On the Power of Finite Automata with both Nondeterministic and Probabilistic States

Anne Condon*
condon@cs.wisc.edu
Department of Computer Sciences
University of Wisconsin
1210 West Dayton St.
Madison, WI 53706

Lisa Hellerstein†
hstein@eecs.nwu.edu
Department of E.E.C.S.
Northwestern University
2145 Sheridan Rd.
Evanston, IL 60208-3118

Samuel Pottle ‡
pottle@cs.wisc.edu
Department of Computer Sciences
University of Wisconsin
1210 West Dayton St.
Madison, WI 53706

Avi Wigderson§
avi@cs.huji.ac.il
Computer Science Department
Hebrew University
Jerusalem
91904, Israel

June 16, 1995

*Condon's research supported by NSF grant CCR-9257241 and by a matching grant from AT&T Bell Labs.

†Hellerstein's research supported in part by NSF grant CCR-9210957.

‡Pottle's research supported by NSF grant CCR-9257241.

§Wigderson's research supported in part by BSF grant 92-00106/1 and a grant from the Wolfson Research Awards.

Abstract

We study finite automata with both nondeterministic and random states (npfa's). We restrict our attention to those npfa's that accept their languages with a small probability of error and run in polynomial expected time. Equivalently, we study Arthur-Merlin games where Arthur is limited to polynomial time and constant space.

Dwork and Stockmeyer asked whether these npfa's accept only the regular languages (this was known if the automaton has only randomness or only nondeterminism). We show that the answer is yes in the case of npfa's with a 1-way input head. We also show that if L is a nonregular language, then either L or $\bar{L}$ is not accepted by any npfa with a 2-way input head.

Toward this end, we define a new measure of the complexity of a language L, called its 1-tiling complexity. For each n, this is the number of tiles needed to cover the 1's in the "characteristic matrix" of L, namely the binary matrix with a row and column for each string of length $\leq n$, where entry $[x, y] = 1$ if and only if the string $xy \in L$. We show that a language has constant 1-tiling complexity if and only if it is regular, from which the result on 1-way input follows. Our main result regarding the general 2-way input tape follows by contrasting two bounds: an upper bound of polylog(n) on the 1-tiling complexity of every language computed by our model, and a lower bound stating that the 1-tiling complexity of a nonregular language or its complement exceeds a function in $2^{\Omega(\sqrt{\log n})}$ infinitely often.

The last lower bound follows by proving that the characteristic matrix of *every* nonregular language has rank n for infinitely many n. This is our main technical result, and its proof extends techniques of Frobenius and Iohvidov developed for Hankel matrices.

1 Introduction

The classical subset construction of Rabin and Scott [25] shows that finite state automata with just nondeterministic states (nfa's) accept exactly the regular languages. Results of Rabin [24], Dwork and Stockmeyer [7] and Kaneps and Freivalds [17] show that the same is true of probabilistic finite state automata which run in polynomial expected time. Here and throughout the paper, we restrict attention to automata which accept languages with error probability which is some constant ϵ less than $1/2$.

However, there has been little previous work on finite state automata which have both probabilistic and nondeterministic states. Such automata are equivalent to the Arthur-Merlin games of Babai and Moran [3], restricted to constant space, with an unbounded number of rounds of communication between Arthur and Merlin. In this paper, we refer to them as npfa's. In the computation of an npfa, each transition from a probabilistic state is chosen randomly according to the transition probabilities from that state, whereas from a nondeterministic state, it is chosen so as to maximize the probability that an accepting state is eventually reached. We let 1NPFA and 2NPFA-polytime denote the classes of languages accepted by npfa's which have a 1-way or a 2-way input head, respectively, and which run in polynomial expected time. Dwork and Stockmeyer [8] asked whether 2NPFA-polytime is exactly the set of regular languages, which we denote by *Regular*.

In this paper, we prove the following two results on npfa's.

Theorem 1.1 *1NPFA = Regular.*

Theorem 1.2 *If L is nonregular, then either L or $\bar{L}$ is not in 2NPFA-polytime.*

Thus, we resolve the question of Dwork and Stockmeyer for npfa's with 1-way head, and in the case of the 2-way head model, we reduce the question to that of deciding whether 2NPFA-polytime is closed under complement. Theorem 1.1 also holds even if the automaton has universal, as well as nondeterministic and probabilistic states. Moreover, Theorem 1.2 holds even for Arthur-Merlin games that use $o(\log \log n)$ space.

In proving the two results, we introduce a new measure of the complexity of a language L called its *1-tiling complexity*. Tiling complexity arguments have been used previously to prove lower bounds for communication complexity (see e.g. Yao [29]). With each language $L \subseteq \Sigma^*$, we associate an infinite binary matrix M_L, whose rows and columns are labeled by the strings of Σ^*. Entry $M_L[x, y]$ is 1 if the string $xy \in L$ and is 0 otherwise. Denote by $M_L(n)$ the finite submatrix of M_L, indexed by strings of length $\leq n$. Then, the 1-tiling complexity of L (and of the matrix $M_L(n)$) is the minimum size of a set of *1-tiles* of $M_L(n)$ such that every 1-valued entry of $M_L(n)$ is in at least one 1-tile of the set. Here, a 1-tile is simply a submatrix (whose rows and columns are not necessarily contiguous) in which all entries have value 1.

In Section 3, we prove the following theorems relating language acceptance of npfa's to tiling

complexity. The proofs of these theorems build on previous work of Dwork and Stockmeyer [8] and Rabin [24].

Theorem [3.1] *A language L is in 1NPFA only if the 1-tiling complexity of L is $O(1)$.*

Theorem [3.3] *A language L is in 2NPFA-polytime only if the 1-tiling complexity of L is bounded by a polynomial in $\log n$.*

What distinguishes our work on tiling is that we are interested in the problem of tiling the matrices $M_L(n)$, which have distinctive structural properties. If L is a unary language, then $M_L(n)$ is a matrix in which all entries along each diagonal from the top right to the bottom left are equal. Such a matrix is known as a *Hankel matrix*. An elegant theory on properties of such Hankel matrices has been developed [15], from which we obtain strong bounds on the rank of $M_L(n)$ if L is unary. In the case that L is not a unary language, the pattern of 0's and 1's in $M_L(n)$ is not as simple as in the unary case, although the matrix still has much structure. Our main technical contribution, presented in Section 4, is to prove new lower bounds on the rank of $M_L(n)$ when L is not unary. Our proof uses techniques of Frobenius and Iohvidov developed for Hankel matrices.

Theorem [4.4] *If L is nonregular, then the rank of $M_L(n)$ is at least $n+1$ infinitely often.*

By applying results from communication complexity relating the rank of a matrix to its tiling complexity, we can obtain a lower bound on the 1-tiling complexity of non-regular languages.

Theorem [4.5] *If L is nonregular, then the 1-tiling complexity of either L or $\bar{L}$ exceeds a function in $2^{\Omega(\sqrt{\log n})}$ infinitely often.*

However, there are nonregular languages, even over a unary alphabet, with 1-tiling complexity $O(\log n)$ (see Section 4). Thus the above lower bound on the 1-tiling complexity of L or $\bar{L}$ does not always hold for L itself. A simpler theorem holds for regular languages.

Theorem [4.1] *The 1-tiling complexity of L is $O(1)$ if and only if L is regular.*

By combining these theorems on the 1-tiling complexity of regular and non-regular languages with the theorems relating 1-tiling complexity to acceptance by npfa's, our two main results (Theorems 1.1 and 1.2) follow as immediate corollaries.

The rest of the paper is organized as follows. In Section 2, we define our model of the npfa, and the tiling complexity of a language. We conclude that section with a discussion of related work on probabilistic finite automata and Arthur-Merlin games. In Section 3, we present Theorems 3.1 and 3.3, which relate membership of a language L in the classes 1NPFA and 2NPFA-polytime to the 1-tiling complexity of L. A similar theorem is presented for the class 2NPFA, in which the underlying automata are not restricted to run in polynomial expected time. In Section 4, we present our bounds on the tiling complexity of both regular and nonregular languages. Theorems 1.1 and 1.2 are immediate corollaries of the main results of Sections 3 and 4. Extensions of these results to alternating automata and to Turing machines with small

space are presented in Section 5. Conclusions and open problems are discussed in Section 6.

2 Preliminaries

We first define our npfa model in Section 2.1. This model includes as special cases the standard models of nondeterministic and probabilistic finite state automata. In Section 2.2 we define our notion of the tiling complexity of a language. Finally, in Section 2.3, we discuss previous work on this and related models.

2.1 Computational Models and Language Classes

A *two-way nondeterministic probabilistic finite automaton* (2npfa) consists of a set of states Q, an input alphabet Σ, and a transition function δ, with the following properties. The states Q are partitioned into three subsets: the *nondeterministic states* N, the *probabilistic* (or *random*) states R, and the *halting states* H. H consists of two states: the *accepting state* q_a and the *rejecting state* q_r. There is a distinguished state q_0, called the *initial state*. There are two special symbols $\not\in, \$ \notin \Sigma$, which are used to mark the left and right ends of the input string, respectively.

The transition function δ has the form

$$\delta : Q \times (\Sigma \cup \{\not\in, \$\}) \times Q \times \{-1, 0, 1\} \to \{0, 1/2, 1\}.$$

For each fixed q in R, the set of random states, and $\sigma \in (\Sigma \cup \{\not\in, \$\})$, the sum of $\delta(q, \sigma, q', d)$ over all q' and d equals 1. The meaning of δ in this case is that if the automaton is in state q reading symbol σ, then with probability $\delta(q, \sigma, q', d)$ the automaton enters state q' and moves its input head one symbol in direction d (left if $d = -1$, right if $d = 1$, stationary if $d = 0$). For each fixed q in N, the set of nondeterministic states, and $\sigma \in (\Sigma \cup \{\not\in, \$\})$, $\delta(q, \sigma, q', d) \in \{0, 1\}$ for all q' and d. The meaning of δ in this case is that if the automaton is in state q reading symbol σ, then the automaton nondeterministically chooses some q' and d such that $\delta(q, \sigma, q', d) = 1$, enters state q' and moves its input head one symbol in direction d. Once the automaton enters state q_a (resp. q_r), the input head moves repeatedly to the right until the right endmarker \$ is read, at which point the automaton halts. In other words, for $q \in \{q_a, q_r\}$, $\delta(q, \sigma, q, 1) = 1$ for all $\sigma \in \Sigma \cup \{\not\in\}$, and $\delta(q, \sigma, q', 1) = 0$ for all $\sigma \in \Sigma \cup \{\not\in\}$ and $q' \neq q$. On a given input, the automaton is started in the *initial configuration*, that is, in the initial state with the head at the left end of the input. If the automaton halts in state q_a on the input, we say that it *accepts* the input, and if it halts in state q_r, we say that it *rejects* the input.

Fix some input string $w = w_0 w_1 w_2 \ldots w_n w_{n+1}$, where $w_0 = \not\in$ and $w_{n+1} = \$$. A *nondeterministic strategy* (or just *strategy*) on w is a function

$$S_w : N \times \{0, \ldots, n+1\} \to Q \times \{-1, 0, 1\}$$

such that $\delta(q, \sigma, q', d) = 1$ whenever $S_w(q, j) = (q', d)$ and $w_j = \sigma$. The meaning of S_w is that if the automaton is in state $q \in N$ reading w_j, then if $S_w(q, j) = (q', d)$, the automaton enters state q' and moves its input head one symbol in direction d. The strategy indicates which nondeterministic choice should be made in each configuration.

