1-way quantum finite automata: strengths, weaknesses and generalizations

Andris Ambainis* UC Berkeley

Rūsiņš Freivalds† University of Latvia

Abstract

We study 1-way quantum finite automata (QFAs). First, we compare them with their classical counterparts. We show that, if an automaton is required to give the correct answer with a large probability (greater than 7/9), then any 1-way QFA can be simulated by a 1-way reversible automaton. However, quantum automata giving the correct answer with smaller probabilities are more powerful than reversible automata.

Second, we show that 1-way QFAs can be very space-efficient. We construct a 1-way QFA that is exponentially smaller than any equivalent classical (even randomized) finite automaton. We think that this construction may be useful for design of other space-efficient quantum algorithms.

Third, we consider several generalizations of 1-way QFAs. Here, our goal is to find a model which is more powerful than 1-way QFAs keeping the quantum part as simple as possible.

1 Introduction

It is quite possible that the first implementations of quantum computers will not be fully quantum mechanical. Instead, they may have two parts: a quantum part and a classical part with a communication between two parts. In this case, the quantum part will be considerably more expensive than the classical part. Therefore, it will be useful to make the quantum part as small as possible even if it leads to some (reasonable) increases in the size of the classical part. This motivates the study of systems with a small quantum mechanical part.

Quantum finite automata (QFA) is a theoretical model for such systems. [12] introduced both 1-way and 2-way QFAs, with emphasis on 2-way automata because they are more

*Address: Computer Science Division, University of California, Berkeley, CA 94720, e-mail: ambainis@cs.berkeley.edu. Supported by Berkeley Fellowship for Graduate Studies. Part of this work done during 1998 I.S.I.-Elsag Bailey research meeting on quantum computing.

†Address: Institute of Mathematics and Computer Science, University of Latvia, Raina bulv. 29, Riga, LV-1459, Latvia, e-mail: rusins@cclu.lv. Supported by Latvia Science Council Grant 96.0282

powerful. However, the model of 2-way QFAs is not quite consistent with the idea of a system with a small quantum mechanical part. [12] allows superpositions where different parts of superposition have the head of QFA at different locations. (Even more, using such superpositions was the main idea in the proof that 2-way QFAs are more powerful than classical finite automata.) This means that the position of the head must be encoded into quantum state. Hence, the number of quantum states necessary to implement a 2-way QFA is not a constant but grows when the size of the input increases. This also makes state transformations more complicated (and more difficult to implement).

Hence, we think that more attention should be given to the study of simpler models like 1-way QFAs. A 1-way quantum automaton is a very reasonable model of computation and it is easy to see how it can be implemented. The finite dimensional state-space of a QFA corresponds to a system with finitely many particles. Each letter has a corresponding unitary transformation on the state-space. A classical device can read symbols from the input and apply the corresponding transformations to the quantum mechanical part.

Results about 1-way QFAs in [12] were quite pessimistic. It was shown that the class of languages recognized by 1-way QFAs is a proper subset of regular languages. We continue the investigation of 1-way QFAs and show that, despite being limited in some situations, they perform well in other situations.

Our first results consider relations between 1-way QFAs and 1-way reversible automata. Clearly, a 1-way reversible automaton is a special case of a QFA and, therefore, cannot recognize all regular languages. It is a natural question whether 1-way QFAs are more powerful than 1-way reversible automata. Interestingly, the answer depends on the accepting probability of a QFA. If a QFA gives a correct answer with a large probability (greater than 7/9), it can be replaced by a 1-way reversible automaton. However, this is not true for 0.68... and smaller probabilities.

Then, we show that QFAs can be much more space-efficient than deterministic and even probabilistic finite automata. Namely, there is a 1-way QFA that can check whether the number of letters received from the input is divisible by a prime p with only $O(\log p)$ states (this is equivalent to $\log\log p$ bits of memory). Any deterministic or probabilistic finite automaton needs p states ($\log p$ bits of memory). We think that this space-efficient quantum algorithm may be interesting for design of other quantum algoritms as well.

Finally, we consider modifications of 2-way quantum automata where the head is always at the same position for all parts of superposition. Modified 2-way QFAs can be implemented with a quantum system of constant size. Several modifications are proposed. In one of our models (1-way QFAs with a probabilistic preprocessing), some non-regular languages can be recognized.

2 Definitions

2.1 Quantum finite automata

We consider 1-way quantum finite automata (QFA) as defined in [12]. Namely, a 1-way QFA is a tuple $M = (Q, \Sigma, \delta, q_0, Q_{acc}, Q_{rej})$ where Q is a finite set of states, Σ is an input alphabet, δ is a transition function, $q_0 \in Q$ is a starting state and $Q_{acc} \subset Q$ and $Q_{rej} \subset Q$ are sets of accepting and rejecting states. The states in Q_{acc} and Q_{rej} are called *halting states* and the states in $Q_{non} = Q - (Q_{acc} \cup Q_{rej})$ are called *non-halting states.* $\not{c}$ and \$ are symbols that do not belong to Σ. We use $\not{c}$ and \$ as the left and the right endmarker, respectively. The *working alphabet* of M is $\Gamma = \Sigma \cup \{\not{c}, \$\}$.

A superposition of M is any element of $l_2(Q)$ (the space of mappings from Q to $\mathbb{C}$ with l_2 norm). For $q \in Q$, $|q\rangle$ denotes the unit vector with value 1 at q and 0 elsewhere. All elements of $l_2(Q)$ can be expressed as linear combinations of vectors $|q\rangle$. We will use ψ to denote elements of $l_2(Q)$.

The transition function δ maps $Q \times \Gamma \times Q$ to $\mathbb{C}$. The value $\delta(q_1, a, q_2)$ is the amplitude of $|q_2\rangle$ in the superposition of states to which M goes from $|q_1\rangle$ after reading a. For $a \in \Gamma$, V_a is a linear transformation on $l_2(Q)$ defined by

$$V_a(|q_1\rangle) = \sum_{q_2 \in Q} \delta(q_1, a, q_2)|q_2\rangle. \tag{1}$$

We require all V_a to be unitary.

The computation of a QFA starts in the superposition $|q_0\rangle$. Then transformations corresponding to the left endmarker $\not{c}$, the letters of the input word x and the right endmarker \$ are applied. The transformation corresponding to $a \in \Gamma$ consists of two steps.

1. First, V_a is applied. The new superposition ψ' is $V_a(\psi)$ where ψ is the superposition before this step.

2. Then, ψ' is observed with respect to the observable $E_{acc} \oplus E_{rej} \oplus E_{non}$ where $E_{acc} = span\{|q\rangle : q \in Q_{acc}\}$, $E_{rej} = span\{|q\rangle : q \in Q_{rej}\}$, $E_{non} = span\{|q\rangle : q \in Q_{non}\}$. This observation gives $x \in E_i$ with the probability equal to the amplitude of the projection of ψ'. After that, the superposition collapses to this projection.

   If we get $\psi' \in E_{acc}$, the input is accepted. If $\psi' \in E_{rej}$, the input is rejected. If $\psi' \in E_{non}$, the next transformation is applied.

We regard these two transformations as reading a letter a.

Another definition of QFAs. Independently of [12], quantum automata were introduced in [13]. There is one difference between these two definitions. In [12], a QFA is observed after reading each letter (after doing each V_a). In [13], a QFA is observed only after all letters have been read. Any language recognized by a QFA according to the definition of [13] is recognized by a QFA according to [12]. The converse is not true. Any finite language can be recognized in the sense of [12]. However, no finite non-empty language can be recognized in the sense of [13]. Everywhere in this paper, we will use the more general definition of [12]. However, our results of section 4.1 which show that 1-way QFAs can be more space-efficient than deterministic or probabilistic automata are true in the more restricted model of [13] as well.

2.2 Example

To explain our notation, we give an example of a 1-way QFA. To keep it simple, we use a one letter alphabet $\Sigma = \{a\}$. The state space is $Q = \{q_0, q_1, q_{acc}, q_{rej}\}$ with the set of accepting states $Q_{acc} = \{q_{acc}\}$ and the set of rejecting states $Q_{rej} = \{q_{rej}\}$. The starting state is q_0.

