Theory Comput. Systems **39**, 165–188 (2006)
DOI: 10.1007/s00224-005-1263-x

Theory of Computing Systems

Algebraic Results on Quantum Automata*

Andris Ambainis,[1] Martin Beaudry,[2] Marats Golovkins,[3]
Arnolds Ķikusts,[4] Mark Mercer,[5] and Denis Thérien[5]

[1]Department of Combinatorics and Optimization
and
Institute for Quantum Computing, University of Waterloo,
200 University Avenue West, Waterloo, Ontario, Canada N2L 2T2
ambainis@math.uwaterloo.ca

[2]Département d'Informatique 2500, Boul. Université,
Sherbrooke, Quebec, Canada J1K 2R1
martin.beaudry@usherbrooke.ca

[3]Computer Science Division, University of California,
Berkeley, CA 94720, USA
marats@cs.berkeley.edu

[4]Institute of Mathematics and Computer Science, University of Latvia,
Raiņa bulv. 29, Riga, Latvia
arnolds_k@one.lv

[5]School of Computer Science, McGill University,
3480 rue University, Montréal, Quebec, Canada H3A 2A7
{jmerce1,denis}@cs.mcgill.ca

Abstract. We use tools from the algebraic theory of automata to investigate the class of languages recognized by two models of Quantum Finite Automata (QFA): Brodsky and Pippenger's end-decisive model (which we call BPQFA) and a new QFA model (which we call LQFA) whose definition is motivated by implementations of quantum computers using nucleo-magnetic resonance (NMR). In particular, we are interested in LQFA since NMR was used to construct the most powerful physical quantum machine to date. We give a complete characterization of the languages recognized by LQFA and by Boolean combinations of BPQFA. It is a surprising consequence of our results that LQFA and Boolean combinations of BPQFA are equivalent in language recognition power.

* The research of A.A. was supported by NSF Grant DMS-0111298. Most of this work was done while he was at the University of Latvia. The research of M.B., M.M., and D.T. was supported by NSERC and FQRNT. The research of M.G. and A.K. was supported by Grant No. 01.0354 from the Latvian Council of Science; European Commission, Contract IST-1999-11234; and the University of Latvia, Kristaps Morbergs fellowship.

1. Introduction

In the classical theory of finite automata, it is unanimously recognized that the algebraic point of view is an essential ingredient in understanding and classifying computations that can be realized by finite state machines, i.e., the regular languages. It is well known that there is a canonical finite monoid associated with every regular language L (namely its syntactic monoid $M(L)$) and unsurprisingly the algebraic structure of this monoid strongly characterizes the combinatorial properties of L. The theory of pseudo-varieties of Eilenberg (which in this paper are called **M**-varieties for short) provides an elegant abstract framework in which these correspondences between monoids and languages can be uniformly discussed.

Finite automata are a natural model for classical computing with finite memory, and likewise *quantum finite automata* (QFA) are a natural model for quantum computers that use a finite-dimensional state space as memory. The more general model of *quantum circuits* [16] gives us an upper bound on the capability of quantum machines, but the fact that several years have passed without the construction of such circuits (despite the efforts of many scientists) suggests that the first quantum machines are not going to be this strong. Thus it is not only interesting but practical to study simpler models alongside the more general quantum circuit model.

There are several models of QFA [14], [12], [7], [4], [8], [6] which differ in which quantum measurements are allowed. Independently, Ciamarra [8] and Bertoni et al. [6] showed that QFAs can recognize all regular languages if they are permitted to make unrestricted transformations and measurements. In contrast, if only one measurement is allowed at the end, the power of QFAs is then equal to that of permutation automata [14], [7] (i.e., they recognize exactly those languages whose syntactic monoid is a group). In intermediate models [12], [7], [4], more than one measurement is allowed but the form of those measurements is restricted. In this case the language recognition power of QFAs lies between those in [14] and those in [8] and [6], but has not been characterized exactly, despite considerable effort [3], [2]. The most general definition of QFAs describes what is achievable in principle according to laws of quantum mechanics while some of the more restricted definitions correspond to what is actually achieved by current implementations of quantum computers.

In view of the enduring success of the algebraic approach to analyze classical finite state devices, it is natural to ask whether the framework can be used in the quantum context as well. The work that we present here answers the question in the affirmative. We analyze two types of QFA: one introduced in [7] (which we call BPQFA) and a new type of QFA (which we call LQFA) whose definition is motivated by the properties of nucleo-magnetic resonance (NMR) quantum computing. Among various physical systems used to implement quantum computing, liquid state NMR has been the most successful so far, realizing physical implementation of a quantum computer with 7 qubits [20]. Liquid state NMR imposes restrictions on what measurements can be performed, and the definition of LQFA reflects this. In both cases we are able to provide an algebraic characterization for the languages that these models can recognize. It turns out that the class of languages recognized by these two models coincides almost exactly (that is, up to Boolean combinations), which is quite surprising considering the differences between the two definitions (for example, the latter allows mixed states while the former does not). It is a pleasant fact that the **M**-variety that turns up in analyzing these QFAs is a natural one that has been extensively studied by algebraists.

In addition to algebra, our arguments are also based on providing new constructions to enlarge the class of languages previously known to be recognizable in these models, as well as proving new impossibility results using subspace techniques (as developed in [3]), information theory (as developed in [15]), and quantum Markov chains (as developed in [1]). In particular, we show that BPQFA cannot recognize the language $a\Sigma^*$ (it is already known [12] that Σ^*a is not recognizable), and that LQFA cannot recognize $a\Sigma^*$ or Σ^*a.

The paper is organized as follows. In Section 2 we give an introduction to the algebraic theory of automata and we give all of the necessary QFA definitions. In the next two sections we give our results on the two models we introduced, and in the last section we outline some open problems.

2. Preliminaries

2.1. *Algebraic Theory of Automata*

A language $L \subseteq \Sigma^*$ is said to be *recognized* [17] by the monoid M if there exists a homomorphism $\varphi\colon \Sigma^* \to M$ and a set $F \subseteq M$ such that $\varphi^{-1}(F) = L$. It can be easily shown that a language is recognized by some finite monoid if and only if it is regular. Given a regular language L, we can construct a canonical finite monoid $M(L)$ recognizing L, which is called the *syntactic monoid*. Let $\sim_L$ be the congruence on Σ^* defined by $v \sim_L w$ if, for all x, y, $xvy \in L \Leftrightarrow xwy \in L$. Then $M(L)$ is the quotient set induced by $\sim_L$.

A monoid M divides N (we write $M \leq N$) if M is a morphic image of a submonoid of N. The division relation is transitive. If M recognizes L and $M \leq N$, it follows that L is also recognized by N. The syntactic monoid is the smallest monoid recognizing L with respect to the division relation.

An **M**-variety is a class of finite monoids which is closed under taking submonoids, surjective homomorphisms, and direct products. Given an **M**-variety **V**, with each finite alphabet Σ we associate the class of regular languages $\mathcal{V}(\Sigma^*) = \{L \subseteq \Sigma^*\colon M(L) \in \mathbf{V}\}$. It can be shown that $\mathcal{V}(\Sigma^*)$ is a Boolean algebra closed under quotients (i.e., if $L \in \mathcal{V}(\Sigma^*)$ then for all $w \in \Sigma^*$ we have $w^{-1}L = \{x\colon wx \in L\} \in \mathcal{V}(\Sigma^*)$ and $Lw^{-1} = \{x\colon xw \in L\} \in \mathcal{V}(\Sigma^*)$) and inverse homomorphisms (i.e., if $\varphi\colon \Sigma^* \to \Sigma^*$ is a homomorphism and $L \in \mathcal{V}(\Sigma^*)$, then $\varphi^{-1}(L) \in \mathcal{V}(\Sigma^*)$). Any class of languages satisfying these closure properties is called a $*$-variety of languages. A theorem of Eilenberg [9] says that there is a 1–1 correspondence between **M**-varieties and $*$-varieties of languages: a driving theme of the research in automata theory has been to find explicit instantiations of this abstract correspondence.

The **M**-variety that plays the key role in our work is the so-called block groups [18], classically denoted **BG**. This variety is ubiquitous: it appears in topological analysis of languages [18], in questions arising in the study of nonassociative algebras [5], and in constraint satisfaction problems [11]. It can be defined by the following algebraic condition: M is a block group iff for any $e = e^2$ and $f = f^2$ in M, $eM = fM$ or $Me = Mf$ implies $e = f$. For any language L, $M(L)$ is a block group iff L is a Boolean combination of languages of the form $L_0 a_1 L_1 \cdots a_k L_k$, where each $a_i \in \Sigma$ and each L_i is a language that can be recognized by a finite group: this class of languages is the largest $*$-variety that does not contain $a\Sigma^*$ or $\Sigma^* a$ for arbitrary alphabet satisfying $|\Sigma| \geq 2$ [18].

2.2. *Models*

We adopt the following conventions. Unless otherwise stated, for any machine M where these symbols are defined, Q is a finite set of states with $|Q| = n$, Σ is the input alphabet, q_0 is the initial state, and $Q_{\text{acc}} \subseteq Q$ ($Q_{\text{rej}} \subseteq Q$) are accepting (rejecting) states. If Q_{acc} and Q_{rej} are defined then we require $Q_{\text{acc}} \cap Q_{\text{rej}} = \emptyset$. Also, each model in this paper uses distinct start and endmarkers, ¢ and \$, respectively. On input w, M processes the characters of ¢w\$ from left to right.

A *superposition* over a finite set Q is a mapping $\psi\colon Q \to \mathbb{C}^n$ that satisfies $\|\psi\|_2 = \sqrt{\sum_q \psi(q)^2} = 1$. We say $\psi(q)$ is the *amplitude* with which ψ is in q. Superpositions can be expressed mathematically as vectors in $\mathbb{C}^n$. For each $q \in Q$ we uniquely associate an element of the canonical basis of $\mathbb{C}^n$, and we denote this element $|q\rangle$. Now the superposition can be written as the vector $\sum_q \psi(q)|q\rangle$.

For all QFAs in this paper, the *state of the machine M* at any given time is a superposition over Q. Note that this is different from the notion of state for DFAs, so some care is required to avoid confusion.

In the case of an LQFA, the state of the machine after reading some prefix is, in general, a random variable. In other words, the state is taken from a probability distribution of superpositions $\{(p_i, \psi_i)\}$, each ψ_i with probability p_i. In this case we say the system is in a *mixed state*. Mixed states can be expressed in terms of *density matrices* [16], and these are usually denoted ρ. If the distribution is trivial, we say that the machine is in a *pure state*. A more detailed discussion of mixed states is given in Section 3.2.1.

A quantum *transformation* is a linear unitary transformation. We say that $A \in \mathbb{C}^{n \times n}$ is *unitary* if $A^* = A^{-1}$, where A^* is the Hermitian conjugate of A and is obtained by taking the conjugate of every element in A^T. Unitary transformations are length preserving, and they are closed under product.

