Dense Quantum Coding and Quantum Finite Automata

ANDRIS AMBAINIS

Institute for Advanced Study, Princeton, New Jersey

ASHWIN NAYAK

California Institute of Technology, Pasadena, California

AMNON TA-SHMA

Tel-Aviv University, Tel-Aviv, Israel

AND

UMESH VAZIRANI

University of California, Berkeley, California

Abstract. We consider the possibility of encoding m classical bits into many fewer n quantum bits (qubits) so that an arbitrary bit from the original m bits can be recovered with good probability. We show that nontrivial quantum codes exist that have no classical counterparts. On the other hand, we show that quantum encoding cannot save more than a logarithmic additive factor over the best classical encoding. The proof is based on an entropy coalescence principle that is obtained by viewing Holevo's theorem from a new perspective.

In the existing implementations of quantum computing, qubits are a very expensive resource. Moreover, it is difficult to reinitialize existing bits during the computation. In particular, reinitialization is impossible in NMR quantum computing, which is perhaps the most advanced implementation of quantum computing at the moment. This motivates the study of quantum computation with restricted memory and no reinitialization, that is, of quantum finite automata. It was known that there are languages that are recognized by quantum finite automata with sizes exponentially smaller than those of corresponding classical automata. Here, we apply our technique to show the surprising result that there are languages for which quantum finite automata take exponentially more states than those of corresponding classical automata.

Preliminary versions of this work appeared as Ambainis et al. [1999] and Nayak [1999b]. A. Ambainis was supported by the Berkeley Fellowship for Graduate Studies and, in part, by NSF grant CCR-9800024; A. Nayak and U. Vazirani were supported by JSEP grant FDP 49620-97-1-0220-03-98 and NSF grant CCR-9800024.

Authors' addresses: A. Ambainis, Institute for Advanced Study, Einstein Dr., Princeton, NJ 08540, e-mail: ambainis@ias.edu; A. Nayak, Computer Science Department and Institute for Quantum Information, Mail Code 256-80, Pasadena, CA 91125, e-mail: nayak@cs.caltech.edu; A. Ta-Shma, Computer Science Department, Tel-Aviv University, Ramat Aviv, Tel Aviv 69978, Israel, e-mail: amnon@post.tau.ac.il; U. Vazirani, Computer Science Division, 671 Soda Hall, University of California, Berkeley, CA, e-mail: vazirani@cs.berkeley.edu.

Journal of the ACM, Vol. 49, No. 4, July 2002, pp. 496–511.

Categories and Subject Descriptors: F.2.0 [**Analysis of Algorithms and Problem Complexity**]: General; F.1.1 [**Computation by Abstract Devices**]: Models of Computation—*automata (e.g. finite, push-down, resource-bounded)*

General Terms: Theory

Additional Key Words and Phrases: Automaton size, communication complexity, encoding, finite automata, quantum communication, quantum computation

1. *Introduction*

The tremendous information processing capabilities of quantum mechanical systems may be attributed to the fact that the state of an n quantum bit (qubit) system is given by a unit vector in a 2^n dimensional complex vector space. This suggests the possibility that classical information might be encoded and transmitted with exponentially fewer qubits. Yet, according to a fundamental result in quantum information theory, Holevo's theorem [Holevo 1973], no more than n classical bits of information can faithfully be transmitted by transferring n quantum bits from one party to another. In view of this result, it is tempting to conclude that the exponentially many degrees of freedom latent in the description of a quantum system must necessarily stay hidden or inaccessible.

However, the situation is more subtle since the recipient of the n-qubit quantum state has a choice of measurement he or she can make to extract information about their state. In general, these measurements do not commute. Thus making a particular measurement will disturb the system, thereby destroying some or all the information that would have been revealed by another possible measurement. This opens up the possibility of quantum *random access* codes, which encode classical bits into many fewer qubits, such that the recipient can choose which bit of classical information he or she would like to extract out of the encoding. We might think of this as a disposable quantum phone book, where the contents of an entire telephone directory are compressed into a few quantum bits such that the recipient of these qubits can, via a suitably chosen measurement, look up any *single* telephone number of his or her choice. Such quantum codes, if possible, would serve as a powerful primitive in quantum communication.

To formalize this, say we wish to encode m bits $b_1, \ldots, b_m$ into n qubits ($m \gg n$). Then a quantum random access encoding with parameters m, n, p (or simply an $m \overset{p}{\mapsto} n$ encoding) consists of an encoding map from $\{0, 1\}^m$ to mixed states with support in $\mathbb{C}^{2^n}$, together with a sequence of m possible measurements for the recipient. The measurements are such that if the recipient chooses the ith measurement and applies it to the encoding of $b_1 \cdots b_m$, the result of the measurement is b_i with probability at least p.

The main point here is that since the m different possible measurements may be noncommuting, the recipient cannot make the m measurements in succession to recover all the encoded bits with a good chance of success. Thus the existence of $m \overset{p}{\mapsto} n$ quantum random access codes with $m \gg n$ and $p > \frac{1}{2}$ does not necessarily violate Holevo's bound. Furthermore, even though $\mathbb{C}^k$ can accommodate only k mutually orthogonal unit vectors, it can accommodate a^k almost mutually orthogonal unit vectors (i.e., vectors such that the inner product of any two has an absolute value less than, say, $\frac{1}{10}$) for some $a > 1$. Indeed, there is no a priori reason

to rule out the existence of codes that represent a^n classical bits in n quantum bits for some constant $a > 1$.

We start by showing that quantum encodings are more powerful than classical ones. We describe a $2 \overset{0.85}{\mapsto} 1$ quantum encoding, and prove that there is no $2 \overset{p}{\mapsto} 1$ classical encoding for any $p > \frac{1}{2}$. Our quantum encoding may be generalized to a $3 \overset{0.78}{\mapsto} 1$ encoding, as was shown by Chuang [1997], and to encodings of more bits into one quantum bit.

The main result in this paper is that (despite the potential of quantum encoding shown by the arguments and results presented above) quantum encoding does not provide much compression. We prove that any $m \overset{p}{\mapsto} n$ quantum encoding satisfies $n \geq (1 - H(p))\, m$, where $H(p) = -p \log p - (1 - p) \log(1 - p)$ is the *binary entropy function*. The main technique in the proof is the use of the *entropy coalescence lemma*, which quantifies the increase in entropy when we take a convex combination of mixed states. This lemma is obtained by viewing Holevo's theorem from a new perspective.

