

Thesis Proposal: Automata on Shift Spaces

Mihir Vahanwala

September 3, 2026

Abstract

A word is a sequence of symbols drawn from a finite set, and is often a device to represent the evolution of a (discretised) system. Words are central to both symbolic dynamics and verification. In the former discipline, words that represent the same system constitute a *shift space*, and are typically studied for their combinatorial structure. In verification, a quintessential objective is to certify that a system satisfies a logical specification by showing that its representative words are accepted by an automaton. In this thesis, we bridge the two perspectives through the following contributions.

1. We describe how the action of a deterministic finite transducer on an infinite word preserves the structure of the word. If the input word x has a recurrence property, then it is preserved for the output word $\mathcal{A}(x)$; using the ergodic theory of shift spaces, we identify a large class of words x , including *Sturmian words*, such that $\mathcal{A}(x)$ is guaranteed to admit factor frequencies.
2. We show that we can determine whether an automaton accepts a word by reading its so-called *S-adic construction*. This allows us, e.g., to decide, given an automaton $\mathcal{A}$, whether $\mathcal{A}$ accepts all Sturmian words, or all d -letter *Arnoux-Rauzy words*.
3. As applications, we study *Monadic Second-Order (MSO)* logic, and show that the MSO theory of the natural numbers remains decidable when expanded with the following sets of predicates: (i) powers of 2 and 3; (ii) powers of 2, 3, and 5, subject to Schanuel's conjecture, (iii) perfect squares and powers of 2, subject to the binary expansion of $\sqrt{2}$ being a *disjunctive word*.

This document outlines the content of the thesis, and proposes examiners, along with dates for submission and defence.

Introduction

Since antiquity, we have had the mathematical instinct to represent a *trace* of observations of a (possibly continuous) system using a discrete sequence of symbols. For instance, consider the word $0^{364}10^{364}10^{364}10^{365}1\dots$ obtained by

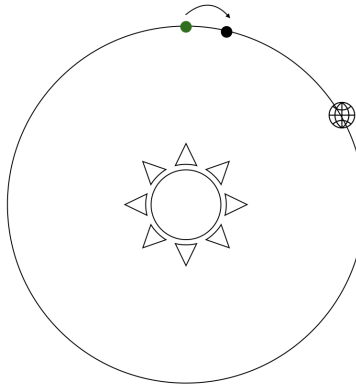


Figure 1: We assume the Earth revolves around the Sun in a perfectly circular, constant orbit. In one day, the Earth revolves the angle between the green and black points. The green point represents the point at which the vernal equinox occurs. The condition to write the letter 1 is equivalent to the Earth being in the arc between the green and black points, including the green point, at midnight.

writing a letter each midnight: the letter 1 if the vernal equinox occurred during the previous day, and the letter 0 otherwise (Fig. 1).

Unless the number of days $365.2422\dots$ in an average tropical year is rational, we will have constructed an aperiodic word; in fact, this would be a *Sturmian word*.

Symbolic dynamics and verification are two modern disciplines that formalise this technique: in today’s parlance, we call the (finite) set of symbols an *alphabet*, and the discrete sequences *words*.

Symbolic dynamics is the study of words that form traces of dynamical systems such as the ones described above. Lind and Marcus [24, Preface] attribute the foundations to Hadamard [20], who, in 1898, introduced the quintessential elements of the discipline while studying more complicated systems called geodesic flows on surfaces of negative curvature.

A central object in symbolic dynamics is the *shift space*, which is the set of all words that do not contain a given set of *forbidden factors*. (A forbidden factor is a finite sequence of letters that may not occur as a contiguous subword of the words under consideration.) In the ensuing decades, it was discovered by mathematicians such as Morse, Hedlund, and others, that in addition to Hadamard’s systems, several other dynamical systems are also such that the respective sets of their symbolic traces constitute shift spaces. Today, shift spaces are considered interesting in their own right, and the study of their attendant word combinatorics motivates active research.

Words were found to play a central role in logic and verification by Büchi, whose seminal 1962 paper [10] showed that the *Monadic Second-Order (MSO)* theory of the structure $\langle \mathbb{N}; < \rangle$ is decidable. The influence of Büchi’s work lies in the following verification paradigm: should a system designer wish to verify that a trace satisfies a specification which can be formulated in MSO logic, they construct an equivalent *automaton* and determine whether it *accepts* the trace.

