# The Quantum Complexity of Set Membership


Jaikumar Radhakrishnan   Pranab Sen   S. Venkatesh*

School of Technology and Computer Science
Tata Institute of Fundamental Research
Homi Bhabha Road, Mumbai 400005, India
email: {jaikumar, pranab, venkat}@tcs.tifr.res.in



**Abstract**

We study the *quantum complexity* of the static set membership problem: given a subset $S$ ($|S| \leq n$) of a universe of size $m$ ($\gg n$), store it as a table, $T : \{0,1\}^r \to \{0,1\}$, of bits so that queries of the form 'Is $x$ in $S$?' can be answered. The goal is to use a small table and yet answer queries using few bitprobes. This problem was considered recently by Buhrman, Miltersen, Radhakrishnan and Venkatesh [BMRV00], who showed lower and upper bounds for this problem in the classical deterministic and randomised models. In this paper, we formulate this problem in the "quantum bitprobe model". We assume that access to the table $T$ is provided by means of a black box (oracle) unitary transform $O_T$ that takes the basis state $|y, b\rangle$ to the basis state $|y, b \oplus T(y)\rangle$. The query algorithm is allowed to apply $O_T$ on any superposition of basis states.

We show tradeoff results between space (defined as $2^r$) and number of probes (oracle calls) in this model. Our results show that the lower bounds shown in [BMRV00] for the classical model also hold (with minor differences) in the quantum bitprobe model. These bounds almost match the classical upper bounds. Our lower bounds are proved using linear algebraic arguments.


## 1 Introduction

In this paper we study the *static membership* problem: Given a subset $S$ of at most $n$ keys, store it so that queries of the form "Is $x$ in $S$?" can be answered quickly. This fundamental data structure problem has been studied earlier in various settings (e.g. by


*Current address: School of Mathematics, Institute for Advanced Study, Einstein Drive, Princeton, NJ 08540, USA. email: venkat@ias.edu

Minsky and Papert [MP69], Yao [Yao81], Yao & Yao [YY97], Fredman, Komlós and Szemerédi [FKS84]) and Pagh [Pag01]. Most of these results were in the classical deterministic model. Recently, this problem was considered by Buhrman, Miltersen, Radhakrishnan and Venkatesh [BMRV00] in the bitprobe model, which was introduced in [MP69]; they studied tradeoffs between space and number of probes in the classical deterministic model, and also showed lower and upper bounds when the query algorithm was randomised. In the classical bitprobe model the storage scheme is deterministic and stores the given set as a string of bits. The query scheme is either deterministic or randomised and answers membership queries probing only one bit of the string at a time.

In this paper, we allow the query algorithm to perform quantum search on the table. For this we formulate the problem in the quantum bitprobe model. The table of $2^r$ bits, $T : \{0,1\}^r \rightarrow \{0,1\}$, is modelled using an oracle unitary transformation $O_T$ that takes the basis state $|y, b\rangle$ to $|y, b \oplus T(y)$. (This is similar to the model used in Grover's [Gro96] algorithm and in the lower bound proofs of Ambainis [Amb00]). If the inputs are restricted to basis states, this oracle reduces to (the reversible version of) the classical table that stores one bit for each address in $\{0,1\}^r$. We, therefore, define the size of the table to be $2^r$.

The main point of departure from the classical model, is in the query algorithm. We allow the algorithm to feed a superposition of basis states to the oracle. Each use of the oracle counts as one probe of the table, even if a superposition is supplied to the oracle, and the output depends on the value of several bits of the underlying table $T$. In the preparation of this superposition and in the processing of the output returned by the oracle, we allow arbitrary unitary transformations. It is known that this form of access often leads to significant improvement over classical algorithms for several problems (e.g. Grover [Gro96]).

Previously the number of probes to the black box as a complexity measure had been studied in the quantum setting (e.g. [BBBV97], [BBC+98], [Amb00], [Gro96]). Both lower bounds and upper bounds for various problems were proved. The notion of number of bits (space) as a complexity measure had been studied implicitly (e.g. [Nay99], [Hol73]) in the quantum information-theoretic sense.

The main contribution of this paper is the study of the tradeoff between space and number of probes in the quantum bitprobe model. For the set membership problem, we show that several limitations of classical computation (shown in [BMRV00]) continue to persist even if quantum query algorithms are allowed.

## 1.1 Our results

**The exact quantum model:** For deterministic (classical) query algorithms [BMRV00] showed that any $(s,t)$-scheme (which uses space $s$ and $t$ bitprobes) satisfies $\binom{m}{n} \leq \binom{s}{nt} 2^{nt}$ (see [BMRV00]). We show a stronger (!) tradeoff result in the quantum bitprobe model.

> **Result 1:** *Suppose there exists an exact quantum scheme for storing subsets of size at most $n$ from a universe* $\mathbf{U}$ *of size $m$ that uses $s$ bits of storage and*

*answers membership queries with t quantum probes. Then*

$$\sum_{i=0}^{n} \binom{m}{i} \leq \sum_{i=0}^{nt} \binom{s}{i}.$$

This has two immediate consequences. First, by setting $t = 1$, we see that if only one probe is allowed than $m$ bits of storage are necessary. (In [BMRV00], for the classical model, this was justified using an ad hoc argument.) Second, it follows that the classical storage scheme of Fredman, Komlós and Szemerédi [FKS84], which uses $O(n \log m)$ bits of storage and answers queries using $O(\log m)$ bitprobes, is optimal even with quantum query algorithms — schemes that use $O(n \log m)$ bits of storage must make $O(\log m)$ probes if $n \leq m^{1-O(1)}$.

Interestingly, the above theorem holds even in the presence of errors, provided the error is restricted to positive instances, that is the query algorithm sometimes (with probability $< 1$) returns the answer 'No' for a query $x$ that is actually in the set $S$.(See Theorem 2 below.) This was not observed earlier even in the classical model, although one can easily modify the proof of the tradeoff result in [BMRV00] to give this.