We turn to upper bounds on compression next, and show that the lower bound is asymptotically tight up to an additive logarithmic term, and can be achieved even with *classical* encoding. For any $p > 1/2$, we give a construction for $m \overset{p}{\mapsto} n$ classical codes with $n = (1 - H(p))\, m + O(\log m)$. Thus, even though quantum random access codes can be more succinct as compared to classical codes, they may be only a logarithmic number of qubits shorter.

In many of the existing quantum computing implementations, the complexity of implementing the system grows tremendously as the number of qubits increases. Moreover, even discarding one qubit and replacing it by a new qubit initialized to $|0\rangle$ (often called a *clean* qubit) while keeping the total number of qubits the same might be difficult or impossible (as in NMR quantum computing [Nielsen and Chuang 2000]). This has motivated a huge body of work on one-way quantum finite automata (QFAs), which are devices that model computers with a small finite memory. During the computation of a QFA, no clean qubits are allowed, and in addition no intermediate measurements are allowed, except to decide whether to accept or reject the input.

We define generalized one-way quantum finite automata (GQFAs) that capture the most general quantum computation that can be carried out with restricted memory and no extra clean qubits. In particular, the model allows arbitrary measurements upon the state space of the automaton as long as the measurements can be carried out without clean qubits. We believe our model accurately incorporates the capabilities of today's implementations of quantum computing.

In Kondacs and Watrous [1997] it was shown that not every language recognized by a classical deterministic finite automaton (DFA) is recognized by a QFA. On the other hand, there are languages that are recognized by QFAs with sizes exponentially smaller than those of corresponding classical automata [Ambainis and Freivalds 1998]. It remained open whether for any language that can be recognized by a one-way finite automaton both classically and quantum-mechanically, a classical automaton can be efficiently simulated by a QFA with no extra clean qubits. We answer this question in the negative.

We apply the entropy coalescence lemma in a computational setting to give a lower bound on the size of (GQFAs). We prove that there is a sequence of languages for which the minimal GQFA has exponentially more states than the minimal DFA.

It may be surprising that despite their quantum power (and irreversible computation, thanks to the intermediate measurements) GQFAs are exponentially less powerful for certain languages than classical DFAs. This lower bound highlights the need for clean qubits for efficient computation.

2. *Preliminaries*

2.1. QUANTUM SYSTEMS. Just as a bit (an element of $\{0, 1\}$) is a fundamental unit of classical information, a qubit is the fundamental unit of quantum information. A qubit is described by a unit vector in the two-dimensional Hilbert space $\mathbb{C}^2$. Let $|0\rangle$ and $|1\rangle$ be an orthonormal basis for this space.[1] In general, the state of the qubit is a linear superposition of the form $\alpha\,|0\rangle + \beta\,|1\rangle$. The state of n qubits is described by a unit vector in the n-fold tensor product $\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2$. An orthonormal basis for this space is now given by the 2^n vectors $|x\rangle$, where $x \in \{0, 1\}^n$. This is often referred to as the *computational basis*. In general, the state of n qubits is a linear superposition of the 2^n computational basis states. Thus the description of an n qubit system requires 2^n complex numbers. This is arguably the source of the astounding information processing capabilities of quantum computers.

The information in a set of qubits may be "read out" by *measuring* it in an orthonormal basis, such as the computational basis. When a state $\sum_x \alpha_x\,|x\rangle$ is measured in the computational basis, we get the outcome x with probability $|\alpha_x|^2$. More generally, a (*von Neumann*) measurement on a Hilbert space $\mathcal{H}$ is defined by a set of orthogonal projection operators $\{P_i\}$. When a state $|\phi\rangle$ is measured according to this set of projection operators, we get outcome i with probability $\|P_i\,|\phi\rangle\|^2$. Moreover, the state of the qubits "collapses" to (i.e., becomes) $P_i\,|\phi\rangle\,/\,\|P_i\,|\phi\rangle\|$, when the outcome i is observed. In order to retrieve information from an unknown quantum state $|\phi\rangle$, it is sometimes advantageous to augment the state with some ancillary qubits, so that the combined state is now $|\phi\rangle \otimes \overline{|0\rangle}$, before measuring them jointly according to a set of operators $\{P_i\}$ as above. This is the most general form of quantum measurement, and is called a *positive operator valued measurement* (POVM).

2.2. DENSITY MATRICES. In general, a quantum system may be in a *mixed state*—a probability distribution over superpositions. For example, such a mixed state may result from the measurement of a *pure* state $|\phi\rangle$.

Consider the mixed state $\{p_i, |\phi_i\rangle\}$, where the superposition $|\phi_i\rangle$ occurs with probability p_i. The behavior of this mixed state is completely characterized by its *density matrix* $\boldsymbol{\rho} = \sum_i p_i\,|\phi_i\rangle\langle\phi_i|$. (The "bra" notation $\langle\phi|$ here is used to denote the conjugate transpose of the superposition (column vector) $|\phi\rangle$. Thus $|\phi\rangle\langle\phi|$ denotes the *outer product* of the vector with itself.) For example, under a unitary transformation U, the mixed state $\{p_i, |\phi_i\rangle\}$ evolves as $\{p_i, U\,|\phi_i\rangle\}$, so that the resulting density matrix is $U\boldsymbol{\rho}U^\dagger$. When measured according to the projection operators $\{P_j\}$, the probability q_j of getting outcome j is $q_j = \sum_i p_i \|P_j\,|\phi_i\rangle\|^2 = \mathrm{Tr}(P_j\boldsymbol{\rho}P_j)$, and the residual density matrix is $P_j\boldsymbol{\rho}P_j/q_j$. Thus, two mixed states with the same density matrix have the same behavior under any physical operation. We will therefore identify a mixed state with its density matrix.

[1] This is Dirac's ket notation. $|\phi\rangle$ is another way of denoting a vector $\vec{\phi}$.

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

(1) $\boldsymbol{\rho}$ is Hermitian, that is, $\boldsymbol{\rho} = \boldsymbol{\rho}^\dagger$;
(2) $\boldsymbol{\rho}$ has unit trace, that is, $\mathrm{Tr}(\boldsymbol{\rho}) = \sum_i \boldsymbol{\rho}(i, i) = 1$;
(3) $\boldsymbol{\rho}$ is positive semidefinite, that is, $\langle\psi|\,\boldsymbol{\rho}\,|\psi\rangle \geq 0$ for all $|\psi\rangle$.

Thus, every density matrix is *unitarily diagonalizable* and has nonnegative real eigenvalues that sum up to 1.

Recall that the amount of randomness (or the uncertainty) in a classical probability distribution may be quantified by its *Shannon entropy*. Doing the same for a mixed state is tricky because all mixed states consistent with a given density matrix are physically indistinguishable, and therefore contain the same amount of "entropy." Before we do this, we recall the classical definitions.