A language $L \subseteq \Sigma^*$ is accepted with *bounded error probability* if for some constant $\epsilon < 1/2$,

1. for all $w \in L$, there exists a strategy S_w on which the automaton accepts with probability $\geq 1 - \epsilon$, and
2. for all $w \notin L$, on every strategy S_w, the automaton accepts with probability $\leq \epsilon$.

Language acceptance could be defined with respect to a more general type of strategy, in which the nondeterministic choice made from the same configuration at different times may be different. It is known (see [4, Theorem 2.6]) that if L is accepted by an npfa with respect to this more general definition, then it is also accepted with respect to the definition above. Hence, our results also hold for such generalized strategies.

A *one-way nondeterministic probabilistic finite automaton* (1npfa) is a 2npfa which can never move its input head to the left; that is, $\delta(q, \sigma, q', -1) = 0$ for all q, q', and σ. Also, a *probabilistic finite automaton* (pfa) and a *nondeterministic finite automaton* (nfa) are special cases of an npfa in which there are no nondeterministic and no probabilistic states, respectively.

We denote by 1NPFA and 2NPFA the classes of languages accepted with bounded error probability by 1npfa's and 2npfa's, respectively. If, on all inputs w and all nondeterministic strategies, the 2npfa halts in polynomial expected time, we say that L is in the class 2NPFA-polytime. The classes 1PFA, 2PFA and 2PFA-polytime are defined similarly, with pfa replacing npfa. Finally, Regular denotes the class of regular languages.

Our model of the 2npfa is equivalent to an Arthur-Merlin game in which Arthur is a 2pfa, and our classes 2NPFA and 2NPFA-polytime are identical to the classes AM(2pfa) and AM(ptime-2pfa), respectively, of Dwork and Stockmeyer [8].

2.2 The Tiling Complexity of a Language

We adapt the notion of the tiling complexity of a function, used in communication complexity theory, to obtain a new measure of the complexity of a language. Given a finite, two-dimensional matrix M, a *tile* is a submatrix of M in which all entries have the same value. A tile is specified by a pair (R, C) where R is a nonempty set of rows and C is a nonempty set of columns. The entries in the tile are said to be *covered* by the tile. A tile is a b-tile if all entries of the submatrix are b. A set of b-tiles is a *b-tiling* of M if every b-valued entry of M is covered by at least one tile in the set. If M is a binary matrix, the union of a 0-tiling and a 1-tiling of M is called a *tiling* of M. Let $T(M)$ be the minimum size of a tiling of M. Let $T^1(M)$ be the minimum size of a 1-tiling of M, and let $T^0(M)$ be the minimum size of a 0-tiling of M. Then,

$T(M) = T^1(M) + T^0(M)$. Note that in these definitions it is permitted for tiles of the same type to overlap.

We can now define the tiling complexity of a language. Associated with a language L over alphabet Σ is an infinite binary matrix M_L. The rows and columns of M_L are indexed (say, in lexicographic order), by the strings in Σ^*. Entry $M_L[x, y] = 1$ if and only if $xy \in L$. Let L_n be the strings of L of length $\leq n$. Let $M_L(n)$ be the finite submatrix of M_L whose rows and columns are indexed by the strings of length $\leq n$. The *1-tiling complexity* of a language L is defined to be the function $T^1_L(n) = T^1(M_L(n))$. Similarly, the *0-tiling complexity* of L is $T^0_L(n) = T^0(M_L(n))$ and the *tiling complexity* of L is $T_L(n) = T(M_L(n))$.

A tiling of a matrix M is *disjoint* if every entry $[x, y]$ of M is covered by exactly one tile. The disjoint tiling complexity of a matrix M, $\tilde{T}(M)$, is the minimum size of a disjoint tiling of M. Also, the disjoint tiling complexity of a language, $\tilde{T}_L(n)$, is $\tilde{T}(M_L(n))$.

Tilings are often used in proving lower bounds in communication complexity. Let $f : X \times Y \to \{0, 1\}$. The function f is represented by a matrix M_f whose rows are indexed by elements of X and whose columns are indexed by elements of Y, such that $M_f[x, y] = f(x, y)$. Let T_f denote $T(M_f)$. Suppose that two cooperating parties, P_1 and P_2, get inputs $x \in X$ and $y \in Y$ respectively, and want to compute $f(x, y)$. They can do so by exchanging information according to some protocol (precise definitions of legal protocols can be found in [13]). If the protocol is deterministic, then the worst case number of bits that need to be exchanged (that is, the deterministic communication complexity) is bounded below by $\log \tilde{T}_f$[29]. If the protocol is non-deterministic, then the lower bound is $\log T_f$ [1]. Finally, if the object of the non-deterministic protocol is only to verify that $f(x, y) = 1$ (if that is indeed the case), then the lower bound on the number of bits exchanged is $\log T^1_f$.

2.3 Related Work

Our work on npfa's builds on a rich literature on probabilistic finite state automata. Rabin [24] was the first to consider probabilistic automata with bounded error probability. He showed that 1PFA = Regular. However, with a 2-way input head, pfa's can recognize nonregular languages. This was shown by Freivalds [10], who constructed a 2pfa for the language $\{0^n1^n \mid n \geq 0\}$. Greenberg and Weiss [12] showed that exponential expected time is required by any 2pfa accepting this language. Dwork and Stockmeyer [7] and independently Kaneps and Freivalds [17] showed that in fact any 2pfa which recognizes a nonregular language must run in exponential expected time. It follows that 2PFA-polytime = Regular.

Roughly, Rabin's proof shows that any language L accepted by a 1pfa has only finitely many equivalence classes. Here, two strings x, x' are equivalent if and only if for all y, $xy \in L \Leftrightarrow x'y \in L$. The Myhill-Nerode theorem [14] states that a language has a finite number of equivalence classes if and only if it is regular. This, combined with Rabin's result, implies that 1PFA = Regular. Two decades later, this idea was extended to 2pfa's. A strengthened version

of the Myhill-Nerode theorem is needed for this extension. Given a language L, we say that two strings x, x' are *pairwise n-inequivalent* if for some y, $xy \in L \Leftrightarrow x'y \notin L$, and furthermore, $|xy|, |x'y| \leq n$. Let $N_L(n)$ (the *nonregularity* of L) be size of the largest set of pairwise n-inequivalent strings. Kaneps and Freivalds [16] showed that $N_L(n) \geq \lfloor (n+3)/2 \rfloor$ for infinitely many n. (It is interesting to note that to prove their bound, Kaneps and Freivalds first showed that $N_L(n)$ equals the number of states of the minimal deterministic 1-way finite automaton that accepts all words of length $\leq n$ that are in L and rejects all words of length $\leq n$ that are not in L. Following Karp [19], we denote the latter measure by $\phi_L(n)$. Karp [19] previously proved that $\phi_L(n) > n/2 + 1$ for infinitely many n. Combining this with the fact that $N_L(n)$ and $\phi_L(n)$ are equal, it follows immediately that $N_L(n) > n/2 + 1$ for infinitely many n. This is stronger (by 1) for even n than Kaneps and Freivalds' lower bound. We also note that Dwork and Stockmeyer [7] obtained a weaker bound on $N_L(n)$ without using $\phi_L(n)$.) Using tools from Markov chain theory, Dwork and Stockmeyer [7] and Kaneps and Freivalds [17] showed that if a language is accepted by a 2pfa in polynomial expected time, then the language has "low" nonregularity. In fact, $N_L(n)$ is bounded by some polynomial in $\log n$. This, combined with the result of Kaneps and Freivalds, implies that 2PFA-polytime = Regular.

Models of computation with both nondeterministic and probabilistic states have been studied intensively since the work of Papadimitriou [23] on games against nature. Babai and Moran [3] defined Arthur-Merlin games to be Turing machines with both nondeterministic and probabilistic states, which accept their languages with bounded error probability. Their work on polynomial time bounded Arthur-Merlin games laid the framework for the remarkable progress on interactive proof systems and their applications (see for example [2] and the references therein). Space bounded Arthur-Merlin games were first considered by Condon and Ladner [6]. Condon [4] showed that AM(log-space), that is, the class of languages accepted by Arthur-Merlin games with logarithmic space, is equal to the class P. However, it is not known whether the class AM(log-space, polytime) — the subclass of AM(log-space) where the verifier is also restricted to run in polynomial time — is equal to P, or whether it is closed under complement. Fortnow and Lund [9] showed that NC is contained in AM(log-space,poly-time).

Dwork and Stockmeyer [8] were the first to consider npfa's, which are Arthur-Merlin games restricted to constant space. They described conditions under which a language is not in the classes 2NPFA or 2NPFA-polytime. The statements of our Theorems 3.2 and 3.3 generalize and simplify the statements of their theorems, and our proofs build on theirs. In communication complexity theory terms, their proofs roughly show that languages accepted by npfa's have low "fooling set complexity". This measure is defined in a manner similar to the tiling complexity of a language, based on the following definition. Define a 1-fooling set of a binary matrix A to be a set of entries $\{[x_1, y_1], [x_2, y_2], \ldots, [x_m, y_m]\}$ such that $A[x_i, y_j] = 1$ if and only if $i = j$. The size of a 1-fooling set of a binary matrix is always at most the 1-tiling complexity of the matrix, because no two distinct entries in the 1-fooling set, $[x_i, y_i]$ and $[x_j, y_j]$, can be in the same tile. However, the 1-tiling complexity may be significantly larger than the 1-fooling set complexity; in fact, for a random $n \times n$ binary matrix, the expected size of the largest 1-fooling set is $O(\log n)$ whereas the expected number of tiles needed to tile the 1-entries is $\Omega(n/\log n)$

[1].

3 NPFA's and Tiling

Three results are presented in this section. For each of the classes 1NPFA, 2NPFA and 2NPFA-polytime, we describe upper bounds on the tiling complexity of the languages in these classes. The proof for 1NPFA's is a natural generalization of Rabin's proof that 1PFA = Regular [24]. The other two proofs build on previous results of Dwork and Stockmeyer [8] on 2npfa's.

3.1 1NPFA and Tiling

Theorem 3.1 *A language L is in 1NPFA only if the 1-tiling complexity of L is $O(1)$.*

Proof: Suppose L is accepted by some 1npfa M with error probability $\epsilon < 1/2$. Let the states of M be $\{1, \ldots, c\}$.

Consider the matrix M_L. For each 1-entry $[x, y]$ of M_L, fix a nondeterministic strategy that causes the string xy to be accepted with probability at least $1-\epsilon$. With respect to this strategy, define two vectors of dimension c. Let $\mathbf{p}_{xy}$ be the *state probability vector* at the step when the input head moves off the right end of x. That is, the i'th entry of the vector is the probability of being in state i at that moment, assuming that the automaton is started at the left end of the input $\not{c}xy\$$ in the initial state. Let $\mathbf{r}_{xy}$ be the column vector whose i'th entry is the probability of accepting the string xy, assuming that the automaton is in state i at the moment that the head moves off the right end of x. Then the probability of accepting the string xy is the inner product $\mathbf{p}_{xy} \cdot \mathbf{r}_{xy}$.

Let $\mu = (1/2 - \epsilon)/c$. Partition the space $[0, 1]^c$ into *cells* of size $\mu \times \mu \times \cdots \times \mu$ (the final entry in the cross product should actually be less than μ if 1 is not a multiple of μ). Associate each 1-entry $[x, y]$ with the cell containing the vector $\mathbf{p}_{xy}$; we say that $[x, y]$ belongs to this cell.