The transition function can be specified in two ways: by specifying δ or by specifying V_x for all letters $x \in \Gamma$. These methods are equivalent: all V_x are determined by δ and equation (1). We shall define the automaton by describing V_x.

$$V_a(|q_0\rangle) = \frac{1}{2}|q_0\rangle + \frac{1}{2}|q_1\rangle + \frac{1}{\sqrt{2}}|q_{rej}\rangle,$$

$$V_a(|q_1\rangle) = \frac{1}{2}|q_0\rangle + \frac{1}{2}|q_1\rangle - \frac{1}{\sqrt{2}}|q_{rej}\rangle,$$

$$V_{\$}(|q_0\rangle) = |q_{rej}\rangle, V_{\$}(|q_1\rangle) = |q_{acc}\rangle.$$

It can be also defined by describing δ. For example, $V_a(|q_0\rangle) = \frac{1}{2}|q_0\rangle + \frac{1}{2}|q_1\rangle + \frac{1}{\sqrt{2}}|q_{rej}\rangle$ would be

$$\delta(q_0, a, q_0) = \frac{1}{2}, \delta(q_0, a, q_1) = \frac{1}{2},$$

$$\delta(q_0, a, q_{acc}) = 0, \delta(q_0, a, q_{rej}) = \frac{1}{\sqrt{2}}.$$

As we see, this is much longer. For this reason, we will mainly use V_x notation.

There are some transitions that we have not described. For example, $V_a(q_{acc})$ has not been specified. These values are not important and can be arbitrary. We need them to be such that V_a is unitary but this is not difficult. As long as all specified $V_a(q_i)$ are orthogonal, the remaining values can be assigned so that the whole V_a is unitary. In the sequel, we will often shorten descriptions of QFAs by leaving out transitions that can be defined arbitrarily.

Next, we show how this automaton works on the word aa.

1. The automaton starts in $|q_0\rangle$. Then, V_a is applied, giving $\frac{1}{2}|q_0\rangle + \frac{1}{2}|q_1\rangle + \frac{1}{\sqrt{2}}|q_{rej}\rangle$. This is observed. Two outcomes are possible. With probability $(1/\sqrt{2})^2 = 1/2$, a rejecting state is observed. Then, the superposition collapses to $|q_{rej}\rangle$, the word is rejected and the computation terminates. Otherwise (with probability 1/2), a non-halting state is observed and the superposition collapses to $\frac{1}{2}|q_0\rangle + \frac{1}{2}|q_1\rangle$. In this case, the computation continues.

2. A simple computation shows that $\frac{1}{2}|q_0\rangle + \frac{1}{2}|q_1\rangle$ is mapped to itself by V_a. After that, a non-halting state is observed. (There are no accepting or rejecting states in this superposition.)

3. Then, the word ends and the transformation $V_\$$ corresponding to the right endmarker \$ is done. It maps the superposition to $\frac{1}{2}|q_{rej}\rangle + \frac{1}{2}|q_{acc}\rangle$. This is observed. With probability $(1/2)^2 = 1/4$, the rejecting state q_{rej} is observed. With probability 1/4, the accepting state q_{acc} is observed.

The total probability of accepting is 1/4, the probability of rejecting is $1/2 + 1/4 = 3/4$.

2.3 Reversible automata

A 1-way reversible finite automaton (RFA) is a QFA with $\delta(q_1, a, q_2) \in \{0, 1\}$ for all q_1, a, q_2. Alternatively, RFA can be defined as a deterministic automaton where, for any q_2, a, there is at most one state q_1 such that reading a in q_1 leads to q_2. We use the same definitions of acceptance and rejection. States are partitioned into accepting, rejecting and non-halting states and a word is accepted (rejected) whenever the RFA enters an accepting (rejecting) state. After that, the computation is terminated. Similarly to quantum case, endmarkers are added to the input word. The starting state is one, accepting (rejecting) states can be multiple. This makes our model different from both [3] (where only one accepting state was allowed) and [14] (where multiple starting states with a non-deterministic choice between them at the beginning were allowed). We define our model so because we want it to be as close to our model of QFAs as possible.

Generally, it's hard to introduce probabilism into finite automata without losing reversibility. However, there are some types of probabilistic choices that are consistent with reversibility. For example, we can choose the starting state probabilistically. The next example shows that such probabilistic choices increase the power of an automaton.

Example. Consider the language $L = \{a^{2n+3} | n \in \mathbb{N}\}$. It cannot be recognized by a 1-way RFA. However, there are 3 1-way RFAs such that each word in the language is accepted by 2 of them and each word not in the language is rejected by 2 out of 3. Hence, if we choose one of these three automata equiprobably, L will be recognized with the probability of correct answer 2/3.

Probabilistic choices of this type can be easily done in our model of QFAs. This may lead to a claim that QFAs are more powerful than classical reversible automata because they can do such probabilistic choices. We wish to avoid such situations and to separate probabilistic choices from real quantum effects.

Therefore, we define 1-way finite automata with probabilistic choices (PRFAs) and compare capabilities of QFAs with them. A PRFA is a probabilistic finite automaton such that, for any state q_1 and any $a \in \Gamma$, there is at most one state q_2 such that the probability of passing from q_2 to q_1 after reading a is non-zero. Definitions of acceptance and rejection are similar to QFAs and RFAs. Now, the probabilistic automaton from the example above becomes a 1-way PRFA.

Theorem 1 *1. If a language is accepted by a 1-way RFA, it is accepted by a 1-way PRFA.*

2. If a language is accepted by a 1-way PRFA, it is accepted by a 1-way QFA with the same probability of correct answer.

Proof: Easy. □

In section 3.2 we will compare the power of 1-way QFAs and PRFAs and show that 1-way quantum automata can actually do more than just probabilistic choices.

3 Capabilities of RFAs and QFAs

3.1 QFAs with probability of correct answer above 7/9

We characterize the languages recognized by 1-way QFAs in terms of their minimal automata. The minimal automaton of a language L is a 1-way deterministic finite automaton recognizing it with the smallest number of states. (Note: the minimal automaton can be non-reversible, even for some languages L that can be recognized by a 1-way RFA. The extreme case of this is our Theorem 12 where the smallest 1-way RFA is exponentially bigger than the minimal nonreversible automaton.) It is well known[9] that the minimal automaton is unique and can be effectively constructed.

Theorem 2 *Let L be a language and M be its minimal automaton. Assume that there is a word x such that M contains states q_1, q_2 satisfying:*

1. *$q_1 \neq q_2$,*
2. *If M starts in the state q_1 and reads x, it passes to q_2,*
3. *If M starts in q_2 and reads x, it passes to q_2, and*
4. *q_2 is neither "all - accepting" state, nor "all - rejecting" state.*

Then L cannot be recognized by a 1-way QFA with probability at least $7/9+\epsilon$ for any fixed $\epsilon > 0$.

Proof: We prove the result for a slightly smaller probability of correct answer $5/6+\epsilon$ (instead of $7/8+\epsilon$). The proof for $7/8+\epsilon$ is technically more complicated.

Let L be a language such that its minimal automaton contains the "forbidden construction" and M be a QFA. We show that, for some word y the probability of M giving the correct answer to "$y \in L$?" is less than $5/6+\epsilon$. This implies that L cannot be recognized with probability of correct answer being $5/6+\epsilon$.

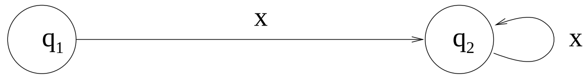

Figure 1: *"The forbidden construction" of Theorem 2.*

For simplicity, we assume that q_1 is the starting state of M. We introduce some notation. Let $P_{non}(\psi)$ be the non-halting part of ψ and $P_{halt}(\psi)$ be the halting part of ψ. $V'_a = P_{non}V_a$ is a transformation that maps ψ to the non-halting part of $V_a(\psi)$. If x is a word consisting of letters $a_1 \dots a_k$, then V'_x denotes $V'_{a_k} \dots V_{a_1}$. ψ_x denotes the non-halting part of the QFA's configuration after reading x. It is easy to see that, for any word x and letter a, $\psi_{xa} = V'_a(\psi_x)$.

We recall that $l_2(Q)$ denotes the state-space of M with l_2 norm $\|\psi\|$. $l_2(Q) = E_{acc} \oplus E_{rej} \oplus E_{non}$.

First, we prove that the state-space of M can be decomposed into two parts with different behavior.

Lemma 1 *There are subspaces E_1, E_2 such that $E_{non} = E_1 \oplus E_2$ and*

(i) If $\psi \in E_1$, then $V_x(\psi) \in E_1$,

(ii) If $\psi \in E_2$, then $\|V'_{x^k}(\psi)\| \to 0$ when $k \to \infty$.