2.3. CLASSICAL ENTROPY AND MUTUAL INFORMATION. The *Shannon entropy* $S(X)$ of a classical random variable X that takes values x in some finite set with probability p_x is defined as

$$S(X) \;=\; -\sum_x p_x \log p_x.$$

The *mutual information* $I(X : Y)$ of a pair of random variables X, Y is defined by

$$I(X : Y) = S(X) + S(Y) - S(XY),$$

where XY denotes the joint random variable with marginals X and Y. It quantifies the amount of correlation between the random variables X and Y.

Fano's inequality asserts that if Y can predict X well, then X and Y have large mutual information. We use a simple form of Fano's inequality, referring only to Boolean variables X and Y.

FACT 2.1 (FANO'S INEQUALITY). *Let X be a uniformly distributed boolean random variable, and let Y be a boolean random variable such that* $\Pr(X = Y) = p$. *Then* $I(X : Y) \geq 1 - H(p)$.

For other properties of these concepts we refer the reader to a standard text (such as Cover and Thomas [1991]) on information theory.

2.4. VON NEUMANN ENTROPY. Consider the mixed state $X = \{p_i, |\phi_i\rangle\}$, where the superposition $|\phi_i\rangle$ occurs with probability p_i. Since the constituent states $|\phi_i\rangle$ of the mixture are not perfectly distinguishable in general, we cannot define the entropy of this mixture to be the Shannon entropy of $\{p_i\}$. Another way to see this is that this mixture is equivalent to any other mixture with the same density matrix, and so should have the same entropy as that mixture. Indeed, a special such equivalent mixture can be obtained by diagonalizing the density matrix—the constituent states of this mixture are orthogonal, and therefore perfectly distinguishable. Now, the entropy of the density matrix can be defined to be the Shannon entropy of these probabilities.

To formalize this, recall that every density matrix $\boldsymbol{\rho}$ is unitarily diagonalizable:

$$\boldsymbol{\rho} = \sum_j \lambda_j\, |\psi_j\rangle\langle\psi_j|,$$

and has nonnegative real eigenvalues $\lambda_j \geq 0$ that sum up to 1, and the corresponding eigenvectors $|\psi_j\rangle$ are all orthonormal. The *von Neumann entropy* $S(\boldsymbol{\rho})$ of the density matrix $\boldsymbol{\rho}$ is then defined as $S(\boldsymbol{\rho}) = -\sum_i \lambda_i \log \lambda_i$. In other words, $S(\boldsymbol{\rho})$ is the Shannon entropy of the distribution induced by the eigenvalues of $\boldsymbol{\rho}$ on the corresponding eigenvectors.

We summarize some basic properties of von Neumann entropy below. For a comprehensive introduction to this concept and its properties, see, for instance, Preskill [1998] and Wehrl [1978].

If the constituent states of a mixture lie in a Hilbert space $\mathcal{H}$, then the corresponding density matrix is said to have *support* in $\mathcal{H}$. A density matrix with support in a Hilbert space of dimension d, has d eigenvalues, and hence the entropy of any such distribution is at most $\log d$. I.e.,

FACT 2.2. *If $\boldsymbol{\rho}$ is a density matrix with support in a Hilbert space of dimension d, then $0 \leq S(\boldsymbol{\rho}) \leq \log d$.*

Quantum mechanics requires that the evolution of the state of an isolated system be unitary, and therefore reversible. This implies that information cannot be erased and entropy is invariant under unitary operations:

FACT 2.3. *For any density matrix $\boldsymbol{\rho}$ and unitary operator U, $S(U\boldsymbol{\rho}U^\dagger) = S(\boldsymbol{\rho})$.*

This is easy to see since the eigenvalues of the resulting matrix $U\boldsymbol{\rho}U^\dagger$ are the same as those of $\boldsymbol{\rho}$.

In the classical world observing a value does not disturb its state, and as a result measurements (or observations) do not change entropy. In the quantum world, however, measurements usually disturb the system, introducing new uncertainties. Thus, the entropy increases. Consider for example a system of one qubit that is with probability 1 in the pure state $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, and thus has 0 entropy. Suppose we measure it in the $|0\rangle$, $|1\rangle$ basis. We get each result with equal probability, and the resulting mixed state of the qubit, disregarding the outcome of the measurement, is $\{(\frac{1}{2}, |0\rangle), (\frac{1}{2}, |1\rangle)\}$ which has entropy 1.

FACT 2.4. *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 von Neumann measurement in $\mathcal{H}$. If $\boldsymbol{\rho}' = \sum_j P_j \boldsymbol{\rho} P_j$ is the density matrix resulting from a measurement of the mixed state with respect to these projections (disregarding the measurement outcome), then $S(\boldsymbol{\rho}') \geq S(\boldsymbol{\rho})$.*

A proof of this fact may be found in Peres [1995], Chapter 9, pp. 262–263.

3. *Holevo's Theorem and the Entropy Coalescence Lemma*

Consider two parties Alice and Bob communicating over a quantum channel, where Alice wishes to transmit some classical information, given by a random variable X, to Bob by encoding it into some number of qubits and sending these qubits to Bob. Holevo's theorem [Holevo 1973] bounds the amount of information Bob can extract from the quantum encoding.

THEOREM 3.1 (HOLEVO). *Let $x \mapsto \boldsymbol{\rho}_x$ be any quantum encoding of bit strings into density matrices. Let X be a random variable with a distribution given*

by $\Pr(X = x) = p_x$, *and let* $\boldsymbol{\rho} = \sum_x p_x \boldsymbol{\rho}_x$ *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*

$$I(X : Y) \;\leq\; S(\boldsymbol{\rho}) - \sum_x p_x S(\boldsymbol{\rho}_x).$$

Viewing Holevo's theorem from a different perspective is the key to the lower-bound results in this paper. We consider the scenario where a mixture $\boldsymbol{\rho}$ is obtained as the convex combination of two mixtures $\boldsymbol{\rho}_0$, $\boldsymbol{\rho}_1$ of equal entropy. When can we say that the combination results in a mixture of higher entropy? This is the content of the *entropy coalescence lemma* below. This lemma can quite easily be generalized to the case where $\boldsymbol{\rho}$ is obtained from a more general mixture of density matrices.