With each cell C, associate the *rectangle* R_C defined as

$$\{x | \text{ there exists } y \text{ such that } [x, y] \text{ belongs to } C\} \\ \times \\ \{y | \text{ there exists } x \text{ such that } [x, y] \text{ belongs to } C\}.$$

This is the minimal submatrix that covers all of the entries associated with cell C.

We claim that R_C is a valid 1-tile — that is, R_C covers only 1-entries. To see this, suppose $[x, y] \in R_C$. If $[x, y]$ belongs to C, then it must be a 1-entry. Otherwise, there exist x' and y' such that $[x, y']$ and $[x', y]$ belong to C; that is, $xy', x'y \in L$ and $\mathbf{p}_{xy'}$ and $\mathbf{p}_{x'y}$ are in the same cell.

We claim that xy is accepted with probability at least $1/2$ on some strategy, namely the strategy that while reading x, uses the strategy for xy', and while reading y, uses the strategy for $x'y$. To see this, note that

$$\begin{aligned}(\mathbf{p}_{x'y} - \mathbf{p}_{xy'}) \cdot \mathbf{r}_{x'y} &= \sum_{i=1}^{c} [\mathbf{p}_{x'y} - \mathbf{p}_{xy'}]_i [\mathbf{r}_{x'y}]_i \\ &\leq \mu \sum_{i=1}^{c} [\mathbf{r}_{x'y}]_i \\ &\leq \mu c \\ &= 1/2 - \epsilon, \text{ by our choice of } \mu.\end{aligned}$$

Hence, the probability that xy is accepted on the strategy described above is

$$\begin{aligned}\mathbf{p}_{xy'}\mathbf{r}_{x'y} &\geq \mathbf{p}_{x'y}\mathbf{r}_{x'y} - (1/2 - \epsilon) \\ &\geq (1 - \epsilon) - (1/2 - \epsilon) \\ &= 1/2 > \epsilon.\end{aligned}$$

Because xy is accepted with probability greater than ϵ on this strategy, it cannot be that $xy \notin L$. Hence, for all $[x, y] \in R_C$, xy must be in L. Therefore R_C is a 1-tile in M_L.

Every 1-entry $[x, y]$ is associated with some cell C, and is covered by the 1-tile R_C that is associated with C. Thus, every 1-entry of M_L is covered by some R_C.

Hence L can be 1-tiled using one tile per cell, which is a total of $\lceil 1/\mu \rceil^c = O(1)$ tiles. □

3.2 2NPFA and Tiling

We next show that if $L \in$ 2NPFA, then $T_L^1(n)$ is bounded by a polynomial.

Theorem 3.2 *A language L is in 2NPFA only if the 1-tiling complexity of L is bounded by a polynomial in n.*

Proof: Suppose L is accepted by some 2npfa M with error probability $\epsilon < 1/2$. Let c be the number of states of M. As in Theorem 3.1, for each 1-entry $[x, y]$ of $M_L(n)$, fix a nondeterministic strategy that causes M to accept the string xy with probability at least $1 - \epsilon$.

We construct a stationary Markov chain H_{xy} that models the computation of M on xy using this strategy.

This Markov chain has $d = 2c + 4$ states. $2c$ of the states are labeled (q, l), where q is a state of M and $l \in \{0, 1\}$. The other states are labeled Initial, Accept, Reject, and Loop. The state $(q, 0)$ of H_{xy} corresponds to M being in state q while reading the rightmost symbol of $\not\in x$. The state $(q, 1)$ of H_{xy} corresponds to M being in state q while reading the leftmost symbol of $y\$$. The state Initial corresponds to the initial configuration of M. The states Accept, Reject, and Loop are sink states of H_{xy}.

A single step of the Markov chain H_{xy} corresponds to running M on input xy (using the fixed nondeterministic strategy) from the appropriate configuration for one or more steps, until M enters a configuration corresponding to one of the chain states (q, l). If M halts in the accepting (resp., rejecting) state before entering one of these configurations, H_{xy} enters the Accept (resp., Reject) state. If M does not halt and never again reads the rightmost symbol of $\not\in x$ or the leftmost symbol of $y\$$, then H_{xy} enters the Loop state. The transition probabilities are defined accordingly.

Consider the transition matrix of H_{xy}. Collect the rows corresponding to the chain states Initial and $(q, 0)$ (for all q) and call this submatrix P_{xy}. Collect the rows corresponding to the chain states $(q, 1)$ and call this submatrix R_{xy}. Then the transition matrix looks like this:

$$H_{xy} = \begin{array}{r|cc} \text{Initial} & & \\ (q,0) & P_{xy} & \\ \hline (q,1) & R_{xy} & \\ \hline \begin{array}{r}\text{Accept}\\ \text{Reject}\\ \text{Loop}\end{array} & \begin{array}{c}0\\0\\0\end{array} & \begin{array}{c}0\\0\\I_3\end{array} \end{array}$$

where I_3 denotes the identity matrix of size 3. (We shall engage in a slight abuse of notation by using H_{xy} to refer both to the transition matrix and to the Markov chain itself.) Note that the entries of P_{xy} depend only on x and the nondeterministic strategy used; these transition probabilities do not depend on y. This assertion appears to be contradicted by the fact that our choice of nondeterministic strategy may depend on y; however, the idea here is that if we replace y with y' while maintaining the same nondeterministic strategy we used for xy, then $P_{xy'}$ will be identical to P_{xy}, because the transitions involved simulate computation of M on the left part of its input only. Similarly, R_{xy} depends only on y and the strategy, and not on x.

We now show that if $|x| \leq n$ and if p is a nonzero element of P_{xy}, then $p \geq 2^{-cn-1}$. Form a second Markov chain $K(\not\in x)$ with states of the form (q, l), where q is a state of M and $1 \leq l \leq |\not\in x| + 1$. The chain state (q, l) with $l \leq |\not\in x|$ corresponds to M being in state q scanning the l'th symbol of $\not\in x$. Transition probabilities from these states are obtained from the transition probabilities of M in the obvious way. Chain states of the form $(q, |\not\in x| + 1)$ are sink states of $K(\not\in x)$ and correspond to the head of M falling off the right end of $\not\in x$ with M in state q. Now consider a transition probability p in P_{xy}. Suppose that, in the Markov chain H_{xy}, p is the transition probability from $(q, 0)$ to $(q', 1)$. Then $p \in \{0, 1/2, 1\}$, since if H_{xy} makes this transition, it must be simulating a single computation step of M. Suppose p is the transition probability from $(q, 0)$ to $(q', 0)$. If $p > 0$, then there must be some path of nonzero probability in $K(\not\in x)$ from state $(q, |\not\in x|)$ to $(q', |\not\in x|)$ that visits no state $(q'', |\not\in x|)$, and since $K(\not\in x)$ has at most cn states that can be on this path, there must be such a path of length at most $cn + 1$. Since $1/2$ is the smallest nonzero transition probability of M, it follows that $p \geq 2^{-cn-1}$. The cases where p is a transition probability from the Initial state are similar.

Similarly, if $|y| \leq n$ and if r is a nonzero element of R_{xy}, then $r \geq 2^{-cn-1}$.

Next we present a lemma which bounds the effect of small changes in the transition probabilities of a Markov chain. This lemma is a slight restatement of a lemma of Greenberg and Weiss [12]. This version is due to Dwork and Stockmeyer [8].

If k is a sink state of a Markov chain R, let $a(k,R)$ denote the probability that R is (eventually) trapped in state k when started in state 1. Let $\beta \geq 1$. Say that two numbers r and r' are *β-close* if either (i) $r = r' = 0$ or (ii) $r > 0, r' > 0$, and $\beta^{-1} \leq r/r' \leq \beta$. Two Markov chains $R = \{r_{ij}\}_{i,j=1}^{s}$ and $R' = \{r'_{ij}\}_{i,j=1}^{s}$ are *β-close* if r_{ij} and r'_{ij} are β-close for all pairs i, j.

Lemma 3.1 *Let R and R' be two s-state Markov chains which are β-close, and let k be a sink state of both R and R'. Then $a(k,R)$ and $a(k,R')$ are β^{2s}-close.*

The proof of this lemma is based on the Markov chain tree theorem of Leighton and Rivest [20], and can be found in [8].

Our approach is to partition the 1-entries of $M_L(n)$ into equivalence classes, as in the proof of Theorem 3.1, but this time we will make entries $[x,y]$ and $[x',y']$ equivalent only if the corresponding Markov chains H_{xy} and $H_{x'y'}$ are β-close, where β will be chosen small enough that we can use Lemma 3.1 to show that xy' and $x'y$ are accepted with high probability by combining the strategies for xy and $x'y'$.

If $[x,y]$ is a 1-entry such that $|x| \leq n$ and $|y| \leq n$, then for any nonzero p of P_{xy} (or r of R_{xy}), $p \in [2^{-cn-1}, 1]$, so $\log_2 p \in [-cn-1, 0]$ (and similarly $\log_2 r \in [-cn-1, 0]$).

By partitioning each coordinate interval $[-cn-1, 0]$ into subintervals of length μ, we divide the space $[-cn-1, 0]^{d^2}$ into at most $\lceil (cn+1)/\mu \rceil^{d^2}$ cells, each of size at most $\mu \times \mu \times \cdots \mu$.

Partition the 1-entries in $M_L(n)$ into equivalence classes by making xy and $x'y'$ equivalent if H_{xy} and $H_{x'y'}$ have the property that for each state transition, if p and p' are the respective transition probabilities, either $p = p' = 0$, or $\log p$ and $\log p'$ are in the same (size μ) subinterval of $[-cn-1, 0]$

Note that the number of equivalence classes is at most $(\lceil (cn+1)/\mu \rceil + 1)^{d^2}$.

We claim that if μ is chosen small enough, these equivalence classes induce a 1-tiling of $M_L(n)$ of size at most the number of equivalence classes. As in Theorem 3.1, we associate with each equivalence class C the rectangle R_C defined by

$$\{x | \text{there exists } y \text{ such that } [x,y] \in C\} \times \{y | \text{ there exists } x \text{ such that } [x,y] \in C\}.$$

We claim that for each $[x,y]$ in R_C, $xy \in L$. That is, all entries in the rectangle are 1, so the rectangle forms a 1-tile. Let $[x,y]$ be in R_C. There must be some y' such that $[x,y'] \in C$ and some x' such that $[x',y] \in C$. Consider the associated Markov chains $H_{xy'}$ and $H_{x'y}$, and in particular, consider the transition submatrices $P_{xy'}$ and $R_{x'y}$. The first is associated with a particular nondeterministic strategy on x, namely one which assumes the input is xy' and tries to cause xy' to be accepted with high probability. The second is associated with a particular

nondeterministic strategy on y, namely one which assumes the input is $x'y$ and tries to cause $x'y$ to be accepted with high probability. The two matrices $P_{xy'}$ and $R_{x'y}$ taken together correspond to a hybrid strategy on xy: while reading x, use the strategy for xy', and while reading y, use the strategy for $x'y$. We will argue that this hybrid strategy causes xy to be accepted with probability $\geq 1/2$.

We construct a hybrid Markov chain H_{xy} using $P_{xy'}$ and $R_{x'y}$. This chain models the computation of M on xy using the hybrid strategy.

