Optimal lower bounds for quantum automata and random access codes [*]

Ashwin Nayak [†]

Computer Science Division
University of California
Berkeley, CA 94720
E-mail: ashwin@cs.berkeley.edu

Abstract

Consider the finite regular language $L_n = \{w0 \mid w \in \{0,1\}^*, |w| \leq n\}$. In [3] it was shown that while this language is accepted by a deterministic finite automaton of size $O(n)$, any one-way quantum finite automaton (QFA) for it has size $2^{\Omega(n/\log n)}$. This was based on the fact that the evolution of a QFA is required to be reversible. When arbitrary intermediate measurements are allowed, this intuition breaks down. Nonetheless, we show a $2^{\Omega(n)}$ lower bound for such QFA for L_n, thus also improving the previous bound. The improved bound is obtained from simple entropy arguments based on Holevo's theorem [8]. This method also allows us to obtain an asymptotically optimal $(1 - H(p))n$ bound for the dense quantum codes (random access codes) introduced in [3]. We then turn to Holevo's theorem, and show that in typical situations, it may be replaced by a tighter and more transparent in-probability bound.

1. Introduction

One-way quantum finite automata (QFA) were defined in [11, 9] and have drawn much interest since because they reflect the capabilities of currently feasible experimental quantum computers. Moreover, their study provides much insight into the nature of quantum computation. Results like those of [2] and [3] show that the laws underlying quantum computation are a mixed blessing. [2] shows how one may use superpositions to design QFA for certain languages that are exponentially more succinct than the corresponding classical FA. In contrast, other results from [3] show that the reversibility requirements of quantum mechanics imposes serious limits on the power of QFA—they show that QFA for certain other languages are exponentially larger than the corresponding DFA. In this paper we consider a different model of QFA (called *enhanced* QFA) where the state of the QFA can be measured while each symbol is processed. In the case of more general models such as quantum Turing machines such intermediate measurements do not increase the power of the model, since measurements can always be replaced by safe storage. However, in the case of QFA, the space limitations inherent in the definition preclude the possibility of similar reasoning. Moreover, in this new model, the evolution of the system is no longer reversible, so the intuition from [9, 3] no longer applies. Indeed, this new model of QFA was suggested by Dorit Aharonov as a more physically appropriate model that might not suffer from unnecessary handicaps resulting from the reversibility property embedded in the definitions from [11, 9].

In this paper, we show that enhanced QFA are also exponentially larger than the corresponding DFA for certain languages. The conceptual framework for our proof is completely different from that in [3]. We consider the evolution of a QFA on a random input string and show that the entropy of the mixed state that it exists in can only increase with each successive symbol read. This holds true even in the presence of intermediate measurements. Moreover, for certain languages, it is possible to bound from below the increase in entropy that results from processing each symbol, by appealing to Holevo's theorem [8]. Finally, we can bound the total information capacity of the QFA in terms of the number of states of the QFA, and therefore obtain a lower bound on the number of states required to correctly recognize strings of the language. The new bound we get is tight, and therefore answers an issue left open in [3].

The paper [3] also introduced the novel possibility of dense quantum codes that seem to violate Holevo's bound by exploiting the fact that in general measurements do not commute. This raised the possibility of (for instance) parsimoniously encoding an entire telephone directory such that any single number could be extracted from it via a suitable measurement. Examples of such *random access codes* were

[*] This work was initiated at the 1998 Elsag-Bailey – I.S.I. Foundation research meeting on quantum computation.

[†] Supported by JSEP grant FDF 49620-97-1-0220-03-98 and NSF grant CCR 9800024.

given in [3] that have no classical counterparts. However, it was also shown that no more than a logarithmic factor compression is achievable. We can use the same conceptual framework as described above to give a linear bound on the number of qubits required for such codes. This bound is optimal up to an additive logarithmic term, as follows from the classical upper bound given in [3]. Thus, quantum encoding offers no asymptotic advantage over classical encoding in this scenario. This resolves an open question from [3].

Finally we turn our attention to Holevo's bound [8] itself. Typically in quantum computation applications (though not in this paper), Holevo's bound is applied by converting, often implicitly, a statement about the probability of correct decoding into a statement in terms of entropy, when a random variable X is transmitted over a quantum channel using m quantum bits. We give a tight bound on this decoding probability by a direct argument which allows us to infer lower bounds for m without resorting to Holevo's theorem. Since the probability bound is tight, the inferred bounds are at least as good as those implied by Holevo's theorem. We also provide an example where it gives a strictly better bound than the latter. We should mention that the proof of Holevo's bound (which is essentially equivalent to the strong subadditivity property of von Neumann entropy) is rather involved, while the proof of the probability bounds is quite transparent.

2. Summary of results

A QFA (as defined in [9]) differs from a DFA in that its state is in general a superposition of the classical (basis) states. It starts in such a state, and when a new input symbol σ is seen, a corresponding unitary operator U_σ is applied to it. The state is then measured to check for acceptance, rejection or continuation. If the result of the measurement is 'continue,' the next symbol is read, otherwise the input is accepted or rejected. A QFA recognizes a language if all the strings in it (or not in it) are accepted (respectively, rejected) with constant probability bounded away from $1/2$. See Section 3.2 for a more precise definition of QFA.