Proof. We define two sequences of subspaces $E_1^1, E_1^2, \dots$ and $E_2^1, E_2^2, \dots$ such that $E_{non} = E_1^i \oplus E_2^i$. Let $E_1^1 = \{\psi | \psi \in E_{non} \text{ and } V_a(\psi) \in E_{non}\}$ (i.e., the subspace of all ψ such that both ψ and $V_a(\psi)$ have only non-halting components). E_2^1 consists of all vectors in E_{non} orthogonal to E_1^1. For $i > 1$, $E_1^i = E_1^{i-1} \cap \{\psi | V_a(\psi) \in E_1^{i-1}\}$ and E_2^i consists of all vectors in E_{non} orthogonal to E_1^i.

Clearly, $E_1^1 \supseteq E_1^2 \supseteq \dots$. If E_1^{i+1} is a proper subspace of E_1^i, then the dimensionality of E_1^{i+1} is smaller than the dimensionality of E_1^i. This can happen only finitely many times because the original E_1^1 is finite-dimensional. Hence, there is i_0 such that $E_1^{i_0} = E_1^{i_0+1}$. We define $E_1 = E_1^{i_0}$, $E_2 = E_2^{i_0}$. Next, we check that both (i) and (ii) are true.

Let $\psi \in E_1$. Then, $V_a(\psi) \in E_1^{i_0+1} = E_1^{i_0} = E_1$ and $V_a(\psi) \in E_{non}$ by $E_1 \subseteq E_1^1$ and the definition of E_1^1. It remains to prove that E_2 satisfies (ii) condition of Lemma 1.

Claim 1 *If $\psi \in E_1^i$, then $P_{halt}(V_a(V'_{a^l}(\psi))) = \overrightarrow{0}$ for all $l \leq i-1$.*

Proof: By induction. If $i = 1$, then $P_{halt}(V_a(\psi)) = \overrightarrow{0}$ by definition of E_1^1. If $i > 1$ and $l = 0$, then $P_{halt}(V_a(\psi)) = \overrightarrow{0}$ because $E_1^i \subseteq E_1^1$.

The only remaining case is $i > 1$ and $l > 1$. Let $\psi' = V_a(\psi)$. By definition of E_1^i, $V_a(\psi) \in E_1^{i-1}$. We also have $V_a(\psi) \in E_{non}$ because $E_1^{i-1} \subseteq E_{non}$. Hence, $V'_a(\psi) = P_{non}(V_a(\psi)) = V_a(\psi) = \psi'$ and

$$P_{halt}(V_a(V'_{a^l}(\psi))) = P_{halt}(V_a(V'_{a^{l-1}}(\psi'))) = \overrightarrow{0}$$

by $\psi' \in E_1^{i-1}$ and inductive assumption. □

Claim 2 *Let $\psi = \psi_1 + \psi_2$, $\psi_1 \in E_1^i, \psi_2 \in E_2^i$. Then, for all $l \leq i-1$,*

$$P_{halt}(V_a(V'_{a^l}(\psi))) = P_{halt}(V_a(V'_{a^l}(\psi_2))).$$

Proof: By linearity of P_{halt}, V_a, V'_a,

$$P_{halt}(V_a(V'_{a^l}(\psi))) = P_{halt}(V_a(V'_{a^l}(\psi_1))) + P_{halt}(V_a(V'_{a^l}(\psi_2))).$$

Claim 1 implies that $P_{halt}(V_a(V'_{a^l}(\psi_1))) = \overrightarrow{0}$. $\square$

Claim 3 *Let $j \in \{1, \ldots, i_0\}$. There is a constant $\delta_j > 0$ such that for any $\psi \in E_2^i$ there is $l \in \{0, \ldots, j-1\}$ such that $\|P_{halt}(V_a(V'_{a^l}(\psi)))\| \geq \delta_j \|\psi\|$.*

Proof: By induction.
Base Case. Consider the linear transformation $T_1 : E_2^1 \to E_{acc} \oplus E_{rej}$ that maps $\psi \in E$ to the halting part of $V_a(\psi)$. $\|T_1\|$ (the norm of T_1) is the minimum of $\|T_1(x)\|$ over all x such that $\|x\| = 1$.

If $\|T_1\| = 0$, there is $x \in E_2^1$ such that $\|x\| = 1$ but $\|T_1(x)\| = 0$. This means that $T_1(x) = \overrightarrow{0}$, implying that $x \in E_1^1$. However, $E_1^1 \cap E_2^1 = \{\overrightarrow{0}\}$, leading to a contradiction. Hence, $\|T_1\| > 0$. Also, $\|T_1\| \leq 1$ because V_a is unitary and projection to the halting subspace can only decrease the norm.

We take $\delta_1 = \|T_1\|$. Then, the halting part of $V_a(\psi)$ is at least $\|T_1\| \|\psi\| = \delta_1 \|\psi\|$.

Inductive Case. We assume that the lemma is true for E_2^i and prove it for E_2^{i+1}.

We consider the transformation T_{i+1} mapping $\psi \in E_2^{i+1}$ to the projection of $V_a(\psi)$ on E_2^i. If $T_{i+1}(\psi) = \overrightarrow{0}$, then $\psi \in E_1^{i+1}$ by the definition of E_1^{i+1}. Similarly to the previous case, $E_1^{i+1} \cap E_2^{i+1} = \{\overrightarrow{0}\}$. Hence, if $T_{i+1}(\psi) = \overrightarrow{0}$ and $\psi \in E_2^{i+1}$, then $\psi = \overrightarrow{0}$. This means that $\|T_{i+1}\| > 0$. We can also prove that $\|T_{i+1}\| \leq 1$.

We define $\delta_{i+1} = \frac{\|T_{i+1}\|}{4} \delta_i$.

Let $E_3^i = \{x \in E_2^{i+1}$ and $x \perp E_2^i\}$. Then, $E_2^{i+1} = E_2^i \oplus E_3^i$. We also note that E_3^i is a subspace of E_1^i. (This follows from definitions of E_1^i and E_3^i.)

To show that one of $P_{halt}(V_a(V'_{a^l}(\psi)))$ is large enough, we represent $\psi \in E_2^{i+1}$ as $\psi_2 + \psi_3$, $\psi_2 \in E_2^i$, $\psi_3 \in E_3^i$. There are two cases:

1. $\|\psi_2\| \geq \frac{\|T_{i+1}\|}{4} \|\psi\|$.

   Then,

   $$P_{halt}(V_a(V'_{a^l}(\psi_2))) \geq \delta_i \|\psi_2\| \geq \delta_i \frac{\|T_{i+1}\|}{4} \|\psi\| = \delta_{i+1} \|\psi\|$$

   for some $l \leq i-1$ because $\psi_2 \in E_2^i$ and we can apply the inductive assumption. Claim 2 implies that this is also true for $P_{halt}(V_a(V'_{a^l}(\psi)))$.

2. $\|\psi_2\| < \frac{\|T_{i+1}\|}{4}\|\psi\|$.

   Then, by triangle inequality,

$$\|\psi_3\| \geq \|\psi\| - \|\psi_2\| \geq (1 - \frac{\|T_{i+1}\|}{4})\|\psi\| \geq \frac{3\|T_{i+1}\|}{4}\|\psi\|.$$

   Let ψ', ψ'_2 and ψ'_3 be the projections of $V_a(\psi)$, $V_a(\psi_2)$, $V_a(\psi_3)$ on E_2^i. Clearly, $\psi' = \psi'_2 + \psi'_3$. Triangle inequality gives us

$$\|\psi'\| \geq \|\psi'_3\| - \|\psi'_2\| \geq \|\psi'_3\| - \|\psi_2\| \geq \frac{3\|T_{i+1}\|}{4}\|\psi\| - \frac{\|T_{i+1}\|}{4}\|\psi\| = \frac{\|T_{i+1}\|}{2}\|\psi\|.$$

   We have $\psi' = P_{non}(\psi') + P_{halt}(\psi')$. Again, we have two cases.

   If $\|P_{halt}(\psi')\| \geq \frac{\|T_{i+1}\|}{4}\|\psi\|$, then $\|P_{halt}(\psi')\| \geq \delta_{i+1}\|\psi\|$ because $\delta_{i+1} = \frac{\|T_{i+1}\|}{4}\delta_i$ and $\delta_i \leq 1$ because all $\|T_i\|$ are at most 1.

   Otherwise, by triangle inequality, $\|P_{non}(\psi')\| \geq \|\psi'\| - \|P_{halt}(\psi')\| \geq \frac{\|T_{i+1}\|}{4}\|\psi\|$. By inductive assumption, there is $l \leq i-1$ such that $\|P_{halt}V_a(V'_{a^l}(\psi'))\| \geq \delta_i\|\psi'\|$. Therefore,

$$\|P_{halt}(V_a(V'_{a^{l+1}}(\psi)))\| \geq \frac{\|T_{i+1}\|}{4}\delta_i\|\psi\| = \delta_{i+1}\|\psi\|.$$

□

Claim 4 *There is c such that $0 < c < 1$ and, for any $\psi \in E_2$, $t \in \mathbb{N}$, $\|V'_{a^{i_0 t}}(\psi)\| \leq c^t\|\psi\|$.*

Proof: We take $c = \sqrt{1-\delta_{i_0}^2}$.