**The $\epsilon$-error model:** In the classical randomised model, there exists a two-sided $\epsilon$-error scheme that uses just *one* bitprobe and uses space $O(\frac{n \log m}{\epsilon^2})$; also, any such scheme must use space $\Omega(\frac{n \log m}{\epsilon \log(1/\epsilon)})$ (see [BMRV00]). Since different sets must be represented by different tables, every scheme, no matter how many bitprobes the query algorithm is allowed, must use $\Omega(n \log(m/n))$ bits of storage, even in the quantum model. However, one might ask if the dependence of space on $\epsilon$ is significantly better in the quantum bitprobe model. We show the following lower bound which implies that a quantum scheme needs significantly more than the information-theoretic optimal space if sub-constant error probabilities are desired.

> **Result 2:** *Let $n/m < \epsilon < 1/8$. Suppose there is a quantum scheme with two-sided error $\epsilon$ which answers membership queries using one quantum probe. It must use space*
>
> $$s = \Omega\left(\frac{n \log(m/n)}{\epsilon^{1/6} \log(1/\epsilon)}\right).$$

The method used to prove this result can be generalised to algorithms that use more probes than one.

> **Result 3:** *For any $p \geq 1$ and $n/m < 4^{p-1}\epsilon^p < 1/8$, suppose there is a quantum scheme with two-sided error $\epsilon$ which answers membership queries using $p$ quantum probes. It must use space*
>
> $$s = \Omega\left(\frac{n \log(m/n)}{\epsilon^{1/6p} \log(1/\epsilon)}\right).$$

Such a tradeoff between space and error probability for multiple probes was not known earlier even in the classical randomised model.

The results described above are inspired by similar results proved earlier in [BMRV00] in the classical model. However, the methods used for classical models, which were based on combinatorial arguments involving set systems, seem to be powerless in giving the results in the quantum model. Our results are based on linear algebraic arguments, involving counting the dimensions of spaces of various operators that arise in the quantum query algorithm.

**Bounds for classical models:** As stated above, Theorem 1 is stronger than what was known earlier even in the classical model. One might wonder if this stronger result is somehow easier to prove in the classical model. We show that the linear algebraic techniques used in the proof of Theorem 1 can be considerably simplified when we assume the classical deterministic model, and give the same inequality as stated in Theorem 1. This is stated as Theorem 5 below. Also, one can easily modify the proof to yield the same tradeoff for randomised schemes where the error is restricted to positive instances. See Theorem 6 below.

The proof in [BMRV00] of the space lower bound for a classical two-sided $\epsilon$-error randomised scheme, namely $s = \Omega(\frac{n \log m}{\epsilon \log(1/\epsilon)})$, involved some tricky use of both upper and lower bounds for $r$-cover-free families shown by Nisan and Wigderson [NW94], Erdős, Frankl and Füredi [EFF85], Dyachkov and Rykov [DR82]. The proof of the analogous bound (Theorem 3) in the quantum model completely avoids these. In fact, we give a proof of a (slightly weaker) lower bound $s = \Omega(\frac{n \log m}{\epsilon^{1/4} \log(1/\epsilon)})$ in the classical randomised model by adapting the ideas used in the proof of Theorem 3 to the classical setting. We first diminish the error probability of the one-probe query algorithm by repetition and then we can, by fixing the random coin tosses, make it a deterministic query algorithm which however, uses more than one probe. We then apply our (classical) deterministic space-time tradeoff equation (Theorem 5) to complete the proof. This approach is inspired by our proof of Theorem 3. See Theorem 7 below. Besides being simpler, this proof has the advantage that it generalises readily to more than one probe. No such result was known earlier in the classical setting.

**Result 4:** *Let $p \geq 1$ and $1/12 > \epsilon > n/m^{1/3p}$. Any two-sided $\epsilon$-error classical randomised scheme which stores subsets of size $n$ from a universe* $\mathbf{U}$ *of size $m$ and answers membership queries using at most $p$ bitprobes must use space*

$$\Omega\left(\frac{n \log m}{\epsilon^{2/5p} \log(1/\epsilon)}\right)$$

This result is proved in a manner similar to the one-probe case. See Theorem 8 below. [BMRV00] prove an upper bound of $O(\frac{n \log m}{\epsilon^2})$ for the single bitprobe case. We note that for $p$ bitprobes, an upper bound of $O(\frac{n \log m}{\epsilon^{4/p}})$ follows by taking the storage scheme of [BMRV00] for error probability $\frac{\epsilon^{2/p}}{4}$, and repeating the single probe query scheme $p$ times. This diminishes the probability of error to $\epsilon$ for suitable values of the various parameters.

### 1.2 Organisation of the paper

In the next section, we describe our quantum bitprobe model formally and give the framework of our proofs. Detailed proofs of all our results (for both quantum and classical models) appear in Section 3. We conclude in Section 4 with some open problems.

# 2 Definitions and Notations

In this section we first describe our *quantum bitprobe model* and then give some formal definitions and notations which will be used in the proofs of the theorems.

### 2.1 The model

Our model is a quantum analogue of the classical bitprobe model which has been extensively used in the past to study data structure problems (see e.g. [MP69, Mil93, YY97, BMRV00, Pag01]).

A static data structure problem is a function $f : D \times Q \rightarrow \{0,1\}$, where $D$ is a finite set called the set of data to be stored and $Q$ is a finite set called the set of queries. A classical $(s,t)$-bitprobe scheme for a static data structure problem consists of a deterministic storage scheme which stores the given data $d$ as a bitstring of length $s$, and a query scheme which given a query $q$ makes at most $t$ bitprobes to the stored string and computes $f(d,q)$. A query scheme can be either deterministic or randomised. For more details about the classical model, see [BMRV00].

A quantum $(s,t)$-bitprobe scheme for a static data structure problem has two components: a storage scheme that stores the data $d$ using $s$ bits, and a query scheme that answers queries by 'probing' at most $t$ times.