We start by showing an exponential lower bound for QFA.

Theorem 2.1 *Let L_n be the language*

$$\{w0 \mid w \in \{0,1\}^*,\ |w| \leq n\}.$$

Then,

1. *L_n is recognized by a DFA of size $O(n)$,*
2. *L_n is recognized by some QFA, and*
3. *Any QFA recognizing L_n with some constant probability greater than $\frac{1}{2}$ has $2^{\Omega(n)}$ states.*

Note that a $2^{\Omega(n)}$ versus $O(n)$ separation is the best possible if only finite languages (or regular languages with sufficiently high probability of acceptance by a QFA) are considered: such languages are recognized by reversible (deterministic) FA that are at most exponentially larger than the corresponding DFA [2].

We then consider enhanced QFA, in which instead of only applying a unitary transformation when a new input symbol is seen, we allow a combination of unitary operators and orthogonal measurements. With the introduction of irreversibility via measurements, it may appear that such automata be at least as powerful as DFA. However, it is not hard to verify (by applying a technique of [14], also used in [9]) that enhanced QFA accept only regular languages. Moreover, we show that the bound of Theorem 2.1 continues to hold.

Theorem 2.2 *The statements of Theorem 2.1 hold also for enhanced QFA.*

It also follows from the proof of this theorem that enhanced QFA accept only a strict subset of the regular languages.

Random access encoding was introduced in [3] as a potentially powerful primitive in quantum information processing. An (n,m,p)-*random access encoding* is a function f that maps n-bit strings to mixed states over m qubits such that, for every $i \in \{1,\ldots,n\}$, there is a measurement $\mathcal{O}_i$ with outcome 0 or 1 that has the property that for all $x \in \{0,1\}^n$,

$$\Pr\left[\mathcal{O}_i(f(x)) = x_i\right] \quad \geq \quad p.$$

Serial encoding was defined similarly, except that the measurement $\mathcal{O}_i$ is allowed to depend on all the subsequent bits $x_{i+1}\cdots x_n$ of the encoded string. The technique used in proving Theorem 2.1 also yields a bound for such encoding. This bound matches the classical upper bound of $(1-H(p))n + O(\log n)$ shown in [3] up to the logarithmic additive term.

Theorem 2.3 *Any (n,m,p)-random access (or serial) encoding has $m \geq (1-H(p))n$.*

To finish, we present a simple alternative to Holevo's bound [8].

Theorem 2.4 *Let X be a random variable over bit strings which are encoded as mixed states over m qubits and let $P(X,d)$ denote the net probability of the d most likely strings in the sample space of X. If Y is any random variable obtained by making some measurement of the encoding of X, then*

1. *there is a decoding procedure $\mathcal{D}_0$ such that*

$$\Pr[\mathcal{D}_0(Y) = X] \quad \geq \quad 2^{-H(X|Y)},$$

where $H(X|Y)$ is the conditional Shannon entropy of X with respect to Y; and

2. *for any decoding function $\mathcal{D}$,*

$$\Pr[\mathcal{D}(Y) = X] \quad \leq \quad P(X, 2^m).$$

In particular, this implies that when X is distributed uniformly, the mutual information $I(X : Y)$ of X and Y is at most m. Typical applications of the Holevo's bound such as that in [10, 3] involve only this weaker form. Our bound thus obviates the need for a translation of in-probability statements into statements about mutual information in these cases, also giving better bounds than Holevo's theorem in the process.

3. Preliminaries

First, in Section 3.1, we review the basic elements of quantum information theory. Then, in Section 3.2, we define enhanced QFA formally using some of the concepts presented there.

3.1. Information theory basics

We use the following notation in this paper. Let X and Y be two random variables. $H(X)$ denotes the *Shannon entropy* of X; $H(X|Y)$, the *conditional* Shannon entropy of X with respect to the variable Y; and $I(X : Y)$, the *mutual information* of the two variables X and Y. We also use $H : [0, 1] \rightarrow [0, 1]$ to denote the binary entropy function. We refer the reader to [7] for the definition and properties of these standard concepts from classical information theory.

The quantum mechanical analogue of a random variable is a probability distribution over superpositions, also called a *mixed state*. Consider the mixed state $\{p_i, |\phi_i\rangle\}$, where the superposition $|\phi_i\rangle$ is drawn with probability p_i. The behaviour of this mixed state is completely characterized by its *density matrix* $\boldsymbol{\rho} = \sum_i p_i \, |\phi_i\rangle\langle\phi_i|$. We will therefore identify a mixed state with its density matrix.

The following properties of density matrices are immediate from the definition. For any density matrix $\boldsymbol{\rho}$,

1. $\boldsymbol{\rho}$ is Hermitian, i.e., $\boldsymbol{\rho} = \boldsymbol{\rho}^\dagger$.

2. $\boldsymbol{\rho}$ has unit trace, i.e., $\mathrm{Tr}(\boldsymbol{\rho}) = \sum_i \boldsymbol{\rho}(i, i) = 1$.