In 1966, shortly after Büchi’s work, Elgot and Rabin [18] used automata theory to show that the MSO theory of the expansion $\langle \mathbb{N}; <, P \rangle$ is decidable when P is a unary predicate with “arithmetic” origins, e.g., the set of squares, the set of powers of two, the set of factorial numbers, and so on. The key property of these sets is that they are ultimately periodic modulo any integer divisor. The techniques and results were later generalised by Carton and Thomas [11] through the notion of (*effective*) *profinite ultimate periodicity*.

The natural question of whether expansions with multiple predicates have decidable MSO theories remains open in general. Examples of such structures include:

1. $\langle \mathbb{N}; <, \{2^n \mid n \in \mathbb{N}\}, \{3^n \mid n \in \mathbb{N}\} \rangle$,
2. $\langle \mathbb{N}; <, \{2^n \mid n \in \mathbb{N}\}, \{3^n \mid n \in \mathbb{N}\}, \{5^n \mid n \in \mathbb{N}\} \rangle$,
3. $\langle \mathbb{N}; <, \{2^n \mid n \in \mathbb{N}\}, \{n^2 \mid n \in \mathbb{N}\} \rangle$,
4. $\langle \mathbb{N}; <, \{n! \mid n \in \mathbb{N}\}, \{n^2 \mid n \in \mathbb{N}\} \rangle$.

Observe that the fourth example subsumes the Brocard-Ramanujan problem of determining whether there are more than three solutions to the equation $n^2 = m! + 1$.

This is a challenging research question, not least because numerically sophisticated dynamics may underlie such expanded structures. It is only recently [6] that substantial progress has been reported: using automata theory and Baker’s (Fields-medal-winning) theory of linear forms of logarithms, we reduce, e.g., the problems of deciding the first three theories respectively to the problems of deciding whether a given automaton accepts the following words, which have a distinctly symbolic-dynamical nature:

1. the Sturmian word $2322323223 \dots$, which records the order in which the powers of 2 and 3 appear,
2. the *cutting sequence* $2325232532 \dots$, which records the order in which the powers of 2, 3, and 5 appear,
3. the word $10110101000 \dots$ formed by the digits of the binary expansion of $\sqrt{2}$.

We shall, of course, duly return to the solutions of these reduced problems. It is worth noting that, at this point, we may ask the broader question that forms the subject of this thesis:

How do automata run on shift spaces?

Semënov’s study of *almost-periodic* words (see, e.g., [26]) is instrumental in answering this question. Chapter 2 of the thesis, based on [8], adds insights to, and extends such results. We show that recurrence properties of a word x also hold for the output $\mathcal{A}(x)$ when x is input to a deterministic finite transducer $\mathcal{A}$.

Chapter 3 of the thesis, also based on [8], is the measure-theoretic counterpart to the word-combinatorial chapter that precedes it. The motivating question is to understand the set of words x for which the transducer output $\mathcal{A}(x)$ is guaranteed to *admit factor frequencies*. This is a profound question, and one intrinsically connected to symbolic dynamics and ergodic theory: we show that a word admits factor frequencies if and only if it is *generic* for an *invariant measure* on the shift it generates. We focus on cases where we can prove that the shift generated by $\mathcal{A}(x)$ is *uniquely ergodic*, i.e., it has only one invariant measure for which *all* points are generic: *Boshernitzan’s condition* is key for our techniques. As alluded to in the abstract, Sturmian words satisfy Boshernitzan’s condition.

Chapter 4 of the thesis, which is based on [7], marks a return to the perspective of automata as recognisers of ω -regular languages. Its main results have the following form: given an ω -regular language L , we can compute an ω -regular language L' such that a word $x \in L$ if and only if its *S-adic directive sequence* $\sigma \in L'$. The point is that directive sequences often elucidate structural information about a word, e.g., directive sequences of Sturmian words indicate their *slopes* and *intercepts*.

Chapter 5, the final technical chapter of the thesis, is based on [6], and applies the developments of the preceding chapters to give (sometimes conditional) decision procedures for the automaton acceptance problems of order words such as the ones mentioned above. We thus obtain (conditional) decidability results for the corresponding MSO theories.

In the rest of this document, we give an overview of the highlights of each chapter, starting with Chapter 1 of the thesis, which records technical preliminaries.

1 Preliminaries

In this section, we record the most important definitions and standard propositions that lay the foundation of the thesis. We use $\mathbb{N}$ to denote the set of natural numbers, and note that this set includes 0.