**The storage scheme:** For the set membership problem, the data to be stored is a subset $S$ of a universe $\mathbf{U}$ ($|S| \leq n$, $|\mathbf{U}| = m$). Let $x = x(S) \in \{0,1\}^s$ be the bitstring that is stored by the storage scheme for recording $S$. The storage scheme is classical deterministic. The difference now, is that this bitstring is made available to the query algorithm in the form of an oracle unitary transform $O_S$. To define $O_S$ formally, we represent the basis states as $|j, b, z\rangle$, where $j \in \{0, \ldots, s-1\}$ is a binary string of length $\log s$, $b$ is either 0 or 1, and $z$ is a binary string of some fixed length.

$O_S$ maps $|j, b, z\rangle$ to $(-1)^{bx_j}|j, b, z\rangle$.

*Remark:* The oracle described in the introduction mapped $|j, b, z\rangle$ to $|j, b \oplus x_j, z\rangle$. It is known [Deu89] that the oracle $O_S$ defined above is equivalent in power (up to constant factors) to this oracle.

Thus, information about the string $x$ appears in the phase of the basis states in the output, and $O_S$ is represented by a diagonal matrix (in the standard observational basis):

the $i$th diagonal entry, where $i \equiv |j, b, z\rangle$), is

$$(O_S)_{i,i} = (-1)^{bx_j}.$$

Here, as before, $x_j$ is the $j$th bit of $x$. We will extend this notation and let $[x]_T \triangleq \sum_{i \in T} x_i \pmod 2$; in particular, $[x]_\emptyset = 0$. Thus, $(O_S)_{i,i} = (-1)^{[x(S)]_{l_i}}$, where $l_i$ is some subset of $[s]$ of size 1 (when $b = 1$, $l_i = \{j\}$) or 0 (when $b = 0$, $l_i = \emptyset$).

*Remark:* Our model for storage does not permit $O_S$ to be any arbitrary unitary transformation. However, this restricted form of the oracle is closer to the way bits are accessed in the classical case. Moreover, in most previous works storage has been modelled using such an oracle (see e.g. [Gro96], [BBC+98], [BBBV97], [Amb00]).

**The query scheme:** The query scheme is a quantum circuit which takes a query as input (together with some ancilla qubits) and returns the answer on one of its output wires. The query scheme can be exact or have error; the error can be one-sided or two-sided. When the query scheme is exact, a measurement at the output wire gives the correct answer with probability 1. If one-sided error $\epsilon$ is allowed, the measurement produces a 0 with probability 1 when the answer is 0, but when the answer is 1, is required to produce a 1 with probability only at least $1 - \epsilon$. If two-sided error $\epsilon$ is allowed, the answer can be wrong, with probability at most $\epsilon$, for both positive and negative instances.

**The framework for the proofs:** We now describe the general framework in which the various proofs are presented and also give some definitions and notations which will be used throughout the paper.

Suppose a subset $S \subseteq \mathbf{U}, |S| \leq n$, has been stored and $x(S) \in \{0,1\}^s$ is the corresponding bitstring. A quantum query scheme with $t$ probes is just a sequence of unitary transformations

$$U_0 \rightarrow O_S \rightarrow U_1 \rightarrow O_S \rightarrow \dots U_{t-1} \rightarrow O_S \rightarrow U_t$$

where $U_j$'s are arbitrary unitary transformations that do not depend on the set $S$ stored. For a query $q, q \in \mathbf{U}$, the computation starts in an observational basis state $|q\rangle|0\rangle$, where we assume that the ancilla qubits are initially in the basis state $|0\rangle$. Then we apply the operators $U_0, O_S, \dots, O_S, U_t$ and measure the final state. The result of the query is the rightmost bit of the state obtained by the measurement.

Define $|\phi_q\rangle \triangleq |q\rangle|0\rangle$. The set of vectors $|\phi_q\rangle, q \in \mathbf{U}$ form an orthogonal system of vectors. They are independent of the set $S$ stored.

Define two Hilbert spaces, $A_0$ and $A_1$, where $A_i$ is the space of all state vectors that can be spanned by basis states having an $i$ at the rightmost bit (i.e. if the state vector lies in $A_i$ then on measuring the rightmost bit at the output, one gets $i$ with probability 1). Then the entire state space $V$ decomposes as an orthogonal direct sum of the spaces $A_0, A_1$.

Define the unitary transformations $\{W_S\}_{S \subseteq \mathbf{U}, |S| \leq n}$ as follows.

$$W_S \triangleq U_t O_S U_{t-1} O_S U_{t-2} O_S \cdots U_2 O_S U_1 O_S U_0$$

Thus when a set $S$ is stored, in the exact quantum case $W_S|\phi_i\rangle, i \in S$ lie in $A_1$ and $W_S|\phi_i\rangle, i \notin S$ lie in $A_0$. In the one-sided $\epsilon$-error case, $W_S|\phi_i\rangle, i \notin S$ lies in $A_0$, but $W_S|\phi_i\rangle, i \in S$ may not lie entirely in $A_1$, but in fact might have a projection onto $A_0$ of length at most $\sqrt{\epsilon}$. In the two-sided $\epsilon$-error case, $W_S$ has to send the vector $|\phi_i\rangle$ "approximately" to the correct space, i.e. the projection of $W_S|\phi_i\rangle$ onto the correct space is of length more than $\sqrt{1-\epsilon}$.

**Notation:** In the proofs we have to take tensor products of vectors and matrices. For any vector $v$ or matrix $M$, we have the following notation,

$$v^{\otimes t} \triangleq \underbrace{v \otimes \cdots \otimes v}_{t \text{ times}}$$

$$M^{\otimes t} \triangleq \underbrace{M \otimes \cdots \otimes M}_{t \text{ times}}$$

We note that since the entire state space $V$ is the orthogonal direct sum of $A_0$ and $A_1$,

$$V^{\otimes t} = A_0^{\otimes t} \oplus (A_0^{\otimes t-1} \otimes A_1) \oplus \cdots \oplus A_1^{\otimes t}$$

and the $2^t$ vector spaces in the above direct sum are pairwise orthogonal.

Below, $A \triangle B$ stands for the symmetric difference of sets $A$ and $B$; $M^\dagger$ stands for the conjugate transpose of the matrix $M$.

# 3 Proofs of Theorems

## 3.1 Quantum Schemes

We first prove our space v/s probes tradeoff result for exact quantum schemes.