Since the 1-entries $[x, y']$ and $[x', y]$ are in the same equivalence class C, it follows that if p and p' are corresponding transition probabilities in the Markov chains $H_{xy'}$ and $H_{x'y}$, then either $p = p' = 0$ or $|\log p - \log p'| \leq \mu$. Therefore, $H_{xy'}$ and $H_{x'y}$ are 2^{μ}-close, and it immediately follows that H_{xy} is 2^{μ}-close to $H_{xy'}$ (and to $H_{x'y}$). Let $a_{xy'}$ be the probability that M accepts input xy' on the strategy for xy', and let a_{xy} be the probability that M accepts input xy using the hybrid strategy. Then $a_{xy'}$ (resp., a_{xy}) is exactly the probability that the Markov chain $H_{xy'}$ (resp., H_{xy}) is eventually trapped in the Accept state, when started in the Initial state. Now $xy' \in L$ implies $a_{xy'} \geq 1 - \epsilon$. Since H_{xy} and $H_{xy'}$ are 2^{μ}-close, Lemma 3.1 implies that

$$\frac{a_{xy}}{a_{xy'}} \geq 2^{-2d\mu}$$

which implies

$$a_{xy} \geq (1 - \epsilon)2^{-2d\mu}.$$

Since ϵ and d are constants, and since $\epsilon < 1/2$, we can choose μ to be a constant so small that $a_{xy} \geq 1/2$. Therefore xy must be in L.

Since each 1-entry $[x, y]$ is in some equivalence class, the matrix $M_L(n)$ can be 1-tiled using at most $(\lceil (cn+1)/\mu \rceil + 1)^{d^2}$ tiles. Therefore,

$$T_L^1(n) \leq (\lceil (cn+1)/\mu \rceil + 1)^{d^2}.$$

Since c, d, and μ are constants independent of n, this shows that $T_L^1(n)$ is bounded by a polynomial in n. □

3.3 2NPFA-polytime and Tiling

We now show that if $L \in$ 2NPFA-polytime, then $T_L^1(n)$ is bounded by a polylog function.

Theorem 3.3 *A language L is in 2NPFA-polytime only if the 1-tiling complexity of L is bounded by a polynomial in* $\log n$.

Proof: Suppose L is accepted by some 2npfa M with error probability $\epsilon < 1/2$ in expected time at most $t(n)$. Let c be the number of states of M. For each 1-entry $[x, y]$ of $M_L(n)$, fix a nondeterministic strategy that causes M to accept the string xy with probability at least $1 - \epsilon$.

We construct the Markov chain H_{xy} just as in Theorem 3.2.

Say that a probability p is *small* if $p < t(n)^{-2}$; otherwise, p is *large*. Note that if p is a large transition probability, then $p \in [t(n)^{-2}, 1]$, so $\log_2 p \in [-2\log_2 t(n), 0]$. When dividing the 1-entries of $M_L(n)$ into equivalence classes, make xy and $x'y'$ equivalent if H_{xy} and $H_{x'y'}$ have the property that for each state transition, if p and p' are the respective transition probabilities, either p and p' are both small, or $\log p$ and $\log p'$ are in the same (size μ) subinterval of $[-2\log_2 t(n), 0]$.

This time the number of equivalence classes is at most $(\lceil 2\log_2 t(n)/\mu \rceil + 1)^{d^2}$.

Model the computation of M on inputs $x'y$, xy', and xy by Markov chains $H_{x'y}$, $H_{xy'}$, and H_{xy}, respectively, as before.

If p and p' are corresponding transition probabilities in any two of these Markov chains, then either p and p' are 2^{μ}-close or p and p' are both small. Let $\mathcal{E}_{x'y}$ be the event that, when $H_{x'y}$ is started in state Initial, it is trapped in state Accept or Reject before any transition labeled with a small probability is taken; define $\mathcal{E}_{xy'}$ and $\mathcal{E}_{xy}$ similarly. Since M halts in expected time at most $t(n)$ on the inputs $x'y$, xy', and xy, the probabilities of these events go to 1 as n increases. Therefore, by changing all small probabilities to zero, we do not significantly change the probabilities that $H_{x'y}$, $H_{xy'}$, and H_{xy} enter the Accept state, provided that n is sufficiently large. A formal justification of this argument can be found in Dwork and Stockmeyer [8].

After these changes, we can argue that

$$a_{xy} \geq (1-\epsilon)2^{-2d\mu}$$

and choose μ so that $a_{xy} \geq 1/2$, as before. It then follows that

$$T_L^1(n) \leq (\lceil 2\log_2 t(n)/\mu \rceil + 1)^{d^2} \tag{1}$$

for all sufficiently large n, establishing the result. □

4 Bounds on the Tiling Complexity of Languages

In this section, we obtain several bounds on the tiling complexity of regular and nonregular languages. In Section 4.1, we prove several elementary results. First, all regular languages have constant tiling complexity. Second, the 1-tiling complexity of all nonregular languages is at least $\log n - O(1)$ infinitely often. We also present an example of a (unary) non-regular language which has 1-tiling complexity $O(\log n)$. In Section 4.2, we use a rank argument to show that for all nonregular languages L, either L or its complement has "high" 1-tiling complexity infinitely often.

4.1 Simple Bounds on the Tiling Complexity of Languages

The following lemma is useful in proving some of the theorems in this section. Its proof is implicit in work of Melhorn and Schmidt [21]; we include it for completeness.

Lemma 4.1 *Any binary matrix A that can be 1-tiled with m tiles has at most 2^m distinct rows.*

Proof: Let A be a binary matrix that can be 1-tiled by m tiles $\{T_1, \ldots, T_m\}$, where $T_j = (R_j, C_j)$. For each row r of A, let $I(r) = \{T_j \mid j \in \{1, \ldots, m\}$ such that $r \in R_j\}$. Suppose r_1 and r_2 are rows such that $I(r_1) = I(r_2)$. We show that in this case, rows r_1 and r_2 are identical. To see this, consider any column c of A. Suppose that entry $[r_1, c]$ has value 1, and is covered by some tile $T_j \in I(r_1)$. Therefore, $c \in C_j$. Since $I(r_1) = I(r_2)$, $T_j \in I(r_2)$ and therefore $r_2 \in R_j$ and $[r_2, c]$ is covered by tile T_j. Hence entry $[r_2, c]$ must have value 1, since T_j is a 1-tile. Hence, if $[r_1, c]$ has value 1, so does $[r_2, c]$. Similarly, if $[r_2, c]$ has value 1, then so does entry $[r_1, c]$. Therefore r_1 and r_2 are identical rows. Since there are only 2^m possible values for $I(r)$, A can have at most 2^m distinct rows. □

Theorem 4.1 *The 1-tiling complexity of L is $O(1)$ if and only if L is regular.*

Proof: By the Myhill-Nerode theorem [14, Theorem 3.6], L is regular if and only if M_L has a finite number of distinct rows.

Suppose L is regular. Then by the above fact there exists a constant k such that M_L has at most k distinct rows. Consider any (possibly infinite) set R of identical rows in M_L. Let C_b be the set of columns which have bit b in the rows of R, for $b = 0, 1$. Then the subset specified by (R, C_b) is a b-tile and covers all the b-valued entries in the rows of R. It follows that the 1-valued entries of R can be covered by a single tile, and hence there is a 1-tiling of $M_L(n)$ of size k. (Similarly, there is a 0-tiling of $M_L(n)$ of size k.)

Suppose L is not regular. Since L is not regular, M_L has an infinite number of distinct rows. It follows immediately from Lemma 4.1 that M cannot be tiled with any constant number of tiles. □

The above theorem uses the simple fact that the 1-tiling complexity $T_L^1(n)$ of a language L is a lower bound on the number of distinct rows of $M_L(n)$. In fact, the number of distinct rows of $M_L(n)$, for a language L, is closely related to a measure that has been previously studied by many researchers. Dwork and Stockmeyer called this measure *non-regularity*, and denoted the non-regularity of L by $N_L(n)$ [7]. $N_L(n)$ is the maximum size of a set of *n-dissimilar* strings of L. Two strings, w and w', are considered n-dissimilar if $|w| \leq n$ and $|w'| \leq n$, and there exists a string v such that $|wv| \leq n$, $|w'v| \leq n$, and $wv \in L$ if and only if $w'v \notin L$. It is easy to show that the number of distinct rows of $M_L(n)$ is between $N_L(n)$ and $N_L(2n)$. Previously, Kaneps and Freivalds [16] showed that $N_L(n)$ is equal to the number of states of the minimal 1-way deterministic finite state automaton which accepts a language L' for which $L'_n = L_n$, where L_n is the set of strings of L of length $\leq n$.

Shallit [28] introduced a similar measure: the *nondeterministic nonregularity* of L, denoted by $NN_L(n)$, is the minimal number of states of a 1-way nondeterministic finite automaton which accepts a language L' for which $L'_n = L_n$. In fact, it is not hard to show that

$$T^1_L(n) \leq NN_L(2n)$$

To see this, suppose that M is an automaton with $NN_L(2n)$ states, which accepts a language L' for which $L'_{2n} = L_{2n}$. We construct a 1-tiling of $M_L(n)$ with one tile T_q per state q of M, where entry $[x, y]$ is covered by T_q if and only if there is an accepting path of M on xy which enters state q as the head falls off the rightmost symbol of x. It is straightforward to verify the set of tiles defined in this way is indeed a valid 1-tiling of $M_L(n)$. A similar argument was used by Schmidt [27] to prove lower bounds on the number of states in an unambiguous nfa.

We next turn to simple lower bounds on the 1-tiling complexity of nonregular languages. From Theorem 4.1, it is clear that if L is nonregular, then $T^1_L(n)$ is unbounded. We now use a known lower bound on the nonregularity of nonregular languages to prove a lower bound for $T^1_L(n)$.

Theorem 4.2 *If L is not regular, then $T^1_L(n) \geq \log_2 n - 1$ for infinitely many n.*

Proof: Kaneps and Freivalds [16] proved that if L is not regular, then $N_L(n) \geq \lfloor (n+3)/2 \rfloor$ for infinitely many n. By the definition of $N_L(n)$, the matrix $M_L(n)$ must have at least $N_L(n)$ distinct rows. Therefore, by Lemma 4.1, $T^1_L(n) \geq \log_2 N_L(n)$. The lemma follows immediately.

□

We next present an example of a unary nonregular language, with 1-tiling complexity $O(\log n)$. Thus, the lower bound of Theorem 4.2 is optimal to within a constant factor.

Theorem 4.3 *Let L be the complement of the language $\{a^{2^k-1} \mid k > 0\}$. Then, L has 1-tiling complexity $O(\log n)$.*

Proof: We show that the 1-valued entries of $M_L(n)$ can be covered with $O(\log n)$ 1-tiles. Let $\lg n$ denote $\lfloor \log_2 n \rfloor + 1$, and let $\lg 0 = 0$. Let x and y be binary numbers, of length at most $\lg n$. Number the bits of these numbers from right to left, starting with 1, so that for example $y = y_{\lg n} \ldots y_2 y_1$. For any binary number q, $\lg q$ is the maximum index i such that $q_i = 1$ ($\lg q = 0$ if $q = 0$).

Clearly if q is equal to $2^k - 1$ for some integer $k > 0$, then for all indices $i, 1 \leq i \leq \lg q$, $q_i = 1$. The next fact follows easily.

Fact: $x + y = 2^k - 1$ for some integer $k > 0$ if and only if for all j such that $j \leq \max\{\lg x, \lg y\}$, $x_j \neq y_j$.