1.1 Words

We will usually use Σ and sometimes use Γ to denote a finite nonempty alphabet. The set of infinite words over Σ (indexed by $\mathbb{N}$) is denoted by Σ^ω , the set of finite words is denoted by Σ^* , the empty word is denoted by ε , and the set of finite non-empty words is denoted by Σ^+ .

For a word x , the notation $x(i)$ denotes the letter in the i -th position of x . A word x is *effective* if, given $i \in \mathbb{N}$, we can compute the letter $x(i)$. The notation $x(i, j)$ denotes the finite subword, or *factor* of x starting in position i and ending at position $j - 1$ (i.e., the length of the factor is $j - i$). If x is an infinite word, the notation $x(i, \infty)$ denotes the suffix of x starting at position i . We shall denote the length of a finite word u as $|u|$.

When $u = x(i, j)$, we say that u is a factor of x with an *occurrence* at index i . The factor complexity of a word x is the function p_x that counts for each n , the number of distinct length- n factors of x . We denote by $\mathcal{L}(x)$ the *factor language* of x , i.e., the set of finite factors u of x . We shall sometimes use $\mathcal{L}_n(x)$ as a convenient shorthand for $\mathcal{L}(x) \cap \Sigma^n$.

A word $x \in \Sigma^\omega$ is said to be *recurrent* if every factor u of x is recurrent, i.e., occurs infinitely often. We can define the set of *return words* to a recurrent factor u of x : these are words r such that ru is a factor of x , and ru has exactly two occurrences of u , once as a prefix, and once as a suffix. A recurrent word x is said to be *uniformly recurrent* if for every factor u , its set of return words $\mathcal{R}_x(u)$ is finite. Similarly, a word x is said to be *almost-periodic* if every recurrent factor u has a finite set of return words.

For a uniformly recurrent word x , we can define a *return-time* (or *recurrence*) function $R_x: \mathbb{N} \rightarrow \mathbb{N}$, such that $R_x(n)$ gives the length of the longest return word to a length- n factor of x . A uniformly recurrent word x is said to be *linearly recurrent* if $R_x(n) \in O(n)$. If for an effective and uniformly (respectively, linearly) recurrent word $x \in \Sigma^\omega$, given any n , we can compute $\mathcal{L}_n(x)$,

we say that x is *effectively* uniformly (respectively, linearly) recurrent. (This condition also defines effectively almost-periodic words.)

The *factor frequency* of a word u of length ℓ in x is defined as

$$f_x(u) = \frac{|\{i \mid x(i, i + \ell) = u\}|}{|x| - |u| + 1} \quad (1)$$

if x is finite, and as

$$f_x(u) = \lim_{N \rightarrow \infty} \frac{1}{N} |\{i < N \mid x(i, i + \ell) = u\}| \quad (2)$$

if x is infinite.

We say that a word $x \in \Sigma^\omega$ *admits factor frequencies* if for every recurrent factor u of x , the above limit exists. If furthermore, given any u , this limit can be computed to arbitrary precision, then we say that x admits *computable* factor frequencies.

We now discuss *primitive morphic* words, the best-known construction of aperiodic words that are effectively linearly recurrent and admit factor frequencies [30]. They are constructed using *substitutions*, which are maps from an alphabet Σ to finite words over an alphabet Γ . A word x is *substitutive* if there exists a substitution σ (which is not the identity) such that $\sigma(x) = x$. It is *primitive substitutive* if the fixing substitution σ is *primitive*, i.e., there exists m such that for every letter a , the word $\sigma^m(a)$ contains all the letters of the alphabet. A (*primitive*) *morphic* word is obtained by applying a substitution τ to a (primitive) substitutive word x .

Morphic words constitute a special class of self-similar words. More generally, given an alphabet Σ , let $\mathcal{S}(\Sigma)$ denote the set of all substitutions from Σ to Σ^+ , and let $S \subseteq \mathcal{S}(\Sigma)$ be a set of substitutions. We say that a word $x \in \Sigma^\omega$ is *S-adic* if there exist a sequence of words $(x^{(n)})_{n \in \mathbb{N}}$ and a sequence of substitutions $\sigma = (\sigma_n)_{n \in \mathbb{N}}$ over S such that $x^{(0)} = x$, and for all n we have $x^{(n)} = \sigma_n(x^{(n+1)})$. We call σ the *directive sequence* and say that x is directed by σ . A directive sequence σ is called *primitive* if for every n , there exists k such that for every letter a , the image $\sigma_n \cdots \sigma_{n+k}(a)$ contains all letters of the alphabet Σ . It is well known that words with primitive directive sequences are uniformly recurrent [4].