By Claim 3, one of $P_{halt}(V_a(V'_{a^l}(\psi)))$ is at least $\delta_{i_0}\|\psi\|$. $P_{non}(V_a(V'_{a^l}(\psi)))$ is orthogonal to this vector. Hence, $P_{non}(V_a(V'_{a^l}(\psi)))$ is at most

$$\sqrt{\|\psi\|^2 - \delta_{i_0}^2\|\psi\|^2} = \sqrt{1-\delta_{i_0}^2}\|\psi\|.$$

$V'_{a^{i_0}}$ can be only smaller because V_a is unitary and P_{non} does not increase vectors.

We have shown that $V'_{a^{i_0}}(\psi) \leq c\|\psi\|$. Repeating this t times, we obtain Claim 4. □

Clearly, $c^t \to 0$ if $t \to \infty$. This completes the proof of Lemma 1. □

Let $\psi_{\not{c}} = \psi_{\not{c}}^1 + \psi_{\not{c}}^2$, $\psi_{\not{c}}^1 \in E_1$, $\psi_{\not{c}}^2 \in E_2$. We consider two cases.

Case 1. $\|\psi_{\not{c}}^2\| > 1/\sqrt{3}$.

Then, $\|\psi_{\not{c}}^1\| < \sqrt{1-(1/\sqrt{3})^2} = 2/\sqrt{3}$. This also means $\|V'_{x^i}(\psi_{\not{c}}^1)\| < 2/\sqrt{3}$. For sufficiently large i, $\|V'_{x^i}(\psi_{\not{c}}^2)\|$ becomes negligible (part (ii) of Lemma 1). Then, $\|\psi_{\not{c}x^i}\| < 2/\sqrt{3}$. The probability of M halting after this moment is less than $2/3$. Hence, M has already halted with probability more than $1/3$ and accepted (or rejected) with probability more than 1/6. This means that M cannot reject (accept) any continuation of x^i with probability 5/6. However, x^i has both continuations in L and continuations not in L. Hence, M does not recognize L.

Case 2. $\|\psi^2_{\not{c}}\| \leq 1/\sqrt{3}$.

q_1 and q_2 are different states in the minimal automaton of L. Therefore, there is a word $y \in \Sigma^*$ such that y leads to acceptance from one of q_1, q_2 but not from the other one. We consider the distributions of probabilities on M's answers "accept" and "reject" on y and $x^i y$. On one of these words, M must accept with probability at least $5/6 + \epsilon$ and reject with probability at most $1/6 - \epsilon$. On the other word, M must accept with probability most $1/6 - \epsilon$ and reject with probability at least $5/6 + \epsilon$. Therefore, both the probabilities of accepting and the probabilities of rejecting must differ by at least $2/3 + 2\epsilon$. This means that the *variational distance* between two probability distributions (the sum of these two distances) must be at least $4/3 + 4\epsilon$. We show that it cannot be so large.

First, we select an appropriate i. Let m be so large that $\|V'_{x^m}(\psi^2_{\not{c}})\| \leq \delta$ for $\delta = \epsilon/4$. $\psi^1_{\not{c}}, V'_x(\psi^1_{\not{c}}), V'_{x^2}(\psi^1_{\not{c}}), \ldots$ is a sequence in a finite-dimensional space. Therefore, it has a limit point and there are i, j such that

$$\|V'_{x^j}(\psi^1_{\not{c}}) - V'_{x^{i+j}}(\psi^1_{\not{c}})\| < \delta.$$

We choose i, j so that $i > m$.

The difference between two probability distributions comes from two sources. The first is difference between $\psi_{\not{c}}$ and $\psi_{\not{c}x^i}$ (the states of M before reading y). The second source is the possibility of M accepting while reading x^i (the only part that is different in two words). We bound the difference created by each of these two sources.

The difference $\psi_{\not{c}} - \psi_{\not{c}x^i}$ can be partitioned into three parts.

$$\psi_{\not{c}} - \psi_{\not{c}x^i} = (\psi_{\not{c}} - \psi^1_{\not{c}}) + (\psi^1_{\not{c}} - \psi^1_{\not{c}x^i}) + (\psi^1_{\not{c}x^i} - \psi_{\not{c}x^i}).$$

The first part is $\psi_{\not{c}} - \psi^1_{\not{c}} = \psi^2_{\not{c}}$ and $\|\psi^2_{\not{c}}\| \leq \frac{1}{\sqrt{3}}$. The second and the third parts are both small. For the second part, notice that V'_x is unitary on E_1 (because V_x is unitary and $V_x(\psi)$ does not contain halting components for $\psi \in E_1$). Hence, V'_x preserves distances on E_1 and

$$\|\psi^1_{\not{c}} - \psi^1_{\not{c}x^i}\| = \|\psi^1_{\not{c}x^j} - \psi^1_{\not{c}x^{i+j}}\| \leq \delta.$$

The third part is $\psi_{\not{c}x^i} - \psi^1_{\not{c}x^i} = \psi^2_{\not{c}x^i}$ and $\|\psi^2_{\not{c}x^i}\| \leq \delta$ because $i > m$.

Next, we state two lemmas relating differences between two superpositions and differences between probability distributions created by observing these superpositions. The first lemma is by Bernstein and Vazirani[4].

Lemma 2 *[4] Let ψ and ϕ be such that $\|\psi\| \leq 1$, $\|\phi\| \leq 1$ and $\|\psi - \phi\| \leq \epsilon$. Then the total variational distance resulting from measurements of ϕ and ψ is at most 4ϵ.*

The second lemma is our improvement of lemma 2.

Lemma 3 *Let ψ^1 and ψ^2 be such that $\psi^1 \perp \psi^2$. Then the total variational distance resulting from measurements of ψ^1 and $\psi^1+\psi^2$ is at most*

$$\|\psi^2\|\sqrt{\|4\psi^1\|^2+\|\psi^2\|^2}.$$

Proof. Omitted. $\Box$

We apply lemma 3 to $\psi^1_{\not\in}$ and $\psi^2_{\not\in}$. This gives that the variational distance between distributions generated by $\psi^1_{\not\in}$ and $\psi^1_{\not\in}+\psi^2_{\not\in}$ is at most 1. Then, we apply lemma 2 to two other parts of $\psi_{\not\in}-\psi_{\not\in x^i}$. Each of them influences the variational distance by at most 4δ. Together, the variational distance between distributions obtained by observing $\psi_{\not\in}$ and $\psi_{\not\in x^i}$ is at most $1+8\delta$.

The probability of M halting while reading x^i is at most $\|\psi^2_{\not\in}\|^2=1/3$. Adding it increases the variational distance by at most $1/3$. Hence, the total variational distance is at most $4/3+8\delta=4/3+2\epsilon$. However, if M distinguishes y and $x^i y$ correctly, it should be at least $4/3+4\epsilon$. Hence, M does not recognize one of these words correctly. $\Box$

Theorem 3 *Let L be a language and M be its minimal automaton. If M does not contain the "forbidden construction" of Theorem 2, then L can be recognized by a 1-way reversible finite automaton.*

Proof. We define a non-reversibility as a tuple $\langle q_1, q_2, q, a\rangle$ where $q_1, q_2, q \in Q$, $a \in \Sigma$, $q_1 \neq q_2$ and reading a in q_1 or q_2 leads to q. Let m be the number of non-reversibilities in M. We show how to modify M so that the number of non-reversibilities decreases. A reversible automaton is obtained by repeating this modification several times.

We define a partial ordering $<$ on non-reversibilities. $\langle q_1, q_2, q, a\rangle < \langle q_1', q_2', q', a'\rangle$ if and only if one of q_1' and q_2' is reachable from q. It is easy to see that $<$ is transitive.