Let $E_1 \oplus \cdots \oplus E_j$ be a partition of $\mathbb{C}^n$ into orthogonal subspaces. A *projective measurement* of a superposition ψ with respect to $E_1 \oplus \cdots \oplus E_j$ has the effect of probabilistically projecting (or *collapsing*) ψ into exactly one E_i, according to the distribution outlined below. For all i, let P_i be the projection operator for E_i. Then the probability of projecting into E_i while measuring with respect to $E_1 \oplus \cdots \oplus E_j$ is $\|P_i \psi\|_2^2$. When such a measurement is made on a quantum state, the index of the projection is communicated to the observer. Measurements are the only way in which an observer can obtain a priori information about a quantum state, thus the output of QFAs are based on the outcome of some measurement.

We consider two modes of acceptance. For a probabilistic machine M, we say M recognizes L with *bounded* (*two-sided*) *error* if M accepts any $w \in L$ and rejects any $w \notin L$ with probability at least $p > \frac{1}{2}$. We say M recognizes L with *bounded positive* *one-sided error* if any $w \in L$ is accepted with probability $p > 0$ and any $w \notin L$ is rejected with probability 1.

We consider three models of QFAs. The first is motivated by liquid state NMR. Liquid state NMR technology has been used to realize physically a 7-qubit quantum computer [20]. NMR uses nuclei of atoms as quantum bits, and the state of the machine is a molecule in which seven different atoms can be individually addressed. One of the features of NMR is that quantum transformations are simultaneously applied to a liquid containing 10^{21} molecules. Thus, we have the same quantum computation carried out by 10^{21} identical quantum computers. Applying a measurement is problematic, however. On different molecules, the measurement can have a different result. We can determine the fraction of molecules that produce each outcome, but we cannot separate the molecules by the measurement outcome. Because of that, the operations performed cannot be conditional on the outcome of a measurement. On the other hand, measurements which do not affect the next transformation are allowed. This situation is reflected in the definition of our new model, given below:

Latvian QFA (*LQFA*). An LQFA is a tuple $M = (Q, \Sigma, \{A_\sigma\}, \{P_\sigma\}, q_0, Q_{\text{acc}})$, where, for each $\sigma \in \Sigma \cup \{¢, \$\}$, A_σ is a unitary matrix and P_σ is a measurement (each P_σ is specified by a partition $E_1 \oplus \cdots \oplus E_j$ of $\mathbb{C}^n$ into orthogonal subspaces). We define $Q_{\text{rej}} = Q \backslash Q_{\text{acc}}$ and we require that $P_\$$ is a measurement with respect to $E_{\text{acc}} \oplus E_{\text{rej}}$, where $E_{\text{acc}} = span\{|q\rangle\colon q \in Q_{\text{acc}}\}$ and $E_{\text{rej}} = span\{|q\rangle\colon q \in Q_{\text{rej}}\}$. Let ψ be the state of M after reading some partial input. On input σ, ψ is transformed by A_σ and then measured with respect to P_σ. Note that the outcome of the measurement is probabilistic, so the state after the measurement is a random variable. At the end of the input, M accepts or rejects according to the outcome of the $P_\$$ measurement. The acceptance mode for LQFA is bounded error.

Also, in [10], a probabilistic automata model related to an LQFA was introduced, which Golovkins and Kravtsev called "1-way probabilistic reversible C-automata" (we abbreviate this to PRA). A PRA is a tuple $M = (Q, \Sigma, \{A_\sigma\}, q_0, Q_{\text{acc}})$, where each A_σ is a *doubly stochastic* matrix. A matrix is doubly stochastic if the sum of the elements in each row and column is 1. The acceptance mode for a PRA is bounded error. The two models are related in the following way: If M is an LQFA such that each P_σ measures with respect to $\bigoplus_{q \in Q} span\{|q\rangle\}$ for every $\sigma \in \Sigma$, then M can be simulated by a PRA. Conversely, a PRA can be simulated by an LQFA if each A_σ of the PRA has a *unitary prototype* [10]. A matrix $U = [u_{ij}]$ is a unitary prototype for $S = [s_{ij}]$ if, for all i, j, $|u_{i,j}|^2 = s_{i,j}$. When S has a unitary prototype it is called unitary stochastic [13]. This relationship between an LQFA and a PRA is helpful in proving that certain languages are recognized by an LQFA.

LQFAs were introduced as QRA-M-C in the classification of QFAs proposed in [10]. A superset of the LQFA model has been studied in [15] and [4].

The first and most studied model of a QFA was defined by Kondacs and Watrous in [12] (we call these KWQFAs). A KWQFA is defined by a tuple $M = (Q, \Sigma, \{A_\sigma\}, q_0, Q_{\text{acc}}, Q_{\text{rej}})$ where each A_σ is unitary. The state sets Q_{acc} and Q_{rej} will be halt/accept and halt/reject states, respectively. We also define $Q_{\text{non}} = Q \backslash (Q_{\text{acc}} \cup Q_{\text{rej}})$ to be the the set of nonhalting states. Lastly, for $\mu \in \{\text{acc, rej, non}\}$ we define $E_\mu = \text{span}\{|q\rangle\colon q \in Q_\mu\}$, and

P_μ to be the projection onto E_μ. Let ψ be the state of M after reading some partial input. On input σ, ψ is transformed by A_σ and the outcome is measured with respect to $E_{\mathrm{acc}} \oplus E_{\mathrm{rej}} \oplus E_{\mathrm{non}}$. If the outcome of the measurement is acc or rej, then M halts and accepts or rejects accordingly. Otherwise, the state becomes $\psi' = P_{\mathrm{non}} A_\sigma \psi / \| P_{\mathrm{non}} A_\sigma \psi \|_2$ and M continues. We require that after reading \$ the state is in E_{non} with probability 0, so that $Pr[M \text{ accepts } w] + Pr[M \text{ rejects } w] = 1$. The acceptance mode for a KWQFA is bounded error.

BPQFAs were introduced in [7] as a natural restriction of the KWQFA model. We define BPQFAs below.

Brodsky–Pippenger QFA (*BPQFA*). A BPQFA M is a KWQFA where M is not permitted to halt in an accepting state until \$ is read, and the acceptance mode is changed to bounded positive one-sided error.

In [7], it was shown that BPQFAs recognize positive Boolean combinations (unions and intersections) of $\Sigma^* a_1 \Sigma^* a_2 \cdots a_k \Sigma^*$. In this paper we generalize these results and give nearly tight upper bounds.

3. Latvian QFA

Our main result for this model is a complete characterization of the languages recognized by LQFAs:

Theorem 1. *LQFAs recognize exactly the class of languages whose syntactic monoid is in* **BG**.

To prove this result, we first show that the class of languages recognized by LQFAs forms a $*$-variety of languages. Then we give tight upper and lower bounds on the languages recognized by LQFAs. Before we begin the proof of this theorem, we establish a few simple properties:

Lemma 1. *Given an LQFA M recognizing L with probability $p > \frac{1}{2}$, we can construct M' recognizing L with probability $1 - \varepsilon$ for any $\varepsilon > 0$.*

Proof. Let $M = (Q, \Sigma, q_0, \{A_\sigma\}, \{P_\sigma\}, Q_{\mathrm{acc}})$. Our boosting strategy is to construct a single machine that simulates m copies of M in parallel using a single machine M' and accepts only if the majority of the copies accept. By a Chernoff argument, we can always find an m that will give the desired probability of acceptance. Let M' have state set Q^m, initial state $(q_0, \ldots, q_0)$, set of transitions $\{\bigotimes_{i=1}^{m} A_\sigma\}_{\sigma \in \Sigma}$, similarly defined measurements, and accepting state set $\{(q_{x_1}, \ldots, q_{x_m}) \colon |\{q_{x_i} \in Q_{\mathrm{acc}}\}| \geq m/2\}$. This machine simulates m trials of M as required. □

Claim 1. *Consider a sequence of l finite transformations and measurements operating on a finite space E. These operations can be simulated by one transformation and one measurement on a* (*possibly larger*) *finite subspace E'.*

Proof. Assume we have a sequence of l unitaries U_i on a space E, each of them followed by a measurement $E_{i1} \oplus \cdots \oplus E_{ik_i}$. Define a new space E' of dimension $(\dim E) \cdot \prod_i k_i$. It is spanned by states $|\psi\rangle|j_1\rangle \cdots |j_l\rangle$, $|\psi\rangle \in E$, $j_i \in \{0, \ldots, (k_i - 1)\}$. Each U_i can be viewed as a transformation on E' that acts only on the $|\psi\rangle$ part of the state. Replace the measurements by unitary transformations V_i defined by

$$V_i|\psi\rangle|j_1\rangle \cdots |j_i\rangle \cdots |j_l\rangle = |\psi\rangle|j_1\rangle \cdots |(j_i + j) \bmod k_i\rangle \cdots |j_l\rangle$$

for $|\psi\rangle \in E_{ij}$. We claim that the unitary operation $V_l U_l \cdots V_1 U_1$, followed by the measurement operation that measures all of $j_1, \ldots, j_l$, will simulate the l transformations and measurements as required. It is sufficient to consider the case where the system is currently in a pure state ψ. Applying the original sequence of operations will produce a mixed state $\{(p_i, \psi_i)\}$, each ψ_i with probability p_i. Now consider our simulation of the original sequence. We start in state $|\psi\rangle|j_1\rangle \cdots |j_l\rangle$ and then move to $\{(p_{j'_1,\ldots,j'_l}, |\psi\rangle|j'_1\rangle \cdots |j'_l\rangle)\}$, where $p_{j'_1,\cdots,j'_l}$ is the probability that, for all i, the ith measurement caused a projection into $E_{i,(j'_i - j_i) \bmod k_i}$. Thus when we restrict our attention to the E part of the state, the behavior of our simulation is actually equivalent to that of the original sequence. □

Now to prove that the languages recognized by LQFAs form a variety, it is sufficient to show:

Theorem 2. *The class of languages recognized by LQFAs is closed under union, complement, inverse homomorphisms, and word quotient.*

Proof. For any LQFA M recognizing L, we can trivially construct $\overline{M}$ recognizing $\overline{L}$ by swapping the accept and reject states. This proves closure under complement.

Next we show closure under union. Let M_1 and M_2 be LQFAs recognizing L_1 and L_2 with probability p_1 and p_2, respectively. Without loss of generality assume $p_1 \geq \frac{3}{4}$ and $p_2 \geq \frac{3}{4}$. Now to compute the union of $L_1 \cup L_2$, construct the tensor product M' of M_1 and M_2 as in Lemma 1 but set $Q'_{\text{acc}} = \{(q_i, q_j)\colon q_i \in Q_{1,acc} \vee q_j \in Q_{2,acc}\}$. It is easy to check that M' recognizes L with probability at least $\frac{9}{16}$.

Now suppose that M recognizes L, and that $h\colon \Sigma^* \to \Sigma^*$ is a homomorphism. Define M' so that, on input $\sigma \in \Sigma$, M' simulates the state transition of M on input $h(\sigma)$. By Claim 1 this simulation can be performed with one unitary transformation and one measurement, so M' is constructible. It is easy to check that M' recognizes $h^{-1}(L)$.