LEMMA 3.2. *Let* $\boldsymbol{\rho}_0$ *and* $\boldsymbol{\rho}_1$ *be two density matrices, and let* $\boldsymbol{\rho} = \frac{1}{2}(\boldsymbol{\rho}_0 + \boldsymbol{\rho}_1)$ *be a uniformly random mixture of these matrices. If* $\mathcal{O}$ *is a measurement with outcome* 0 *or* 1 *such that making the measurement on* $\boldsymbol{\rho}_b$ *yields the bit* b *with probability at least* p, *then*

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

PROOF. We view $\boldsymbol{\rho}_b$ as an encoding of the bit b. If X is an unbiased random variable over $\{0, 1\}$, then $\boldsymbol{\rho}$ 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, $\Pr(Y = X) \geq p$. Thus, by Fano's inequality—Fact 2.1:

$$I(X : Y) \;\geq\; 1 - H(p).$$

Also, by Holevo's Theorem 3.1:

$$I(X : Y) \;\leq\; S(\boldsymbol{\rho}) - \frac{1}{2}\,[S(\boldsymbol{\rho}_0) + S(\boldsymbol{\rho}_1)].$$

Rearranging, $S(\boldsymbol{\rho}) \geq \frac{1}{2}[S(\boldsymbol{\rho}_0) + S(\boldsymbol{\rho}_1)] + (1 - H(p))$ as desired. □

4. *Random Access Encodings*

We first define random access encodings.

Definition 4.1. A $m \overset{p}{\mapsto} n$ quantum random access encoding is a function $f : \{0, 1\}^m \times R \mapsto \mathbb{C}^{2^n}$ (here R is the set of random choices in the encoding) such that for every $1 \leq i \leq m$, there is a measurement $\mathcal{O}_i$ that returns 0 or 1 and has the property that

$$\forall b \in \{0, 1\}^m : \; \Pr_r(\mathcal{O}_i \mid f(b, r)\rangle = b_i) \geq p.$$

We call f the encoding function, and $\mathcal{O}_i$ the decoding function. We say the encoding is classical if f is a mapping into $\{0, 1\}^n$.

4.1. A QUANTUM ENCODING WITH NO CLASSICAL COUNTERPART. We begin by constructing a random access encoding of two classical bits into one qubit. This encoding was first used by Bennett et al. [1982] in the context of quantum cryptography and was independently rediscovered by the authors of this paper in the context of coding.

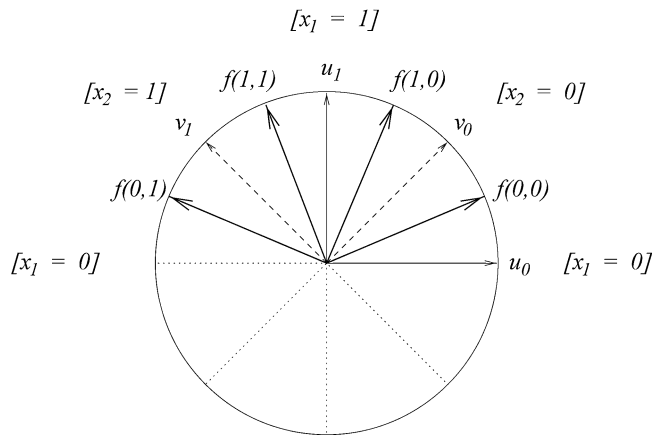

FIG. 1. A two-into-one quantum encoding with probability of success ≈ 0.85.

LEMMA 4.1. *There is a* $2 \overset{0.85}{\mapsto} 1$ *quantum encoding.*

PROOF. Let $|u_0\rangle = |0\rangle$, $|u_1\rangle = |1\rangle$, and $|v_0\rangle = \frac{1}{\sqrt{2}}(|1\rangle + |0\rangle)$, $|v_1\rangle = \frac{1}{\sqrt{2}}(|1\rangle - |0\rangle)$. Define $f(x_1, x_2)$, the encoding of the string x_1x_2, to be $|u_{x_1}\rangle + |v_{x_2}\rangle$ normalized, unless $x_1x_2 = 01$, in which case it is $-|u_0\rangle + |v_1\rangle$ normalized. The four vectors $f(0, 0), \ldots, f(1, 1)$ appear in Figure 1.

The decoding functions are defined as follows: for the first bit x_1, we measure the message qubit according to the u basis and associate $|u_0\rangle$ with $x_1 = 0$ and $|u_1\rangle$ with $x_1 = 1$. Similarly, for the second bit, we measure according to the v basis, and associate $|v_0\rangle$ with $x_2 = 0$ and $|v_1\rangle$ with $x_2 = 1$. See Figure 1.

For all four code words, and for any $i = 1, 2$, the angle between the code word and the correct subspace is $\pi/8$. Hence the success probability is $\cos^2(\pi/8) \approx 0.85$. $\square$

This example was further refined into a $3 \overset{0.78}{\mapsto} 1$ quantum encoding by Chuang [1997].

The next lemma shows that such classical codes are not possible.

LEMMA 4.2. *No* $2 \overset{p}{\mapsto} 1$ *classical encoding exists for any* $p > \frac{1}{2}$.

PROOF. Let there be a classical $2 \overset{p}{\mapsto} 1$ encoding for some p. Let $f : \{0, 1\}^2 \times R \mapsto \{0, 1\}$ be the corresponding probabilistic encoding function and $V_i : \{0, 1\} \times R' \mapsto \{0, 1\}$ the probabilistic decoding functions.

We first give a geometric characterization of the decoding functions. Each V_i depends only on the encoding, which is either 0 or 1. Define the point P^j (for $j = 0, 1$) in the unit square $[0, 1]^2$ as $P^j = (a_1^j, a_2^j)$, where $a_i^j = \Pr_{r'}(V_i(j, r') = 1)$. The point P^0 characterizes the decoding functions when the encoding is 0, and P^1 characterizes the decoding functions when the encoding is 1. For example, $P^1 = (1, 1)$ means that given the encoding 1, the decoding functions return $y_1 = 1$ and $y_2 = 1$ with certainty, and $P^0 = (0, 1/4)$ means that given the encoding 0, the decoding functions return $y_1 = 1$ with probability zero and $y_2 = 1$ with probability $1/4$.

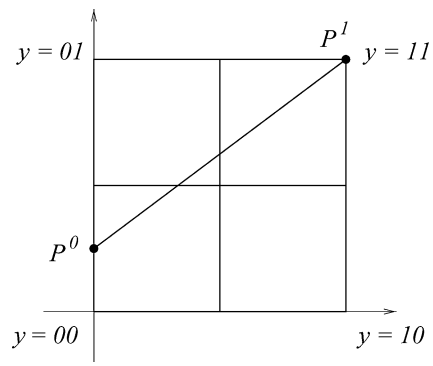

FIG. 2. A geometric characterization of the probabilistic decoding functions of Lemma 4.2.