Lemma 4 *$<$ is anti-reflexive.*

Proof. For a contradiction, assume there is $\langle q_1, q_2, q, a\rangle$ such that $\langle q_1, q_2, q, a\rangle < \langle q_1, q_2, q, a\rangle$. We also assume that q_2 is reachable from q by reading a word y. (Otherwise, q_1 is reachable from q and we can just exchange q_1 and q_2.) Then, reading $x=ay$ leads from q_1 to q_2 and from q_2 to q_2. This contradicts our assumption that M does not contain such q_1, q_2. $\Box$

Hence, there is a tuple $\langle q_1, q_2, q, a\rangle$ that is maximal with respect to $<$. We create two copies for state q and all states reachable from q. If M reads a in q_1, it passes to one copy of q, if it reads a in q_2, it passes to the second copy. This eliminates this non-reversibility. Other non-reversibilities are not dublicated because they are not reachable from q. Hence, the number of non-reversibilities is decreased. $\Box$

Corollary 1 *A language can be recognized by a 1-way QFA with probability $7/9 + \epsilon$ if and only if it can be recognized by a 1-way reversible finite automaton.*

Proof. Clearly, a RFA is a special case of a QFA. The other direction follows from Theorems 2 and 3. □

This immediately implies the same result about 1-way reversible automata with probabilistic choices. For this type of automata, a stronger result can be proved.

Theorem 4 *A language can be recognized by a 1-way PRFA with probability $2/3 + \epsilon$ (for arbitrary $\epsilon > 0$) if and only if it can be recognized by a 1-way reversible finite automaton.*

Proof: Omitted. □

The example in Section 2.3 shows that Theorem 4 is tight.

3.2 QFAs with probability of correct answer below 7/9

For smaller probabilities, QFAs are slightly more powerful than RFAs or even PRFAs.

Theorem 5 *The language a^*b^* can be recognized by a 1-way QFA with the probability of correct answer $p = 0.68...$ where p is the root of $p^3 + p = 1$.*

Proof. We describe a 1-way QFA M accepting this language. The automaton has 4 states: q_0, q_1, q_{acc} and q_{rej}. $Q_{acc} = \{q_{acc}\}$, $Q_{rej} = \{q_{rej}\}$. The initial state is $\sqrt{1-p}|q_0\rangle + \sqrt{p}|q_1\rangle$. The transition function is

$$V_a(|q_0\rangle) = (1-p)|q_0\rangle + \sqrt{p(1-p)}|q_1\rangle + \sqrt{p}|q_{rej}\rangle,$$

$$V_a(|q_1\rangle) = \sqrt{p(1-p)}|q_0\rangle + p|q_1\rangle - \sqrt{1-p}|q_{rej}\rangle,$$
$$V_b(|q_0\rangle) = |q_{rej}\rangle, V_b(|q_1\rangle) = |q_1\rangle,$$
$$V_\$(|q_0\rangle) = |q_{rej}\rangle, V_\$(|q_1\rangle) = |q_{acc}\rangle.$$

Case 1. The input is $x = a^*$.

It is straightforward that δ maps $\sqrt{1-p}|q_0\rangle + \sqrt{p}|q_1\rangle$ to itself while it receives a from the input. Hence, after reading a^* the state remains $\sqrt{1-p}|q_0\rangle + \sqrt{p}|q_1\rangle$ and, after reading the right endmarker, it becomes $\sqrt{1-p}|q_{rej}\rangle + \sqrt{p}|q_{acc}\rangle$. This means that the automaton accepts with probability p.

Case 2. The input is $x = a^*b^+$.

Again, the state remains $\sqrt{1-p}|q_0\rangle + \sqrt{p}|q_1\rangle$ while input contains a. Reading the first b changes it to $\sqrt{1-p}|q_{rej}\rangle + \sqrt{p}|q_1\rangle$. The non-halting part of this state is $\sqrt{p}|q_1\rangle$. It is left unchanged by next bs and mapped to $|q_{acc}\rangle$ after reading the right endmarker. Again, the accepting probability is p.

Case 3. The input is $x \notin a^*b^*$.

Then, the initial segment of x is $a^*b^+a^+$. After reading the first b, the state is $\sqrt{1-p}|q_{rej}\rangle + \sqrt{p}|q_1\rangle$. The automaton rejects at this moment with probability $(1-p)$. The non-halting part $\sqrt{p}|q_1\rangle$ is mapped to $p\sqrt{1-p}|q_0\rangle + (1-p)\sqrt{p}|q_1\rangle - \sqrt{p(1-p)}|q_{rej}\rangle$ by the next V_a. Then, the automaton rejects with probability $p(1-p)$. The non-halting part $p\sqrt{1-p}|q_0\rangle + (1-p)\sqrt{p}|q_1\rangle$ is unchanged by as. However, either b or right endmarker follows as and then q_0 is mapped to $|q_{rej}\rangle$ and the automaton rejects with probability $p^2(1-p)$. We add the probabilities of rejecting at different moments together and get that M rejects $x \notin a^*b^*$ with probability at least

$$(1-p) + p(1-p) + p^2(1-p) = (1+p+p^2)(1-p) =$$
$$\frac{1-p^3}{1-p}(1-p) = 1-p^3 = p.$$

□

It is easy to see that the minimal automaton of a^*b^* contains the "forbidden construction" of Theorem 2. Therefore, we have

Corollary 2 *There is a language that can be recognized by a 1-QFA with probability* 0.68... *but not with probability* $7/9+\epsilon$.

Proof: Follows from Theorems 2 and 5. □

For probabilistic computation, the property that the probability of correct answer can be increased arbitrarily is considered evident. Hence, it was not surprising that [12] wrote "with error probability bounded away from 1/2" about QFAs, thinking that all such probabilities are equivalent. However, mixing reversible (quantum computation) and nonreversible (measurements after each step) components in one model makes it impossible for QFAs. It is open whether a counterpart of Corollary 2 is true for 2-way QFAs.

Corollary 3 *There is a language that can be recognized by a 1-QFA with probability* 0.68... *but not by a classical 1-way reversible FA.*

This corollary can be improved by showing that even a 1-way probabilistic reversible automaton cannot recognize this language (and even with probability $1/2+\epsilon$).

Theorem 6 *Let L be a language and M be its minimal automaton. Assume that there are words x, y and M's states q_1, q_2 such that*

1. *none of q_1 and q_2 is "all-accepting" or "all-rejecting" state;*
2. *reading x in q_1 leads to q_1;*
3. *reading y in q_1 leads to q_2;*
4. *reading y in q_2 leads to q_2;*

5. *there is no* $i > 0$ *such that reading* x^i *leads from* q_2 *to* q_2.

Then L *cannot be recognized by a 1-way PRFA with probability* $1/2+\epsilon$, *for any* $\epsilon > 0$.

Proof. Without the loss of generality, we assume that q_1 is the starting state of M. Let M_p be a 1-way probabilistic reversible automaton. We are going to show that, for some word x, the probability of M_p giving the right answer on the input x is less than $1/2+\epsilon$.

Lemma 5 *For any state* q *and* $a \in \Sigma^+$, *there is* k *such that* $0 < k \leq |Q|$ *and, for any sequence of probabilistic choices, one of the following happens:*

1. *After reading* a^k *in state* q, M_p *returns to* q;

2. *After reading* $a^{|Q|+1}$ *in state* q, M_p *accepts or rejects.*

Proof. Let $q_0, q_1, \ldots$ be any sequence of non-halting states such that $q_0 = q$ and the probability that reading a causes M_p to go from q_i to q_{i+1} is non-zero. If the length of the sequence is greater than $|Q|$, then some state appears twice in this sequence. We consider the first state which appears twice. If it is not q_0, then it has two preceding states: the state preceding it when it appears in the sequence for the first time and the state preceding it when it appears in the sequence for the second time. This contradicts the definition of a probabilistic reversible automaton. We have shown that q_0 is the first state which appears twice.

Next, assume we have two such sequences: $q_0, q_1, \ldots$ and $q'_0, q'_1, \ldots$. Let k_1, k_2 be the smallest numbers such that $k_1 > 0$, $q_{k_1} = q_0$ and $k_2 > 0$, $q'_{k_2} = q'_0$, respectively. We show that $k_1 = k_2$. For a contradiction, assume that $k_1 > k_2$ ($k_2 > k_1$ case is similar.). Then, $q_{k_1-1} = q'_{k_2-1}$ (because the state $q_{k_1} = q'_{k_2}$ cannot have two preceding states), $q_{k_1-2} = q'_{k_2-2}$ and so on, $q_{k_1-k_2} = q'_{k_2-k_2} = q'_0 = q_0$. This contradicts the assumption that k_1 is the smallest number such that $k_1 > 0$ and $q_{k_1} = q_0$. Hence, $k_1 = k_2$. □

Lemma 6 *For any state* q *and* $a \in \Sigma^+$, *one of the following happens:*

1. *There is* k *such that, after reading* a^k *in state* q, M_p *returns to* q *for any sequence of probabilistic choices;*

2. *The probability of halting after reading* a^k *in state* q *tends to 1 when* $k \to \infty$.