**Theorem 1** *Suppose there exists an exact quantum scheme for storing subsets of size at most $n$ from a universe $\mathbf{U}$ of size $m$ that uses $s$ bits of storage and answers membership queries with $t$ quantum probes. Then*

$$\sum_{i=0}^{n} \binom{m}{i} \leq \sum_{i=0}^{nt} \binom{s}{i}.$$

**Proof**: We use the notation of Section 2. For any subset $S \subseteq \mathbf{U}$, $|S| \leq n$, let us define

$$W_S \triangleq U_t O_S U_{t-1} O_S U_{t-2} O_S \cdots U_2 O_S U_1 O_S U_0$$

**Claim 1** $\{W_S^{\otimes n}\}_{S \in \binom{\mathbf{U}}{\leq n}}$ *are linearly independent.*

**Proof**: Suppose there is a nontrivial linear combination

$$\sum_{S \in \binom{\mathbf{U}}{\leq n}} \alpha_S W_S^{\otimes n} = 0.$$

Let $T$ be a set of largest cardinality such that $\alpha_T \neq 0$ and let $T = \{i_1, \ldots, i_k\}$, $k \leq n$. We define a vector

$$|\phi_T\rangle \triangleq |\phi_{i_1}\rangle^{\otimes(n-k+1)} \otimes |\phi_{i_2}\rangle \otimes \cdots \otimes |\phi_{i_k}\rangle.$$

Applying $|\phi_T\rangle$ to the linear combination above, we have

$$\sum_{S \in \binom{\mathbf{U}}{\leq k}, S \neq T} \alpha_S W_S^{\otimes n} |\phi_T\rangle + \alpha_T W_T^{\otimes n} |\phi_T\rangle = 0. \tag{1}$$

For any set $S$,

$$W_S^{\otimes n} |\phi_T\rangle = (W_S |\phi_{i_1}\rangle)^{\otimes(n-k+1)} \otimes W_S |\phi_{i_2}\rangle \otimes \cdots \otimes W_S |\phi_{i_k}\rangle.$$

- If $S = T$, $W_T |\phi_{i_l}\rangle \in A_1$ for all $l$, $1 \leq l \leq k$. Hence $W_T^{\otimes n} |\phi_T\rangle \in A_1^{\otimes n}$.
- If $S \neq T$, there exists an element $i_j$ in $T - S$ (by choice of $T$). $W_S |\phi_{i_j}\rangle \in A_0$. Hence $W_S^{\otimes n} |\phi_T\rangle \notin A_1^{\otimes n}$. In fact, $W_S^{\otimes n} |\phi_T\rangle$ is orthogonal to $A_1^{\otimes n}$.

Hence, in the above linear combination (equation 1), the only vector which has a nontrivial projection along $A_1^{\otimes n}$ is $W_T^{\otimes n} |\phi_T\rangle$. Hence, $\alpha_T = 0$ leading to a contradiction. ■

**Claim 2** $\{W_S^{\otimes n}\}_{S \in \binom{\mathbf{U}}{\leq n}}$ *lie in a vector space of dimension at most* $\sum_{i=0}^{nt} \binom{s}{i}$.

**Proof**: By definition, for any set $S$, $|S| \leq n$,

$$W_S \triangleq U_t O_S U_{t-1} O_S U_{t-2} O_S \cdots U_2 O_S U_1 O_S U_0$$

where $U_0, \ldots, U_t$ are unitary transformations (matrices) independent of the set stored.

For any pair of indices $(i, j)$,

$$\begin{aligned} (W_S)_{i,j} &= \sum_{k_{t-1}, \ldots, k_0} (U_t)_{i,k_{t-1}} (O_S)_{k_{t-1},k_{t-1}} (U_{t-1})_{k_{t-1},k_{t-2}} (O_S)_{k_{t-2},k_{t-2}} \\ &\qquad \cdots (U_1)_{k_1,k_0} (O_S)_{k_0,k_0} (U_0)_{k_0,j} \\ &= \sum_{k_{t-1}, \ldots, k_0} (U_t)_{i,k_{t-1}} (-1)^{[x(S)]_{l_{k_{t-1}}}} (U_{t-1})_{k_{t-1},k_{t-2}} (-1)^{[x(S)]_{l_{k_{t-2}}}} \\ &\qquad \cdots (U_1)_{k_1,k_0} (-1)^{[x(S)]_{l_{k_0}}} (U_0)_{k_0,j} \end{aligned}$$

where, recalling the notation of Section 2, $x(S)$ is the string stored by the storage scheme for set $S$ and $l_i$ is either the single location in the string corresponding to index $i$ or the empty set.

Therefore, if we define $T_{k_{t-1},\ldots,k_0} \triangleq l_{k_{t-1}} \triangle l_{k_{t-2}} \triangle \cdots \triangle l_{k_0}$ and $[x(S)]_T$ to be the parity of the bits stored in $x(S)$ at the locations of $T$, we have

$$\begin{aligned}(W_S)_{i,j} &= \sum_{k_{t-1},\ldots,k_0} (-1)^{[x(S)]_{T_{k_{t-1},\ldots,k_0}}} (U_t)_{i,k_{t-1}} (U_{t-1})_{k_{t-1},k_{t-2}} \cdots (U_1)_{k_1,k_0} (U_0)_{k_0,j} \\ &= \sum_{T \in \binom{[s]}{\leq t}} (-1)^{[x(S)]_T} \sum_{\substack{k_{t-1},\ldots,k_0 \\ T_{k_{t-1},\ldots,k_0} = T}} (U_t)_{i,k_{t-1}} (U_{t-1})_{k_{t-1},k_{t-2}} \cdots (U_1)_{k_1,k_0} (U_0)_{k_0,j}\end{aligned}$$

Let us define for every set $T \subseteq [s]$, $|T| \leq t$, a matrix $A_T$ as follows:

$$(A_T)_{i,j} \triangleq \sum_{\substack{k_{t-1},\ldots,k_0 \\ T_{k_{t-1},\ldots,k_0} = T}} (U_t)_{i,k_{t-1}} (U_{t-1})_{k_{t-1},k_{t-2}} \cdots (U_1)_{k_1,k_0} (U_0)_{k_0,j}$$