3. $\boldsymbol{\rho}$ is positive semi-definite, i.e., $\langle\psi| \, \boldsymbol{\rho} \, |\psi\rangle \;\geq\; 0$ for all $|\psi\rangle$.

Thus, every density matrix is *unitarily diagonalizable* and has non-negative real eigenvalues that sum up to 1. The *von Neumann entropy* $S(\boldsymbol{\rho})$ of a density matrix $\boldsymbol{\rho}$ is defined as $S(\boldsymbol{\rho}) = -\sum_i \lambda_i \log \lambda_i$, where $\{\lambda_i\}$ is the multiset of all the eigenvalues of $\boldsymbol{\rho}$. In other words, $S(\boldsymbol{\rho})$ is the Shannon entropy of the distribution induced by the eigenvalues of $\boldsymbol{\rho}$ on the corresponding eigenvectors. For a comprehensive introduction to this concept and its properties, see, for instance, [15, 12, 13].

The density matrix corresponding to a mixed state with superpositions drawn from a Hilbert space $\mathcal{H}$ is said to have *support* in $\mathcal{H}$. First, we note the following.

Fact 3.1 *If $\boldsymbol{\rho}$ is a density matrix with support in a Hilbert space of dimension d, then $S(\boldsymbol{\rho}) \leq \log d$.*

This is because the probability distribution induced by the eigenvalues of $\boldsymbol{\rho}$ has support of size at most d. The Shannon entropy of any such distribution is at most $\log d$.

When a unitary operator U is applied to a mixed state, the corresponding density matrix $\boldsymbol{\rho}$ is transformed to $U\boldsymbol{\rho}U^\dagger$. Since the eigenvalues of $U\boldsymbol{\rho}U^\dagger$ are the same as those of $\boldsymbol{\rho}$, we conclude that entropy is invariant under unitary operations.

Fact 3.2 *For any density matrix $\boldsymbol{\rho}$ and unitary operator U, we have $S(U\boldsymbol{\rho}U^\dagger) = S(\boldsymbol{\rho})$.*

On the other hand, when we make an orthogonal measurement on a mixed state, the the entropy of the system can only increase.[1] If a mixed state $\boldsymbol{\rho}$ is measured according to an orthogonal set of projections $\{P_j\}$, it is easily verified that the resulting density matrix is given by $\sum_j P_j \boldsymbol{\rho} P_j$.

Fact 3.3 *Let $\boldsymbol{\rho}$ be the density matrix of a mixed state in a Hilbert space $\mathcal{H}$ and let the set of orthogonal projections $\{P_j\}$ define a measurement in $\mathcal{H}$. Further, let $\rho' = \sum_j P_j \boldsymbol{\rho} P_j$ be the density matrix resulting from a measurement of the mixed state with respect to this observable. Then $S(\boldsymbol{\rho}') \geq S(\boldsymbol{\rho})$.*

It is not hard to see that this is in fact a consequence of the property of density matrices that the entropy of any random variable obtained by making a measurement on a mixed state is at least as much as the entropy of its density matrix. A proof of this property may be found in [12, Chapter 9, pp. 262–263].

3.2. Enhanced one-way quantum finite automata

An enhanced one-way quantum finite automaton (QFA) is a theoretical model for a quantum computer with finite workspace. Models for such space-restricted quantum computers were first considered by [11, 9]. However, these models did not include the full range of operations allowed

[1] This fact may appear to be counterintuitive at first, since the entropy of a system is usually understood to quantify our ignorance of the state of the system, and making a measurement reveals some information about its state. However, it should be noted that the increase in entropy is not claimed in the state of the system conditioned on the state of the observer, but in the state of the system with the state of the observer traced out.

by the laws of quantum mechanics. In particular, the model of [11] does not include measurements as intermediate steps in a computation, and the model of [9] allows only measurements that check for acceptance, rejection or continuation. The model we describe below rectifies this situation by allowing any *orthogonal* measurement as a valid intermediate computational step. Our model may be seen as a finite memory version of the mixed state quantum computers defined in [1]. Note that we do not allow the more general "positive operator valued measurements" because the implementation of such measurements involves the joint unitary evolution of the state of the automaton with a fresh set of ancilla qubits, which runs against the (fixed finite workspace) spirit of the model.

In abstract terms, we may define an enhanced QFA as follows. It has a finite set of basis states Q, which consists of three parts: accepting states, rejecting states and non-halting states. The sets of accepting, rejecting and non-halting basis states are denoted by $Q_{\rm acc}, Q_{\rm rej}$ and $Q_{\rm non}$, respectively. One of the states, q_0, is distinguished as the starting state.

Inputs to a QFA are words over a finite alphabet Σ. We shall also use the symbols '$\cent$' and '$' that do not belong to Σ to denote the left and the right end-marker, respectively. The set $\Gamma = \Sigma \cup \{\cent, \$\}$ denotes the working alphabet of the QFA. For each symbol $\sigma \in \Gamma$, an enhanced QFA has a corresponding "superoperator" $\mathcal{U}_\sigma$ which is given by a composition of a finite sequence of unitary transformations and orthogonal measurements on the space $\mathbb{C}^Q$. An enhanced QFA is thus defined by describing $Q, Q_{\rm acc}, Q_{\rm rej}, Q_{\rm non}, q_0, \Sigma$, and $\mathcal{U}_\sigma$ for all $\sigma \in \Gamma$.