Proof. Let k be as in Lemma 5. If M_p always returns to q after reading a^k, Lemma 6 is true. It remains to consider the case if there is a sequence of probabilistic choices for which M_p does not return to q. Then, by Lemma 5, this sequence causes M_p to halt. Let p be the probability of returning to q after reading a^k. Then, the probability of returning to q after reading a^{ik} is p^i. With probability $1-p^i$, M_p does not returns to q at some moment and (this is the only alternative) terminates after reading $a^{ik+|Q|}$ (or some its prefix). Clearly, $1-p^i \to 1$, if $i \to \infty$. □

We note that one can use the same $k = |Q|!$ for all q and a. (For any q, a, $k \leq |Q|$ and $|Q|!$ is a multiple of any such k.) We shall call the states of the first type *return states* for a.

Let p_i be the probability of non-halting after reading x^i and $p = \lim_{i\to\infty} p_i$. We select i so that $|p - p_i| < \epsilon$. Let p'_j be the probability of non-halting after reading $x^i y^j$ and $p' = \lim_{j\to\infty} p'_j$. We select j so that j is a multiple of $|Q|!$ and $|p' - p'_j| < \epsilon$.

Next, we compare the behaviour of M_p on $x^i y^j$ and $x^i y^j x^{|Q|!}$. These words correspond to different states in the minimal automaton. Hence, there is a continuation z such that exactly one of $x^i y^j z$ and $x^i y^j x^{|Q|!} z$ is in L.

If M_p had accepted or rejected after $x^i y^j$ (without seeing the right endmarker), it accepts (rejects) both $x^i y^j z$ and $x^i y^j x^{|Q|!} z$. It remains to consider the sequences of probabilistic choices where M_p does not accept until $x^i y^j$.

Let q_x be the state of M_p after reading x^i. We consider three cases:

1. q_x is not a return state for x.

   Then, reading more x's cause M_p to halt with probability 1. However, the probability of halting after reading more than i x's is less than ϵ (by the definition of i). Hence, the probability of this case is less than ϵ.

2. q_x is not a return state for y.

   Then, reading y's cause M_p to halt with probability 1. If it does not happen before reading y^j, it happens later with probability 1. The definition of j implies that, if M does not halt before reading y^j, then the probability of it halting later is less than ϵ. Hence, the probability that q_x is not a return state and M_p does not halt before reading $x^i y^j$ is less than ϵ.

3. q_x is return state for both x and y.

   Then, reading y^j causes M_p to return to q_x because j is a multiple of $|Q|!$ and reading $x^{|Q|}$ causes it to return to q_x as well. In both cases, it is in the same state after reading $x^i y^j$ and after reading $x^i y^j x^{|Q|}$ and, hence, does the same thing on both $x^i y^j z$ and $x^i y^j x^{|Q|!} z$.

We see that the third case causes M_p to react similarly on $x^i y^j z$ and $x^i y^j x^{|Q|!} z$ and the probability of the other two cases together is less than 2ϵ. Hence, the probabilities of accepting these two words differ by less than 2ϵ. However, one of them is in L and must be accepted with probability $1/2 + \epsilon$ and the second is not in L and must be accepted with probability at most $1/2 - \epsilon$. This means that M_p does not recognize L with probability $1/2 + \epsilon$. $\square$

The "forbidden construction" of Theorem 6 is also present in the minimal automaton of $a^* b^*$. Therefore, we have

Corollary 4 *There is a language that can be recognized by a 1-QFA with probability* 0.68... *but cannot be recognized by a 1-PRFA with probability* $1/2 + \epsilon$*, for any* $\epsilon > 0$.

We do not know whether all languages with minimal automata not containing the construction in Theorem 6 can be recognized by 1-way PRFAs. Another open question is characterizating the languages recognized by 1-way QFAs in terms of "forbidden constructions".

4 Complexity

4.1 Divisibility by a prime

All previous work on 1-way QFAs ([12, 13] and the previous sections of this paper) considers the question what languages can be recognized by quantum automata. However, there is another interesting and important question: how efficient are QFAs compared to their classical counterparts?

For 1-way finite automata, the most natural complexity measure is the number of states in the automaton. We can follow the proof in [12] that any language recognized by a 1-way QFA is regular step by step and add complexity bounds to it. Then, we get

Theorem 7 *Let L be a language recognized by a 1-way QFA with n states. Then it can be recognized by a 1-way deterministic automaton with $2^{O(n)}$ states.*

So, transforming a QFA into a classical automaton can cause an exponential increase in its size. Our next results show that, indeed, 1-way QFAs can be exponentially smaller than their classical counterparts.

Let p be a prime. We consider the language $L_p = \{a^i | i \text{ is divisible by } p\}$. It is easy to see that any deterministic 1-way finite automaton recognizing L_p has at least p states. However, there is a much more efficient QFA!

Theorem 8 *For any $\epsilon > 0$, there is a QFA with $O(\log p)$ states recognizing L_p with probability $1 - \epsilon$.*

Proof. First, we construct an automaton accepting all words in L with probability 1 and accepting words not in L with probability at most 7/8. Later, we will show how to increase the probability of correct answer to $1 - \epsilon$ for an arbitrary constant $\epsilon > 0$.

Let U_k, for $k \in \{1, \ldots, p-1\}$ be a quantum automaton with a set of states $|Q| = \{q_0, q_1, q_{acc}, q_{rej}\}$, a starting state $|q_0\rangle$, $Q_{acc} = \{q_{acc}\}$, $Q_{rej} = \{q_{rej}\}$. The transition function is defined as follows. Reading a maps $|q_0\rangle$ to $\cos\phi|q_0\rangle + i\sin\phi|q_1\rangle$ and $|q_1\rangle$ to $i\sin\phi|q_0\rangle + \cos\phi|q_1\rangle$ where $\phi = \frac{2\pi k}{p}$. (It is easy to check that this transformation is unitary.) Reading the right endmarker \$ maps $|q_0\rangle$ to $|q_{acc}\rangle$ and $|q_1\rangle$ to $|q_{rej}\rangle$.

Lemma 7 *After reading a^j, the state of U_k is*

$$\cos\left(\frac{2\pi jk}{p}\right)|q_0\rangle + i\sin\left(\frac{2\pi jk}{p}\right)|q_1\rangle.$$

Proof. By induction. □

If j is divisible by p, then $\frac{2\pi jk}{p}$ is a multiple of 2π, $\cos(\frac{2\pi jk}{p})=1$, $\sin(\frac{2\pi jk}{p})=0$, reading a^j maps the starting state $|q_0\rangle$ to $|q_0\rangle$ and the right endmarker \$ maps it to $|q_{acc}\rangle$. Therefore, all automata U_k accept words in L with probability 1.

For a word $a^j \notin L$, call U_k "good" if U_k rejects a^j with probability at least 1/2.

Lemma 8 *For any $a^j \notin L$, at least $(p-1)/2$ of all U_k are "good".*

Proof. The superposition of U_k after reading a^j is $\cos(\frac{2\pi jk}{p})|q_0\rangle + \sin(\frac{2\pi jk}{p})|q_1\rangle$. This is mapped to $\cos(\frac{2\pi jk}{p})|q_{acc}\rangle + \sin(\frac{2\pi jk}{p})|q_{rej}\rangle$ by the right endmarker. Therefore, the probability of U_k accepting a^i is $\cos^2(\frac{2\pi jk}{p})$. $\cos^2(\frac{2\pi jk}{p}) \leq 1/2$ if and only if $|\cos(\frac{2\pi jk}{p})| \leq 1/\sqrt{2}$. This happens if and only if $\frac{2\pi jk}{p}$ is in $[2\pi l+\pi/4, 2\pi l+3\pi/4]$ or in $[2\pi l+5\pi/4, 2\pi l+7\pi/4]$ for some $l \in \mathbb{N}$.

$\frac{2\pi(jk \bmod p)}{p} \in [\pi/4, 3\pi/4]$ if and only if $\frac{2\pi jk}{p} \in [2\pi l+\pi/4, 2\pi l+3\pi/4]$ for some l. p is a prime and j is relatively prime with p. Therefore, $j \bmod p$, $2j \bmod p$, …, $(p-1)j \bmod p$ are just 1, 2, …, $p-1$ in different order. Hence, it is enough to count k such that $\frac{2\pi k}{p} \in [\pi/4, 3\pi/4]$ or $\frac{2\pi k}{p} \in [5\pi/4, 7\pi/4]$.