As examples, consider Sturmian words over the alphabet $\{0, 1\}$. A Sturmian word x is parametrised by its irrational *slope* $\eta \in (0, 1)$ and *intercept* $\chi \in [-\eta, 1 - \eta]$, and satisfies one of the following.

$$x(i) = \lfloor (i + 2)\eta + \chi \rfloor - \lfloor (i + 1)\eta + \chi \rfloor \text{ for all } i, \quad (3)$$

$$x(i) = \lceil (i + 2)\eta + \chi \rceil - \lceil (i + 1)\eta + \chi \rceil \text{ for all } i. \quad (4)$$

These words have primitive directive sequences with substitutions from the set $S = \{\lambda_0, \lambda_1, \rho_0, \rho_1\}$ of Sturmian morphisms, where $\lambda_i(i) = \rho_i(i) = i$, and $\lambda_i(j) = ij, \rho_i(j) = ji$ [5].

1.2 Automata

In the thesis, automata will serve as a means to recognise (ω -)regular languages, and also be viewed as an interesting model of computation in their own right. The automata we consider are deterministic unless specified otherwise.

A (*deterministic*) *automaton* $\mathcal{A}$ is given by $(Q, q_{\text{init}}, \Sigma, \delta)$, where Q is a finite non-empty set of states, q_{init} is the initial state, Σ is the alphabet it runs over, and $\delta : Q \times \Sigma \rightarrow Q$ is the transition function which prescribes the next state, given the current state and the letter being read. We naturally extend the domain of δ to $Q \times \Sigma^*$ by defining $\delta(q, \varepsilon) = q$ for all q , and $\delta(q, ua) = \delta(\delta(q, u), a)$.

We say that an automaton is a *permutation automaton* if for every letter a , the transition function on a is a permutation; likewise we say that an automaton is a *reset automaton* if for every letter a , the transition function on a is a constant function or the identity function. *Permutation-reset automata* are those in which for every letter a , the transition function on a is either a permutation or a constant function.

For a word $x = a_0a_1a_2 \cdots$ over the alphabet Σ , the *run* $\mathcal{A}(x)$ of $\mathcal{A}$ on x is the word $(q_0, a_0)(q_1, a_1)(q_2, a_2) \cdots$ over the alphabet $Q \times \Sigma$, where $q_0 = q_{\text{init}}$, and for all n we have $\delta(q_n, a_n) = q_{n+1}$. We similarly define “the run v of $\mathcal{A}$ on u , starting in state q ” in the obvious way. This is useful when u is a factor of x , and we are interested in the fragment of the run produced while reading u .

For nondeterministic automata, q_{init} is replaced by a set Q_{init} of initial states, and δ is replaced by a transition relation $\Delta \subseteq Q \times \Sigma \times Q$. For a word $x = a_0a_1a_2 \cdots$ over the alphabet Q , a word $(q_0, a_0)(q_1, a_1)(q_2, a_2) \cdots$ is a run of $\mathcal{A}$ on x if $q_0 \in Q_{\text{init}}$, and for all n we have that $(q_n, a_n, q_{n+1}) \in \Delta$.

An automaton $\mathcal{A}_0 = (Q, q_{\text{init}}, \Sigma, \delta)$ may be extended into a *transducer* $\mathcal{A}$ by specifying an output alphabet Γ and an output function $\tau : Q \times \Sigma \rightarrow \Gamma^*$. The transducer, upon reading a letter a in the state q , prints $\tau(q, a)$. The output $\mathcal{A}(x)$ of the transducer on input x is then $\tau(\mathcal{A}_0(x))$.

Recall that a language $L \subseteq \Sigma^*$ is *regular* if and only if there exists an automaton $(Q, q_{\text{init}}, \Sigma, \delta)$ and a set of states $F \subseteq Q$ such that $u \in L$ if and only if $\delta(q_{\text{init}}, u) \in F$. A language $L \subseteq \Sigma^\omega$ is ω -*regular* if and only if it is a finite union of languages of the form UV^ω , where $U \subseteq \Sigma^*$ and $V \subseteq \Sigma^+$ are regular languages. Büchi [10] established the equivalence of MSO logic on the natural numbers and ω -regular languages, and showed that the latter are recognised by nondeterministic automata with the Büchi acceptance condition. A nondeterministic Büchi automaton $(Q, Q_{\text{init}}, \Sigma, \Delta)$ is additionally equipped with a set $F \subseteq Q$ of accepting states, and accepts a word x if and only if the automaton has a run on x that visits some state in F infinitely often.