Finally, we prove closure under quotient. Let M be an LQFA recognizing L. We can construct M' recognizing $w^{-1}L$ as follows. Use Claim 1 to simulate $A'_{¢w}$ with one unitary transformation and measurement, and make these the new $A_¢$ and $P_¢$ operation. Make all other operations in M' the same as in M. Clearly M' will recognize $\{x\colon wx \in L\}$. Right quotient is similar. □

3.1. *LQFA Lower Bounds*

We now proceed to show that LQFAs recognize any language whose syntactic monoid is in **BG**. We begin with the following simpler result:

Theorem 3. *LQFAs can recognize languages of the form* $\Sigma^* a_1 \Sigma^* \cdots a_k \Sigma^*$.

Proof. To prove this result, we first give a construction of a PRA that recognizes $\Sigma^* a_1 \Sigma^* \cdots a_k \Sigma^*$ with probability $((n-1)/n)^k$, for any $n \in \mathbb{N}$. Then we show that the transitions of this PRA can be simulated by an LQFA.

We construct our PRA inductively on the length of the subword. For $k = 1$ we construct $M^{(1)} = (Q^{(1)}, q_0, \Sigma, \{A_\sigma^{(1)}\}, Q_{\text{acc}}^{(1)})$ as follows. Let $Q^{(1)} = \{q_0, q_2, \ldots, q_n\}$, $A_{a_1}^{(1)} = (1/n)\mathbf{1}$ (where $\mathbf{1}$ is an $n \times n$ matrix of all ones), $A_\sigma^{(1)} = I$ for all $\sigma \neq a_1$, and $Q_{\text{acc}}^{(1)} = Q^{(1)} \backslash \{q_0\}$. It is easy to check that this machine accepts any $w \in \Sigma^* a_1 \Sigma^*$ with probability $((n-1)/n)$ and rejects any $w \notin \Sigma^* a \Sigma^*$ with probability 1.

Assume we have a machine $M^{(i-1)} = (Q^{(i-1)}, q_0, \Sigma, \{A_\sigma^{(i-1)}\}, Q_{\text{acc}}^{(i-1)})$ recognizing inputs containing the subword $a_1 \cdots a_{i-1}$ with probability $((n-1)/n)^{i-1}$, we construct $M^{(i)} = (Q^{(i)}, q_0, \Sigma, \{A_\sigma^{(i)}\}, Q_{\text{acc}}^{(i)})$ recognizing inputs containing the subword $a_1 \cdots a_i$ with probability $((n-1)/n)^i$. Our augmentation will proceed as follows. First let $Q_{\text{acc}}^{(i)}$ be a set of $(n-1)^i$ new states all distinct from $Q^{(i-1)}$, and let $Q^{(i)} = Q^{(i-1)} \cup Q_{\text{acc}}^{(i)}$. For each $q \in Q_{\text{acc}}^{(i-1)}$ we uniquely associate $n-1$ states $q_2, \ldots, q_n \in Q_{\text{acc}}^{(i)}$. We leave q_0 unchanged.

It remains to define the $A_\sigma^{(i)}$ transitions. Define $\tilde{A}_\sigma^{(i-1)}$ to be the transformation that acts as $A_\sigma^{(i-1)}$ on $Q^{(i-1)} \subset Q^{(i)}$ and as the identity elsewhere. We let $A_\sigma^{(i)} = \tilde{A}_\sigma^{(i-1)} B_\sigma^{(i)}$, where $B_\sigma^{(i)}$ is an additional transformation that will process the a_i character (note that the matrices are applied from right to left). For all $\sigma \neq a_i$ we define $B_\sigma^{(i)} = I$. For $\sigma = a_i$ we define $B_\sigma^{(i)}$ so that, independently for each $q \in Q_{\text{acc}}^{(i-1)}$, the transformation $(1/n)\mathbf{1}$ is applied to $\{q, q_2, q_3, \ldots, q_n\}$. At the end we have a machine $M = M^{(k)}$ that recognizes $\Sigma^* a_1 \Sigma^* \cdots a_k \Sigma^*$.

To simplify notation, we define $Q^{(0)} = Q_{\text{acc}}^{(0)} = \{q_0\}$ and $B_\sigma^{(1)} = A_\sigma^{(1)}$ for all σ. The correctness of the construction follows from this lemma:

Lemma 2. *Let w be any word. As we process w with M, for all $0 \leq i < k$ the total probability of M being in one of the states of $Q^{(i)}$ is nonincreasing.*

Proof. For any $S \subseteq Q$, denote by $P(S)$ the sum probability of being in one of the states of S. Every nontrivial A_σ matrix can be decomposed into a product of $B_{a_i}^{(i)}$ matrices operating on different parts of the state space. All of these matrices operate on the machine state in such a way that for all j and for any $\{q, q'\} \subseteq Q_{\text{acc}}^{(j)}$, at any time there is an equal probability of being in state q or q'. Thus the distribution of the state at any time can be completely specified by $P(Q_{\text{acc}}^{(0)}), \ldots, P(Q_{\text{acc}}^{(k)})$.

For all $0 \leq i < k$ the machine can only move from $Q^{(i)}$ to $Q \backslash Q^{(i)}$ when $B_{a_{i+1}}^{(i+1)}$ is applied, and this matrix has the effect of averaging the likelihood of being in any given state of $Q_{\text{acc}}^{(i)} \cup Q_{\text{acc}}^{(i+1)}$. Since $|Q_{\text{acc}}^{(i+1)}| = (n-1)|Q_{\text{acc}}^{(i)}|$, it follows that a $B_{a_{i+1}}^{(i+1)}$ operation will not increase $P(Q^{(i)})$ unless $P(Q_{\text{acc}}^{(i+1)}) > (n-1)P(Q_{\text{acc}}^{(i)})$. It can easily be shown by induction on the sequence of $B_{a_j}^{(j)}$ matrices forming the transitions of M that this condition is never satisfied. Thus $P(Q^{(i)})$ is nonincreasing for all i. □

We are now ready to prove that M recognizes $L = \Sigma^* a_1 \Sigma^* \cdots a_k \Sigma^*$. First we show that any $w \notin L$ is rejected with certainty. The transitions are constructed in such a way that M can only move from $Q^{(i-1)}$ to $Q^{(i)}$ upon reading a_i, and M cannot move from $Q^{(i-1)}$ to $Q^{(i+1)}$ in one step (even if $a_i = a_{i+1}$). Next we show that any $w \in L$ is accepted with probability $((n-1)/n)^k$. After reading the first a_1, $P(Q_{\text{acc}}^{(1)}) \geq ((n-1)/n)$ and by Lemma 2 this remains satisfied until a_2 is read, at which point M will satisfy $P(Q_{\text{acc}}^{(2)}) \geq ((n-1)/n)^2$. Inductively after reading subword a, M satisfies $P(Q_{\text{acc}}) \geq ((n-1)/n)^k$. Thus M indeed recognizes $\Sigma^* a_1 \Sigma^* \cdots a_k \Sigma^*$.

All that remains is to show that we can simulate each A_σ using LQFA transformations. Recall that each A_σ is a product of $B_{a_i}^{(i)}$ matrices operating on different parts of the state space. If each $B_{a_i}^{(i)}$ has a unitary prototype, then each $B_{a_i}^{(i)}$ can be simulated by a single transformation followed by a single measurement, and each A_σ could be simulated using the series of l transformations and measurements. Thus, by Claim 1, it is sufficient to show that each $B_{a_i}^{(i)}$ has a unitary prototype.

Observe that any block diagonal matrix such that all of the blocks have unitary prototypes is itself a unitary prototype, and that unitary prototypes are trivially closed under permutations. Each $B_{a_i}^{(i)}$ can be written as a block diagonal matrix, where each block is the 1×1 identity matrix or the $(1/n)\mathbf{1}$ matrix, so it remains to show that there is a unitary prototype for $(1/n)\mathbf{1}$ matrices. Coincidentally the quantum Fourier transform matrix [16], which is the basis for most efficient quantum algorithms, is a unitary prototype for $(1/n)\mathbf{1}$. Thus, A_σ can be simulated by an LQFA. This completes the proof of Theorem 3. □

We can generalize Theorem 3 as follows:

Theorem 4. *LQFAs recognize any language whose syntactic monoid is in* **BG**.

Proof. We give a PRA construction recognizing the language L defined by $w \in L$ if and only if $w = w_0 a_1 w_1 \cdots a_k w_k$ where, for each i, $w_0 a_1 w_1 \cdots w_i \in L_i$ for some prespecified group languages $L_0, \ldots, L_k$. By the cancellative law of groups, it is sufficient to show that PRA recognize any language of the form $L_0 a_1 L_1 \cdots a_k L_k$. We will see that each transition matrix has a unitary prototype, thus there is an LQFA recognizing this language as well. This along with the closure properties of an LQFA is sufficient to prove that any language whose syntactic monoid is in **BG** is recognized by an LQFA.

For all i let $G_i = M(L_i)$. Also let $\varphi_i\colon \Sigma^* \to G_i$ and F_i be such that $\varphi_i^{-1}(F_i) = L_i$. We compose these groups into a single group $G = G_0 \times \cdots \times G_k$ with identity $1 = (1, 1, \ldots, 1)$.

Let $M = (Q, q_0, \Sigma, \{A_\sigma\}, Q_{\text{acc}})$ be a PRA recognizing the subword $a_1 \cdots a_k$ constructed as in Theorem 3. From M we construct an LQFA $M' = (Q', q_0', \Sigma, \{A_\sigma'\}, Q_{\text{acc}}')$ recognizing L. We set $Q' = Q \times G$, $q_0' = (q_0, 1)$, $Q_{\text{acc}}' = Q_{\text{acc}} \times (G_1 \times \cdots \times G_{k-1} \times F_k)$, and $A_¢ = A_\$ = I$. For each $\sigma \in \Sigma$ define A_σ' as follows. Let P_σ be the permutation matrix that maps (q, g) to $(q, g\sigma)$ for each $q \in Q$ and $g \in G$. For each $1 \leq i \leq k$ let $A_{\sigma i}'$ be the matrix that, for each $f \in G_1 \times \cdots \times F_{i-1} \times G_i \times \cdots \times G_k$, acts as the transformation $B_\sigma^{(i)}$ on $Q^{(i)} \times \{f\}$ and as the identity everywhere else. Finally, $A_\sigma' = P_\sigma A_{\sigma 1}' \cdots A_{\sigma k}'$.

The A_σ' are constructed so that M' keeps track of the current group element at every step. If M is in state (q, g), then after applying $A_1', \ldots, A_k'$ it remains in $Q \times \{g\}$ with probability 1. The P_σ matrix "translates" all of the transition probabilities from $Q \times \{g\}$ to $Q \times \{g\sigma\}$. Initially M is in $Q \times \{1\}$, so after reading any partial input w, M will be in $Q \times \{1w\}$ with probability 1. In this way M will always keep track of the current group element.

Each A'_σ matrix refines A_σ from the $\Sigma^* a_1 \Sigma^* a_2 \cdots a_k \Sigma^*$ construction in such a way that, on input σ after reading w, we do not move from $Q^{(i-1)}$ to $Q^{(i)}$ (the action performed by $B^{(i)}_{a_i}$) unless $\sigma = a_i$ and $w \in F_{i-1}$. This is exactly what we need to recognize L. The transition matrices can be simulated by LQFAs by the same argument as in Theorem 3.