We do the counting for $p=8m+1$. (Other cases are similar.) Then $\frac{2\pi k}{p} \in [\pi/4, 3\pi/4]$ if and only if $m+1 \leq k \leq 3m$ and $\frac{2\pi k}{p} \in [5\pi/4, 7\pi/4]$ if and only if $5m+1 \leq k \leq 7m$. Together, this gives us $4m=(p-1)/2$ "good" k's. □

Next, we consider sequences of $\lceil 8\ln p \rceil$ k's. A sequence is *good* for a^j if at least 1/4 of all its elements are good for a^j.

Lemma 9 *There is a sequence of length $\lceil 8\ln p \rceil$ which is good for all $a^j \notin L$.*

Proof. First, we show that at most $1/p$ fraction of all sequences is not good for any fixed $a^j \notin L$.

We select a sequence randomly by selecting each of its elements uniformly at random from $\{1, \ldots, p-1\}$. The probability of selecting a good k in each step is at least 1/2. By Chernoff bounds, the probability that less than $1/4 = 1/2 - 1/4$ fraction of all elements is good is at most

$$e^{-2(1/4)^2 8\ln p} = \frac{1}{p}.$$

Hence, the fraction of sequences which are bad for at least one $j \in \{1, 2, \ldots, p-1\}$ is at most $(p-1)/p$ and there is a sequence which is good for all $j \in \{1, \ldots, p-1\}$. This sequence is good for $a^j \notin L$ with $j > p$ as well because any U_k returns to the starting state after reading a^p and, hence, works in the same way on a^j and $a^{j \bmod p}$. □

Next, we use a good sequence $k_1, \ldots, k_{\lceil 8\ln p \rceil}$ to construct a quantum automaton recognizing L_p. The automaton consists of U_{k_1}, U_{k_2}, …, $U_{k_{\lceil 8\ln p \rceil}}$ and a distinguished starting state. Upon reading the left endmarker ¢, it passes from the starting state to a superposition where $|q_0\rangle$ states of all U_{k_l} have equal amplitudes.

Words in L are always accepted because all U_k accept them. For any $a^j \notin L$, at least $1/4$ of the sequence is good. This means that at least $1/4$ of all U_{k_l} reject it with probability at least $1/2$ and the total probability of rejecting any $a^j \notin L$ is at least $1/8$.

Finally, we sketch how to increase the probability of correct answer to $1-\epsilon$ for an arbitrary $\epsilon > 0$. We do it by increasing the probability of correct answer for each U_k.

Namely, we consider an automaton U'_k with 2^d non-halting states where d is a constant depending on the required probability $1-\epsilon$. The states are labelled by strings of 0s and 1s of length d: $q_{0\ldots 00}, q_{0\ldots 01}$ and so on. The starting state is the state $q_{0\ldots 00}$ corresponding to the all-0 string. The transition function is defined by

$$\delta(q_{x_1\ldots x_d}, a, q_{y_1\ldots y_d}) = \prod_{j=1}^{d} \delta(q_{x_1}, a, q_{y_1}).$$

It is easy to see that this is just the tensor product of d copies of U_k. The automaton also has one accepting state and $2^d - 1$ rejecting states. After reading the right endmarker, the automaton passes to the accepting state from $q_{0\ldots 00}$ and to a rejecting state from any other state $q_{x_1\ldots x_d}$. (To ensure unitarity, one-to-one correspondence between $q_{x_1\ldots x_d}$ and rejecting states is established.) A counterpart of Lemma 7 is

Lemma 10 *The state of U'_k after reading a^j is*

$$\underbrace{\left(\cos\left(\frac{2\pi jk}{p}\right)|q_0\rangle + i\sin\left(\frac{2\pi jk}{p}\right)|q_1\rangle\right) \otimes \ldots \otimes \left(\cos\left(\frac{2\pi jk}{p}\right)|q_0\rangle + i\sin\left(\frac{2\pi jk}{p}\right)|q_1\rangle\right)}_{d\ times}.$$

The amplitude of $|q_0\rangle \otimes \ldots \otimes |q_0\rangle = q_{0\ldots 0}$ in this superposition is $\cos^d(\frac{2\pi jk}{p})$. If j is a multiple of p, then this is 1, meaning that words in L_p are always accepted. For $a^j \notin L_p$, we call U'_k δ-good if it rejects a^j with probability at least $1-\delta$. We formulate a counterpart of Lemma 8.

Lemma 11 *For a suitable constant d, at least $1-\delta$ of all U'_k are δ-good.*

Then, we define a δ-good sequence of automata as a sequence such that, for any $a^j \notin L$, at least $1-2\delta$ of all automata in the sequence are δ-good. Similarly to Lemma 9, we show that there is a δ-good sequence U'_{k_1}, U'_{k_2}, $\ldots$ of length $O(\log n)$. Then, we consider an automaton consisting of U'_{k_1}, U'_{k_2}, $\ldots$ and a distinguished starting state. Upon reading the left endmarker $\not{c}$, it passes from the starting state to a superposition where $|q_0\rangle$ states of all U'_{k_l} have equal amplitudes. Again, it accepts $a^j \in L_p$ with probability 1 because all U'_k accept $a^j \in L_p$. Words $a^j \notin L_p$ are rejected by at least $1-2\delta$ of U'_{k_l} with probability $1-\delta$. Therefore, the probability of rejecting any $a^j \notin L_p$ is at least $(1-2\delta)(1-\delta) > 1-3\delta$. Taking $\delta = \epsilon/3$ and choosing d so that it satisfies Lemma 11 completes the proof. $\square$

We have shown an exponential gap between deterministic and quantum 1-way finite automata. Next, we compare quantum and probabilistic finite automata. Generally, probabilistic finite automata can recognize some languages with the number of states being close to the logarithm of the number of states needed by a deterministic automaton[1, 7]. However, this is not the case with L_p. Here, adding probabilism does not help to decrease the number of states at all.

Theorem 9 *Any 1-way probabilistic finite automaton recognizing L_p with probability $1/2+\epsilon$, for a fixed $\epsilon > 0$, has at least p states.*

Proof. Assume that there is a 1-way probabilistic finite automaton with less than p states recognizing L_p with probability $\frac{1}{2}+\epsilon$, for a fixed $\epsilon > 0$. Since the language L_p is in a single-letter alphabet, the automaton can be described as a Markov chain. We use the classification of Markov chains described in Section 2 of [10] . According to this classification, the states of the Markov chain (the automaton) are divided into ergodic and transient states. An ergodic set of states is a set which cannot be left once it is entered. A transient set of states is a set in which every state can be reached from every other state, and which can be left. An ergodic state is an element of an ergodic set. A transient state is an element of a transient set.

If a Markov chain has more than one ergodic set, then there is absolutely no interaction between these sets. Hence we have two or more unrelated Markov chains lumped together. These chains may be studied seperately. If a Markov chain consists of a single ergodic set, the chain is called an ergodic chain. According to results in Section 2 of [10] , every ergodic chain is either regular or cyclic.

If a Markov chain is regular, then sufficiently high powers of the state transition matrix P of the Markov chain are with all positive elements. Thus no matter where the process starts, after sufficient lapse of time it can be in any state. Moreover, by Theorem 4.2.1 of [10] there is limiting vector of probabilities of being in the states of the chain, not dependent of the initial state.

If a Markov chain is cyclic, then the chain has a period d, and its states are subdivided into d cyclic sets ($d > 1$). For a given starting position, it moves through the cyclic sets in a definite order, returning to the set of the starting state after d steps. Hence the d-th power of the state transition matrix P describes a regular Markov chain.

We have assumed that p is prime, and the automaton has less than p states. Hence for every cyclic state of the automaton the value of d is strictly less than p, and because of primality of p, d is relatively prime to p. By D we denote the least common multiple of all such values d. Hence D is relatively prime to p, and so is any positive degree D^n of D. Since $1^{D^n} \notin M_p$ but $1^{D^n p} \in M_p$, the total of the probabilities to be in an accepting state exceeds $\frac{1}{2}+\epsilon$ for 1^{D^n} and is less than $\frac{1}{2}-\epsilon$ for $1^{D^n p}$. Contradiction with Theorem 4.2.1 of [10]. □

Corollary 5 *For the language L_p, the number of states needed by a classical (deterministic or probabilistic) 1-way automaton is exponential in the number of states of a 1-way QFA.*

Proof: Follows from Theorems 8 and 9. □

4.2 Equality

Divisibility by a prime is quite natural problem and we expect that our algorithm can be used as a subroutine, making other quantum algorithms more space-efficient. Here, we show how to use our quantum automaton for another problem as well. This problem is checking whether the length of the input word is equal to some constant n.