It is possible, and often desirable, to recognise ω -regular languages using deterministic automata. We use (deterministic) parity automata for this purpose [28]. We equip an automaton $(Q, q_{\text{init}}, \Sigma, \delta)$ with a colouring function $\text{col} : Q \rightarrow \mathbb{N}$, and declare that $\mathcal{A}$ accepts x if and only if the maximum colour seen infinitely often along the run $\mathcal{A}(x)$ is even.

Indeed, in Chapter 5 of the thesis, we formulate our MSO-logic decidability results in terms of the following decision problem Acc_x parametrised by *charac-*

teristic words x of structures: given a deterministic parity automaton $\mathcal{A}$, decide whether $\mathcal{A}$ accepts x .

The following simple observation is often helpful.

Lemma 1.1. *Let $\mathcal{A} = (Q, q_{\text{init}}, \Sigma, \delta_Q, \Gamma, \tau)$ be a transducer, and let x, y be infinite words such that $y = \mathcal{A}(x)$. We have that Acc_y reduces to Acc_x .*

In Chapter 4 of the thesis, we will rely not only on automata but also on ω -semigroups (see [27, Sec. 7]) to recognise ω -regular languages [27, Thm. 7.6]. Formally, an ω -semigroup is a “two-sorted” algebra equipped with the following operations:

1. A binary operation defined on M_f and denoted multiplicatively,
2. A mapping $M_f \times M_\omega \rightarrow M_\omega$, called the *mixed product*, also denoted multiplicatively,
3. An infinite product π that maps infinite sequences over M_f to an element of M_ω .

These operations must satisfy the following associativity properties:

1. M_f , equipped with the binary operation, is a semigroup.
2. For every $m_1, m_2 \in M_f$ and $m_3 \in M_\omega$, we have that $(m_1 m_2) m_3 = m_1 (m_2 m_3)$.
3. For every sequence $(m_n)_{n \in \mathbb{N}}$ over M_f , and every strictly increasing sequence $(k_n)_{n \in \mathbb{N}}$ of indices, we have that $\pi(m_0, m_1, \dots) = \pi(m_0 \cdots m_{k_0}, m_{k_0+1} \cdots m_{k_1}, \dots)$.
4. For every $m \in M_f$ and every sequence $(m_n)_{n \in \mathbb{N}}$ over M_f , we have that $m \cdot \pi(m_0, m_1, \dots) = \pi(m, m_0, m_1, \dots)$.

A set $L \subseteq \Sigma^+ \cup \Sigma^\omega$ is recognised by a morphism h from Σ^ω into an ω -semigroup M if there exists a subset $H \subseteq M_f \cup M_\omega$ such that $L = h^{-1}(H)$.

We use the equivalence of finite ω -semigroups and automata through the following lemma.

Lemma 1.2. *Let $M = (M_f, M_\omega)$ be a finite ω -semigroup, and $x \in M_f \cup M_\omega$. Extend M_f into a monoid M'_f by adjoining a distinguished neutral element 1_M : for all $m \in M'_f$, $1_M \cdot m = m \cdot 1_M = m$. We can construct an automaton $\mathcal{A}_x$ over the alphabet M'_f that accepts $m_0 m_1 \cdots \in (M'_f)^\omega$ if and only if the infinite product $m_0 m_1 \cdots$ (defined as the possibly finite product of the subsequence $m_{i_0} m_{i_1} \cdots$ obtained by discarding all 1_M terms) equals x .*

Chapters 2 and 3 of the thesis view automata and transducers as input-output devices as opposed to recognisers of languages. They present results of the following form: if a word x has combinatorial property P , then so does the word $\mathcal{A}(x)$ for every transducer $\mathcal{A}$. The strategy to prove such results is to

“decompose” $\mathcal{A}$ into simpler automata. The Krohn-Rhodes theorem [23] (see [25] for an automata theoretic formulation) achieves this decomposition. For almost all applications in the thesis, however, we can use a version that is easier to prove.

Proposition 1.3 (Lightweight Krohn-Rhodes). *Let $\mathcal{A} = (Q, q_{\text{init}}, \Sigma, \delta)$ be an automaton. We can compute permutation-reset automata $\mathcal{B}_1, \dots, \mathcal{B}_k$, and a substitution σ such that for any word x , we have that $\mathcal{A}(x) = \sigma \circ \mathcal{B}_k \circ \dots \circ \mathcal{B}_1(x)$.*

The full Krohn-Rhodes theorem imposes further algebraic requirements on the elements $\mathcal{B}_i$ of the cascade.