Lemma 3. *Let w be any word. As we process the characters of w in M, for all $0 \le i < k$ the total probability of being in one of the states of $Q^{(i)} \times G$ is nonincreasing.*

Proof. Same argument as in Lemma 2 holds. □

We claim that the machine M constructed above rejects every $w \notin L = L_0 a_1 L_1 \cdots a_k L_k$ with certainty. The PRA M does not move out of $Q^{(0)} \times G$ unless some prefix $w_0 a_1$ with $w_0 \in L_0$ is read. Inductively, we do not move into Q_{acc} unless we have read each subword letter on the correct context and the current state corresponds to a group element $f \in F_k$.

Now suppose $w \in L$. Rewrite w as $w_0 a_1 \cdots a_k w_k$. Clearly M does not move out of $Q^{(0)} \times G$ while reading w_0. The character a_1 is now read, and M moves to $(Q^{(1)} \times G) \backslash (Q^{(0)} \times G)$ with probability $(n-1)/n$. By the previous lemma, this probability does not decrease while reading w_1. So now after reading $w_0 a_1 w_1$ we will be in $Q^{(1)}_{\text{acc}} \times G$ with probability $(n-1)/n$. If a_2 is read we move to $Q^{(2)}$ with probability $((n-1)/n)^2$. By induction after reading $w_0 a_1 \cdots w_{k-1} a_k$ we move to $(Q^{(k)} \times G) \backslash (Q^{(k-1)} \times G)$ with total probability at least $((n-1)/n)^k$. Finally, after reading w_k we move to Q'_{acc} with total probability at least $((n-1)/n)^k$, and so we accept any $w \in L$ with this probability. By choosing a suitable n we can recognize L with arbitrarily high probability. This completes the proof of Theorem 4. □

3.2. *LQFA Upper Bounds*

Next, to prove that LQFAs cannot recognize any language whose syntactic monoid is not in **BG**, we need to show that LQFAs cannot recognize $\Sigma^* a$ or $a\Sigma^*$. We note that LQFAs are a special case of Nayak's EQFA model [15], and EQFAs cannot recognize $\Sigma^* a$. The proof that LQFAs cannot recognize $a\Sigma^*$ is considerably harder.

Theorem 5. *LQFAs cannot recognize $a\Sigma^*$.*

The focus of the remainder of the section is the proof of this result.

3.2.1. *Mixed States, Density Matrices, and CPSOs.* This section provides definitions and properties needed for the proof of Theorem 5 which we give in the next section. For more information, see [16].

Mixed states: A mixed state $\{(p_i, |\psi_i\rangle)\}$, $0 \le p_i \le 1$, $\sum_i p_i = 1$, is a classical probability distribution over quantum states $|\psi_i\rangle$ (which are called *pure states*). The quantum system described by a mixed state is in the state $|\psi_i\rangle$ with probability p_i.

Density matrices: A density matrix of a pure state $|\psi\rangle$ is $|\psi\rangle\langle\psi|$. A density matrix of a mixed state is $\rho = \sum_i p_i|\psi_i\rangle\langle\psi_i|$. We often identify the mixed state with its density matrix.

Unitary transformations and measurements: Definitions of unitary transformations and measurements extend naturally to mixed states. For example, a unitary transformation U maps a mixed state $\{(p_i, |\psi_i\rangle)\}$ to $\{(p_i, U|\psi_i\rangle)\}$.

This can be described in terms of density matrices. If, before U, the system was in a mixed state with a density matrix ρ, the state after the transformation is the mixed state with the density matrix $U\rho U^\dagger$.

If we measure a state with density matrix ρ with respect to $E_1 \oplus \cdots \oplus E_k$, the result is i with probability $\mathrm{Tr}\, P_i\rho$, Tr being the trace of a matrix (the sum of its diagonal entries). The remaining state is $P_i\rho P_i / \mathrm{Tr}\, P_i\rho$.

Completely positive superoperators: Transformations allowed by quantum mechanics are various combinations of unitary transformations and measurements. Any such transformation E has the following properties:

1. Let ρ be a $d \times d$ density matrix and let $E\rho$ be the density matrix of the state which results if we apply E. The transformation $\rho \to E\rho$ is a linear transformation on the d^2-dimensional space of $d \times d$ matrices.
2. E is trace-preserving: $\mathrm{Tr}\, E\rho = \mathrm{Tr}\, \rho$.
3. E is *completely positive*, i.e., if H is the space on which E operates, then for any additional space H' the transformation $E \otimes I$ is a positive map on $H \otimes H'$.

A transformation satisfying these requirements is called a trace-preserving CPSO (completely positive superoperator). Any trace-preserving CPSO can be constructed from unitary transformations and measurements [16]. Therefore, these properties can be taken as an alternative definition of the transformations permitted by quantum mechanics.

Kraus decomposition: For any trace-preserving CPSO A there exists k matrices $A_1, \ldots, A_k$ such that $\sum_{i=1}^k A_i A_i^\dagger = I$ and $A\rho = \sum_{i=1}^k A_i\rho A_i^\dagger$.

Distance between density matrices: A natural measure of the distance between two density matrices is the *trace distance*. The trace norm $\|A\|_t$ of a matrix A is defined as $\mathrm{Tr}|A|$, where $|A|$ is the positive matrix square root of $AA^\dagger$. The trace distance between ρ_0 and ρ_1 is just the trace norm of $\rho_0 - \rho_1$. We use the following properties of the trace distance:

1. The trace distance describes the distinguishability of quantum states. For any ρ_0 and ρ_1 there is a measurement that, given an unknown $\rho_i \in \{\rho_0, \rho_1\}$, produces i with probability at least $\frac{1}{2} + \|\rho_0 - \rho_1\|_t/4$.
2. The trace distance is nonincreasing. For any CPSO A, we have

$$\|A\rho_0 - A\rho_1\|_t \leq \|\rho_0 - \rho_1\|_t.$$

3.2.2. *Proof of Theorem* 5. We start with a proof outline. During this outline, we state three lemmas (Lemmas 5–7) and prove the theorem, assuming these lemmas. Then we prove the lemmas.

Let E be a sequence $U_1, P_1, U_2, P_2, \ldots, U_l, P_l$, with the U_i's being unitary transformations and the P_i's being measurements (for example, E could be the unitary transformation + measurement corresponding to reading a letter or it could be a sequence of unitaries and measurements corresponding to reading a word). We view E as one operation mapping (mixed) quantum state ρ to (mixed) quantum state $E\rho$. E is a particular case of the CPSOs.

In our case we have an additional constraint on E. Not every CPSO can be represented as a sequence U_1, P_1, U_2, P_2, ..., U_l, P_l. For example, a mapping that replaces any quantum state by a fixed state (say, $|0\rangle$) is a CPSO. However, it cannot be represented as a sequence U_1, P_1, U_2, P_2, ..., U_l, P_l (and is not allowed in NMR implementations of quantum computing as well). This constraint is nicely captured by a quantity called the *Von Neumann entropy*. (The Von Neumann entropy $S(\rho)$ is defined as $-\sum_i \lambda_i \log_2 \lambda_i$, with λ_i being the eigenvalues of ρ. For this proof, three properties of S are sufficient. These properties are given by Lemmas 4, 8, and 9.)

Lemma 4 [4]. *Let E be a sequence U_1, P_1, U_2, P_2, ..., U_l, P_l, with U_i being unitary transformations and P_i being measurements. Then, for any ρ, $S(E\rho) \geq S(\rho)$.*

From this moment, we assume that the transformation corresponding to each letter x is a CPSO E with the property that $S(E\rho) \geq S(\rho)$.

We study the effect of repeatedly applying E to a (mixed) quantum state ρ. We would like to study the sequence ρ, $E\rho$, $E^2\rho$, ... However, this sequence might not converge (for example, if E is a unitary transformation

$$U = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

this sequence is periodic with period 2). To avoid this problem, define E' as an operation consisting of applying E with probability $\frac{1}{2}$ and applying identity otherwise.

Note 1. This is similar to making a periodic Markov chain aperiodic by adding self-loops.

Note 2. A similar periodicity problem comes up in quantum walks [1]. There, it is solved by a different approach (Cesaro limit). We think our approach (introducing E') gives results that are similar to the Cesaro limit. In this paper we choose to introduce E' instead of using the Cesaro limit because this seems to make the analysis of our problem simpler.

Lemma 5.

1. *For any CPSO E such that $S(E\rho) \geq S(\rho)$ and any mixed state ρ, the sequence $E'\rho$, $(E')^2\rho$, ..., $(E')^i\rho$, ... converges.*
2. *Let $E_{\lim}$ be the map $\rho \to \lim_{i\to\infty}(E')^i\rho$. Then $E_{\lim}$ is a CPSO and $S(E_{\lim}\rho) \geq S(\rho)$ for any density matrix ρ.*

Lemma 6. *Let A, B be two sequences of unitary transformations and measurements. Let $C = A_{\lim}B_{\lim}$ and $D = B_{\lim}A_{\lim}$. Then $C_{\lim} = D_{\lim}$.*

Assume that we are given an LQFA M. We show that M does not recognize the language $a\Sigma^*$.

Let A, B be the transformations corresponding to reading letters a, b. We also consider A_{lim}, B_{lim}, $C = A_{\text{lim}}B_{\text{lim}}$, $D = B_{\text{lim}}A_{\text{lim}}$, C_{lim} and D_{lim}.

Intuitively, A_{lim} (B_{lim}) corresponds to reading a long sequence of letters a (b), with the length being a random variable. C_{lim} (D_{lim}) corresponds to a long sequence of a^i and b^j alternating with a^i at the beginning (b^j at the beginning). If M is correct, it must accept if C_{lim} is applied to the starting state and reject if D_{lim} is applied. However, by Lemma 6, $C_{\text{lim}} = D_{\text{lim}}$ which causes a contradiction.

More formally, let ρ_x be the (mixed) state after reading the word x. We consider two sets of mixed states Q_a and Q_b. The set Q_a (Q_b) consists of all probabilistic combinations of states ρ_{ax} (ρ_{bx}). Let $\overline{Q_a}$, $\overline{Q_b}$ be closures of Q_a and Q_b.

Lemma 7. *Let ρ be the state after reading the start marker ¢. Then $C_{\text{lim}}\rho \in \overline{Q_a}$ and $D_{\text{lim}}\rho \in \overline{Q_b}$.*

We consider applying the right endmarker and the final measurement to the state $C_{\text{lim}}\rho = D_{\text{lim}}\rho$. This state belongs to $\overline{Q_a}$. Therefore, it is a limit of a sequence $\rho_1, \rho_2, \ldots$ with each ρ_i being a probabilistic combination of final states of M on words which belong to $a\Sigma^*$. If M accepts $a\Sigma^*$, applying the right endmarker and the final measurement to any such ρ_i must cause acceptance with probability at least p. Therefore, M must accept with probability at least p. On the other hand, since $C_{\text{lim}}\rho = D_{\text{lim}}\rho$ also belongs to $\overline{Q_b}$, M must reject with probability at least p as well. This is a contradiction, proving that M does not recognize $a\Sigma^*$.

To prove the theorem, it remains to prove Lemmas 5–7.