Now fix the decoding functions V_1, V_2. They define two points P^0 and P^1 in $[0, 1]^2$ and the line connecting them:

$$P(q) = (1 - q)P^0 + qP^1.$$

We divide $[0, 1]^2$ to four quadrants, and we associate each quadrant with its corner $(x_1, x_2) \in \{0, 1\}^2$ (see Figure 2). The connecting line $P(q)$ cannot strictly pass through all four quadrants. To see that, let us assume without loss of generality that the point $P(\frac{1}{2})$ is at or above the center $(\frac{1}{2}, \frac{1}{2})$. If the line is monotone increasing, then the line must miss the bottom right quadrant, while if it is monotone decreasing, it must miss the left bottom quadrant. If the line misses the quadrant associated with (x_1, x_2), we say the decoding functions miss (x_1, x_2).

We now look at the encoding. We know the decoding functions miss some (x_1, x_2) and without loss of generality let us say that they miss $(1, 0)$. Given the input $x = (1, 0)$, the encoder can choose (based on r) whether to encode x as 0 or 1. Let us say that he or she encodes x as 1 with probability q_x. Let us denote by $P^x = (a_1(x), a_2(x))$ the point with $a_i(x) = \Pr_{r,r'}(V_i(f(x, r), r'))$. Then,

$$P^x = (1 - q_x)P^0 + q_x P^1 = P(q_x).$$

In particular it lies on the line connecting P^0 and P^1 and therefore it is not in the interior of the bottom right quadrant. Thus, either $a_1(x)$ is at most $\frac{1}{2}$ or $a_2(x)$ is at least $\frac{1}{2}$. It follows that either the first bit ($x_1 = 1$) or the second bit ($x_2 = 0$) is decoded correctly with probability $p \le \frac{1}{2}$. $\square$

5. *The Asymptotic of Random Access Codes*

5.1. THE LOWER BOUND. We now prove a lower bound on the number of qubits required for quantum random access codes.

THEOREM 5.1. *Let $\frac{1}{2} < p \le 1$. Any quantum (and hence any classical) $m \overset{p}{\mapsto} n$ encoding satisfies $n \ge (1 - H(p))\, m$.*

PROOF. Let ρ_x denote the density matrix corresponding to the encoding of the m-bit string x, and let ρ be the density matrix corresponding to picking x

uniformly from $\{0,1\}^m$ and encoding it. Then,

$$\rho = \frac{1}{2^m}\sum_x \rho_x.$$

Furthermore, for any $y \in \{0,1\}^k$, where $0 \le k \le m$, let

$$\rho_y = \frac{1}{2^{m-k}}\sum_{z\in\{0,1\}^{m-k}} \rho_{zy}$$

be the mixture corresponding to picking x uniformly from all strings in $\{0,1\}^m$ with postfix y and encoding it. We prove by downward induction on k that $S(\rho_y) \ge (1-H(p))(m-k)$.

Base case. Assume $k=m$ and $y \in \{0,1\}^m$. We need to prove that $S(\rho_y) \ge 0$. Indeed, the von Neumann entropy of any mixed state is nonnegative.

Induction step. Suppose the claim is true for $k+1$. We have $\rho_y = \frac{1}{2}(\rho_{0y}+\rho_{1y})$. By hypothesis,

$$S(\rho_{by}) \ge (1-H(p))(m-k-1),$$

for $b=0,1$. Moreover, ρ_{by} is a mixture arising for encoding strings with b in the $(m-k)$th bit. In particular, the measurement $\mathcal{O}_{m-k}$ when applied to the density matrix ρ_{by} returns b with probability at least p. Thus, by the entropy coalescence lemma (Lemma 3.2), we get

$$S(\rho_y) \;\ge\; \frac{1}{2}(S(\rho_{0y})+S(\rho_{1y})) + (1-H(p)) \;\ge\; (1-H(p))(m-k).$$

In particular, for $k=0$ we get that $S(\rho) \ge (1-H(p))\,m$. On the other hand, ρ is defined over a Hilbert space of dimension 2^n (as the encoding uses only n qubits) and Fact 2.2 implies that $S(\rho) \le n$. Together we see that $n \ge (1-H(p))\,m$ as desired. $\square$

5.2. A MATCHING UPPER BOUND. We now present a (nonconstructive) classical encoding scheme that asymptotically matches the lower bound derived in the previous section.

THEOREM 5.2. *For any $p > \frac{1}{2}$ there is a classical $m \overset{p}{\mapsto} n$ encoding with $n = (1-H(p))\,m + O(\log m)$.*

PROOF. If $p > 1-\frac{1}{m}$, $H(p) \le \frac{\log m+2}{m}$ and there is a trivial encoding—the identity map. So we turn to the case where $p \le 1-\frac{1}{m}$.

We use a code $S \subseteq \{0,1\}^m$ such that, for every $x \in \{0,1\}^m$, there is a $y \in S$ within Hamming distance $(1-p-\frac{1}{m})\,m$. It is known (see, e.g., Cohen et al. [1997], Theorem 12.1.2) that there is such a code S, called a *covering code*, of size

$$|S| \le 2^{(1-H(p+\frac{1}{m}))\,m+2\log m} \le 2^{(1-H(p))\,m+4\log m}.$$

For explicit constructions of covering codes, we refer the reader to Cohen et al. [1997]. (The explicit constructions, however, do not achieve the bound we seek.)

Let $S(x)$ denote the code word in the covering code S as above closest to x. One possibility is to encode a string x by $S(x)$. This would give us an encoding of the right size. Further, for every x, at least $(p+\frac{1}{m})\,m$ out of the m bits would be correct.

This means that the probability (over all bits i) that $x_i = S(x)_i$ is at least $p + \frac{1}{m}$. However, for our encoding we need this probability to be at least p for *every* bit, not just on average over all bits. So we introduce the following modification:

Let r be an m-bit string, and π be a permutation of $\{1, \ldots, m\}$. For a string $x \in \{0, 1\}^m$, let $\pi(x)$ denote the string $x_{\pi(1)}x_{\pi(2)} \cdots x_{\pi(m)}$.

Consider the encoding $S_{\pi,r}$ defined by $S_{\pi,r}(x) = \pi^{-1}(S(\pi(x + r))) + r$. We show that if π and r are chosen uniformly at random, then for any x and any index i, the probability that the ith bit in the encoding is different from x_i is at most $1 - p - \frac{1}{m}$. First, note that if i is also chosen uniformly at random, then this probability is bounded by $1 - p - \frac{1}{m}$. So all we need to do is to show that this probability is independent of i.