Then we have,

$$W_S = \sum_{T \in \binom{[s]}{\leq t}} (-1)^{[x(S)]_T} A_T.$$

Hence,

$$\begin{aligned}(W_S)^{\otimes n} &= \left( \sum_{T_1 \in \binom{[s]}{\leq t}} (-1)^{[x(S)]_{T_1}} A_{T_1} \right) \otimes \cdots \otimes \left( \sum_{T_n \in \binom{[s]}{\leq t}} (-1)^{[x(S)]_{T_n}} A_{T_n} \right) \\ &= \sum_{\substack{T_i \in \binom{[s]}{\leq t} \\ 1 \leq i \leq n}} (-1)^{[x(S)]_{T_1}} \cdots (-1)^{[x(S)]_{T_n}} (A_{T_1} \otimes \cdots \otimes A_{T_n}) \\ &= \sum_{\tilde{T} \in \binom{[s]}{\leq nt}} (-1)^{[x(S)]_{\tilde{T}}} B_{\tilde{T}}\end{aligned}$$

where for $\tilde{T} \in \binom{[s]}{\leq nt}$,

$$B_{\tilde{T}} \triangleq \sum_{\substack{T_1 \triangle \cdots \triangle T_n = \tilde{T} \\ T_i \in \binom{[s]}{\leq t}, 1 \leq i \leq n}} A_{T_1} \otimes \cdots \otimes A_{T_n}$$

Hence, we see that $\{B_{\tilde{T}}\}_{\tilde{T} \in \binom{[s]}{\leq nt}}$ span $\{W_S^{\otimes n}\}_{S \in \binom{\mathcal{U}}{\leq n}}$. Therefore, $\{W_S^{\otimes n}\}_{S \in \binom{\mathcal{U}}{\leq n}}$ lie in a vector space of dimension at most $\sum_{i=0}^{nt} \binom{s}{i}$. ∎

Now the theorem is an easy consequence of the two claims made above. ∎

The space-time tradeoff equation for the exact quantum case holds for the one-sided error case too, as shown below.

**Theorem 2** *The tradeoff result of Theorem 1 also holds for quantum schemes where the query scheme may err with probability less than* 1 *on the positive instances (i.e. if an element is present it may be erroneously reported absent), but not on the negative instances (i.e. if an element is absent it has to be reported absent).*

**Proof**: Essentially the same proof of Theorem 1 goes through. Since the query scheme can make an error only if the element is present, we observe that the only vector in the linear combination (equation 1) that has a non-zero projection on the space $A_1^{\otimes n}$ is the vector $W_T^{\otimes n}|\phi_T\rangle$. Hence $\alpha_T = 0$ and the operators $\{W_S\}_{S\subseteq \mathbf{U},|S|\leq n}$ continue to be linearly independent. Hence the same tradeoff equation holds in this case too. ■

We now prove the lower bound on the space used by a 2-sided $\epsilon$-error 1-probe quantum scheme.

**Theorem 3** *Let $n/m < \epsilon < 1/8$. Suppose there is a quantum scheme with two-sided error $\epsilon$ which answers membership queries using one quantum probe. It must use space*

$$s = \Omega\left(\frac{n\log(m/n)}{\epsilon^{1/6}\log(1/\epsilon)}\right).$$

**Proof**: Since we are looking at a one probe quantum scheme, $W_S = U_1 O_S U_0$. We start by picking a family $F$ of sets, $F = \{S_1, \ldots, S_k\}$, $S_i \subseteq \mathbf{U}$, $|S_i| = n$ and $|S_i \cap S_j| \leq n/2$ for all $i \neq j$. By picking the sets greedily [EFF85, NW94], one obtains a family $F$ with

$$|F| \geq \frac{\binom{m}{n}}{\binom{n}{\frac{n}{2}}\binom{m-\frac{n}{2}}{\frac{n}{2}}} \geq \left(\frac{m}{4n}\right)^{n/2}.$$

Let $t \triangleq \left\lceil \frac{4\log|F|}{n\log(1/(4\epsilon))}\right\rceil$. Since, $m/n \geq 1/\epsilon$, $\frac{4\log|F|}{n\log(1/(4\epsilon))} \geq 2$; therefore,

$$t \leq \frac{6\log|F|}{n\log(1/(4\epsilon))}. \tag{2}$$

**Claim 3** *$\{W_S^{\otimes nt}\}_{S\in F}$ are linearly independent.*

**Proof**: Suppose there is a non-trivial linear combination

$$\sum_{S\in F} \alpha_S W_S^{\otimes nt} = 0.$$

Fix a $T \in F$. Let $T = \{i_1, \ldots, i_n\}$. Define

$$|\phi_T\rangle \triangleq (|\phi_{i_1}\rangle \otimes |\phi_{i_2}\rangle \otimes \cdots \otimes |\phi_{i_n}\rangle)^{\otimes t}.$$

Applying $\phi_T$ to the above linear combination, we get

$$\sum_{S\in F} \alpha_S W_S^{\otimes nt}|\phi_T\rangle = 0.$$

$$\Rightarrow \sum_{S\in F} \alpha_S (W_S|\phi_{i_1}\rangle \otimes \cdots \otimes W_S|\phi_{i_n}\rangle)^{\otimes t} = 0.$$

Taking inner product of the above combination with the vector

$$W_T^{\otimes nt}|\phi_T\rangle = (W_T|\phi_{i_1}\rangle \otimes \cdots \otimes W_T|\phi_{i_n}\rangle)^{\otimes t},$$

we get

$$\sum_{S \in F} \alpha_S \langle (W_S|\phi_{i_1}\rangle \otimes \cdots \otimes W_S|\phi_{i_n}\rangle)^{\otimes t} | (W_T|\phi_{i_1}\rangle \otimes \cdots \otimes W_T|\phi_{i_n}\rangle)^{\otimes t}\rangle = 0$$

$$\Rightarrow \sum_{S \in F} \alpha_S (\langle \phi_{i_1}|W_S^\dagger W_T|\phi_{i_1}\rangle \cdots \langle \phi_{i_n}|W_S^\dagger W_T|\phi_{i_n}\rangle)^t = 0. \quad (3)$$