1.3 Shift Spaces

The *shift operator* T maps a word $x \in \Sigma^\omega$ to $x(1, \infty)$, i.e., it deletes the starting letter. We write $T^n x$ for the n -fold application of T to $x \in \Sigma^\omega$. A *shift system* (also called a *shift*) is the dynamical system (X, T) where $X \subseteq \Sigma^\omega$ is closed and satisfies $TX \subseteq X$. When the dynamics is clear from the context, we also refer to X as a shift.

Given a word x , the *shift generated by x* is the topological closure of the orbit $\{T^n x \mid n \in \mathbb{N}\}$ of x under T . For a shift X , we define $\mathcal{L}(X) = \{u \mid u \in \mathcal{L}(x'), x' \in X\}$, and as before, we declare $\mathcal{L}_n(X)$ to be shorthand for $\mathcal{L}(X) \cap \Sigma^n$. We have that if X is the shift generated by x , then $\mathcal{L}(X) = \mathcal{L}(x)$, and $X = \{x' \mid \mathcal{L}(x') \subseteq \mathcal{L}(x)\}$ [30, Prop. 4.6]. A shift X is said to be *minimal* if it does not have a nonempty proper closed subset Y such that $TY \subseteq Y$. Therefore, the shift generated by any x' in a minimal shift X must be X itself. We have that a word x generates a minimal shift if and only if it is uniformly recurrent [30, Prop. 4.7].

A *cylinder* of $X \subseteq \Sigma^\omega$ is a set of the form $\{ux' : x' \in \Sigma^\omega\} \cap X$ where $u \in \Sigma^*$. The cylinder generated by u will be denoted by $[u]_X$; we omit the superscript X when it is clear from the context. The cylinders are *clopen* (i.e., both closed and open, and in particular are the open balls of X), and a subset of X is clopen if and only if it is a finite union of cylinders. Since there are countably many cylinders, any open set is a countable union of cylinders.

We refer the reader to [17, App. B.5] or a standard text such as [3] for the basic concepts of measure spaces. Given a topological space X , its collection of *Borel-measurable sets*, or simply *Borel sets*, is the σ -algebra generated by the open subsets of X . If X is a shift space, the σ -algebra can equivalently be generated by the cylinders of X .

A (*Borel probability*) *measure* μ on X maps each Borel set B to $\mu(B) \in [0, 1]$, satisfies $\mu(X) = 1$, and is countably additive, i.e., for any countable collection $(B_n)_{n=0}^\infty$ of disjoint Borel sets we have $\mu(\bigcup_n B_n) = \sum_n \mu(B_n)$.

An *invariant measure* μ of a compact dynamical system (X, T) is a (Borel probability) measure μ such that $\mu(T^{-1}B) = \mu(B)$ for every measurable $B \subseteq X$. By the Krylov-Bogolyubov theorem [17, 3.8.4], every (X, T) has an invariant measure. By the Carathéodory extension theorem [17, Thm. B.5.1], such a

measure is defined by its valuations on cylinders, which must all satisfy the following compatibility conditions:

$$\mu([u]_X) = \sum_a \mu([ua]_X) = \sum_a \mu([au]_X). \quad (5)$$

We say that $x \in X$ is a *generic point* of an invariant measure μ (or simply, x is generic for μ) if for every continuous $h \in C(X, \mathbb{R})$ (where $C(X, \mathbb{R})$ denotes the set of all continuous functions from X to $\mathbb{R}$),

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=0}^{N-1} h(T^i x) = \int_X h d\mu.$$

We refer the reader to [17, App. B.5.2] for the precise formal definition of the integral.

2 Automata on Recurrent Words

Semënov showed that if a word $x \in \Sigma^\omega$ is effectively almost-periodic then for any transducer $\mathcal{A}$, the word $\mathcal{A}(x)$, if infinite, is also effectively almost-periodic (see [26, Sec. 3] for an exposition of that result in English). In particular, if x is effectively almost-periodic, then Acc_x is decidable. Using this result and techniques from the proof, Pritykin showed that if a word $y \in \Sigma^\omega$ has a uniformly recurrent suffix x , then for any transducer $\mathcal{A}$, the word $\mathcal{A}(y)$, if infinite, also has a uniformly recurrent suffix [29]. Around the same time, Delvenne et al. independently developed a purely symbolic-dynamic proof [15, Prop. 9]. We dedicate Chapter 2 of the thesis to extensions of these results.