If π and r are uniformly random, then $\pi(x + r)$ is uniformly random as well. Furthermore, for a fixed $y = \pi(x + r)$, there is exactly one r corresponding to any permutation π that gives $y = \pi(x + r)$. Hence, if we condition on $y = \pi(x + r)$, all π (and, hence, all $\pi^{-1}(i)$) are equally likely. This means that the probability that $x_i \neq S_{\pi,r}(x)_i$ (or, equivalently, that $\pi(x + r)_{\pi^{-1}(i)} \neq (S(\pi(x + r))_{\pi^{-1}(i)})$ for random π and r is just the probability of $y_j \neq S(y)_j$ for random y and j. This is independent of i (and x).

Finally, we show that there is a small set of permutation-string pairs such that the desired property continues to hold if we choose π, r uniformly at random from *this* set, rather than the entire space of permutations and strings. We employ the probabilistic method to prove the existence of such a small set of permutation-string pairs.

Let $\ell = m^3$, and let the strings $r_1, \ldots, r_\ell \in \{0, 1\}^m$ and permutations $\pi_1, \ldots, \pi_\ell$ be chosen independently and uniformly at random. Fix $x \in \{0, 1\}^m$ and $i \in \{1, \ldots, m\}$. Let X_j be 1 if $x_i \neq S_{\pi_j,r_j}(x)_i$ and 0 otherwise. Then $\sum_{j=1}^{\ell} X_j$ is a sum of ℓ independent Bernoulli random variables, the mean of which is at most $(1 - p - \frac{1}{m})\,\ell$. Note that $\frac{1}{\ell}\sum_{j=1}^{\ell} X_j$ is the probability of encoding the ith bit of x erroneously when the permutation-string pair is chosen uniformly at random from the set $\{(\pi_1, r_1), \ldots (\pi_\ell, r_\ell)\}$. By the Chernoff bound, the probability that the sum $\sum_{j=1}^{\ell} X_j$ is at least $(1-p-\frac{1}{m})\,\ell + m^2$ (i.e., that the error probability $\frac{1}{\ell}\sum_{j=1}^{\ell} X_j$ mentioned above is at least $1 - p$) is bounded by $\mathrm{e}^{-2m^4/\ell} = \mathrm{e}^{-2m}$. Now, the union bound implies that the probability that the ith bit of x is encoded erroneously with probability more than $1 - p$ for *any* x or i is at most $m2^m\mathrm{e}^{-2m} < 1$. Thus, there is a combination of strings $r_1, \ldots, r_\ell$ and permutations $\pi_1, \ldots, \pi_\ell$ with the property we seek. We fix such a set of ℓ strings and permutations.

We can now define our random access code as follows: To encode x, we select $j \in \{1, \ldots, \ell\}$ uniformly at random and compute $y = S_{\pi_j,r_j}(x)$. This is the encoding of x. To decode the ith bit, we just take y_i. For this scheme, we need $\log(\ell|S|) \leq \log \ell + \log |S| = (1 - H(p))\,m + 7 \log m$ bits. This completes the proof of the theorem. □

6. *One-Way Quantum Finite Automata*

In this section, we define generalized one-way quantum finite automata, and use the techniques developed above to prove size lower bounds on GQFAs. We first introduce the model.

6.1. THE ABSTRACT MODEL. A one-way quantum finite automaton is a theoretical model for a quantum computer with finite work space. QFAs were first considered by Moore and Crutchfield [2000] and Kondacs and Watrous [1997]. These models do not allow intermediate measurements, except to decide whether to accept or reject the input. The model we describe below allows the full range of operations permitted by the laws of quantum physics, subject to a space constraint. In particular, we allow any *orthogonal* (or von Neumann) 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 Aharonov et al. [1998]. We have to take care to formulate the model to properly account for all the qubits that are used in the computation. Thus any clean qubits must be accounted for explicitly in the finite memory of the automaton. For example, performing a general "positive operator valued measurement" on the state of the automaton would require a joint measurement of the state with a fresh set of ancilla qubits. Once these ancillary qubits are explicitly included in the accounting, the same effect can be achieved by a von Neumann measurement.

In abstract terms, we may define a GQFA as follows: A GQFA has a finite set of basis states Q, which consists of three parts: accepting states, rejecting states, and nonhalting states. The sets of accepting, rejecting and nonhalting basis states are denoted by Q_{acc}, Q_{rej}, and Q_{non}, respectively. One of the states, q_0, is distinguished as the starting state.

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

At any time, the state of a GQFA 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 ¢, 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 the 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 operators $\{P_{\mathsf{acc}}, P_{\mathsf{rej}}, P_{\mathsf{non}}\}$, where the P_i are orthogonal projections on the spaces E_i defined as follows: $E_{\mathsf{acc}} = \mathsf{span}\{|q\rangle \mid q \in Q_{\mathsf{acc}}\}$, $E_{\mathsf{rej}} = \mathsf{span}\{|q\rangle \mid q \in Q_{\mathsf{rej}}\}$, and $E_{\mathsf{non}} = \mathsf{span}\{|q\rangle \mid q \in Q_{\mathsf{non}}\}$. The probability of observing $i \in \{\mathsf{acc}, \mathsf{rej}, \mathsf{non}\}$ is equal to $\mathrm{Tr}(P_i \boldsymbol{\rho}')$. If we observe acc (or rej), the input is accepted (or rejected). Otherwise, the computation continues (with the state $P_{\mathsf{non}} \boldsymbol{\rho}' P_{\mathsf{non}} / \mathrm{Tr}(P_{\mathsf{non}} \boldsymbol{\rho}')$), and the next transformation, if any, is applied.

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

A GQFA 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.

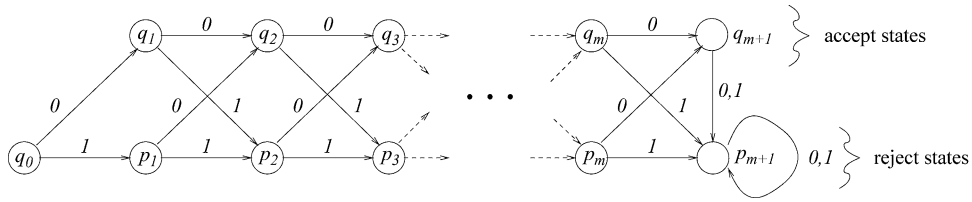

FIG. 3. *A DFA that accepts the language* $L_m = \{w0 \mid w \in \{0,1\}^*, |w| \leq m\}$.