At any time, the state of a QFA can be described by a density matrix with support in $\mathbb{C}^Q$. The computation starts in the state $|q_0\rangle\langle q_0|$. Then transformations corresponding to the left end marker '$\cent$,' the letters of the input word x and the right end marker '$' are applied in succession to the state of the automaton, unless a transformation results in acceptance or rejection of the input. A transformation corresponding to a symbol $\sigma \in \Gamma$ consists of two steps:

1. First, $\mathcal{U}_\sigma$ is applied to $\boldsymbol{\rho}$, the current state of the automaton, to obtain the new state $\boldsymbol{\rho}'$.

2. Then, $\boldsymbol{\rho}'$ is measured with respect to the observable $E_{\rm acc} \oplus E_{\rm rej} \oplus E_{\rm non}$, where $E_{\rm acc} = {\rm span}\{|q\rangle \mid q \in Q_{\rm acc}\}$, $E_{\rm rej} = {\rm span}\{|q\rangle \mid q \in Q_{\rm rej}\}$, $E_{\rm non} = {\rm span}\{|q\rangle \mid q \in Q_{\rm non}\}$. The probability of observing E_i is equal to ${\rm Tr}(P_i \boldsymbol{\rho}')$, where P_i is the orthogonal projection onto E_i. If we observe $E_{\rm acc}$ (or $E_{\rm rej}$), the input is accepted (or rejected). Otherwise, the computation continues (with the state $P_{\rm non}\boldsymbol{\rho}' P_{\rm non}$), and the next transformation, if any, is applied.

We regard these two steps together as reading the symbol σ.

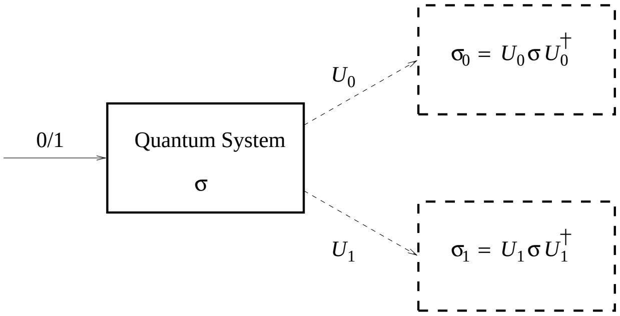
0/1
Quantum System
σ
U_0
U_1
$\sigma_0 = U_0 \sigma U_0^\dagger$
$\sigma_1 = U_1 \sigma U_1^\dagger$

Figure 1. *A stream of random bits determining the evolution of a quantum system.*

A QFA M is said to *accept* (or *recognize*) a language L with probability $p > \frac{1}{2}$ if it accepts every word in L with probability at least p, and rejects every word not in L with probability at least p.

The *size* of a finite automaton is defined as the number of (basis) states in it. The "space used by the automaton" refers to the number of (qu)bits required to represent an arbitrary automaton state.

The model of QFA as defined in [9] differs from this model in that the superoperators $\mathcal{U}_\sigma$ are all required to be given by unitary transformations U_σ.

4. The automata and coding lower bounds

In this section, we prove the first three theorems of Section 2. They are all based on a common framework which we present in Section 4.1.

4.1. The conceptual framework

Consider the evolution of the a quantum system under a random sequence of unitary transformations (V_i), where each V_i is either U_0 or U_1 (see Figure 1). Now suppose that the transformations U_0 and U_1 are distinguishable in the sense that for every superposition $|\phi\rangle$ of the system, $U_0\,|\phi\rangle$ can be distinguished from $U_1\,|\phi\rangle$ with success probability, say, $2/3$ by some fixed measurement. At each step, the system gains some information about the transformation applied to it, and we expect the entropy of the system to increase accordingly. In general, we could apply one of two arbitrary but distinguishable quantum operations on the system, and we would expect the same increase in entropy. This is the essential content of our key lemma:

Lemma 4.1 *Let $\boldsymbol{\sigma}_0$ and $\boldsymbol{\sigma}_1$ be two density matrices, and let $\boldsymbol{\sigma} = \frac{1}{2}(\boldsymbol{\sigma}_0 + \boldsymbol{\sigma}_1)$ be a random mixture of these matrices. If $\mathcal{O}$ is a measurement with outcome 0 or 1 such that* *making the measurement on $\boldsymbol{\sigma}_b$ yields the bit b with average probability p, then*

$$S(\boldsymbol{\sigma}) \quad \geq \quad \frac{1}{2}[S(\boldsymbol{\sigma}_0) + S(\boldsymbol{\sigma}_1)] + (1 - H(p)).$$

This lemma is a simple corollary of the classic Holevo theorem [8] from quantum information theory which bounds the amount of information we can extract from a quantum encoding of classical bits.