Proof of Lemma 5. Let $H(p) = -p\log_2 p - (1-p)\log_2(1-p)$ be the usual Shannon entropy and let $S(\rho)$ be the Von Neumann entropy of a mixed quantum state ρ.

Lemma 8 [4]. *Let τ_0, τ_1 be two density matrices and let $\tau = \frac{1}{2}\tau_0 + \frac{1}{2}\tau_1$. If there is a measurement that, given τ_i, outputs i correctly with probability at least p, then*

$$S(\tau) \geq \tfrac{1}{2}(S(\tau_0) + S(\tau_1)) + 1 - H(p).$$

Lemma 9 [16, Theorem 11.8]. *For any mixed state ρ of dimension d, $S(\rho) \leq \log_2 d$, with the equality if and only if ρ is a d-dimensional completely mixed state.*

Lemma 10 [16, Theorem 11.6]. *Let τ_0, τ_1 be two density matrices of dimension d and let $\varepsilon = \|\tau_0 - \tau_1\|_t$, $\varepsilon < \frac{1}{3}$. Then*

$$|S(\tau_0) - S(\tau_1)| \leq \varepsilon \log_2 d - \varepsilon \log_2 \varepsilon.$$

Let ρ_0 be the initial state and let $\rho_{i+1} = E'\rho_i$ be the sequence we are studying. Since $S(E\rho_i) \geq S(\rho_i)$, Lemma 8 implies $S(E'\rho_i) \geq S(\rho_i)$. Consider the sequence of numbers $s_i = S(\rho_i)$. This is an nondecreasing sequence and, by Lemma 9, is bounded from above by $\log_2 d$. Therefore, it converges to a value s_{lim}.

Moreover, $\rho_0, \rho_1, \ldots$ is a sequence in a closed bounded subset of a finite-dimensional space (the set of all $d \times d$ density matrices). Therefore, it must have a limit point ρ, i.e., ρ such that, for every $\varepsilon > 0$, there exists i satisfying $\|\rho - \rho_i\|_t \leq \varepsilon$. It follows from the continuity of S that $S(\rho) = s_{\lim}$. We will show that the sequence converges to ρ.

To show that, it suffices to show $E\rho = \rho$. If this is the case, then $E'\rho = \rho$. Therefore,

$$\|\rho_{i+1} - \rho\|_t = \|E'\rho_i - E'\rho\|_t \leq \|\rho_i - \rho\|_t.$$

This means that if $\|\rho_i - \rho\|_t \leq \varepsilon$, then $\|\rho_j - \rho\|_t \leq \varepsilon$ for all $j \geq i$. Therefore, ρ is the limit of ρ_i.

It remains to show $E\rho = \rho$. Assume that this is not true. Then $\|E\rho - \rho\|_t = \delta > 0$. Define $\rho' = \frac{1}{2}(\rho + E\rho)$. Since the trace distance describes distinguishability, Lemma 8 implies $S(\rho') \geq S(\rho) + 1 - H(\frac{1}{2} + \delta/4)$. We now choose $\varepsilon > 0$ so that $\varepsilon \log_2 d - \varepsilon \log_2 \varepsilon < H(\frac{1}{2} + \delta/4)$. Since ρ is a limit point, there exists i such that $\|\rho - \rho_i\|_t \leq \varepsilon$. Then $\|\rho' - \rho_{i+1}\|_t \leq \varepsilon$. By Lemma 10, $S(\rho_{i+1}) \geq S(\rho') - \varepsilon \log_2 d + \varepsilon \log_2 \varepsilon$. This implies $S(\rho_{i+1}) > s_{\lim}$.

However, this is not possible. Let $\delta = S(\rho_{i+1}) - s_{\lim}$ and pick ε so that $\varepsilon \log_2 d - \varepsilon \log_2 \varepsilon < \delta$ Then there exists i such that $\|\rho - \rho_i\| \leq \varepsilon$. We have $S(\rho_i) > S(\rho) - \delta = s_{\lim}$ which contradicts $S(\rho_i)$ being a nondecreasing sequence that converges to $s_{\lim}$. Therefore, it must be the case that $E\rho = \rho$. This completes the proof of the first part of Lemma 5.

To see the second part, notice that the limit of a sequence of linear maps on $d \times d$ matrices is a linear map on $d \times d$ matrices. Furthermore, if each map is trace preserving and positive, the limit is trace preserving and positive. Finally, $S(E_{\lim}\rho) = s_{\lim} \geq S(\rho_i)$. □

Proof of Lemma 6.

Proposition 1. *For a mixed state ρ, $C_{\lim}\rho = \rho$ if and only if $D_{\lim}\rho = \rho$.*

Proof. It suffices to prove that $C_{\lim}\rho = \rho$ implies $D_{\lim}\rho = \rho$ since both directions are similar.

$C_{\lim}\rho = \rho$ implies $C\rho = \rho$. Otherwise, by Lemma 8, $S(C'\rho) > S(\rho)$ and, since $S((C')^i\rho) \geq S(C'\rho)$ (Lemma 4), we have $S(C_{\lim}\rho) > S(\rho)$ and $C_{\lim}\rho \neq \rho$.

We can rewrite $C\rho = \rho$ as $A_{\lim}B_{\lim}\rho = \rho$. Definition of $B_{\lim}$ implies that $B_{\lim} = B_{\lim}B'$. Therefore, $A_{\lim}B_{\lim}B'\rho = \rho$. Similarly to previous paragraph, this implies $S(B'\rho) = S(\rho)$ and $B\rho = \rho$. Therefore, $B'\rho = \rho$, $B_{\lim}\rho = \rho$, and $A_{\lim}\rho = A_{\lim}B_{\lim}\rho = \rho$.

This implies $D\rho = B_{\lim}A_{\lim}\rho = B_{\lim}\rho = \rho$ and $D_{\lim}\rho = \rho$. □

Proposition 2. *Let A be an arbitrary CPSO. Assume that ρ is such that $A\rho = \rho$. Let H be the support of ρ (subspace spanned by pure states from which ρ consists). Then $A(H) \subseteq H$.*

Proof. By contradiction, assume that ρ' is a state in H which is not mapped to H by A. We can represent ρ as a probabilistic combination $\varepsilon\rho' + (1 - \varepsilon)\rho''$ where ρ'' is some other density matrix. This implies that ρ is not mapped to H either and $\rho \neq A\rho$. □

We now use these two claims to show that, for any ρ, $C_{\mathrm{lim}}\rho = D_{\mathrm{lim}}\rho$.

Let $\rho_{\mathrm{diff}} = C_{\mathrm{lim}}\rho - D_{\mathrm{lim}}\rho$. We would like to show that $\rho_{\mathrm{diff}} = 0$. $C_{\mathrm{lim}}\rho$ is fixed by C_{lim} and, by Proposition 1, by D_{lim} as well. Similarly, $D_{\mathrm{lim}}\rho$ is fixed by both D_{lim} and C_{lim}. Therefore, the difference of these two density matrices is fixed by both C and D as well: $C_{\mathrm{lim}}\rho_{\mathrm{diff}} = D_{\mathrm{lim}}\rho_{\mathrm{diff}} = \rho_{\mathrm{diff}}$.

We decompose $\rho_{\mathrm{diff}} = \rho_+ - \rho_-$, with ρ_+ being the state formed by eigenvectors of ρ_{diff} with positive eigenvalues and ρ_- being the state formed by eigenvectors with negative eigenvalues. Then we must have $C_{\mathrm{lim}}\rho_+ = D_{\mathrm{lim}}\rho_+ = \rho_+$ and $C_{\mathrm{lim}}\rho_- = D_{\mathrm{lim}}\rho_- = \rho_-$.

Let H_+ and H_- be the subspaces spanned by states forming ρ_+ and ρ_- respetively. By Proposition 2, H_+ and H_- are fixed by C_{lim} and D_{lim}.

We consider a measurement which measures a state ρ with respect to H_+ and its complement. The probability of obtaining H_+ is equal to $\mathrm{Tr}\, P_{H_+}\rho$ where P_{H_+} is a projection to H_+ and Tr is the trace of a matrix.

Proposition 3. *Let E be a CPSO such that $S(E\rho) \geq S(\rho)$. Let H be such that $E(H) \subseteq H$. Then, for any ρ,* $\mathrm{Tr}\, P_H\rho = \mathrm{Tr}\, P_H E\rho$.

Proof. First we show that $E(H) \subseteq H$ implies $E(H^\perp) \subseteq H^\perp$. To see that, let $|\psi_1\rangle$, ..., $|\psi_k\rangle$ be a basis for H and let $|\psi'_1\rangle, \ldots, |\psi'_l\rangle$ be a basis for $H^\perp$. Let ρ_1 be the mixed state that is $|\psi_i\rangle$, $i \in \{1, \ldots, k\}$, with probability $1/k$. Let ρ_2 be the mixed state that is $|\psi'_i\rangle$, $i \in \{1, \ldots, l\}$, with probability $1/l$. Let $\rho = (k/(k+l))\rho_1 + (l/(k+l))\rho_2$. Then $S(\rho_1) = \log_2 k$ and $S(\rho) = \log_2(k+l)$, so, $S(E\rho_1) \geq \log_2 k$ and $S(E\rho) \geq \log_2(k+l)$. By Lemma 9 and the assumption, this means $E\rho_1 = \rho_1$ and $E\rho = \rho$. Therefore, $E(\rho_2) = E(\rho - \rho_1) = \rho - \rho_1 = \rho_2$. By Proposition 2, this means that $H^\perp$ is fixed by E.

Next we show $\mathrm{Tr}\, P_H\rho = \mathrm{Tr}\, P_H E\rho$ for any ρ. It suffices to show this for pure states $\rho = |\psi\rangle\langle\psi|$. We write $|\psi\rangle = \sqrt{\alpha}|\psi_1\rangle + \sqrt{1-\alpha}|\psi_2\rangle$, $|\psi_1\rangle \in H$, $|\psi_2\rangle \in H^\perp$. Then the density matrix of $|\psi\rangle$ is

$$|\psi\rangle\langle\psi| = \alpha\rho_1 + (1-\alpha)\rho_2 + \sqrt{\alpha(1-\alpha)}\rho_3,$$

$$\rho_1 = |\psi_1\rangle\langle\psi_1|, \qquad \rho_2 = |\psi_2\rangle\langle\psi_2|,$$

$$\rho_3 = |\psi_1\rangle\langle\psi_2| + |\psi_2\rangle\langle\psi_1|,$$

$P_H\rho = \alpha|\psi_1\rangle\langle\psi_1|$, and $\mathrm{Tr}\, P_H\rho = \alpha$. Since H and $H^\perp$ are mapped to themselves by E, the states ρ_1 and ρ_2 are mapped to mixed states in H and $H^\perp$. To complete the proof, it suffices to show that $\mathrm{Tr}\, P_H\rho_3 = 0$.

