, 1973.

[15] J. Körner. Fredman-Komlós bounds and information theory. *SIAM Journal on Alg. Disc. Meth.*, 7:560–570, 1986.

[16] I. Newman, P. Ragde, and A. Wigderson. Perfect hashing, graph entropy and circuit complexity. In *Proc. 5th Annual IEEE Symposium on Structure in Complexity Theory*, pages 91–100. IEEE, 1990.

[17] J. Radhakhrishnan. Better bounds for threshold formulas. In *Proc. 32nd Annual IEEE Symposium on Foundations of Computer Science*, pages 314–323. IEEE, 1991.

[18] Y. Shi. Entropy lower bounds of quantum decision tree complexity. *Information Processing Letters*, 81(1):23–27, 2002.

[19] Y. Shi. Quantum lower bounds for the collision and the element distinctness problems. In *Proc. 43rd Annual IEEE Symposium on Foundations of Computer Science*, pages 513–519. IEEE, 2002.

[20] G. Simonyi. Perfect graphs and graph entropy: An updated survey. Chapter 13 in *Perfect Graphs*, edited by J. Ramirez-Alfonsin and B. Reed, pages 293–328. John Wiley and Sons, 2001.

[21] R. Stanley. Two poset polytopes. *Discrete Computational Geometry*, 1:9–23, 1986.

# APPENDIX

## Proof of Lemma 9

Pick $n$ random real numbers $x_j$ uniformly and independently from the interval $[0,1]$, and sort them into ascending order $x_{i_1} \leq x_{i_2} \leq \cdots \leq x_{i_n}$. Let $y = (y_1, y_2, \cdots, y_n)$ where $y_1 = x_{i_1}$, and $y_j = x_{i_j} - x_{i_{j-1}}$, $2 \leq j \leq n$. It is easily verified that the generated $y$ is a random point uniformly distributed over $Q_n$.

Consider the random variable

$$Y_n(x_1, x_2, \cdots, x_n) = -\sum_{1\leq j\leq n} \ln y_j. \tag{10}$$

To prove Lemma 9, we need to show that

$$\Pr\{Y_n \geq n\ln n + (\mu \ln 2)n\} \leq e^{-\mu n/4}. \tag{11}$$

Let us generate $x_1, x_2, \cdots, x_n$ sequentially. When we have generated $k$ random numbers $x_1, x_2, \cdots, x_k$, let $Y_k$ denote the random variable defined as in equation (10) above, except with $n$ replaced by $k$. Let $Y_0$ be the constant random variable 0.

**Lemma A1** Let $1 \leq k \leq n$, and let $x_1, x_2, \cdots, x_{k-1}$ be distinct. Then

$$Y_k(x_1, x_2, \cdots, x_k) \leq Y_{k-1}(x_1, x_2, \cdots, x_{k-1}) + \ln(2/\delta),$$

where $\delta \geq 0$ is the minimum distance between $x_k$ and any numbers in the set $\{0, x_1, x_2, \cdots, x_{k-1}\}$.

**Proof of Lemma A1** The numbers $x_1, x_2, \cdots, x_{k-1}$ divide the interval $[0,1]$ into $k$ sub-intervals. Suppose that $x_k$ falls into a sub-interval $I$ of length $u$, and splits it into two parts of length $\lambda u$ and $(1-\lambda)u$. There are two cases. If $I$ is not the rightmost sub-interval, then

$$\begin{aligned} & Y_k(x_1, x_2, \cdots, x_k) - Y_{k-1}(x_1, x_2, \cdots, x_{k-1}) \\ = & -\ln(\lambda u) - \ln((1-\lambda)u) + \ln u \\ \leq & \ln(\frac{2}{\delta}). \end{aligned}$$

If $I$ is the rightmost interval, then

$$\begin{aligned} & Y_k(x_1, x_2, \cdots, x_k) - Y_{k-1}(x_1, x_2, \cdots, x_{k-1}) \\ = & -\ln(\lambda u) \\ \leq & \ln(\frac{1}{\delta}). \end{aligned}$$

This proves Lemma A1. *Q.E.D.*

Let us regard $Y_0 = 0, Y_1, Y_2, \cdots, Y_n$ as a sequence of random variables where $Y_k$ depends only on $x_1, x_2, \ldots, x_k$. It follows immediately from Lemma A1 that,

$$\Pr\{Y_k - Y_{k-1} \geq t\} \leq 4ke^{-t} \tag{12}$$

for all $t \geq 0$.

Consider the probability distribution $\rho_k$ over $[0, \infty]$:

$$\rho_k(t) = \begin{cases} 0 & t \leq \ln(4k) \\ 4ke^{-t} & t > \ln(4k). \end{cases}$$

Let $A_k$ be a real-valued random variable defined on some probability space such that $\rho_k$ is the density function for the distribution of the value of $A_k$, i.e.,

$$\Pr\{A_k \geq t\} = \int_t^\infty \rho_k(\tau)d\tau. \tag{13}$$

It is easily verified from equations (12) and (13) that $\Pr\{Y_k - Y_{k-1} \geq t\} \leq \Pr\{A_k \geq t\}$. That is, $Y_k - Y_{k-1}$ is stochastically dominated by $A_k$. Now note that $Y_n = \sum_{1\leq k\leq n}(Y_k - Y_{k-1})$. This means

$$\Pr\{Y_n \geq T\} \leq \Pr\{A \geq T\}, \tag{14}$$

where $A = \sum_{1\leq k\leq n} A_k$.

**Lemma A2** Let $T = n\ln n + (\mu \ln 2)n$, then

$$\Pr\{A \geq T\} \leq e^{-\mu n/4}.$$

**Proof** Let $T' = T - \ln(4^n n!)$. Then

$$\begin{aligned} & \Pr\{A \geq T\} \\ = & \int_{\substack{t=(t_1,\cdots,t_n) \\ \sum_k t_k \geq T}} dt \prod_{1\leq k\leq n} \rho_k(t_k) \\ = & (\prod_{1\leq k\leq n}(4k)) \int_{\substack{t=(t_1,\cdots,t_n) \\ t_k\geq\ln(4k), \sum_k t_k \geq T}} e^{-\sum_k t_k} dt \\ = & \int_{\substack{s=(s_1,\cdots,s_n) \\ s_k\geq 0, \sum_k s_k \geq T'}} e^{-\sum_k s_k} ds \\ = & \int_{T'}^\infty e^{-v} \frac{v^{n-1}}{(n-1)!} dv. \end{aligned}$$

By assumption $T' = T - \ln(4^n n!) \geq ((\mu \ln 2) - 2)n$, we have for all $v \geq T'$,

$$\frac{v^{n-1}}{(n-1)!} \leq e^{v/2}.$$

It follows that

$$\begin{aligned} \Pr\{A \geq T\} &\leq \int_{T'}^{\infty} e^{-v/2} dv \\ &= 2e^{-T'/2}. \end{aligned}$$

This implies

$$\Pr\{A \geq T\} \leq e^{-\mu n/4},$$

and proves Lemma A2. *Q.E.D.*

Equation (11) follows from Lemma A2 and equation (14). This completes the proof of Lemma 9.