Roughly, we construct a 1-tiling of $M_L(n)$, corresponding to the following nondeterministic communication protocol. The party P_1 guesses an index j and sends j and x_j to P_2. Also P_1

sends P_2 one bit indicating whether or not $j \leq \lg x$. If $j \leq \lg x$, then P_2 checks that $y_j = x_j$. If $j > \lg x$, P_2 checks that $j \leq \lg y$ and that $y_j = x_j$, or equivalently, that $y_j = 0$. In either case, P_2 can conclude that $y_j = x_j$, and so entry $[a^x, a^y]$ of $M_L(n)$ is 1. The number of bits sent from P_1 to P_2 is $\lg \lg n + 2$.

We now describe the 1-tiling corresponding to this protocol. It is the union of two sets of tiles. The first set has one tile $T_{j,b}$ for each j, b such that $\lg n \geq j \geq 0$ and $b \in \{0, 1\}$, where

$$T_{j,b} = \{a^x \mid 0 \leq x \leq n, \lg x \geq j, x_j = b\} \times \{a^y \mid 0 \leq y \leq n, y_j = b\}.$$

The second set of tiles has one tile $S_{j,0}$, for all j such that $\lceil \log n \rceil \geq j \geq 1$.

$$S_{j,0} = \{a^x \mid 0 \leq x \leq n, \lg x < j, x_j = 0\} \times \{a^y \mid 0 \leq y \leq n, \lg y \geq j, y_j = 0\}.$$

To see that all the 1's in the matrix are covered by one of these tiles, note that if entry $[a^x, a^y]$ of the matrix is 1, then by the Fact, there exists an index j such that $j \leq \max\{\lg x, \lg y\}$ and either $x_j = y_j = 1$, or $x_j = y_j = 0$. So, for example, if $\lg x \geq \lg y$, and j is such that $j \leq \lg x$ and $x_j = y_j = 0$, then entry $[a^x, a^y]$ is covered by tile $T_{j,0}$. □

The nondeterministic communication protocol in the above proof is a slight variation of a simple (and previously known) protocol for the complement of the set distinctness problem. In the set distinctness problem, the two parties each hold a subset of $\{1, \ldots, m\}$, and they must determine whether the subsets are distinct. In our application, the problem is to determine, for $m = \max\{\lg x, \lg y\}$, whether the subset of $\{1, \ldots, m\}$ whose corresponding values in x are 0, is distinct from the subset of $\{1, \ldots, m\}$ whose corresponding values in y are 1.

4.2 Lower Bounds on the Tiling Complexity of Nonregular Languages

In this section we prove that if a language L is nonregular, then the 1-tiling complexity of either L or $\bar{L}$ is "high" infinitely often. To prove this, we first prove lower bounds on the rank of M_L when L is nonregular. We then apply theorems from communication complexity relating rank to tiling complexity.

The proofs of the lower bounds on the rank of M_L are heavily dependent on distinctive structural properties of M_L. Consider first the case where L is a unary language over the alphabet $\Sigma = \{a\}$. In this case, for all i, j where $j > 1$, $a^i a^j = a^{i+1} a^{j-1}$, and therefore $M_L[a^i, a^j] = M_L[a^{i+1}, a^{j-1}]$. It follows that for every n, $M_L(n)$ is such that its auxiliary diagonal (the diagonal from the top right to the bottom left) consists of equal elements, as do all diagonals parallel to that diagonal. An example is shown in Figure 1. Such matrices are classically known as *Hankel matrices*, and have been extensively studied [15]. In fact, a direct application of known results on the rank of Hankel matrices shows that if L is nonregular, then $\text{rank}(M_L(n)) \geq n + 1$ infinitely often. This was first proved by Iohvidov (see [15, Theorem 11.3]), based on previous work of Frobenius [11].

	ϵ	a^1	a^2	a^3	a^4	a^5	a^6
ϵ	1	0	0	1	0	0	1
a^1	0	0	1	0	0	1	0
a^2	0	1	0	0	1	0	0
a^3	1	0	0	1	0	0	1
a^4	0	0	1	0	0	1	0
a^5	0	1	0	0	1	0	0
a^6	1	0	0	1	0	0	1

Figure 1: The Hankel matrix $M_L(6)$ for $L = \{a^i | i \equiv 0 \bmod 3\}$.

If L is a non-unary language, then M_L does not have the simple diagonal structure of a Hankel matrix. Nevertheless, M_L still has structural properties that we are able to exploit. In fact, the term Hankel matrix has been extended from its classical meaning to refer to matrices M_L of non-unary languages (see [26]). In what follows, we generalize the results on the rank of classical Hankel matrices, and prove that for any nonregular language L, over an arbitrary alphabet, $\text{rank}(M_L(n)) \geq n + 1$ infinitely often.

4.2.1 Notation and basic facts

Let L be a language over an arbitrary alphabet, and let $M = M_L$.

Consider a row of M indexed by a string w. This row corresponds to strings that have the prefix w. For any string s, row ws corresponds to strings with the prefix ws. Thus the entries in row ws can be determined by looking at those entries in row w whose columns are indexed by strings beginning with s (see Figure 2). In what follows, we consider this relationship between the rows of M more formally.

Let $M(n, m)$ denote the set of vectors (finite rows) of M which are indexed by strings x of length $\leq n$ and whose columns are indexed by strings of length $\leq m$. Let $\hat{M}(n, m)$ denote the subset of vectors of $M(n, m)$ which are indexed by strings x of length exactly n. If v' is row x of $M(n, m + i)$, where $i > 0$ and v is row x of $M(n, m)$, then v' is called an *extension* of v.

Suppose $v \in M(n, m)$. Let s be a string over Σ of length $\leq m$ (possibly the empty string, ϵ). Define $\text{split}^{(s)}(v)$ to be the subvector formed from v by selecting exactly those columns whose labels have s as a prefix. Also, relabel the columns of $\text{split}^{(s)}(v)$ by removing the prefix s. Note that $\text{split}^{(\epsilon)}(v) = v$. Note also that if Σ is unary, say $\{\sigma\}$, then $\text{split}^{(\sigma)}(v)$ is v with the first column removed. Let $|v|$ denote the dimension (number of entries) of vector v. If Σ is binary and $\sigma \in \Sigma$, then

$$|\text{split}^{(\sigma)}(v)| = (|v| - 1)/2.$$

	ϵ	0	1	00	01	10	11	000	001	010	011	100	101	110	111
ϵ	1	1	1	1	0	0	1	1	0	1	0	0	1	0	1
0	1	1	0	1	0	1	0	1	0	0	0	0	0	1	0
1	1	0	1	0	1	0	1	0	1	0	0	0	0	0	1
00	1	1	0	1	0	0	0	1	0	0	0	1	0	0	0
01	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0
10	0	0	1	0	1	0	0	0	1	0	0	0	1	0	0
11	1	**0**	1	**0**	**0**	0	1	**0**	**0**	**0**	**1**	0	0	0	1
000	1	1	0	1	0	0	0	1	0	0	0	0	0	0	0
001	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0
010	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0
011	0	1	0	0	0	1	0	0	0	0	0	0	0	1	0
100	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0
101	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0
110	**0**	**0**	**0**	**0**	**0**	**0**	**1**	0	0	0	1	0	0	0	0
111	1	0	1	0	0	0	1	0	0	0	0	0	0	0	1

Figure 2: The matrix $M(3)$ for $L = \{w \in \{0,1\}^* | w \text{ is a palindrome}\}$. The bold entries in row 110 are determined by the bold entries in row 11. The bold entries in row 110 comprise $\text{split}^{(0)}(11)$ for $M(2,3)$.

More generally, if $|\Sigma| = c > 1$ and $\sigma \in \Sigma$, then

$$|v| = \frac{c^{m+1} - 1}{c - 1}, \text{ and}$$

$$|\text{ split}^{(\sigma)}(v)| = \frac{|v| - 1}{c} = \frac{c^m - 1}{c - 1}.$$

Also, the vector v consists of the first entry (indexed by the empty string, ϵ), plus an "interleaving" of the entries of $\text{split}^{(\sigma)}(v)$, for each $\sigma \in \Sigma$. More precisely, we have the following fact:

Fact 4.1 *Let $j', s, j \in \Sigma^*$, where $j' = sj$. Then, $v[j'] = \text{split}^{(s)}(v)[j]$.*

We generalize the definition of the split function to sets of vectors. If V is a set of vectors in $M(n,m)$, and $|s| \leq m$, let $\text{split}^{(s)}(V) = \{ \text{split}^{(s)}(v) \mid v \in V\}$. Then we have the following.

Fact 4.2 $\cup_{|s|=i} \text{split}^{(s)}(\hat{\text{M}}(n,m)) = \hat{\text{M}}(n+i, m-i)$. *Thus,*

(a) $\hat{\text{M}}(n+i, m-i) \subseteq \cup_{|s|=i} \text{split}^{(s)}(M(n,m))$, *and*

(b) $\cup_{|s|=i}\text{split}^{(s)}(M(n,m)) = M(n+i, m-i)$.

In what follows, the vectors we consider are assumed to be elements of vector spaces over an arbitrary field **F** (e.g. our proofs will hold if **F** is taken to be the field of rationals **F**). All references to rank, span, and linear independence apply to vector spaces over **F**.

Lemma 4.2 *Suppose that* $b_1, \ldots, b_p \in M(n,m)$ *and that*

$$v = \alpha_1 b_1 + \ldots + \alpha_p b_p,$$

where the α_i *are in the field* **F**. *Suppose that for* $1 \leq k \leq p$, b'_k *is an extension in* $M(n, m+1)$ *of* b_k *and that* v' *is an extension of* v *to the same length as the* b'_k.

Suppose also that for some $i, 0 \leq i \leq m+1$, *it is the case that for all* s *of length* i,

$$\text{split}^{(s)}(v') = \alpha_1 \text{split}^{(s)}(b'_1) + \ldots + \alpha_p \text{split}^{(s)}(b'_p).$$

Then, $v' = \alpha_1 b'_1 + \ldots + \alpha_p b'_p$.

Proof: Clearly, $v'[j] = \alpha_1 b'_1[j] + \ldots + \alpha_p b'_p[j]$, if j is a string of length $\leq m$. Consider a string j' of length $m+1$. Let $j' = sj$, where $|s| = i$. By Fact 4.1,

$$v'[j'] = \text{split}^{(s)}(v')[j].$$

Also,

$$b'_k[j'] = \text{split}^{(s)}(b'_k)[j], \text{ for } 1 \leq k \leq p.$$

By the hypothesis of the lemma,

$$\text{split}^{(s)}(v')[j] = \alpha_1 \text{ split}^{(s)}(b'_1)[j] + \ldots + \alpha_p \text{ split}^{(s)}(b'_p)[j].$$

Putting the last three equalities together, $v'[j'] = \alpha_1 b'_1[j'] + \ldots + \alpha_p b'_p[j']$, as required. □

Let $\text{rank}(M(n,m))$ be the rank of the set of vectors $M(n,m)$ and let $\text{span}(M(n,m))$ be the vector space generated by the vectors in $M(n,m)$. The next lemma follows immediately from the definitions.

Lemma 4.3 *If* $v' \in \text{span}(M(n,m)), m > 0$ *and* $v = \text{split}^{(\sigma)}(v')$, *where* $\sigma \in \Sigma$, *then*

$$v \in \text{span}(\text{split}^{(\sigma)}(M(n,m))).$$

4.2.2 A Lower Bound on the Rank of $M(n)$ when L is Nonregular

A trivial lower bound on the rank of $M(n)$ is given by the following fact.

Fact 4.3 L *is nonregular if and only if there is an infinite sequence of integers* p_r *satisfying* $\text{rank}(M(p_r)) \geq r+1$ *for all integers* r.