Let A_1, ..., A_m be the Kraus decomposition of E. Now consider the state $E(|\psi_1\rangle\langle\psi_1|)$. We have

$$E(|\psi_1\rangle\langle\psi_1|) = \sum_{i=1}^{m} A_i|\psi_1\rangle\langle\psi_1|A_i^\dagger.$$

Remember that E maps H to itself. This is only possible if all $A_i|\psi_1\rangle$ are in H. Similarly, $A_i|\psi_2\rangle \in H^\perp$. Therefore, $E\rho_3$ is a sum of $|\varphi\rangle\langle\varphi'|$, with one of $|\varphi\rangle$ and $|\varphi'\rangle$ in H and the other in $H^\perp$. For each such matrix, $\mathrm{Tr}\, P_H|\varphi\rangle\langle\varphi'| = 0$. Therefore, $\mathrm{Tr}\, P_H\rho_3 = 0$. □

By this proposition, $\mathrm{Tr}\, Pr_{H_+} C_{\lim}\rho = \mathrm{Tr}\, Pr_{H_+} \rho = \mathrm{Tr}\, Pr_{H_+} D_{\lim}\rho$. This implies

$$\mathrm{Tr}\, \rho_+ = \mathrm{Tr}\, Pr_{H_+} \rho_{\mathrm{diff}} = \mathrm{Tr}\, Pr_{H_+}(C_{\lim}\rho - D_{\lim}\rho) = 0.$$

By definition, ρ_+ is the part of ρ_{diff} with positive eigenvalues. Therefore, $\mathrm{Tr}\, \rho_+ = 0$ iff $\rho_+ = 0$. Similarly, $\rho_- = 0$ and we get $\rho_{\mathrm{diff}} = 0$ and $C_{\lim}\rho = D_{\lim}\rho$. □

Proof of Lemma 7.

Proposition 4. *$A(\overline{Q_a}) \subseteq \overline{Q_a}$; $B(\overline{Q_a}) \subseteq \overline{Q_a}$; $A(\overline{Q_b}) \subseteq \overline{Q_b}$; and $B(\overline{Q_b}) \subseteq \overline{Q_b}$.*

Proof. A maps ρ_{ax} to ρ_{axa}. Therefore, a probabilistic combination of states ρ_{ax} gets mapped to a probabilistic combination of states ρ_{axa} and $A(Q_a) \subseteq Q_a$. This implies $A(\overline{Q_a}) \subseteq \overline{A(Q_a)} \subseteq \overline{Q_a}$. Other inclusions are similar. □

Proposition 5. *$A_{\lim}(\overline{Q_a}) \subseteq \overline{Q_a}$; $A_{\lim}(\overline{Q_b}) \subseteq \overline{Q_b}$; $B_{\lim}(\overline{Q_a}) \subseteq \overline{Q_a}$; and $B_{\lim}(\overline{Q_b}) \subseteq \overline{Q_b}$.*

Proof. Since $A(\overline{Q_a}) \subseteq \overline{Q_a}$ and A' is a probabilistic combination of A and identity, $A'(\overline{Q_a}) \subseteq \overline{Q_a}$. Therefore, $(A')^i(\overline{Q_a}) \subseteq \overline{Q_a}$. $A_{\lim}$ is the limit of $(A')^i$. Since $\overline{Q_a}$ is closed, $A_{\lim}(\overline{Q_a}) \subseteq \overline{Q_a}$. Again, other inclusions are similar. □

Proposition 6. *Let ρ be the state of M after reading the left endmarker. Then $\rho_A = A_{\lim}\rho \in \overline{Q_a}$ and $\rho_B = B_{\lim}\rho \in \overline{Q_b}$.*

Proof. It suffices to prove the first part. Let $\rho_i = (A')^i\rho$. This state is a probabilistic combination of $A^j\rho$, for $j \in \{0, \ldots, i\}$. All of those, except for $A^0\rho = \rho$ are in Q_a. Therefore, $(A')^i\rho = (1/2^i)\rho + (1 - 1/2^i)\rho'_i$, $\rho'_i \in Q_a$.

Let $\rho_A = \lim_{i\to\infty} \rho_i$. Then $\rho_A = \lim_{i\to\infty} \rho'_i$. Since $\rho'_i \in Q_a$, we have $\rho_A \in \overline{Q_a}$. □

Furthermore, by Proposition 5, $C\rho = B_{\lim}A_{\lim}\rho = B_{\lim}\rho_A \in \overline{Q_a}$. By applying Proposition 5 repeatedly, we get $C^i\rho = (B_{\lim}A_{\lim})^i\rho \in \overline{Q_a}$. The closure of $\overline{Q_a}$ gives us $C_{\lim}\rho \in \overline{Q_a}$. Similarly, from $\rho_B \in \overline{Q_b}$, we get $D\rho \in \overline{Q_b}$ and then $D_{\lim}\rho \in \overline{Q_b}$.

Theorem 1 now follows from Theorems 2 and 4 and by the upper bound results of this section.

Theorem 4 is proved by showing that probabilistic reversible automata recognize any language in **BG**. On the other hand, in [10] it is proved that any other regular language is not recognizable by PRAs. Hence an easy corollary is that LQFAs and PRAs recognize exactly the same class of languages.

4. Results for BPQFAs

Our main result for BPQFAs is given below:

Theorem 6. *The language L has its syntactic monoid in* **BG** *iff it is a Boolean combination of languages recognized by BPQFAs.*

Similar to the LQFA case, we first show that Boolean combinations of languages recognized by BPQFAs form a $*$-variety of languages, and then we give tight upper and lower bounds for the languages contained in this variety.

The fact that this class of languages forms a $*$-variety follows from this theorem:

Theorem 7 [7]. *The class of languages recognized by BPQFAs is closed under inverse homomorphisms and word quotient.*

For the remainder of the section, we use a technique introduced in [12] to analyze BPQFAs. Let ψ be an unnormalized state vector of M. Define $A'_\sigma = P_{\text{non}} A_\sigma$, and for any word $w = w_1 \cdots w_k$ let $A'_w = A'_{w_k} \cdots A'_{w_1}$. If $\psi = A'_{¢}|q_0\rangle$, then the vector $\psi_w = A'_w \psi$ completely describes the probabilistic behavior of M on input w, since M halts while reading w with probability $1 - \|\psi_w\|_2^2$ and continues in state $\psi_w / \|\psi_w\|_2$ with probability $\|\psi_w\|_2^2$.

4.1. *BPQFA Lower Bounds*

Theorem 8. *Any language whose syntactic monoid is in* **BG** *is a Boolean combination of languages recognized by BPQFAs.*

Proof. A general construction for $\Sigma^* a_1 \Sigma^* a_2 \cdots a_k \Sigma^*$ was given in [7]. We augment this construction so that it recognizes L defined by $w \in L$ iff $w = w_0 a_1 w_1 \cdots a_k w_k$, where for each i, $w_0 a_1 w_1 \cdots w_i \in L_i$ for some prespecified group language L_i. By the cancellative law of groups and the closure properties, this is sufficient to prove the theorem. We present Brodsky and Pippenger's construction here in full with minor modifications. As above, we adopt the point of view that the state vector is unnormalized.

The key to their construction is what they call a trigger chain. A trigger chain recognizing $a_1, \ldots, a_k$ is constructed out of interleaved 3-tuples of states, one for each a_i with $i \geq 2$. A link in the chain is activated by the following transition:

$$T = \begin{bmatrix} \frac{1}{2} & \frac{1}{\sqrt{2}} & \frac{1}{2} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{2} & -\frac{1}{\sqrt{2}} & \frac{1}{2} \end{bmatrix}.$$

Whenever the middle element is 0 in a three-element vector, T has the following effect:

$$T(\alpha, 0, \beta)^T = \left(\frac{\alpha}{2} + \frac{\beta}{2}, \frac{\alpha}{\sqrt{2}} - \frac{\beta}{\sqrt{2}}, \frac{\alpha}{2} + \frac{\beta}{2} \right)^T.$$

Thus if α and β are positive reals, then T averages the amplitude between the first and the third element, and places any excess amplitude into the middle state. If $\alpha = \beta$, then the trigger will have no effect. In the construction, the middle state will correspond to a rejecting state and so its amplitude will always be 0 at the beginning of every transition. Also define T_i to be the matrix that acts as T on states i, $i+1$, and $i+2$, and as the identity everywhere else.

Now a machine $M = (Q, \Sigma, q_0, \{A_\sigma\}, Q_{\text{acc}}, Q_{\text{rej}})$ is constructed to recognize $\Sigma^* a_1 \Sigma^* \cdots \Sigma^* a_k \Sigma^*$ using $2k+3$ states as follows:

$$Q = \{q_0, q_2, \ldots, q_{2k+2}\},$$
$$Q_{\text{rej}} = \{q_1, q_3, \ldots, q_{2k-3}\} \cup \{q_{2k+1}, q_{2k+2}\},$$
$$Q_{\text{acc}} = \{q_{2k-1}\}.$$

To simplify the construction of the transitions, we define I_m to be the $m \times m$ identity matrix, and

$$R = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

For each character $\sigma \in \Sigma$, we define $A_\sigma = U_{\sigma 1} \cdots U_{\sigma k}$, where for each i,

$$U_{\sigma i} = \begin{cases} \begin{bmatrix} R & \\ & I_{2k+1} \end{bmatrix} & \text{if} \quad i = 0 \quad \text{and} \quad \sigma = a_1, \\ T_{2i-4} & \text{if} \quad 2 \le i \le k \quad \text{and} \quad \sigma = a_1, \\ I_{2k+3} & \text{otherwise.} \end{cases}$$

We define the initial transition $A_{¢}$ such that $A_{¢}|q_0\rangle = \sum_{i=0}^{2k}(1/\sqrt{2k+1})|q_{2i}\rangle$, and finally we define $A_{\$} = FT_{2k-2}$, where

$$F = \begin{bmatrix} I_{2(k-1)} \otimes R & & & & & \\ & 0 & 0 & 0 & 0 & 1 \\ & 0 & 1 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 1 & 0 \\ & 0 & 0 & 1 & 0 & 0 \\ & 1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Here is an outline of the proof of correctness given in [7]. Initially, after reading $¢$ the amplitude is distributed among the nonhalting states. When a_1 is read, the amplitude of q_0 becomes 0 and M halts and rejects with small probability. If a_2 is now read then states q_2 and q_4 are averaged, causing a bounded decrease in amplitude of q_4. Inductively, there will be a bounded amount of amplitude in the accepting state q_{2k-1} if and only if $a_1, \ldots, a_k$ and the endmarker were read in sequence. The $I_{2(k-1)} \otimes R$ submatrix serves to channel all the unused amplitude into the rejecting states.

Now given M we construct $M' = (Q', \Sigma, q_0', \{A_\sigma\}, Q'_{\text{acc}}, Q'_{\text{rej}})$ to recognize L. For all i let $G_i = M(L_i)$. Also let $\varphi_i\colon \Sigma^* \to G_i$ and let F_i be such that $\varphi_i^{-1}(F_i) = L_i$. We can compose these groups into a single group $G = G_0 \times \cdots \times G_k$ with identity $1 = (1, 1, \ldots, 1)$ and $|G| = m$.