Theorem 4.2 (Holevo) *Let $x \mapsto \boldsymbol{\sigma}_x$ be any quantum encoding of bit strings, let X be a random variable with a distribution given by $\Pr[X = x] = p_x$, and let $\boldsymbol{\sigma} = \sum_x p_x \boldsymbol{\sigma}_x$ be the state corresponding to the encoding of the random variable X. If Y is any random variable obtained by performing a measurement on the encoding, then*

$$I(X\!:\!Y) \quad \leq \quad S(\boldsymbol{\sigma}) - \sum_x p_x S(\boldsymbol{\sigma}_x).$$

Proof of Lemma 4.1: Consider $\boldsymbol{\sigma}_b$ to be an encoding of the bit b. If X is an unbiased boolean random variable, then $\boldsymbol{\sigma}$ represents the encoding of X. Let Y be the outcome of the measurement of this encoding according to $\mathcal{O}$. By the hypothesis of the lemma, $\Pr[Y = X] = p$. It is easy to see from the concavity of the entropy function that

$$I(X\!:\!Y) \quad \geq \quad 1 - H(p)$$

(cf. Fano's inequality [7]). The lemma now follows from Theorem 4.2. ■

4.2. The case of quantum automata

We now prove Theorem 2.1 using this framework. The first two parts of the theorem are easy; we turn to part 3. We need the following definition from [3].

Definition 4.1 *An* r-restricted *one-way QFA for a language L is a one-way QFA that recognizes the language with probability $p > \frac{1}{2}$, and which halts with non-zero probability before seeing the right end-marker only after it has read r letters of the input.*

We first prove a bound of $2^{(1-H(p))n}$ for the number of basis states in any n-restricted QFA M for L_n. Note that the evolution of M on reading stream of random bits corresponds exactly to that of the quantum system described in Section 4.1 during the first n steps. So, at the end of reading a random n-bit string, the state of M has entropy at least $(1 - H(p))n$. However, this entropy is bounded by $\log|Q|$ by Fact 3.1 above, where Q is the set of basis states of M. This gives us the above bound. Since we will refer to this argument later, we formalize it below.

Let $\boldsymbol{\rho}_k$ be the state of the QFA M after the kth symbol of a random n-bit input has been read ($0 \leq k \leq n$).

Claim 4.3 $S(\boldsymbol{\rho}_k) \geq (1 - H(p))k$.

Proof: Let U_σ be the unitary operator of M corresponding to the symbol σ. Let E_0 be the span of the accepting basis states of M and let E_1 be the subspace orthogonal to it. Define the measurement $\mathcal{O}$ as applying the transformation $U_\$$ (recall that '\$' is the right end-marker) and then measuring with respect to the observable $E_0 \oplus E_1$. We can now prove the claim by induction.

For $k = 0$, the state of the automaton is pure, so $S(\boldsymbol{\rho}_0) = 0$. Now assume that $S(\boldsymbol{\rho}_{k-1}) \geq (1 - H(p))(k-1)$. After the kth random input symbol is read, the state of M becomes

$$\boldsymbol{\rho}_k \quad = \quad \frac{1}{2}(U_0 \boldsymbol{\rho}_{k-1} U_0^\dagger + U_1 \boldsymbol{\rho}_{k-1} U_1^\dagger).$$

By the definition of M, measuring $U_b \boldsymbol{\rho}_{k-1} U_b^\dagger$ according to $\mathcal{O}$ yields b with probability at least $p > \frac{1}{2}$. So by Lemma 4.1, we have

$$S(\boldsymbol{\rho}_k) \quad \geq \quad \frac{1}{2} \sum_{b=0,1} S(U_b \boldsymbol{\rho}_{k-1} U_b^\dagger) + (1 - H(p)). \quad (1)$$

But the entropy of a mixed state is preserved by unitary transformations (Fact 3.2), so

$$S(U_b \boldsymbol{\rho}_{k-1} U_b^\dagger) \quad = \quad S(\boldsymbol{\rho}_{k-1}) \geq (1 - H(p))(k-1).$$

Inequality (1) now gives us the claimed bound. ■

To pass from a bound on restricted QFA to one for general QFA for the language, we now invoke the following lemma from [3].

Lemma 4.4 *Let M be a one-way QFA with $|Q|$ states recognizing a language L with probability p. Then there is an r-restricted one-way QFA M' with $O(r\,|Q|)$ states that recognizes L with probability p.*

Thus, any general QFA for L_n using $|Q|$ basis states yields an n-restricted QFA that uses $O(n\,|Q|)$ states. By the lower bound derived above, we then have

$$|Q| \quad \geq \quad 2^{(1-H(p))n - \log n - O(1)},$$

the bound stated in Theorem 2.1.

4.3. Robustness of the automata lower bound

As mentioned in Section 1, QFA in which general intermediate measurements are allowed (which we call enhanced QFA), were suggested as a way of overcoming the restriction of reversible evolution that leads to the exponential lower bound shown in [3] (and in the previous section). Theorem 2.2 rules out this possibility. We prove this next.

Armed with the formalism of density matrices, it is not hard to verify (by using a technique of [14], which is also used in [9]) that enhanced QFA accept only regular languages. Moreover, the lower bound of Theorem 2.1 continues to hold for such QFA, as we show below. This essentially follows from the fact that the entropy of a quantum system *cannot decrease* under the action of a sequence of unitary operations and orthogonal measurements.