This is easily shown using the Myhill-Nerode theorem. Clearly, such a sequence exists if and only if the rank of $M(n)$ (as n increases) is unbounded. Moreover, the rank of $M(n)$ is unbounded if and only if the number of distinct rows in $M(n)$ is unbounded. The Myhill-Nerode theorem states that the number of equivalence classes of L (equivalently, the number of distinct rows of M) is finite if and only if L is regular. It follows that L is nonregular if and only if the rank of $M(n)$ is unbounded. This conclusion has already been noted (see Sections II.3 and II.5 of the book by Salomaa and Soittola [26], which describes results from the literature on rational power series and regular languages).

The above lower bound is very weak. In what follows, we significantly improve it by using the special structure of $M(n)$. Namely, we show that there is an infinite sequence of values of n such that $\text{rank}(M(n)) \geq n+1$. We define the first value of n in our sequence to be the length of the shortest word in L (clearly $\text{rank}(M(n)) \geq n+1$ in this case). To construct the remainder of the sequence, we show (in Lemma 4.5) that because L is nonregular, for any value of n, there is some $m \geq n$ such that $\text{rank}(M(n+1, m+1)) > \text{rank}(M(n, m+1))$. We then prove (in Lemma 4.6 and the proof of Theorem 4.4) that if n is such that $\text{rank}(M(n)) \geq n+1$, and we choose the smallest $m \geq n$ such that $\text{rank}(M(n+1, m+1)) > \text{rank}(M(n, m+1))$, then in fact $\text{rank}(M(m+1)) \geq m+2$.

We begin with the following useful lemma.

Lemma 4.4 *Let $n \geq 0, m \geq 1$. Suppose that $M(n+1, m) \subseteq \text{span}(M(n, m))$. Then, for all $i, 1 \leq i \leq m$, $M(n+i, m-i+1) \subseteq \text{span}(M(n, m-i+1))$.*

Proof: By induction on i. The result is true by hypothesis of the lemma in the case $i = 1$. Suppose $1 < i \leq m$ and that the lemma is true for $i-1$.

It follows from the induction hypothesis that if $v \in M(n+i-1, m-i+2)$, then also $v \in \text{span}(M(n, m-i+2))$. Hence, it must also be the case that if $v \in M(n+i-1, m-i+1)$, then $v \in \text{span}(M(n, m-i+1))$. It remains to consider the vectors in $\hat{M}(n+i, m-i+1)$. By Fact 4.2 (a), each such vector v is of the form $\text{split}^{(\sigma)}(v')$, where $v' \in M(n+i-1, m-i+2)$, for some $\sigma, |\sigma| = 1$. By the inductive hypothesis, $v' \in \text{span}(M(n, m-i+2))$. Hence, by Lemma 4.3, $v \in \text{span}(\text{split}^{(\sigma)}(M(n, m-i+2)))$.

Then, by Fact 4.2 (b), all of the vectors in $\text{split}^{(\sigma)}(M(n, m-i+2))$ are in $M(n+1, m-i+1)$. Hence, $v \in \text{span}(M(n+1, m-i+1))$. Finally, by the hypothesis of the lemma, $\text{span}(M(n+1, m-i+1)) = \text{span}(M(n, m-i+1))$. Hence, $v \in \text{span}(M(n, m-i+1))$, as required. □

Corollary 4.1 *For any $n \geq 0$, if $\text{rank}(M(n+1, 2p)) = \text{rank}(M(n, 2p)) \leq r$ then $\text{rank}(M(p)) \leq r$.*

Proof: If $n \geq p$ then $M(p)$ is a submatrix of $M(n, 2p)$ so the result follows trivially. Otherwise, choose i so that $n+i = p$. Then $M(p)$ is a submatrix of $M(n+i, 2p-i+1)$, and hence by Lemma 4.4, the rows of $M(p)$ are contained in $\text{span}(M(n, p))$. Thus again $\text{rank}(M(p)) \leq r$. □

The following lemma shows the existence of an $m \geq n$ such that $\text{rank}(M(n+1, m+1)) > \text{rank}(M(n, m+1))$.

Lemma 4.5 *Let L be a nonregular language. Then for any n, there exists an $m \geq n$ such that* $\text{rank}(M(n+1, m+1)) > \text{rank}(M(n, m+1))$.

Proof: Let r be the number of strings of length $\leq n$. Clearly, $\text{rank}(M(n, m)) \leq r$ for all m, since there are r rows in $M(n, m)$. Let $p = p_r$ as in Fact 4.3, that is, $\text{rank}(M(p)) \geq r + 1$. Hence, by Corollary 4.1, it must be the case that $\text{rank}(M(n+1, 2p)) > \text{rank}(M(n, 2p))$. Thus, $2p$ is one possible value of m that satisfies the lemma. □

It remains to show that if n is such that $\text{rank}(M(n)) \geq n+1$, and m is the smallest number such that $m \geq n$ and $\text{rank}(M(n+1, m+1)) > \text{rank}(M(n, m+1))$, then $\text{rank}(M(m+1)) \geq m+2$. This is clearly true if for all $i \in [0..m-n]$, $\text{rank}(M(n, m-i)) < \text{rank}(M(n, m-i+1))$, because in this case $\text{rank}(M(n, m+1)) \geq m+2$. The difficult case is when there exist values of i such that $\text{rank}(M(n, m-i)) = \text{rank}(M(n, m-i+1))$, To help deal with this case, we prove the following lemma.

Lemma 4.6 *Suppose that the following properties hold:*

1. $M(n+1, n+1) \subseteq \text{span}(M(n, n+1))$
2. *m is the smallest number $> n$ such that* $M(n+1, m+1) \not\subseteq \text{span}(M(n, m+1))$
3. *i is a number in the range $[0, \ldots, m-n]$ such that*

$$\text{rank}(M(n, m-i)) = \text{rank}(M(n, m-i+1)).$$

Then, there is some vector in $M(n+i+1, m-i+1)$ which is not in $\text{span}(M(n, m-i+1))$.

Proof: Let $v' \in M(n+1, m+1) - \text{span}(M(n, m+1))$, where v' is the extension of some $v \in M(n+1, m)$.

Then, we claim that for some $s, |s| = i$, $\text{split}^{(s)}(v') \notin \text{span}(M(n, m-i+1))$. Since $\text{split}^{(s)}(v') \in M(n+i+1, m-i+1))$ by Fact 4.2 (b), this is sufficient to prove the lemma.

Suppose to the contrary that for all s of length i, $\text{split}^{(s)}(v') \in \text{span}(M(n, m-i+1))$.

Let $\{b_1, \ldots, b_p\}$ be a basis of $M(n, m)$. Let $\{b'_1, \ldots, b'_p\}$ be an extension of this basis in $M(n, m+1)$. By Properties 1 and 2 of the lemma, v is in $\text{span}(M(n, m))$. Let $v = \alpha_1 b_1 + \ldots + \alpha_p b_p$. Then, applying Fact 4.1, we see that for all $s, |s| = i$,

$$\text{split}^{(s)}(v) = \alpha_1 \text{ split}^{(s)}(b_1) + \ldots + \alpha_p \text{ split}^{(s)}(b_p). \tag{2}$$

We want to show that for all s of length i,

$$\text{split}^{(s)}(v') = \alpha_1 \text{ split}^{(s)}(b'_1) + \ldots + \alpha_p \text{ split}^{(s)}(b'_p).$$

It follows from this and from Lemma 4.2 that

$$v' = \alpha_1 b'_1 + \ldots + \alpha_p b'_p,$$

contradicting the fact that $v' \notin \text{span}(M(n, m+1))$.

Consider the vectors $\text{split}^{(s)}(b'_k)$. These are in $M(n+i, m-i+1)$, by Fact 4.2 (b). If $i = 0$, this is clearly in $\text{span}(M(n, m+1))$. If $0 < i \leq m-n$, by Lemma 4.4 and by Property 2 of this lemma, these vectors are in $\text{span}(M(n, m-i+1))$. Let $c_1, \ldots, c_l$ be a basis for $\text{span}(M(n, m-i))$, and for $1 \leq k \leq l$, let c'_k be an extension in $M(n, m-i+1)$ of c_k. Clearly the set $\{c'_1, \ldots, c'_l\}$ is also linearly independent, and since $\text{rank}(M(n, m-i)) = \text{rank}(M(n, m-i+1))$, this set is a basis for $\text{span}(M(n, m-i+1))$. Let

$$\text{split}^{(s)}(b'_k) = \gamma^{(s)}_{k,1} c'_1 + \ldots + \gamma^{(s)}_{k,l} c'_l. \tag{3}$$

Then, also

$$\text{split}^{(s)}(b_k) = \gamma^{(s)}_{k,1} c_1 + \ldots + \gamma^{(s)}_{k,l} c_l. \tag{4}$$

Also, since $v \in M(n+1, m)$, from Fact 4.2 (b) it must be that the vectors $\text{split}^{(s)}(v)$ are in $M(n+i+1, m-i)$. Hence, again by Property 2 of this lemma, and by Lemma 4.4, these vectors are in $\text{span}(M(n, m-i))$.

Since $c_1, \ldots, c_l$ is a basis for $\text{span}(M(n, m-i))$ it follows that there exists a unique sequence of coefficients $\tau_1, \ldots, \tau_l$ such that

$$\text{split}^{(s)}(v) = \tau_1 c_1 + \tau_2 c_2 + \ldots + \tau_l c_l.$$

Also, by combining Equation 2 with Equation 4, we see that

$$\begin{aligned} \text{split}^{(s)}(v) \quad = \quad & \alpha_1[\gamma^{(s)}_{1,1} c_1 + \ldots + \gamma^{(s)}_{1,l} c_l] \\ + \quad & \alpha_2[\gamma^{(s)}_{2,1} c_1 + \ldots + \gamma^{(s)}_{2,l} c_l] \\ + \quad & \ldots \\ + \quad & \alpha_p[\gamma^{(s)}_{p,1} c_1 + \ldots + \gamma^{(s)}_{p,l} c_l]. \end{aligned}$$

Thus $\tau_k = \alpha_1 \gamma^{(s)}_{1,k} + \ldots + \alpha_p \gamma^{(s)}_{p,k}$ for all $k \in [1, \ldots, l]$.

We claim

$$\begin{aligned} \text{split}^{(s)}(v') \quad = \quad & \alpha_1[\gamma^{(s)}_{1,1} c'_1 + \ldots + \gamma^{(s)}_{1,l} c'_l] \\ + \quad & \alpha_2[\gamma^{(s)}_{2,1} c'_1 + \ldots + \gamma^{(s)}_{2,l} c'_l] \\ + \quad & \ldots \\ + \quad & \alpha_p[\gamma^{(s)}_{p,1} c'_1 + \ldots + \gamma^{(s)}_{p,l} c'_l]. \end{aligned}$$

We now justify the claim. By our initial assumption, $\text{split}^{(s)}(v')$ is in $\text{span}(M(n, m-i+1))$. Thus for some unique coefficients $\tau'_1, \ldots, \tau'_l$,

$$\text{split}^{(s)}(v') = \tau'_1 c'_1 + \tau'_2 c'_2 + \ldots + \tau'_l c'_l.$$

Each c'_k is an extension of c_k, and there is a unique linear combination of $c_1, c_2, \ldots c_l$ that is equal to $\text{split}^{(s)}(v)$. It follows that each $\tau'_k = \tau_k$. This proves the claim.