A *reversible finite automaton* (RFA) is a GQFA such that, for any $\sigma \in \Gamma$ and $q \in Q$, $\mathcal{U}_\sigma |q\rangle\langle q| = |q'\rangle\langle q'|$ for some distinct $q' \in Q$. In other words, the operator $\mathcal{U}_\sigma$ is a permutation over the basis states.

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.

6.2. GQFA FOR CHECKING EVENNESS.

THEOREM 6.1. *Let* $L_m = \{w0 \mid w \in \{0,1\}^*, |w| \leq m\}$, $m \geq 1$, *define a family of regular languages. Then,*

(1) L_m *is recognized by a one-way deterministic automaton of size* $O(m)$,
(2) L_m *is recognized by a one-way quantum finite automaton, and*
(3) *any generalized one-way quantum automaton recognizing* L_m *with some constant probability greater than* $\frac{1}{2}$ *has* $2^{\Omega(m)}$ *states.*

Theorem 6.1 compares classical and quantum automata for checking if a given input is a small even number (an even number less than 2^{m+1}). The proof of the first two parts of Theorem 6.1 is easy. Figure 3 shows a DFA with $2m+3$ states for the language L_m. Also, since each L_m is a finite language, there is a one-way reversible finite automaton (as defined in Section 6.1), and hence a one-way QFA that accepts it. What then remains to be shown is the lower bound on the size of a one-way GQFA accepting the language.

Define an r*-restricted* one-way GQFA for a language L as a one-way GQFA that recognizes the language with probability $p > \frac{1}{2}$, and which halts with nonzero probability before seeing the right end marker only *after* it has read r letters of the input. We first show a lower bound on the size of m-restricted GQFAs that accept L_m.

Let M be any m-restricted GQFA accepting L_m with constant probability $p > \frac{1}{2}$. Note that the evolution of M on reading a random input bit corresponds exactly to that of the situation in Lemma 3.2, where we get a probabilistic mixture of two distinguishable quantum states. So, at the end of reading the entire m-bit input string, the state of M can be shown to have entropy of at least $(1 - H(p))\,m$. However, this entropy is bounded by $\log |Q|$ by Fact 2.2, where Q is the set of basis states of M. This gives us the claimed bound, as explained in detail below.

Let ρ_k be the density matrix of the GQFA M after the kth symbol of a uniformly random m-bit input has been read ($0 \leq k \leq m$).

CLAIM 6.2. $S(\boldsymbol{\rho}_k) \geq (1 - H(p))k$.

PROOF. We prove the claim by induction.

For $k = 0$, we have $S(\boldsymbol{\rho}_0) \geq 0$. Now assume that $S(\boldsymbol{\rho}_{k-1}) \geq (1 - H(p))(k - 1)$. After the kth random input symbol is read, and the unitary transformation $\mathcal{U}_\sigma$ is applied, the state of M becomes $\boldsymbol{\rho}_k = \frac{1}{2}(\mathcal{U}_0\boldsymbol{\rho}_{k-1} + \mathcal{U}_1\boldsymbol{\rho}_{k-1})$.

By the definition of M, if we now get to see the right end marker, we can learn the value of the last bit b: there is a local measurement $\mathcal{O}$ that yields b with probability at least $p > \frac{1}{2}$. So by Lemma 3.2, we have

$$S(\boldsymbol{\rho}_k) \geq \frac{1}{2}(S(\mathcal{U}_0\boldsymbol{\rho}_{k-1}) + S(\mathcal{U}_1\boldsymbol{\rho}_{k-1})) + (1 - H(p)). \qquad (*)$$

But the entropy of a mixed state is preserved by unitary transformations, and may not decrease when subjected to a von Neumann measurement (Facts 2.3 and 2.4), so $S(\mathcal{U}_b\boldsymbol{\rho}_{k-1}) \geq S(\boldsymbol{\rho}_{k-1}) \geq (1 - H(p))(k - 1)$. Inequality $(*)$ now gives us the claimed bound. $\square$

It only remains to show that the lower bound on the size of restricted GQFAs obtained above implies a lower bound on the size of arbitrary GQFAs accepting L_m. We do this by showing that we can convert *any* one-way GQFA to an r-restricted one-way GQFA which is only $O(r)$ times as large as the original GQFA. It follows that the $2^{\Omega(m)}$ lower bound on number of states of m-restricted GQFAs recognizing L_m continues to hold for arbitrary GQFAs for L_m, exactly as stated in Theorem 6.1.

The idea behind the construction of a restricted GQFA, given an arbitrary GQFA, is as follows: We carry the halting parts of the state of the original automaton as "distinguished" nonhalting parts of the state of the new automaton till at least r more symbols of the input have been read since the halting part was generated, or until the right end marker is encountered. We then map the distinguished parts of the state to accepting or rejecting subspaces appropriately.

LEMMA 6.3. *Let M be a one-way GQFA with S states recognizing a language L with probability p. Then there is an r-restricted one-way GQFA M' with $O(rS)$ states that recognizes L with probability p.*

PROOF. Let M be a GQFA with Q as the set of basis states, Q_{acc} as the set of accepting states, Q_{rej} as the set of rejecting states, and q_0 as the starting state. Let M' be the automaton with basis state set

$$\begin{aligned} &Q \cup (Q_{\mathsf{acc}} \times \{0, 1, \ldots, r + 1\} \times \{\mathsf{acc}, \mathsf{non}\}) \\ &\quad \cup (Q_{\mathsf{rej}} \times \{0, 1, \ldots, r + 1\} \times \{\mathsf{rej}, \mathsf{non}\}). \end{aligned}$$

Let $Q_{\mathsf{acc}} \cup (Q_{\mathsf{acc}} \times \{0, 1, \ldots, r + 1\} \times \{\mathsf{acc}\})$ be its set of accepting states, let $Q_{\mathsf{rej}} \cup (Q_{\mathsf{rej}} \times \{0, 1, \ldots, r + 1\} \times \{\mathsf{rej}\})$ be the set of rejecting states, and let q_0 be the starting state.