We now sketch how the proof of Theorem 2.2 may be completed. We proceed as in the previous section by first showing the bound for *restricted* enhanced QFA, which are defined analogously. Lemma 4.4, which extends easily to enhanced QFA, then gives us the claimed bound.

As before, we consider the state of a restricted automaton for L_n with acceptance probability p after a random n-bit input has been read. Its entropy is bounded by $\log|Q|$, where Q is the set of its basis states. Following Lemma 4.3, we argue that the entropy of the automaton state increases by at least $1 - H(p)$ every time a new random input symbol is read. Claim 4.3 extends easily to this case as well: initially, $S(\boldsymbol{\rho}_0) \geq 0$, and we need only prove that $S(\mathcal{U}_b \boldsymbol{\rho}_{k-1}) \geq S(\boldsymbol{\rho}_{k-1})$ for $b = 0, 1$, where $\mathcal{U}_b$ is the superoperator corresponding to the bit b, and $\boldsymbol{\rho}_i$ is the density matrix of the automaton state after i input symbols have been read. Since $\mathcal{U}_b$ is the composition of a finite sequence of unitary operators and orthogonal measurements, this is immediate from the monotonicity property of density matrices implied by Facts 3.2 and 3.3.

As a simple consequence, we obtain:

Theorem 4.5 *The regular language* $\{0,1\}^*0$ *cannot be accepted by any enhanced QFA with probability bounded away from* $\frac{1}{2}$.

To see this, we note that any enhanced QFA that supposedly recognizes this language also correctly recognizes all words of length at most n of the language L_n, for every n. The proof of Theorem 2.2 now tells us that the number of states in the QFA is $2^{\Omega(n)}$ for every n, which is a contradiction.

4.4. Random access codes

We now prove Theorem 2.3. Consider any random access encoding with parameters n, m, p. Let $\boldsymbol{\rho}_x$ denote the density matrix corresponding to the encoding of the n-bit string x. The density matrix of a random codeword is given by $\boldsymbol{\rho} = \frac{1}{2^n} \sum_x \boldsymbol{\rho}_x$. We can bound the entropy of $\boldsymbol{\rho}$ by m by Fact 3.1. Using Lemma 4.1, we can also prove a lower bound for the entropy of $\boldsymbol{\rho}$, and hence obtain a lower bound on m.

For any $y \in \{0,1\}^k$, where $0 \leq k \leq n$, let

$$\boldsymbol{\rho}_y \;=\; \frac{1}{2^{n-k}} \sum_{z \in \{0,1\}^{n-k}} \boldsymbol{\rho}_{zy}.$$

We claim that

Claim 4.6 $S(\boldsymbol{\rho}_y) \geq (1 - H(p))(n - k)$.

Proof: The proof is by downward induction on k. The base case $k = n$ is satisfied easily: $S(\boldsymbol{\rho}_y) \geq 0$ for all n-bit strings y.

Suppose the claim is true for $k + 1$. We have

$$\boldsymbol{\rho}_y \;=\; \frac{1}{2}(\boldsymbol{\rho}_{0y} + \boldsymbol{\rho}_{1y}).$$

By hypothesis,

$$S(\boldsymbol{\rho}_{by}) \;\geq\; (1 - H(p))(n - k - 1),$$

for $b = 0, 1$. Moreover, since the two density matrices are mixtures arising from strings that differ in the $(n-k)$th bit, the measurement $\mathcal{O}_{n-k}$ distinguishes them correctly with probability p. Thus, by Lemma 4.1, we get

$$S(\boldsymbol{\rho}_y) \;\geq\; \frac{1}{2}(S(\boldsymbol{\rho}_{0y}) + S(\boldsymbol{\rho}_{1y})) + (1 - H(p)),$$

which gives us the claimed bound. ■

Theorem 2.3 now follows by combining the claim (with y chosen to be the empty string) and the upper bound of m on the entropy. Notice that we could allow the measurement $\mathcal{O}_i$ to depend on the subsequent bits of the encoded string in the argument above. This means that the bound holds for serial codes as well.

We conclude this section by observing that the bound of Theorem 2.3 also gives a communication lower bound for the problem of information-theoretically secure private information retrieval with one database (see, e.g., [5]). The problem may be described as the following communication game. One party, Alice, has as input an n-bit string x (the database) and the second party, Bob, has an index $i \in \{1, \ldots, n\}$. Bob wishes to learn the value of the ith entry in the database x_i (with probability $p > \frac{1}{2}$) without revealing any information about i to Alice. The privacy condition translates to the fact that in any (quantum) protocol for this problem, Bob's computation and communication are independent of his input. We may also assume (by the principle of safe storage) that no intermediate measurements are made during the quantum protocol. A lemma due to [10] (based on a technique from [16]) tells us that whenever Bob's actions in a protocol are oblivious to his input, his state lies in a fixed subspace of dimension 2^m independent of Alice's input, if m qubits were exchanged during the protocol. Since his state at the end of an information retrieval protocol is independent of i, Bob may extract *any* bit x_j from the state by making a suitable measurement. Thus, an m-qubit protocol defines a random access code over m qubits, which implies that $m \geq (1 - H(p))n$.