Let $Q'_{\text{acc}} = Q_{\text{acc}} \times (G_1 \times \cdots \times G_{k-1} \times F_k)$. For the endmarkers, define $A'_{¢} = (A_{¢} \otimes I_m)$ and $A'_{\$} = (A_{\$} \otimes I_m)$. For each $\sigma \in \Sigma$, we define $A'_\sigma = P_\sigma U'_{\sigma 1}, \ldots, U'_{\sigma k}$. Each $U'_{\sigma i}$ is the matrix that acts as $U_{\sigma i}$ on $Q \times \{f\}$ for each $f \in G_1 \times \cdots \times G_{i-2} \times F_{i-1} \times G_i \times \cdots \times G_k$ and as the identity everywhere else. Finally P_σ is a permutation matrix such that $P_\sigma|q, g\rangle = |q, g\sigma\rangle$ for all $|q, g\rangle$.

The transition matrices are constructed so that, after reading any partial input w, the state vector will be in the subspace $E = span\{|q, 1w\rangle\colon q \in Q\}$.

The construction contains $k + 1$ triggers. If a series of such triggers are activated in sequence, then as the last trigger is applied there will be a bounded amount of amplitude sent to the middle state of the last trigger. For $1 \le i \le k$ the ith trigger is activated when a_i is read and the current group element is in the set F_{i-1}. When the right endmarker is read, the last trigger is activated. This places amplitude into (q_{2i-1}, g) (where g is the current group element) if and only if $a_1, \ldots, a_k$ are read in the correct context in order. Finally, we accept only if the current group element g is in F_k. M' rejects with probability 1 any word not in the language, and accepts any word in the language with bounded probability, thus M' recognizes L. □

4.2. *BPQFA Upper Bounds*

As discussed at the beginning of the section, we take the states of the BPQFA to be unnormalized. The following lemma nicely characterizes the behavior of the operation $A'_\sigma = P_{\text{non}} A_\sigma$, and thus is very useful for showing upper bounds on the BPQFA:

Lemma 11 [3]. *Let $\{x, y\} \subseteq \Sigma^+$. Then there are subspaces E_1, E_2 such that $E_{\text{non}} = E_1 \oplus E_2$ and*

- *if $\psi \in E_1$, then $A'_x(\psi) \in E_1$, $A'_y(\psi) \in E_1$, and $\|A'_x(\psi)\|_2 = \|A'_y(\psi)\|_2 = \|\psi\|_2$;*
- *if $\psi \in E_2$, then for any $\varepsilon > 0$, and for any word $t \in \{x, y\}^*$ there exists a word $t' \in \{x, y\}^*$ such that $\|A'_{tt'}(\psi)\|_2 < \varepsilon$.*

Theorem 9. *The languages $a\Sigma^*$ and $\Sigma^* a$ cannot be expressed as Boolean combinations of languages recognized by BPQFAs.*

Proof. We begin with the $a\Sigma^*$ result. By closure under inverse homomorphisms, it is sufficient to show this for $\Sigma = \{a, b\}$. Let M be a BPQFA that recognizes L with probability p, and let $\psi = A'_{¢}(|q_0\rangle)$. The first step is to show that for any two prefixes $v, w \in \{a, b\}^+$ and any $\varepsilon > 0$, there exists $v', w' \in \{a, b\}^*$ such that $\|A'_{vv'}\psi - A'_{ww'}\psi\|_2 < \varepsilon$. Thus if $\varepsilon \leq \sqrt{p}$, it follows that M cannot distinguish between vv' and ww' with sufficiently large probability. As in Lemma 11, separate E_{non} into two subspaces E_1 and E_2 with respect to the words $x = a$ and $y = b$. Then we can rewrite ψ as $\psi = \psi_1 + \psi_2$, where $\psi_i \in E_i$. By the lemma, and since A'_a and A'_b act unitarily on E_1, for any ε' there exists v' and w' such that $\|A'_{vv'}\psi - \psi_1\|_2^2 < \varepsilon'$ and $\|A'_{ww'}\psi - \psi_1\|_2^2 < \varepsilon'$. For any fixed ε, we can find a sufficiently small ε' so that $\|A'_{vv'}\psi - A'_{ww'}\psi\|_2^2 < \varepsilon$.

Suppose L is a Boolean combination of m languages $L_1, \ldots, L_m$, with each L_i recognized by a BPQFA M_i. As above, we can construct inductively on the languages L_i, two words $v = v_1 v_2 \cdots v_m \in \{a, b\}^*$ and $w = w_1 w_2 \cdots w_m \in \{a, b\}^*$ such that av and bw are indistinguishable for every M_i. Thus we must have either $\{av, bw\} \subseteq L$ or $L \cap \{av, bw\} = \emptyset$, and in either case $L \neq a\Sigma^*$. For the construction, we first choose v_1 and w_1 so that, for all v' and w', av_1v' and bw_1w' are indistinguishable by M_1. Given that, for all v' and w', $av_1 \cdots v_{i-1}v'$ and $bw_1 \cdots w_{i-1}w'$ are not distinguishable by any of $M_1, \ldots, M_{i-1}$, we construct v_i and w_i so that, for all v' and w', $av_1 \cdots v_i v'$ and $bw_1 \cdots w_i w'$ are indistinguishable by M_i.

A similar analysis can be used to show the $\Sigma^* a$ result. Consider a single BPQFA M recognizing L with probability p. Let $\psi = A'_{¢}|q_0\rangle$ be the initial state. Let $b \in \Sigma \backslash \{a\}$, and let E_1 and E_2 be as in Lemma 11 with $x = a$ and $y = b$. We can uniquely split ψ into $\psi_1 + \psi_2$, where $\psi_1 \in E_1$ and $\psi_2 \in E_2$.

Suppose L is a Boolean combination of m languages $L_1, \ldots, L_m$ where each L_i is recognized by some BPQFA M_i with probability p_i. For any ε, we can construct a word $w = w_1 \cdots w_m$ such that, for all w', the condition $\|A'_{ww'}\psi - A'_{ww'}\psi_1\|_2 < \varepsilon$ is met by each M_i. If we choose $\varepsilon < \sqrt{\min\{p_i\}}$, then there is a k such that for all i, machine M_i satisfies $\|A'_{ww'ab^k}\psi - A'_{ww'a}\|_2 < p_i$. Thus we must have either $\{ww'ab^k, ww'a\} \subseteq L$ or $\{ww'ab^k, ww'a\} \cap L = \emptyset$, and in either case $L \neq \Sigma^* a$. □

Theorem 6 now follows from Theorems 7–9.

Note that in our characterization we have to take Boolean combinations because BPQFAs are not closed under complement. This follows from the theorem below:

Theorem 10. *For any $a \neq b$ and for any Σ satisfying $\{a, b\} \subseteq \Sigma$, BPQFAs cannot recognize $\overline{\Sigma^* b \Sigma^* a \Sigma^*}$.*

By closure under inverse homomorphisms it is sufficient to prove the result for $\Sigma = \{a, b\}$. In this case, $\overline{\Sigma^* b \Sigma^* a \Sigma^*} = a^* b^*$. Our proof makes frequent use of the following corollary to Lemma 11:

Corollary 1. *For any KWQFA (or BPQFA) M and word w we can define subspaces $E_1^w \oplus E_2^w = E_{\text{non}}$ such that $\psi_1 \in E_1^w$ implies $(A'_w)^i(\psi_1) = (A_w)^i(\psi_1)$ for all i, and $\psi_2 \in E_2^w$ implies $\lim_{i\to\infty} \|(A'_w)^i \psi_2\|_2 = 0$.*

Any $\varphi \in E_{\text{non}}$ can be uniquely decomposed into $\varphi_1 + \varphi_2$ so that $\varphi_1 \in E_1^w$ and $\varphi_2 \in E_2^w$. The components φ_1 and φ_2 are called the *ergodic* and *transient* parts of φ, respectively.

We now establish a relationship between projection operations and idempotents. An orthogonal projection P is a Hermitian operator satisfying $P^2 = P$. We define the following subclass of operations:

Definition 1. We say that an orthogonal projection P is an E_{non}*-projection* if $P(E_{\text{non}}) \subseteq E_{\text{non}}$.

Thus if we restrict our attention to vectors in E_{non}, then P will behave exactly as an orthogonal projection. This is relevant to our situation since the state ψ of M after reading some partial input must satisfy $\psi \in E_{\text{non}}$.

Claim 2. *Any E_{non}-projection P can be simulated by a unitary transformation U and the BPQFA measurement.*

Proof. Assume without loss of generality that $|Q_{\text{rej}}| \geq |Q \backslash Q_{\text{rej}}|$ (if this is not the case, then we can simply augment M with the required number of Q_{rej} states). Let $S = \{P\psi\colon \psi \in E_{\text{non}}\}$. Note that S is a subspace. Let $\overline{S}$ be the subspace such that $S \oplus \overline{S} = E_{\text{non}}$. Now to simulate the E_{non}-projection, we choose U to be the operation that rotates $\overline{S}$ into E_{rej}. Any amplitude that was in $\overline{S}$ will be removed when the BPQFA measurement is applied. □

Let L be a language recognized by a BPQFA M with probability p, and let $\varphi\colon \Sigma^* \to M(L)$ be the syntactic morphism. Clearly, if A'_a is an E_{non}-projection, then $\varphi(a)$ must be idempotent (i.e. $\varphi(a) = e = e^2$). We claim that the following converse is also true:

Claim 3. *Let L, M, p, and φ be as above, and let $\varphi(a)$ be an idempotent. Let M' be the machine constructed by replacing each A'_a with an E_{non}-projection onto E_1^a. Then M' also recognizes L with probability p.*

Proof. Suppose that M' does not recognize L with probability p. Thus, either M' accepts some word $w \in L$ with probability $p_w < p$, or M' accepts some word $w \notin L$ with probability $p_w > 0$. We consider the former case, the latter is similar.

Define ε so that $\sqrt{p} = \sqrt{p_w} + \varepsilon$. Let k be the number of occurrences of a in w. Note that $k > 0$, otherwise M and M' would accept w with the same probability. Let $w = w_0 a w_1 \cdots w_{k-1} a w_k$ with $w_i \in (\Sigma \backslash \{a\})^*$. Let U be a unitary matrix such that U' is the E_{non}-projection onto E_1^a. We set j to be such that $\|(A'_a)^j \varphi - U'\varphi\|_2 = \varepsilon' < \varepsilon/k$ for all $\varphi \in E_{\text{non}}$ (we know by Corollary 1 that such a j exists). Now consider:

$$w' = w_0 a^j w_1 \cdots w_{k-1} a^j w_k.$$

We have $w' \in L$ since $\varphi(a)$ is idempotent. Let $\psi = |q_0\rangle$ be the initial state of M. Note that, for all φ, $(A'_a)^j A'_{w_0} \varphi = U'_a A'_{w_0} \varphi + \xi$ for some ξ satisfying $\|\xi\|_2 < \varepsilon'$. So there exists a vector ξ_1 such that $\|\xi_1\|_2 < \varepsilon'$ and

$$\begin{aligned} A'_{w_k}(A'_a)^j \cdots A'_{w_1}(A'_a)^j A'_{w_0}\psi &= A'_{w_k}(A'_a)^j \cdots A'_{w_1}(U' A'_{w_0}\psi + \xi) \\ &= A'_{w_k}(A'_a)^j \cdots A'_{w_1} U' A'_{w_0}\psi + \xi_1. \end{aligned}$$

In general there exist vectors ξ_i, $1 \le i \le k$, such that $\|\xi_i\|_2 \le \varepsilon'$ for all i, and

$$A'_{w_k}(A'_a)^j \cdots A'_{w_1}(A'_a)^j A'_{w_0}\psi = A'_{w_k} U' \cdots A'_{w_1} U' A'_{w_0}\psi + \sum_{i=1}^{k} \xi_i,$$

and so

$$\begin{aligned} \mathbf{P}[M \text{ accepts } w'] &= \left\| P_{\text{acc}} A'_\$ \left(A'_{w_k} U' \cdots A'_{w_1} U' A'_{w_0}\psi + \sum \xi_i \right) \right\|_2^2 \\ &\le \left(\|P_{\text{acc}} A'_\$ (A'_{w_k} U' \cdots A'_{w_1} U' A'_{w_0}\psi)\|_2 + \sum \|\xi_i\|_2 \right)^2 \\ &< \left(\sqrt{p_w} + \varepsilon\right)^2 = p. \end{aligned}$$

The original M accepts w' with probability strictly less than p, a contradiction. □

Proof of Theorem 10. Suppose M is a BPQFA that recognizes $\overline{\Sigma^* b \Sigma^* a \Sigma^*}$ with probability p. We show that such an M cannot exist. By Claim 3, we can assume without loss of generality that A'_a and A'_b are E_{non}-projections.