2.1 Preservation of Recurrence

We use the lightweight version of the Krohn-Rhodes theorem (Prop. 1.3) to prove (in an arguably simpler and more insightful manner) the following strengthening of the results of Semënov and Pritykin, which is the headline result of the chapter. We remark that the Krohn-Rhodes theory is crucially needed in the case where the recurrence is not necessarily uniform. The case of recurrent words is novel; see also [22, Chap. 3] for a treatment of the almost-periodic case.

Theorem 2.1. *Let x be an infinite word, and let $\mathcal{A}$ be a transducer such that $\mathcal{A}(x)$ is also an infinite word. If x possesses one of the following properties, then $\mathcal{A}(x)$ also has that property.*

1. *has a recurrent suffix*
2. *has a uniformly recurrent suffix*
3. *has a linearly recurrent suffix*

4. is (effectively) almost-periodic

Furthermore, if the suffix of x is effectively (uniformly, linearly) recurrent and its starting index in x can be computed, then the corresponding suffix of $\mathcal{A}(x)$ also enjoys these properties.

2.2 Prefix-Independent Verification

We present a variant of Thm. 2.1 for (not necessarily effective) almost-periodic words and *prefix-independent* ω -regular languages. Such languages have been considered for the verification of dynamical systems in [1, 2]. Formally, a language $L \subseteq \Sigma^\omega$ is prefix-independent if for all $u, v \in \Sigma^*$, and $x \in \Sigma^\omega$, we have that $ux \in L$ if and only if $vx \in L$. Clearly, if a nonempty language L is prefix-independent, then it is dense in Σ^ω . (In verification parlance, we refer to languages that are dense in Σ^ω as *liveness properties*.) However, the dense language $(\Sigma\Sigma)^*(a\Sigma)^\omega$ is not a prefix-independent language.

We record two structural properties of prefix-independent languages that might be of independent interest.

Proposition 2.2. *Let $L \subseteq \Sigma^\omega$ be a prefix-independent ω -regular language. Either all disjunctive words over Σ are contained in L , or all disjunctive words are excluded from L . Furthermore, we can determine which is the case.*

Proposition 2.3. *An ω -regular language $L \subseteq \Sigma^\omega$ is prefix-independent if and only if $L = \bigcup_{i=1}^d \Sigma^* V_i^\omega$, where $V_1, \dots, V_d$ are regular languages that do not contain the empty word.*

The following is a corollary of a slight generalisation of [1, Sec. 5].

Corollary 2.4. *Let $X \subseteq \Sigma^\omega$ be the set of all recurrent words x such that we have $\mathcal{L}(x) = \mathcal{L}_*$. Let $L \subseteq \Sigma^\omega$ be a prefix-independent ω -regular language. We have that either $X \subseteq L$, or $X \cap L = \{\}$. We can determine which is the case if we can furthermore decide, given any regular language $L' \subseteq \Sigma^+$, whether $\mathcal{L}_* \cap L' = \{\}$.*

The prerequisites on the effectiveness of the factor language are optimal in view of the following result.

Proposition 2.5. *There exists a recurrent word x such that (i) x is effective, (ii) $\mathcal{L}(x)$ is recursive, but (iii) it is undecidable, given a prefix-independent ω -regular language L , whether $x \in L$.*

We then formulate a definition of *effective recurrence* that mirrors the one of effective uniform recurrence in both letter and consequence.

2.3 Effective Recurrence Revisited

To define effective recurrence, we draw inspiration from Colcombet's proof [13] of Semënov's MSO decidability theorem, and prove the following using elementary arguments.

Proposition 2.6. *Let $x \in \Sigma^\omega$ be a recurrent word which is moreover effective. The following conditions are equivalent:*

- *We can decide, given any regular language L , whether $\mathcal{L}(x) \cap L = \{\}$.*
- *We can decide Acc_x .*

If these conditions hold, we say that x is effectively recurrent.

Interestingly, there is a prefix-independent version of this result which does not require x to be effective.