5. An alternative to Holevo's theorem

In this section, we prove Theorem 2.4. We first prove the lower bound on the decoding probability.

Consider random variables X and Y as in the statement of Theorem 2.4. We describe a natural decoding procedure $\mathcal{D}_0$ and then show that it satisfies the requirement of the theorem. On input y, the decoding algorithm outputs x such that $p_{x|y} = \max_{x'} \ p_{x'|y}$, where $p_{x|y} = \Pr[X = x|Y = y]$. Let $p_y^{\max}$ denote this probability and let x_y denote the corresponding x.

Claim 5.1 *The procedure $\mathcal{D}_0$ described above decodes correctly with probability at least $2^{-H(X|Y)}$.*

Proof: The probability of correct decoding is equal to

$$\begin{aligned}
&\Pr[\mathcal{D}_0(Y) = X] \\
&\quad = \sum_y \Pr[X = x_y|Y = y] \cdot \Pr[Y = y] \\
&\quad = \mathrm{E}\left[p_Y^{\max}\right].
\end{aligned}$$

Now, $H(X|Y = y) = -\sum_x p_{x|y} \log p_{x|y} \geq -\log p_y^{\max}$. So $p_y^{\max} \geq 2^{-H(X|Y=y)}$. Taking expectation over Y, and noting that $2^{-(\cdot)}$ is a convex function, we have

$$\begin{aligned}
\mathrm{E}\left[p_Y^{\max}\right] &\geq \mathrm{E}\left[2^{-H(X|Y=y)}\right] \\
&\geq 2^{-\mathrm{E}[H(X|Y=y)]} \\
&= 2^{-H(X|Y)},
\end{aligned}$$

which gives us the claimed lower bound on the decoding probability. ■

We now turn to the upper bound on the probability of correct decoding. Consider any encoding of strings $x \mapsto \{q_{x,i}, |\phi_{x,i}\rangle\}$ into mixed states over m qubits, and any decoding procedure $\mathcal{D}$. The output of $\mathcal{D}$ may be viewed as the outcome of a measurement given by orthogonal projections $\{P_x\}$ in the Hilbert space of the encoding augmented with some ancilla. The probability may then be bounded as

$$\begin{aligned}
&\Pr[\mathcal{D}(Y) = X] \\
&\quad = \sum_x \Pr[\mathcal{D}(Y) = x] \cdot \Pr[X = x] \\
&\quad = \sum_x p_x \sum_i q_{x,i} \left\| P_x \left|\phi_{x,i}\right\rangle \right\|^2 \\
&\quad \leq \sum_x p_x \left\| P_x \left|\phi_x\right\rangle \right\|^2, \qquad (2)
\end{aligned}$$

where $p_x = \Pr[X = x]$, and $|\phi_x\rangle$ is the pure state $|\phi_{x,i}\rangle$ that maximizes the probability $\| P_x |\phi_{x,i}\rangle \|^2$ of obtaining the correct outcome x when its encoding is measured. (In all the expressions in this section, the ancilla qubits used in the measurement have been suppressed for ease of notation.) We can now bound the decoding probability by using the following claim.

Claim 5.2 $\sum_x \| P_x |\phi_x\rangle \|^2 \leq 2^m$.

Proof: Let E be the subspace spanned by the codewords $|\phi_x\rangle$, and let Q be the projection onto E. Since the codes are over m qubits, E has dimension at most 2^m. Let $\{|e_i\rangle\}$ be an orthonormal basis for E. Let $\{|\hat{e}_{x,j}\rangle\}$ be an orthonormal basis for the range of P_x. The union of all these bases $\{|\hat{e}_{x,j}\rangle\}$ is an orthonormal basis for the entire decoding Hilbert space. Now,

$$\begin{aligned}
\left\| P_x \left|\phi_x\right\rangle \right\|^2 &= \sum_j \left| \left\langle \hat{e}_{x,j} \middle| \phi_x \right\rangle \right|^2 \\
&\leq \sum_j \left\| Q \left|\hat{e}_{x,j}\right\rangle \right\|^2.
\end{aligned}$$

The last inequality follows because the length of the projection of any vector onto a space W is at least the length of its projection onto a subspace V of W. Observe that $\| Q |\hat{e}_{x,j}\rangle \|^2 = \sum_i |\langle e_i | \hat{e}_{x,j}\rangle|^2$. So,

$$\begin{aligned}
\sum_x \left\| P_x \left|\phi_x\right\rangle \right\|^2 &\leq \sum_i \sum_{x,j} \left| \left\langle e_i \middle| \hat{e}_{x,j} \right\rangle \right|^2 \\
&\leq \sum_i \left\| e_i \right\|^2 \\
&\leq 2^m,
\end{aligned}$$

since the orthonormal basis $\{|e_i\rangle\}$ for E has size at most 2^m, which is a bound on the dimension of E. ■

By (2), the probability of correct decoding is at most $\sum_x p_x \| P_x |\phi_x\rangle \|^2$. From the claim above, this expression is equal to $\sum_x p_x \lambda_x$, where $0 \leq \lambda_x \leq 1$ and $\sum_x \lambda_x \leq 2^m$. The maximum over all such $\{\lambda_x\}$ of this quantity may easily be seen to be bounded by the sum of the 2^m largest probability masses p_x, i.e., by $P(X, 2^m)$. Moreover, for any given X and m, there is a natural pair of encoding and decoding functions that achieves this bound. This shows that the bound is tight.