- For any $i_j \in T$, $W_T|\phi_{i_j}\rangle = v_0 + v_1$ where $1 \geq \|v_1\| \geq \sqrt{1-\epsilon}$ and $\|v_0\| \leq \sqrt{\epsilon}$, $v_0 \in A_0$ and $v_1 \in A_1$.
- For any $i_j \in S \cap T$, $|\langle \phi_{i_j}|W_S^\dagger W_T|\phi_{i_j}\rangle| \leq 1$.
- For any $i_j \in T - S$, $W_S|\phi_{i_j}\rangle = u_0 + u_1$ where $1 \geq \|u_0\| \geq \sqrt{1-\epsilon}$ and $\|u_1\| \leq \sqrt{\epsilon}$ and $u_0 \in A_0$ and $u_1 \in A_1$. Hence $|\langle \phi_{i_j}|W_S^\dagger W_T|\phi_{i_j}\rangle| \leq 2\sqrt{\epsilon} \triangleq \delta$.

We now note that for every $T \in F$, we have a linear combination as in equation 3 above. We can write the linear combination in the matrix form as $\alpha M = 0$ where $\alpha = (\alpha_S)_{S \in F}$ and $M$ is a $|F| \times |F|$ matrix whose rows and columns are indexed by members of $F$. For $S, T \in F$,

$$M(S,T) = (\langle \phi_{i_1}|W_S^\dagger W_T|\phi_{i_1}\rangle \cdots \langle \phi_{i_n}|W_S^\dagger W_T|\phi_{i_n}\rangle)^t.$$

where $|\phi_T\rangle$ is as defined above. The diagonal entries of $M$, $M(T,T)$, are 1. The non-diagonal entries satisfy $|M(S,T)| \leq (\delta)^{(n-|S\cap T|)t} \leq \delta^{nt/2}$.

For our choice of $t$, $|F|(\delta)^{tn/2} \leq 1$. Hence $(|F|-1)(\delta)^{tn/2} < 1$. This implies that $M$ is non-singular. So, $\alpha_S = 0$ for all $S \in F$. Hence $\{W_S^{\otimes nt}\}_{S \in F}$ are linearly independent. ■

**Claim 4** $\left\{W_S^{\otimes nt}\right\}_{S \in F}$ *lie in a vector space of dimension at most* $\sum_{j=0}^{nt} \binom{s}{j}$.

**Proof**: Similar to proof of Claim 2 in Theorem 1. ■

Using the two claims above,

$$|F| \leq \sum_{j=0}^{nt} \binom{s}{j} \leq \binom{s+nt}{nt}.$$

Using the bound on $t$ from (2), we get

$$\left(\frac{1}{4\epsilon}\right)^{1/6} \leq \frac{e(s+nt)}{nt}.$$

For the given bounds on the parameters and recollecting also that $\Omega(n \log(m/n))$ is always a lower bound for the storage space, we get

$$s = \Omega\left(\frac{n \log(m/n)}{\epsilon^{1/6} \log(1/\epsilon)}\right)$$

■

We now show how to extend the above argument for 2-sided $\epsilon$-error quantum schemes which make $p$ probes.

**Theorem 4** *For any $p \geq 1$ and $n/m < 4^{p-1}\epsilon^p < 1/8$, suppose there is a quantum scheme with two-sided error $\epsilon$ which answers membership queries using $p$ quantum probes. It must use space*

$$s = \Omega\left(\frac{n\log(m/n)}{\epsilon^{1/6p}\log(1/\epsilon)}\right).$$

**Proof**: The proof of this theorem is similar to the proof of Theorem 3. The difference is that one has to pick a family $F$ of sets, $F = \{S_1, \ldots, S_k\}$, $S_i \subseteq \mathbf{U}$, $|S_i| = n$ and $|S_i \cap S_j| \leq n/2$ for all $i \neq j$ such that $|F| \geq \left(\frac{m}{4n}\right)^{\frac{n}{2p}}$. One can prove that $\left\{W_S^{\otimes nt}\right\}_{S \in F}$, $t \triangleq \left\lceil \frac{4\log|F|}{n\log(1/(4\epsilon))} \right\rceil$, are linearly independent in exactly the same fashion as Claim 3 in Theorem 3 was proved. Now $\left\{W_S^{\otimes nt}\right\}_{S \in F}$ lie in a vector space of dimension at most $\sum_{j=0}^{pnt}\binom{s}{j}$ instead of $\sum_{j=0}^{nt}\binom{s}{j}$. This statement can be proved just as Claim 2 in Theorem 1 was proved. Therefore we get a lower bound

$$\Omega\left(\frac{n\log(m/n)}{\epsilon^{1/6p}\log(1/\epsilon)}\right)$$

■

## 3.2 Classical Schemes

We now give the proof for the space-time tradeoff equation in the classical deterministic case.

**Theorem 5** *Suppose there exists a classical deterministic scheme for storing subsets of size at most $n$ from a universe $\mathbf{U}$ of size $m$ which uses $s$ bits of storage and answers membership queries with $t$ classical bit probes. Then,*

$$\sum_{i=0}^{n}\binom{m}{i} \leq \sum_{i=0}^{nt}\binom{s}{i}$$

**Proof**: For $1 \leq i \leq m$, let $f_i : \{0,1\}^s \to \Re$ denote the function for query $i$, which maps bit strings of length $s$ to $\{0,1\} \subset \Re$ i.e. $f_i$ maps $x \in \{0,1\}^s$ to 1 iff the query scheme given query $i$ and bitstring $x$ evaluates to 1. Consider a mapping $\Phi : \binom{\mathbf{U}}{\leq n} \to (\{0,1\}^s \to \Re)$ i.e. $\Phi$ takes a subset of the universe of size at most $n$ to a function from bit strings of length $s$ to the reals. $\Phi$ is defined as follows

$$\Phi(\{\}) \triangleq \text{constant 1 function}$$

$$\Phi(S) \triangleq f_{i_1} \cdots f_{i_k}, \quad S = \{i_1, \cdots i_k\},\ S \neq \{\}$$