For any M and w, we can define $E_{w,\text{rej}}$ to be the set of all vectors $\psi \in E_{\text{non}}$ such that $A'_w \psi \in E_{\text{rej}}$ (if M halts with certainty before w is processed then $A'_w \psi = \vec{0} \in E_{\text{rej}}$).

It is easy to show by linearity that $E_{w,\mathrm{rej}}$ is a subspace. For shorthand, define

$$E_\alpha = \bigcap_{w,x,y\in\Sigma^*} E_{wbxay\$,\mathrm{rej}}, \qquad E_\beta = \bigcap_{x,y\in\Sigma^*} E_{xay\$,\mathrm{rej}}, \qquad E_\gamma = \bigcap_{y\in\Sigma^*} E_{y\$,\mathrm{rej}}.$$

Observe that $E_\alpha \supseteq E_\beta \supseteq E_\gamma$. At all times, the state vector of M must be contained in the subspace E_α in order to recognize the language $\overline{\Sigma^* b\Sigma^* a\Sigma^*}$, since all words containing the subword ba must be rejected with certainty. When the first b is read, the state vector must fall into the subspace E_β, since by definition $\varphi \in E_\alpha$ implies $A'_b\varphi \in E_\beta$. If an a is read while the state vector is in the subspace E_β, the state vector must fall into the subspace E_γ, and the state vector must remain here until the end of the computation. We argue that any vector in $\psi \in E_\alpha$ will fall into E_γ reading an a followed by a b, thus the word ab is rejected with certainty, a contradiction.

Define $\overline{E_\beta}$ to be the subspace such that $E_\beta \oplus \overline{E_\beta} = \mathbb{C}^n$. The vector $\psi_a = A'_a\psi$ can be uniquely decomposed into $\psi_\alpha + \psi_\beta$, where $\psi_\alpha \in E_\alpha \cap \overline{E_\beta}$ and $\psi_\beta \in E_\beta$. We claim that $\psi_\beta \in E_\gamma$. Observe that $A'_a\psi = A'_aA'_a\psi$, so $\psi_\alpha + \psi_\beta = A'_a(\psi_\alpha + \psi_\beta)$. Let $P_{\overline{\beta}}$ be the projection operator onto $\overline{E_\beta}$. Now,

$$\begin{aligned} \psi_\alpha + \psi_\beta = A'_a(\psi_\alpha + \psi_\beta) \quad &\Longrightarrow \quad P_{\overline{\beta}}(\psi_\alpha + \psi_\beta) = P_{\overline{\beta}}(A'_a(\psi_\alpha + \psi_\beta)) \\ &\Longleftrightarrow \quad \psi_\alpha = P_{\overline{\beta}}(A'_a\psi_\alpha) \\ &\Longrightarrow \quad \psi_\alpha = A'_a\psi_\alpha \end{aligned}$$

and so

$$\begin{aligned} \psi_\alpha + \psi_\beta = A'_a(\psi_\alpha + \psi_\beta) \quad &\Longleftrightarrow \quad \psi_\alpha + \psi_\beta = \psi_\alpha + A'_a\psi_\beta \\ &\Longleftrightarrow \quad \psi_\beta = A'_a\psi_\beta. \end{aligned}$$

From $\psi_\beta \in E_\beta$ it follows that $A'_a\psi_\beta \in E_\gamma$, and thus $\psi_\beta \in E_\gamma$. Now consider $\psi_{ab} = A'_b(\psi_\alpha + \psi_\beta)$. Since $A'_b(\psi_\alpha + \psi_\beta) \in E_\beta$ and $A'_b\psi_\beta \in E_\beta$, we must have $A'_b\psi_\alpha \in E_\beta$. However, $\psi_\alpha \perp E_\beta$ and A'_b is an E_{non}-projection, so we must have $A'_b\psi_\alpha = \vec{0}$. Thus $\psi_{ab} = A'_b\psi_\alpha + A'_b\psi_\beta = A'_b\psi_\beta \in E_\gamma$. Thus, ab is rejected with certainty, as we wanted to show. □

5. Conclusion

In this paper we have produced algebraic characterizations for the languages recognized by a new model which we called Latvian Quantum Finite Automata, and for the Boolean closure of languages recognized by Brodsky–Pippenger Quantum Finite Automata. A somewhat surprising consequence of our results is that the two models are equivalent in power, up to Boolean combinations. It has been shown that a language L is recognizable by an LQFA iff its syntactic monoid is a block group; hence membership in the class is decidable. The situation is more complicated for a BPQFA since the corresponding class of languages is not closed under complement. The good news is that we have shown that the class forms what is known as a *positive* $*$-variety and thus is amenable to algebraic description through the mechanism of *ordered monoids* [19]. We know that this positive $*$-variety strictly contains the regular languages that are open in the group topology and a precise characterization seems to be within reach.

Another open problem is to find an algebraic characterization of the Kondacs–Watrous model. It is an easy consequence of our results on BPQFAs that KWQFAs can recognize any language whose syntactic monoid is in **BG**. However, outside of **BG** the question of language recognition is still unresolved.

The class of languages recognized by KWQFAs is known not be closed under union [3], hence does not form a $*$-variety. It is nevertheless meaningful to ask for an algebraic description of the $*$-variety generated by those languages. We conjecture that the right answer involves replacing block groups by a one-sided version **V** of this **M**-variety defined by the following condition: for any $e = e^2$ and $f = f^2$ in M, $eM = fM$ implies $e = f$. The corresponding variety of languages can be described as the largest variety that does not contain $\Sigma^* a$ for $|\Sigma| \geq 2$.

References

[1] Dorit Aharonov, Andris Ambainis, Julia Kempe, and Umesh Vazirani. Quantum walks on graphs. In *Proceedings of the* 33*rd Annual ACM Symposium on Theory of Computing*: *Hersonissos*, *Crete*, *Greece*, *July* 6–8, 2001, pages 50–59. ACM Press, New York, 2001.

[2] Andris Ambainis and Arnolds Ķikusts. Exact results for accepting probabilities of quantum automata. *Theoretical Computer Science*, 295(1–3):3–25, February 2003.

[3] Andris Ambainis, Arnolds Ķikusts, and Māris Valdats. On the class of languages recognizable by 1-way quantum finite automata. In *Proceedings of the* 18*th Annual Symposium on Theoretical Aspects of Computer Science*, pages 75–86. Volume 2010 of Lecture Notes in Computer Science. Springer-Verlag, Berlin, 2001.

[4] Andris Ambainis, Ashwin Nayak, Amnon Ta-Shma, and Umesh Vazirani. Dense quantum coding and quantum finite automata. *Journal of the ACM*, 49(4):496–511, July 2002.

[5] Martin Beaudry, François Lemieux, and Denis Thérien. Finite loops recognize exactly the regular open languages. In Pierpaolo Degano, Roberto Gorrieri, and Alberto Marchetti-Spaccamela, editors, *Automata*, *Languages and Programming*, 24*th International Colloquium*, Bologna, Italy, 7–11 July 1997, pages 110–120. Volume 1256 of Lecture Notes in Computer Science. Springer-Verlag, Berlin, 1997.

[6] Alberto Bertoni, Carlo Mereghetti, and Beatrice Palano. Quantum computing: 1-way quantum automata. In *Developments in Language Theory*. Volume 2710 of Lecture Notes in Computer Science. Springer-Verlag, Berlin, 2003.

[7] Alex Brodsky and Nicholas Pippenger. Characterizations of 1-way quantum finite automata. *SIAM Journal on Computing*, 31(5):1456–1478, October 2002.

[8] Massimo Pika Ciamarra. Quantum reversibility and a new model of quantum automaton. *Fundamentals of Computation Theory*, 13:376–379, 2001.

[9] Samuel Eilenberg. *Automata*, *Languages*, *and Machines*. Academic Press, New York, first edition, 1976.

[10] Marats Golovkins and Maksim Kravtsev. Probabilistic reversible automata and quantum automata. In *Computing and Combinatorics*, pages 574–583. Volume 2387 of Lecture Notes in Computer Science. Springer-Verlag, 2002.

[11] Peter Jeavons, David Cohen, and Marc Gyssens. Closure properties of constraints. *Journal of the ACM*, 44(4):527–548, June, 1997.

[12] Attila Kondacs and John Watrous. On the power of quantum finite state automata. In *Proceedings of the* 38*th Annual Symposium on Foundations of Computer Science*, 20–22 October 1997, pages 66–75. IEEE Computer Society Press, Los Alamitos, CA, 1997.

[13] Albert Marshal and Ingram Olkin. *Theory of Majorization and Its Applications*. Academic Press, New York, 1979.

[14] Cris Moore and Jim Crutchfield. Quantum automata and quantum grammars. *Theoretical Computer Science*, 237(1–2):275–306, 2000.

[15] Ashwin Nayak. Optimal lower bounds for quantum automata and random access codes. In *Proceedings of the* 40*th Annual Symposium on Foundations of Computer Science* (*FOCS* '99), pages 369–377, October 1999.

[16] Michael Nielsen and Isaac Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, 2000.

[17] Jean-Eric Pin. *Varieties of Formal Languages*. North Oxford, London, 1986.

[18] Jean-Eric Pin. BG = PG, a success story. In John Fountain, editor, *NATO Advanced Study Institute Semigroups*, *Formal Languages*, *and Groups*, pages 33–47. Kluwer, Dordrecht, 1995.

[19] Jean-Eric Pin. A variety theorem without complementation. *Russian Mathematics*, 39:80–90, 1995.

[20] Lieven Vandersypen, Matthias Steffan, Gregory Breyta, Costantino Yannoni, Mark Sherwood, and Isaac Chuang. Experimental realization of Shor's quantum factoring algorithm using nuclear magnetic resonance. *Nature*, 414:883–887, 2001.

Received June 28, 2004, *and in final form May* 16, 2005. *Online publication November* 29, 2005.