The above bound on decoding probability can give us sharper bounds on the number of qubits used in an encoding than an application of Holevo's theorem. We illustrate this with an example encoding of n-bits into $n+1$ orthogonal states $|i\rangle$. Half the strings are encoded as $|0\rangle$, a fourth as $|1\rangle$, an eighth as $|2\rangle$, and so on. A random codeword from this code can be decoded with probability exactly $(n+1)2^{-n}$, which yields the correct answer for the number of qubits used by invoking our bound. On the other hand, the mutual information $I(X:Y)$ between the encoded string and its decoding is

$$\frac{1}{2} + \frac{2}{2^2} + \frac{3}{2^3} + \cdots + \frac{n}{2^n} + \frac{n}{2^n},$$

which sums up to $2 - 2^{-(n-1)}$. This gives us a lower bound of at most 2 when combined with Holevo's theorem.

Note that Theorem 2.4 may be applied in a communication complexity context as well, when combined with the lemma due to [16, 10] mentioned in Section 4.4. This implies that if after the exchange of m quantum bits, n classical bits are transferred with success probability at least $\delta > 0$, then $m \geq n - \log \frac{1}{\delta}$. An application of Holevo's theorem along with Fano's inequality [7] would result in the bound $m \geq \delta n - H(\delta)$. This lower bound is a crucial ingredient in proving the quantum communication complexity of the inner product function [10]. Our result gives a bound similar to that shown in [4] for computing Inner Product, but does not seem to generalize to the case of entanglement assisted communication considered in [6].

Acknowledgements

I would like to thank Dorit Aharonov for suggesting the possibility of enhanced models of quantum finite automata, which motivated much of this work, Andris Ambainis for useful comments on the paper, Umesh Vazirani for many discussions that lead to crucial insights and for help with the presentation of the results, and the referees for their feedback on the paper.

References

[1] D. Aharonov, A. Kitaev and N. Nisan. Quantum circuits with mixed states. *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computation,* 1997, pp. 20–30.

[2] A. Ambainis and R. Freivalds. 1-way quantum finite automata: strengths, weaknesses and generalizations. *Proceedings of the 39th IEEE Symposium on Foundations of Computer Science,* 1998, pp. 332–341.

[3] A. Ambainis, A. Nayak, A. Ta-Shma and U. Vazirani. Dense quantum coding and a lower bound for 1-way quantum automata. *Proceedings of the Thirty-First Annual ACM Symposium on the Theory of Computing*, 1999.

[4] A. Ambainis, L.J. Schulman, A. Ta-Shma, U. Vazirani and A. Wigderson. The quantum communication complexity of sampling. *Proceedings of the 39th IEEE Symposium on Foundations of Computer Science,* 1998, pp. 342–351.

[5] B. Chor, O. Goldreich, E. Kushelivitz and M. Sudan. Private information retrieval. *Proceedings of the 36th IEEE Symposium on Foundations of Computer Science,* 1995, pp. 41–50. To appear in *Journal of the ACM.*

[6] R. Cleve, W. van Dam, M. Nielsen and A. Tapp. Quantum entanglement and the communication complexity of the inner product function. *Proceedings of the 1st International Conference on Quantum Computing and Quantum Communication, Lecture Notes in Computer Science* **1509**, 1998.

[7] T.M. Cover and J.A. Thomas. *Elements of information theory*. Wiley, New York, 1991.

[8] A.S. Holevo. Some estimates of the information transmitted by quantum communication channels. *Problemy Peredachi Informatsii* **9**, 1973, pp. 3–11. English translation in *Problems of Information Transmission* **9**, 1973, pp. 177–183.

[9] A. Kondacs and J. Watrous. On the power of quantum finite state automata. *Proceedings of the 38th IEEE Symposium on Foundations of Computer Science,* 1997, pp. 66–75.

[10] I. Kremer. *Quantum communication*. Master's thesis, The Hebrew University of Jerusalem, 1995.

[11] C. Moore and J. Crutchfield. Quantum automata and quantum grammars. Santa-Fe Institute Working Paper 97-07-062, 1997. Also available at the LANL Quantum Physics e-Print Archive at http://xxx.lanl.gov/archive/quant-ph/9707031.

[12] A. Peres. *Quantum theory: concepts and methods.* Kluwer Academic Publishers, Dordrecht, The Netherlands, 1995.

[13] J. Preskill. Lecture notes for Physics 229: Advanced mathematical methods of Physics, California Institute of Technology, 1998. Available at http://www.theory.caltech.edu/people/preskill/ph229.

[14] M.O. Rabin. Probabilistic automata. *Information and Control* **6**, 1963, pp. 230–245.

[15] A. Wehrl. General properties of entropy. *Reviews of Modern Physics* **50**(2), 1978, pp. 221–260.

[16] A.C.-C. Yao. Quantum circuit complexity. *Proceedings of the 34th IEEE Symposium on Foundations of Computer Science,* 1993, pp. 352–361.