**Claim 5** $\{\Phi(S)\}_{S \in \binom{\mathbf{U}}{\leq n}}$ *are linearly independent over $\Re$.*

**Proof**: Suppose there exists a non-trivial linear combination

$$\sum_{S \in \binom{\mathbf{U}}{\le n}} \alpha_S \Phi(S) = 0$$

Pick a set $T$ of smallest cardinality such that $\alpha_T \neq 0$. Let $x(T) \in \{0,1\}^s$ be the string stored by the storage scheme. Applying $x(T)$ to the above linear combination we get,

$$\sum_{S \in \binom{\mathbf{U}}{\le n}} \alpha_S \Phi(S) x(T) = 0$$

If $S \neq T$ then there exists an element $i \in \mathbf{U}$ such that $i \in S - T$. Then $f_i(x(T)) = 0$ and hence $\Phi(S)(x(T)) = 0$. If $S = T$, then $\Phi(S)(x(T)) = \Phi(T)(x(T)) = 1$. Hence $\alpha_T = 0$ which is a contradiction. Hence the claim is proved. ■

**Claim 6** $\{\Phi(S)\}_{S \in \binom{\mathbf{U}}{\le n}}$ *lie in a vector space of dimension at most* $\sum_{i=0}^{nt} \binom{s}{i}$.

**Proof**: Since the query scheme is deterministic and makes at most $t$ (classical) bitprobes, it is modelled by a decision tree of depth at most $t$. Hence $f_i$ can be represented over $\Re$ as a sum of products of at most $t$ linear functions, where the linear functions are either a variable $y_j$ (representing the value stored at location $j$ in the bit string) or $1 - y_j$ (representing the negation of the value stored at location $j$). Note that for any $x \in \{0,1\}^s$ at most one of these products evaluates to 1. Such a function can be represented as a multilinear polynomial in $y_1, y_2, \ldots, y_s$ of degree at most $t$. A product of at most $n$ such functions can be represented as a multilinear polynomial of degree at most $nt$. Hence $\{\Phi(S)\}_{S \in \binom{\mathbf{U}}{\le n}}$ lie in the span of at most $\sum_{i=0}^{nt} \binom{s}{i}$ functions from $\{0,1\}^s$ to $\Re$. From this the claim follows. ■

From the above two claims, the theorem follows. ■

In fact the tradeoff result can be extended to the one-sided error classical case too.

**Theorem 6** *The tradeoff result of Theorem 5 also holds for classical schemes where the query scheme may err with probability less than* 1 *on the positive instances (i.e. if an element is present it might report it to be absent), but not on the negative instances (i.e, if an element is absent it has to be reported as absent). In fact the tradeoff result holds for nondeterministic query schemes too.*

**Proof**: A proof very similar to that of Theorem 5 goes through. We just observe that now the query scheme is a probability distribution over deterministic query schemes. Note that if the query element is present in the set stored, then there is a decision tree in this family that outputs 1. If the query element is not present in the set stored, then all the decision trees output 0. Let us denote the family of decision trees corresponding to query element $i$, $1 \le i \le m$ by $F_i$. For any decision tree $D$ in $F_i$, let $g_D : \{0,1\}^s \to \{0,1\}$ be the function it evaluates.

Let us now define $f_i \triangleq \sum_{D \in F_i} g_D$. Then

$$f_i(x[T]) \begin{cases} \geq 1 & \text{if } i \in T \\ = 0 & \text{otherwise} \end{cases}$$

With this choice of $f_i$, the rest of the proof is the same as in the deterministic case. ■

Now we give a simple proof of the lower bound for the space used by a classical randomised scheme which answers membership queries with two-sided error at most $\epsilon$ and uses only one bitprobe.

**Theorem 7** *Let $1/12 > \epsilon > n/m^{1/3}$. Any two-sided $\epsilon$-error classical randomised scheme which stores subsets of size $n$ from a universe $\mathbf{U}$ of size $m$ and answers membership queries using one bitprobe must use space*

$$\Omega\left(\frac{n \log m}{\epsilon^{2/5} \log(1/\epsilon)}\right)$$

**Proof**: Suppose there is a two-sided $\epsilon$-error classical randomised scheme which stores subsets of size $n$ from an universe of size $m$ and answers membership queries using one bitprobe and uses $s$ bits of storage. Define $k \triangleq \left\lceil \frac{4\log(27m)}{3\log(1/4e\epsilon)} \right\rceil$. Since $m^{1/3} > 1/\epsilon$, $\frac{4\log(27m)}{3\log(1/4e\epsilon)} \geq 4$; therefore, $k \leq \frac{5\log(27m)}{3\log(1/4e\epsilon)}$. We repeat the randomised scheme $k$ times and accept only if more than $3k/4$ trials accept. Then the probability of making an error on a positive instance (i.e. the query element is present in the set stored) is bounded by

$$\binom{k}{k/4} \epsilon^{k/4} \leq (4e\epsilon)^{k/4}$$

The probability of making an error on a negative instance (i.e. the query element is not present in the set stored) is bounded by

$$\binom{k}{3k/4} \epsilon^{3k/4} \leq \left(\frac{4e\epsilon}{3}\right)^{3k/4} \leq (4e\epsilon)^{3k/4}$$

From the given bounds on the parameters, we get

$$\begin{aligned} \Pr[\text{Error on a positive instance}] &\leq \frac{1}{(27m)^{1/3}} \leq \frac{1}{3n} \\ \Pr[\text{Error on a negative instance}] &\leq \frac{1}{27m} \end{aligned}$$