Combining the claim with Equation 3 yields

$$\text{split}^{(s)}(v') = \alpha_1 \text{ split}^{(s)}(b'_1) + \ldots + \alpha_p \text{ split}^{(s)}(b'_p),$$

as desired. □

We now prove the lower bound.

Theorem 4.4 *If L is nonregular, then* $\text{rank}(M(n)) \geq n+1$ *infinitely often.*

Proof: The base case is n such that the shortest word in the language is of length n.

Suppose that $\text{rank}(M(n)) \geq n+1$ for some fixed n. Let m be the smallest number $\geq n$ such that $\text{rank}(M(n+1, m+1)) > \text{rank}(M(n, m+1))$. By Lemma 4.5 there is such an m. We claim that $\text{rank}(M(m+1)) \geq m+2$.

If $m = n$, then the claim is clearly true. Suppose $m > n$.

Let B_k be a basis for $M(n, k)$, $n \leq k \leq m+1$, where the extensions of all vectors in B_k are in B_{k+1}. Let B'_{k-1} denote the subset of B_k which are extensions of vectors in B_{k-1}.

We construct a set of $m+2$ linearly independent vectors in $M(m+1)$ as follows. For k from n to $m+1$, we define a linearly independent set C_k of vectors in $M(m+1, k)$, of size at least $k+1$. Then, C_{m+1} is the desired set.

Let $C_n = B_n$. This is by definition a linearly independent set, and it has size $\geq n+1$ because (by our initial assumption) $\text{rank}(M(n)) \geq n+1$. Suppose that $n \leq k < m+1$ and that C_k is already constructed and is linearly independent. Construct C_{k+1} as follows.

(i) Let C'_k be the set of extensions in $M(m+1, k+1)$ of the vectors in C_k. Add C'_k to C_{k+1}.

(ii) Add B_{k+1} to C_{k+1}. (Thus, C_{k+1} is expanded to contain those vectors in B_{k+1} which are not in B'_k.)

(iii) Finally, suppose nothing is added to C_{k+1} in step (ii); that is, $\text{rank}(M(n,k)) = \text{rank}(M(n, k+1))$. If i is such that $k = m-i$, then this is equivalent to: $\text{rank}(M(n, m-i)) = \text{rank}(M(n, m-i+1))$. Thus, we can apply Lemma 4.6 to obtain a vector $v' \in M(n+i+1, m-i+1)$ which is not in $\text{span}(M(n, m-i+1))$. (Thus, $v' \in M(n+m+1-k, k+1)$ but is not in $\text{span}(B'_k)$.) Add v' to C_{k+1}.

We claim that the vectors in C_{k+1} are linearly independent. Clearly the set C'_k is linearly independent. Consider each vector u' added to C_{k+1}, which is not in C'_k. By the construction, u' is not in span(B'_k). Let u' be the extension of vector u in $M(m+1,k)$. We claim that the vector u must be linearly dependent on the set B_k. This is true if u' is added in step (ii), since in this case u is in $M(n,k)$ and B_k is a basis for $M(n,k)$. It is also true in the case that $u' = v'$, the vector added in step (iii), since then by Lemma 4.4, $u = v \in$ span(B_k).

Hence, $u \in$ span(C_k), since $B_k \subseteq C_k$. Moreover, u can be expressed as a unique linear combination of the vectors of C_k, with non-zero coefficients only on those vectors in B_k.

If u' were in span(C'_k), then since it is an extension of u, it would also be expressible as a unique linear combination of the vectors of C'_k, with non-zero coefficients only on those vectors in B'_k. But that contradicts the fact that $u' \notin$ span(B'_k). □

4.2.3 The Tiling Complexity Lower Bound

Theorem 4.5 *If L is nonregular, then the 1-tiling complexity of either L or $\bar{L}$ is at least $2^{\sqrt{\log n}-2} - 1$ infinitely often.*

Proof: Melhorn and Schmidt, and independently Orlin, showed that for any binary matrix A, $rank(A) \leq \tilde{T}(A)$ [21, 22]. Their result holds for A over any field. Halstenberg and Reischuk, refining a proof of Aho et. al., showed that $\lceil \log \tilde{T}(A) \rceil \leq \lceil \log T^1(A) \rceil (\lceil \log(T^0(A)+1) \rceil + 2) + 1$ [1, 13]. Let $T^*(A) = \max(T^1(A), T^0(A))$. Then $\lceil \log \text{ rank}(A) \rceil \leq (\lceil \log(T^*(A)+1) \rceil + 1)^2$.

By Theorem 4.4, if L is nonregular, then the rank of $M(n)$ is at least $n+1$ infinitely often. It follows that for infinitely many n, $T^*(M(n)) = \max(T^1_L(n), T^0_L(n)) \geq 2^{\sqrt{\log n}-2} - 1$. □

5 Variations on the Model

In this section, we discuss extensions of our main results to other related models.

We first show that Theorem 1.1 also holds for the following "alternating probabilistic" finite state automaton model. In this model, which we call a 2apfa, the nondeterministic states N are partitioned into two subsets, N_E and N_U of existential and universal states, respectively. Accordingly, for a fixed input, there are two types of strategy, defined as follows for a fixed input string $w = w_0 w_1 w_2 \ldots w_n w_{n+1}$. An *existential (universal) strategy* on w is a function

$$E_w : N_E \times \{0, \ldots, n+1\} \to Q \times \{-1, 0, 1\}$$

$$(U_w : N_U \times \{0, \ldots, n+1\} \to Q \times \{-1, 0, 1\})$$

such that $\delta(q, \sigma, q', d) = 1$ whenever $E_w(q,j) = (q', d)$ $(U_w(q,j) = (q', d))$ and $w_j = \sigma$.

A language $L \subseteq \Sigma^*$ is accepted with *bounded error probability* if for some constant $\epsilon < 1/2$,

1. for all $w \in L$, there exists an existential strategy E_w on which the automaton accepts with probability $\geq 1 - \epsilon$ on all universal strategies U_w, and

2. for all $w \notin L$, on every existential strategy E_w, the automaton accepts with probability $\leq \epsilon$ on some universal strategy U_w.

The complexity classes 1APFA, 1APFA-polytime, and so on, are defined in the natural way, following our conventions for the npfa model.

Theorem 5.1 *1APFA = Regular.*

Proof: As in Theorems 1.1 and 3.1, we show that if L is a language accepted by a 1APFA, then the tiling complexity of L is bounded. We first extend the notation of Theorem 3.1.

If E is an existential strategy on xy and U is a universal strategy on xy, let $\mathbf{p}_{xy}(E, U)$ be the state probability (row) vector at the step when the input head moves off the right end of x, on the strategies E, U. Let $\mathbf{r}_{xy}(E, U)$ be the column vector whose i'th entry is the probability of accepting the string xy, assuming that the automaton is in state i at the moment that the head moves off the right end of x, on the strategies E, U. For each 1-entry $[x, y]$ of M_L, fix an existential strategy E_{xy}, that causes xy to be accepted with probability at least $1 - \epsilon$, for all universal strategies.

Partition the space $[0, 1]^c$ into cells of size $\mu \times \mu \times \ldots \times \mu$, as before. Let $\mathcal{C}$ be a nonempty subset of the cells. We say that entry $[x, y]$ of M_L belongs to $\mathcal{C}$ if $xy \in L$, and $\mathcal{C}$ is the smallest set of cells which contain all the vectors $\mathbf{p}_{xy}(E_{xy}, U)$, for all universal strategies U.

With each nonempty subset $\mathcal{C}$ of the cells, associate a rectangle $R_{\mathcal{C}}$ defined as follows.

$$\{x \mid \text{there exists } y \text{ such that } [x, y] \text{ belongs to } \mathcal{C}\}$$
$$\times$$
$$\{y \mid \text{there exists } x \text{ such that } [x, y] \text{ belongs to } \mathcal{C}\}.$$

Then, $R_{\mathcal{C}}$ is a valid 1-tile. To see this, suppose that $[x, y] \in R_{\mathcal{C}}$. If $[x, y]$ belongs to $\mathcal{C}$, then it must be a 1-entry. Otherwise, there exist x' and y' such that $[x, y']$ and $[x', y]$ belong to $\mathcal{C}$.

Consider the strategy E that while reading x, uses the strategy $E_{xy'}$, and while reading y, uses the strategy $E_{x'y}$. We claim that xy is accepted with probability at least $1/2$ on existential strategy E and any universal strategy U on xy. The probability that xy is accepted on strategies E, U is

$$\mathbf{p}_{xy}(E, U)\mathbf{r}_{xy}(E, U) = \mathbf{p}_{xy'}(E_{xy'}, U)\mathbf{r}_{x'y}(E_{x'y}, U).$$

Since $[x, y']$ and $[x', y]$ belong to the same set of cells $\mathcal{C}$, $\mathbf{p}_{xy'}(E_{xy'}, U)$ and $\mathbf{p}_{x'y}(E_{x'y}, U')$ are in the same cell, for some universal strategy U'. Moreover,

$$\mathbf{p}_{x'y}(E_{x'y}, U')\mathbf{r}_{x'y}(E_{x'y}, U) \geq 1 - \epsilon.$$

This is because this quantity is the probability that $x'y$ is accepted on existential strategy $E_{x'y}$ and a universal strategy which is a hybrid of U and U'; also by definition of $E_{x'y}$, the probability that $x'y$ is accepted with respect to $E_{x'y}$ and any universal strategy is $\geq 1-\epsilon$. Hence,

$$\begin{aligned}(\mathbf{p}_{x'y}(E_{x'y},U') - \mathbf{p}_{xy'}(E_{xy'},U)) \quad & \mathbf{r}_{x'y}(E_{x'y},U) \\ & = \textstyle\sum_{i=1}^{c}[\mathbf{p}_{x'y}(E_{x'y},U') - \mathbf{p}_{xy'}(E_{xy'},U))]_i[\mathbf{r}_{x'y}(E_{x'y},U)]_i \\ & \leq \mu \textstyle\sum_{i=1}^{c}[\mathbf{r}_{x'y}(E_{x'y},U)]_i \\ & \leq \mu c \\ & = 1/2-\epsilon, \text{ by our choice of } \mu.\end{aligned}$$

Hence, the probability that xy is accepted on the strategies E, U is

$$\begin{aligned}\mathbf{p}_{xy'}(E_{xy'},U)\mathbf{r}_{x'y}(E_{x'y},U) \quad &\geq \quad \mathbf{p}_{x'y}(E_{x'y},U')\mathbf{r}_{x'y}(E_{x'y},U) - (1/2-\epsilon) \\ &\geq \quad (1-\epsilon) - (1/2-\epsilon) \\ &= \quad 1/2 > \epsilon.\end{aligned}$$

Since U is arbitrary, it follows that there is an existential strategy E such that on all strategies U, the probability that xy is accepted on the strategies E, U is greater than ϵ, and so it cannot be that $xy \notin L$. Hence, for all $[x,y] \in R_{\mathcal{C}}$, xy must be in L. Therefore $R_{\mathcal{C}}$ is a 1-tile in M_L.

The proof is completed as in Theorem 3.1. □

In the same way, Theorem 3.3 can also be extended to obtain the following.

Theorem 5.2 *A language L is in 2APFA-polytime only if the 1-tiling complexity of L is bounded by $2^{polylog(n)}$.*

Thus, for example, the language *Pal*, consisting of all strings over $\{0,1\}^*$ which read the same forwards as backwards, is not in the class 2APFA-polytime. To see this, consider the submatrix of $M_L(n)$, consisting of all rows and columns labeled by strings of length exactly n. This matrix contains a fooling set of size 2^n; hence a 1-tiling of $M_L(n)$ requires at least 2^n tiles.