Theorem 10 *[7] Let L'_n be a language consisting of one word a^n in a single-letter alphabet.*

1. *Any deterministic automaton that recognizes L'_n has at least n states.*
2. *For any $\epsilon > 0$, there is a probabilistic automaton with $O(\log^2 n)$ states recognizing L'_n with probability $1 - \epsilon$.*

The first part is evident. To prove the second part, Freivalds[7] used the following construction. $O(\frac{\log n}{\log\log n})$ different primes are employed and $O(\log n)$ states are used for every employed prime. At first, the automaton randomly chooses a prime p, and then the remainder modulo p of the length of the input word is found and compared with the standard. Additionally, once in every p steps a transition to a rejecting state is made with a "small" probability $\frac{const\, p}{n}$. The number of used primes suffices to assert that, for every input of length less than n, most of primes p give remainders different from the remainder of n modulo p. The "small" probability is chosen to have the rejection probability high enough for every input length N such that both $N \neq n$ and an ϵ-fraction of all the primes used have the same remainders $mod\, p$ as n.

This 1-way probabilistic automaton is reversible according to the definition of section 2. We can use Theorem 1 to transform it into quantum automaton with the number of states increasing at most twice. Then, we obtain a counterpart of Theorem 10 for quantum case.

However, we can do better by counting modulo prime as in Theorem 8. For that, we need $O(\log p)$ states for each prime p (instead of p states in the probabilistic case). Each prime p is $O(\log n)$ and there are $O(\log n/\log\log n)$ of them. Therefore, the number of states in quantum case will be

$$O(\frac{\log n}{\log\log n}\log p) = O(\frac{\log n}{\log\log n}\log\log n) = O(\log n).$$

We have shown

Theorem 11 *L'_n can be recognized by a 1-way QFA with $O(\log n)$ states.*

Again, the QFA is exponentially smaller than the corresponding deterministic automaton.

4.3 Are QFAs always space-efficient?

Subsections 4.1 and 4.2 showed cases when 1-way QFAs are more space-efficient than their classical counterparts. There can be examples of different kind where deterministic finite automata are exponentially smaller than 1-way QFAs. The construction of theorem 3 which transforms the minimal automaton into a 1-way RFA can increase the size of the automaton exponentially. The next theorem shows that this is inevitable.

Theorem 12 *Let* $L_m = (xy|zy)^m \cup \{(xy|zy)^i xx | 0 \leq i \leq m-1\}$. *Then,*

1. L_m *can be recognized by a 1-way deterministic finite automaton with* $3m+2$ *states;*
2. L_m *can be recognized by a 1-way reversible automaton but it requires at least* $3(2^m - 1)$ *states.*

After first version of this paper appeared, Ambainis, Nayak and Vazirani[2] showed that, for a different language, the number of states needed by a 1-way QFA is almost exponentially bigger than the number of states of a 1-way deterministic finite automaton.

5 Modifications of 2-way QFAs

The advantage of 1-way quantum automata is the simplicity of this model. However, we saw that 1-way automata are quite limited in several situations (despite being good in others) while [12] shows that 2-way QFAs are strictly more powerful than classical finite automata. It would be interesting to come up with a model having both advantages, i.e. being both powerful and simple. In the remainder, we propose several modifications of quantum automata which are intermediate between 1-way QFAs and 2-way QFAs. Quantum part is kept finite in all of these models. Questions about exact power of these models are mostly open but we have shown that, in most of these models, all regular languages can be recognized and, in at least one of them, non-regular languages can be recognized as well.

5.1 Scanning the tape multiple times

The simplest modification is to allow a 1-way QFA to scan its input tape several times (after the right endmarker it goes to the left endmarker and so on). This is enough to make the proof from [12] that 1-way QFAs recognize only regular languages fail. If we allow the automaton to reject words by non-halting, a nonregular language can be recognized.

Theorem 13 *Let* $L = \{a^n b^n | n \in \mathbb{N}\}$. *There is a 1-way QFA* M *scanning tape several times such that*

1. *If* $x \notin L$, M *stops with probability 1 after* $O(|x|)$ *scans of the tape.*

2. If $x \in L$, M never stops.

If we require M to stop in a rejecting state for rejection, a similar question is still open. It is also open whether multiple scans of the tape can be used by a 1-way QFA to recognize an arbitrary regular language. (However, known proofs that 1-way QFAs do not recognize some regular languages also fail in this case.)

5.2 Passing information back to environment

Another possibility is introducing more complicated observables. We can partition all non-halting states into 2 or 3 classes: moving-left states, moving-right states and (may be) non-moving states. Then, after each step we observe whether the automaton is in accepting, rejecting, moving-right or moving-left state. If it is in a halting state, we terminate the computation. If it is in a moving-right state, we feed it the next letter (do the transformation on the quantum system corresponding to the next letter). If it is in a moving-left state, we feed it the previous letter.

The model of section 5.1 is a special case of this model where all non-halting states are classified as moving-right states.

5.3 Preprocessing the input word

In this model, we have two automata M_1 and M_2 instead of one. M_1 is a 2-way deterministic (or probabilistic) finite automaton with output and M_2 is a 1-way QFA. The input word is given to M_1 and M_2 is run on the output of M_1. (This can be viewed as M_1 preprocessing the input word.) Again, the model of section 5.1 can be viewed as a special case of this model where M_1 moves from left to right all the time and outputs all letters that it reads.

Any regular language can be recognized in a trivial way because we can recognize it by M_1 and give the result as an input to M_2. If the preprocessing is done by a probabilistic automaton, we can do more.

Theorem 14 *For any $\epsilon > 0$, there is a 2-way probabilistic finite automaton M_1 and a 1-way QFA M_2 such that, with probability at least $1 - \epsilon$,*

1. *M_1 stops in time quadratic in the length of the input and,*
2. *M_2 accepts the output of M_1 if and only if $x \in \{a^n b^n | n \in \mathbb{N}\}$.*

Any 2-way probabilistic automaton that recognizes a non-regular language has an exponential expected running time[5, 6, 8, 11]. So, neither polynomial time 2-way probabilistic finite automata nor 1-way QFAs can recognize non-regular languages. However, their combination can do that!

Acknowledgments. We thank Isaac Chuang, Ashwin Nayak, Alistair Sinclair, Amnon Ta-Shma and Umesh Vazirani for useful discussions.

References

[1] A. Ambainis, The complexity of probabilistic versus deterministic finite automata. *Proceedings of the International Symposium on Algorithms and Computation(ISAAC'96)*, Lecture Notes in Computer Science, 1178:233-239, 1996.

[2] A. Ambainis, A. Nayak, U. Vazirani, *On the space-efficiency of 1-way quantum finite automata*, LANL e-print quant-ph/9804043.

[3] D. Angluin, Inference of reversible languages. *Journal of the ACM*, 29:741-765, 1982.

[4] E. Bernstein, U. Vazirani, Quantum complexity theory. *SIAM Journal on Computing,* 26:1411-1473, 1997.

[5] C.Dwork and L.Stockmeyer. On the power of 2-way probabilistic finite state automata. *Proceedings of the 30th IEEE Conference on Foundations of Computer Science(FOCS)*, 1989, pp. 480-485.

[6] R.Freivalds. Probabilistic two-way machines. *Lecture Notes in Computer Science*, 188:33-45, 1981

[7] R. Freivalds, On the growth of the number of states in result of determinization of probabilistic finite automata. *Automatic Control and Computer Sciences*, 1982, no. 3, 39-42

[8] A.G.Greenberg and A.Weiss. A lower bound for probabilistic algorithms for finite state machines. *Journal of Computer and System Sciences*, 33:88-105, 1986.

[9] J. Hopcroft, J. Ullman, *An Introduction to Automata Theory, Languages and Computation.* Addison-Wesley, 1979.

[10] J. Kemeny, J. L. Snell. *Finite Markov chains.* Van Nostrand, Princeton, N.J. , 1960

[11] J. Kaņeps and R. Freivalds. Running time to recognize nonregular languages by 2-way probabilistic automata. Lecture Notes in Computer Science, 510:174-185, 1991.

[12] A. Kondacs, J. Watrous, On the power of quantum finite state automata. *Proceedings of the 38th IEEE Conference on Foundations of Computer Science,* 66-75, 1997.

[13] C. Moore, J. Crutchfield. Quantum automata and quantum grammars. Santa-Fe Institute Working Paper 97-07-062, 1997. Also quant-ph/9707031.

[14] J. Pin. On reversible automata. *Proceedings of Latin American Symposium on Theoretical Informatics(LATIN'92)*, Lecture Notes in Computer Science, 583:401-415, 1992.