Hence the probability that a random sequence of coin tosses gives the wrong answer on some query $q \in \mathbf{U}$ and a particular set $S$ stored, is at most

$$\frac{1}{3n} \times n + \frac{1}{27m} \times (m - n) < \frac{1}{2}$$

Call a sequence of coin tosses bad for a set $S$ if when $S$ is stored, there is one query $q \in \mathbf{U}$ for which the query scheme with these coin tosses gives the wrong answer. Thus at most half of the coin toss sequences are bad for a fixed set $S$. By an averaging argument, there exists a sequence of coin tosses which is bad for at most half of the sets $S \in \binom{\mathbf{U}}{n}$. By setting the coin tosses to that sequence, we now get a deterministic scheme which answers membership queries correctly for at least half the sets $S \in \binom{\mathbf{U}}{n}$ and uses $k$ bitprobes. From the proof of Theorem 5 we have that

$$\frac{1}{2}\binom{m}{n} \leq \sum_{i=0}^{nk} \binom{s}{i}$$

$$\Rightarrow \frac{1}{2}\left(\frac{m}{n}\right)^n \leq \binom{s+nk}{nk} \leq \left(\frac{e(s+nk)}{nk}\right)^{nk}$$

From the given bounds on the parameters, we get

$$s = \Omega\left(\frac{n \log m}{\epsilon^{2/5} \log(1/\epsilon)}\right)$$

■

We can extend the classical randomised two-sided error space lower bound above to the case of multiple bitprobes.

**Theorem 8** *Let $p \geq 1$ and $1/12 > \epsilon > n/m^{1/3p}$. Any two-sided $\epsilon$-error classical randomised scheme which stores subsets of size $n$ from a universe $\mathbf{U}$ of size $m$ and answers membership queries using at most $p$ bitprobes must use space*

$$\Omega\left(\frac{n \log m}{\epsilon^{2/5p} \log(1/\epsilon)}\right)$$

**Proof**: The proof of this theorem is similar to the proof of Theorem 7 above. The difference is that here we only look at subsets of size $n$ from a sub-universe of size $m^{1/p}$. We repeat the randomised scheme $k \triangleq \left\lceil \frac{4\log(27m^{1/p})}{3\log(1/4e\epsilon)} \right\rceil$ times and accept only if more than $3k/4$ trials accept. The rest of the proof now follows in the same fashion as the proof of Theorem 7 above. ■

# 4 Conclusion and open problems

In this paper we introduce the quantum bitprobe model and study the complexity of the static membership problem in this model. We study the problem in the exact and bounded (one-sided and two-sided) error versions of the model and give lower bounds which almost match the corresponding classical upper bounds [BMRV00, Pag01]. We also give stronger/simplified proofs of lower bounds for the problem in the classical setting.

The paper of Buhrman et al. [BMRV00] also considers classical schemes for the static membership problem where the error is bounded and restricted only to negative instances (i.e. when the query element is not a member of the stored set). For such schemes they give almost matching upper and lower bounds. But for negative one-sided error quantum schemes, we can only prove similar lower bounds as for two-sided error quantum schemes. Thus there is a quantum/classical gap in this case, and resolving it is an open problem.

The classical bitprobe model has been used in the past to study other static data structure problems (e.g. perfect hashing, element containment [Pag01, Mil93]). The complexity of these problems in the quantum bit probe model is an important open problem.

# References


[Amb00] A. Ambainis. Quantum lower bounds by quantum arguments. In *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing*, pages 636–643, 2000. Also quant-ph/0002066.

[BBBV97] C. Bennett, E. Bernstein, G. Brassard, and U. Vazirani. Strengths and weaknesses of quantum computation. *SIAM Journal of Computing*, 26(3):1510–1523, 1997. Also quant-ph/9701001.

[BBC+98] R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. de Wolf. Quantum lower bounds by polynomials. In *Proceedings of the 39th Annual IEEE Symposium on Foundations of Computer Science*, pages 352–361, 1998. Also quant-ph/9802049.

[BMRV00] H. Buhrman, P. Miltersen, J. Radhakrishnan, and S. Venkatesh. Are bitvectors optimal? In *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing*, pages 449–458, 2000.

[Deu89] D. Deutsch. Quantum computational networks. In *Proceedings of the Royal Society of London A*, volume 425, pages 73–90, 1989.

[DR82] A.G. Dyachkov and V.V. Rykov. Bounds on the length of disjunctive codes. *Problemy Peredachi Informatsii*, 18(3):7–13, 1982. (In Russian).

[EFF85] P. Erdős, P. Frankl, and Z. Füredi. Families of finite sets in which no set is covered by the union of $r$ others. *Israel Journal of Mathematics*, 51:79–89, 1985.

[FKS84] M.L. Fredman, J. Komlós, and E. Szemerédi. Storing a sparse table with $O(1)$ worst case access time. *Journal of the Association for Computing Machinery*, 31(3):538–544, 1984.

[Gro96] L. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the 28th Annual ACM Symposium on Theory of Computing*, pages 212–219, 1996. Also quant-ph/9605043.

[Hol73] A. Holevo. Bounds for the quantity of information transmitted by quantum a communication channel. *Problemy Peredachi Informatsii*, 9(3):3–11, 1973. (In Russian) English translation in *Problems of Information Transmission*, vol. 9, 1973, pages 177–183.

[Mil93] P.B. Miltersen. The bitprobe complexity measure revisited. In *Proceedings of the 10th Annual Symposium on Theoretical Aspects of Computer Science*, pages 662–671, 1993.

[MP69] M. Minsky and S. Papert. *Perceptrons*. MIT Press, Cambridge, Mass., USA, 1969.

[Nay99] A. Nayak. Optimal lower bounds for quantum automata and random access codes. In *Proceedings of the 40th Annual IEEE Symposium on Foundations of Computer Science*, pages 369–376, 1999. Also quant-ph/9904093.

[NW94] N. Nisan and A. Wigderson. Hardness vs randomness. *Journal of Computer and System Sciences*, 49:149–167, 1994.

[Pag01] R. Pagh. On the cell probe complexity of membership and perfect hashing. In *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing*, 2001. (to appear).

[Yao81] A.C-C. Yao. Should tables be sorted? *Journal of the Association for Computing Machinery*, 28(3):615–628, 1981.

[YY97] A.C-C. Yao and F.F. Yao. Dictionary look-up with one error. *Journal of Algorithms*, 25(1):194–202, 1997.