We next extend Theorem 1.2 to automata with $o(\log\log n)$ space. We refer to these as Arthur-Merlin games, since this is the usual notation for such automata which are not restricted to a finite number of states [7]. The definition of an Arthur-Merlin game is similar to that of an npfa, except that the machine has a fixed number of read/write worktapes. The Arthur-Merlin game runs within space $s(n)$ if on any input w with $|w| \leq n$, at most $s(n)$ tape cells are used on any worktape. Thus, the number of different configurations of the Arthur-Merlin game is $2^{O(s(n))}$.

Theorem 5.3 *Let M and $\bar{M}$ be Arthur-Merlin games which recognize a nonregular language L and its complement $\bar{L}$, respectively, within space $o(\log\log n)$. Suppose that the expected running time of both M and $\bar{M}$ is bounded by $t(n)$. Then, for all $b < 1/2$, $\log\log t(n) \geq (\log n)^b$. In particular, $t(n)$ is not bounded by any polynomial in n.*

Proof: The proof of Theorem 1.2 can be extended to space bounded Arthur-Merlin games, to yield the following generalization of Equation 1. Let $c(n)$ be an upper bound on the number of different configurations of M on inputs of length n, and let $d(n) = 2c(n) + 4$. Then, for sufficiently large n, the number of 1-tiles needed to cover $M_L(n)$ is at most

$$T_L^1(n) \leq (\lceil 2 \log_2 t(n)/\mu \rceil + 1)^{d^2(n)} = 2^{\Theta(d^2(n) \log\log t(n))}.$$

Since M uses $o(\log\log n)$ space, for any constant $c > 0$, $d(n) \leq (\log n)^c$, for sufficiently large n.

Now, suppose to the contrary that for some $b < 1/2$, $\log\log t(n) < (\log n)^b$ for sufficiently large n. Then,

$$d^2(n) \log\log t(n) = o(\sqrt{\log n}).$$

Hence, the number of tiles needed to cover the 1-valued entries of $M_L(n)$ is $2^{o(\sqrt{\log n})}$. The same argument for $\bar{M}$ shows that also for for sufficiently large n, the number of tiles needed to cover the 1-valued entries of $M_{\bar{L}}(n)$ is $2^{o(\sqrt{\log n})}$.

Hence, by Theorem 4.5 L must be regular, contradiction. □

Finally, we consider a restriction of the 2npfa model, which, given polynomial time, can only recognize regular languages. A *restricted 2npfa* is a 2npfa for which there is some $\epsilon < 1/2$ such that on all inputs w and strategies S_w, the probability that the automaton accepts is either $\geq 1 - \epsilon$ or $< \epsilon$.

Theorem 5.4 *Any language accepted by a restricted 2npfa with bounded error probability in polynomial time is regular.*

Proof: Let L be accepted by a 2npfa M with bounded error probability in polynomial expected time. Let Σ be the alphabet, δ the transition function, $Q = \{q_1, q_2, \ldots, q_{|Q|}\}$ the set of states and $N \subset Q$ the set of nondeterministic states of M. Without loss of generality, let $N = \{q_1, \ldots, q_{|N|}\}$.

We first define a representation of strategies as strings over a finite alphabet. Let $\Sigma' = (N \times Q \times \{-1, 0, 1\})^{|N|}$. Without loss of generality, assume that $\Sigma \cap \Sigma' = 0$. A string $S_0 S_1 \ldots S_{n+1}$ corresponds to a strategy on $\not\vdash w\$$, where $\not\vdash w\$ = \sigma_0 \sigma_1 \ldots \sigma_{n+1}$, if for $0 \leq j \leq n+1$, S_j is of the form

$$S_j = ((q_1, q_1', d_1), (q_2, q_2', d_2), \ldots, (q_{|N|}, q_{|N|}' d_{|N|})).$$

and $\delta(q_i, \sigma_j, q_i', d_i) = 1$.

Define L' to be the set of strings of the form $\sigma_0 S_0 \sigma_1 S_1 \ldots \sigma_{n+1} S_{n+1}$, where each σ_i is in the alphabet Σ, each S_i is in the alphabet Σ', and furthermore, $S = S_0 S_1 \ldots S_{n+1}$ corresponds to a strategy of M on input $w = \sigma_0 \sigma_1 \ldots \sigma_{n+1}$, which causes w to be accepted.

Then, L' is accepted by a 2pfa with bounded error probability in polynomial time. Thus, L' is regular [7]. Moreover, note that a string of the form $w = \sigma_0 \sigma_1 \ldots \sigma_{n+1}$ is in L if and only if for some choice of $S_0, S_1, \ldots, S_{n+1}$, $\sigma_0 S_0 \sigma_1 S_1 \ldots \sigma_{n+1} S_{n+1}$ is in L'. Let M' be a one-way

deterministic finite state automaton for L', and assume without loss of generality that the set of states in which M' can be when the head is at an even position, is disjoint from the set of states in which M' can be when the head is at an odd position. Then, from M' we can construct a one-way nondeterministic finite state automaton for L, by replacing the even position states by nondeterministic states. Hence, L is regular. $\Box$

6 Conclusions

We have introduced a new measure of the complexity of a language, namely its tiling complexity, and have proved a gap between the tiling complexity of regular and nonregular languages. We have applied these results to prove limits on the power of finite state automata with both probabilistic and nondeterministic states.

An intriguing question left open by this work is whether the class 2NPFA-polytime is closed under complement. If it is, we can conclude that 2NPFA-polytime = Regular. Recall that the class 2NPFA does contain nonregular languages, since it contains the class 2PFA, and Freivalds [10] showed that $\{0^n1^n \mid n \geq 0\}$ is in this class. However, Kaneps [18] showed that the class 2PFA does not contain any nonregular unary language. Another open question is whether the class 2NPFA contains any nonregular unary language. It is also open whether there is a nonregular language in 2APFA-polytime.

There are several other interesting open problems. Can one obtain a better lower bound on the tiling complexity of nonregular languages than that given by Theorem 4.5, perhaps by an argument that is not based on rank? We know of no nonregular language with tiling complexity less than $\Omega(n)$ infinitely often, so the current gap is wide.

References

[1] A. V. Aho, J. D. Ullman and M. Yannakakis. On notions of information transfer in VLSI circuits, Proc. of the Fifteenth Annual ACM Symposium on Theory of Computing, 1983, 133–139.

[2] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and hardness of approximation problems, Proc. of the 33rd IEEE Symposium on Foundations of Computer Science, 1992, 14–23.

[3] L. Babai and S. Moran. Arthur-Merlin games: A randomized proof system and a hierarchy of complexity classes, *J. Comput. System Sci.*, 36 (1988), 254–276.

[4] A. Condon. *Computational Models of Games*, MIT Press, 1989.

[5] On the Power of finite automata with both nondeterministic and probabilistic states, Proc. of the Twenty Sixth Annual ACM Symposium on Theory of Computing, 1994, to appear.

[6] A. Condon and R. Ladner. Probabilistic game automata, *J. Comput. Sys. Sci.*, 36(3) (1988), 452–489.

[7] C. Dwork and L. Stockmeyer. A time-complexity gap for two-way probabilistic finite state automata, *SIAM J. Comput.*, 19 (1990), 1011–1023.

[8] C. Dwork and L. Stockmeyer. Finite state verifiers I: the power of interaction, *J. ACM*, 39(4) (1992), 800–828.

[9] L. Fortnow and C. Lund. Interactive proof systems and alternating time-space complexity, Proc. of the 8th Annual Symposium on Theoretical Aspects of Computer Science, 1991, 263–274.

[10] R. Freivalds. Probabilistic two-way machines, Proc. of the International Symposium on Mathematical Foundations of Computer Science Springer-Verlag Lecture Notes in Computer Science, 188 (1981), 33–45.

[11] G. Frobenius. Über das Trägheitsgesetz der quadratischen Formen, *Sitzungsber. der Königl. Preuss. Akad. der Wiss.* (1894), 407–431.

[12] A. G. Greenberg and A. Weiss. A lower bound for probabilistic algorithms for finite state machines, *J. Comput. Syst. Sci.*, 33 (1986), 88–105.

[13] B. Halstenberg and R. Reischuk. On different modes of communication, Proc. of the Twentieth Annual ACM Symposium on the Theory of Computing, 1988, 162–172.

[14] J. E. Hopcroft and J. D. Ullman. *Introduction to Automata Theory, Languages, and Computation*, Addison Wesley, 1979.

[15] I. S. Iohvidov. *Hankel and Toeplitz Matrices and Forms: Algebraic Theory*, Edited by I. Gohberg, Translated by G. Philip and A. Thijsse, Birkhäuser, Boston, 1982.

[16] J. Kaŋeps and R. Freivalds. Minimal nontrivial space complexity of probabilistic one-way Turing machines, Proc. of the Conference on Mathematical Foundations of Computer Science, Springer Verlag Lecture Notes in Computer Science, 452 (1990), 355–361.

[17] J. Kaŋeps and R. Freivalds. Running Time to Recognize Nonregular Languages by 2-Way Probabilistic Automata, Proc. of the 18th International Colloquium on Automata, Languages, and Programming, Springer-Verlag, New York, 1991, 174-185.

[18] J. Kaŋeps. Regularity of one-letter languages acceptable by 2-way finite probabilistic automata, Proc. of Fundamentals of Computation Theory, Springer Verlag Lecture Notes in Computer Science, 529 (1991), 287–296.

[19] R. M. Karp. Some bounds on the storage requirements of sequential machines and Turing machines, *J. ACM*, 14(3) (1967), 478–489.

[20] F. T. Leighton and R. L. Rivest. The Markov chain tree theorem, Rep. MIT/LCS/TM-249, Laboratory for Computer Science, MIT, Cambridge, Mass., 1983. Also in IEEE Transactions on Information Theory, IT-37(6), (1986) 733-742.

[21] K. Melhorn and E. M. Schmidt. Las Vegas is better than determinism in VLSI and distributed computing, Proc. of the Fourteenth Annual ACM Symposium on Theory of Computing, 1982, 330–337.

[22] J. Orlin. Contentment in Graph Theory: Covering Graphs with Cliques. *Proc. Koniklijke Nederlandse Akademie van Wetenschappen* Amsterdam Series A, 80(5) (1977), 406–424.

[23] C. Papadimitriou. Games against nature, *J. Comput. System Sci.*, 31 (1985), 288–301.

[24] M. O. Rabin. Probabilistic automata, *Inf. Contr.* **6** (1963), 230–245.

[25] M. O. Rabin and D. Scott. Finite automata and their decision problems, *IBM J. Research*, 3(2) (1959), 115–125.

[26] A. Salomaa and M. Soittola. *Automata-theoretic aspects of formal power series*, Texts and Monographs in Computer Science, Springer-Verlag, New York, 1978.

[27] E. M. Schmidt. Succinctness of description of context free, regular and unambiguous languages, Ph.D. thesis, Cornell University, 1978.

[28] J. Shallit. Automaticity: properties of a measure of descriptional complexity, Proc. of the 11th Annual Symposium on Theoretical Aspects of Computer Science, February 1994.

[29] A. C. Yao. Some complexity questions related to distributed computing, Proc. of the Eleventh Annual ACM Symposium on Theory of Computing, 1979, 209–213.

[30] A. C. Yao. Lower bounds by probabilistic arguments, Proc. of the 24th IEEE Symposium on Foundations of Computer Science, 1983, 420–428.