The superoperators for the new GQFA M' are constructed as follows. Consider a superoperator $\mathcal{U}_\sigma$ in M. We first extend it to the state space of M' by tensoring it with identity. Next, we compose it with a unitary operator that acts as the identity

on $\mathbb{C}^{Q_{\mathsf{non}}}$ and has the following additional transitions if $\sigma \neq \$$:

$$|q\rangle \mapsto |q, 0, \mathsf{non}\rangle \quad \text{if} \quad q \in Q_{\mathsf{acc}} \cup Q_{\mathsf{rej}}$$

$$|q, i, \mathsf{non}\rangle \mapsto \begin{cases} |q, i+1, \mathsf{non}\rangle & \text{if} \quad i < r \\ |q, i+1, \mathsf{acc}\rangle & \text{if} \quad q \in Q_{\mathsf{acc}} \quad \text{and} \quad i = r \\ |q, i+1, \mathsf{rej}\rangle & \text{if} \quad q \in Q_{\mathsf{rej}} \quad \text{and} \quad i = r \end{cases}$$

If the symbol $\sigma = \$$, then the unitary operator we use in the composition acts as the identity on the space $\mathbb{C}^Q$, and has the following additional transitions:

$$|q, i, \mathsf{non}\rangle \mapsto \begin{cases} |q, i, \mathsf{acc}\rangle & \text{if} \quad q \in Q_{\mathsf{acc}} \quad \text{and} \quad i \leq r \\ |q, i, \mathsf{rej}\rangle & \text{if} \quad q \in Q_{\mathsf{rej}} \quad \text{and} \quad i \leq r \end{cases}$$

This gives us the superoperator for the symbol σ in the new GQFA M'.

It is not difficult to verify that M' is an r-restricted one-way QFA (of size $O(rS)$) accepting the same language as M, and with the same probability. $\square$

7. *Later Work*

Our bounds were slightly generalized (to the case of interactive communication with prior entanglement) in Nayak [1999a]. Buhrman and de Wolf [2001] observed that our results imply an $\Omega(m)$ lower bound for the single-round communication complexity of determining whether two subsets of $\{1, \ldots, m\}$ are disjoint. There is an $O(\sqrt{m}\ \log m)$ qubit protocol with $O(\sqrt{m})$ rounds of communication for this problem [Buhrman et al. 1998], so we see that greater interaction leads to a decrease in the communication required to solve certain problems.

As noted in Nayak [1999a], our results imply a stronger dependence of communication complexity on the number of rounds. Suppose there are two players Alice and Bob. Alice holds an m bit string $x \in \{0, 1\}^m$ and Bob holds $i \in \{1, \ldots, m\}$. Bob would like to know the value x_i. If we allow two rounds of interaction, Bob can send i to Alice, who can respond with the value x_i, and the overall communication cost is $\log(m) + 1$. On the other hand, if the players are limited to sending one message, then our result shows that $\Omega(m)$ qubits of communication are necessary. This was further extended in Klauck et al. [2001], showing an exponential separation between quantum communication complexity with k and $k+1$ rounds of message exchange, for any constant k.

ACKNOWLEDGMENTS. We would like to thank Ike Chuang for showing us the 3-into-1 quantum encoding; Dorit Aharonov, Ike Chuang, Michael Nielsen, Steven Rudich, and Avi Wigderson for many interesting discussions; and the anonymous referees for their helpful comments.

REFERENCES

AHARONOV, D., KITAEV, A., AND NISAN, N. 1998. Quantum circuits with mixed states. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*. ACM Press, New York, 20–30.

AMBAINIS, A., AND FREIVALDS, R. 1998. 1-way quantum finite automata: Strengths, weaknesses and generalizations. In *Proceedings of the 39th Annual IEEE Symposium on Foundations of Computer Science*. IEEE Computer Society Press, Los Alamitos, Calif., pp. 332–341.

AMBAINIS, A., NAYAK, A., TA-SHMA, A., AND VAZIRANI, U. 1999. Dense quantum coding and a lower bound for 1-way quantum automata. In *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing*. ACM Press, New York, pp. 376–383.

BENNETT, C., BRASSARD, G., BREIDBART, S., AND WIESNER, S. 1982. Quantum cryptography, or unforgeable subway tokens. In *Advances in Cryptology: Proceedings of Crypto'82* (1983). D. Chaum, R. L. Rivest, and A. T. Sherman, Eds. Plenum Press, New York, NY, pp. 267–275.

BUHRMAN, H., CLEVE, R., AND WIGDERSON, A. 1998. Quantum vs. classical communication and computation. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*. ACM Press, New York, pp. 63–68.

BUHRMAN, H., AND DE WOLF, R. 2001. Communication complexity lower bounds by polynomials. In *Proceedings of the 16th Annual IEEE Conference on Computational Complexity*. IEEE Computer Society Press, Los Alamitos, Calif., pp. 120–130.

CHUANG, I. 1997. Personal communication.

COHEN, G., HONKALA, I., LITSYN, S., AND LOBSTEIN, A. 1997. *Covering Codes*. North-Holland Mathematical Library, vol. 54. Elsevier, Amsterdam, The Netherlands.

COVER, T. M., AND THOMAS, J. A. 1991. *Elements of Information Theory*. Wiley Series in Telecommunications. John Wiley & Sons, New York.

HOLEVO, A. 1973. Some estimates of the information transmitted by quantum communication channels. *Probl. Inform. Trans. 9*, 3, 177–183.

KLAUCK, H., NAYAK, A., TA-SHMA, A., AND ZUCKERMAN, D. 2001. Interaction in quantum communication and the complexity of Set Disjointness. In *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing*. ACM Press, New York, pp. 124–133.

KONDACS, A., AND WATROUS, J. 1997. On the power of quantum finite state automata. In *Proceedings of the 38th Annual IEEE Symposium on Foundations of Computer Science*. IEEE Computer Society Press, Los Alamitos, Calif., pp. 66–75.

MOORE, C., AND CRUTCHFIELD, J. 2000. Quantum automata and quantum grammars. *Theor. Comput. Sci. 237*, 1-2, 275–306.

NAYAK, A. 1999a. *Lower Bounds for Quantum Computation and Communication*. Ph.D. thesis, University of California, Berkeley, Calif.

NAYAK, A. 1999b. Optimal lower bounds for quantum automata and random access codes. In *Proceedings of the 40th Annual IEEE Symposium on Foundations of Computer Science*. IEEE Computer Society Press, Los Alamitos, Calif., pp. 369–376.

NIELSEN, M., AND CHUANG, I. 2000. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, England, Chapter 7.7, pp. 324–343.

PERES, A. 1995. *Quantum Theory: Concepts and Methods*. Kluwer Academic Publishers, Norwell, Mass.

PRESKILL, J. 1998. Lecture notes. Available online at http://www.theory.caltech.edu/people/preskill/ph229/.

WEHRL, A. 1978. General properties of entropy. *Rev. Mod. Phys. 50*, 2, 221–260.

RECEIVED JULY 2000; REVISED MAY 2002; ACCEPTED MAY 2002