Proposition 2.7. *Let $x \in \Sigma^\omega$ be such that we can decide, given any regular language $L \subseteq \Sigma^+$, whether $x \in (\Sigma^*L)^\omega$. We can then decide, given any prefix-independent ω -regular language $L' \subseteq \Sigma^\omega$, whether $x \in L'$.*

3 Preservation of Factor Frequencies

In this chapter of the thesis, we use the ergodic theory of shift spaces to study words that admit factor frequencies. The central consideration of this chapter is as follows: for which words x does the infinite word $\mathcal{A}(x)$ admit factor frequencies for all transducers $\mathcal{A}$? We shall begin by relating the existence of factor frequencies with the ergodic-theoretic notion of genericity. This will motivate us to consider *uniquely ergodic* shifts. We shall then observe that unique ergodicity is not necessarily preserved by the action of a transducer, and thus focus on shifts that satisfy *Boshernitzan's condition*, which implies unique ergodicity. We will finally prove that this stronger condition is preserved by the action of a transducer.

3.1 Factor Frequencies and Genericity

The following lemma states the connection between factor frequencies and ergodic theory.

Lemma 3.1. *A word $x \in \Sigma^\omega$ admits factor frequencies if and only if x is generic for an invariant measure μ on the shift X that it generates. Should this be the case, we have $f_x(u) = \mu([u]_X)$ for every word $u \in \Sigma^*$.*

Ergodic measures are those for which “almost” all points are generic. An invariant measure μ is called ergodic when the following condition is met: if a Borel set B is such that $T^{-1}B = B$, then $\mu(B)$ is either 0 or 1. Since X is a compact metric space, (X, T) is guaranteed to have an ergodic measure [17, Prop. 3.8.9].

The exception of non-generic points does not arise in dynamical systems (X, T) that are *uniquely ergodic*, i.e., there exists only one invariant measure μ , which is guaranteed to be ergodic [17, Cor. 3.8.10]. In this case, Oxtoby's theorem [30, Thm. 4.3] tells us that all points $x \in X$ are generic for μ .

Lemma 3.2. *If an effectively uniformly recurrent word x generates a uniquely ergodic shift X with invariant measure μ , then x admits computable factor frequencies.*

3.2 Unique Ergodicity is Perishable

The existence of factor frequencies is very susceptible to failure when words are drawn from minimal shift spaces that are not uniquely ergodic. In particular, if (X, T) is not uniquely ergodic, it will have at least two distinct ergodic measures.

Lemma 3.3 (Thm. 1.1 in [16]). *Let X be a minimal shift space which is not uniquely ergodic. There exists a dense set $B \subseteq X$ of words that are not generic for any invariant measure, and hence do not admit factor frequencies.*

As alluded to earlier, transducers do not preserve unique ergodicity [12].

Lemma 3.4 (Veech, Sataev, Chaika). *There exists a uniformly recurrent word $x \in \{0, 1\}^\omega$ that admits factor frequencies by virtue of generating a uniquely ergodic shift, but the uniformly recurrent word $\mathcal{A}(x)$ does not, where $\mathcal{A}$ is the two-state automaton that changes the state upon reading 1, and stays in the same state upon reading 0.*

The natural questions that follow are: can one identify when a shift is uniquely ergodic? Is there a stronger condition on X which precludes the misbehaviour of the above counterexample? Boshernitzan [9] gave a sufficient (but not necessary) condition in answer to the former. In this chapter, we shall show that this condition answers the latter in the affirmative.

Definition 3.5 (Condition (B)). *Let $X \subseteq \Sigma^\omega$ be a minimal shift space. For an invariant measure μ , define $\kappa: \mathbb{N} \rightarrow \mathbb{R}$ as*

$$\kappa(n) = \min_{u \in \mathcal{L}_n(X)} (n \cdot \mu([u]_X)).$$

The minimal shift space $X \subseteq \Sigma^\omega$ satisfies Boshernitzan's condition (also referred to as Condition (B)) if it admits an invariant measure μ such that the corresponding sequence κ satisfies $\limsup_{n \rightarrow \infty} \kappa(n) > 0$. A uniformly recurrent word $x \in \Sigma^\omega$ satisfies Boshernitzan's condition if it generates a shift space X that satisfies Boshernitzan's condition.

Lemma 3.6 (Boshernitzan [9]). *A minimal shift space $X \subseteq \Sigma^\omega$ that satisfies Boshernitzan's condition is uniquely ergodic. If a word $x \in \Sigma^\omega$ satisfies Boshernitzan's condition, then it admits uniform factor frequencies.*

Sturmian words satisfy Boshernitzan's condition, as do linearly recurrent words and hence primitive morphic words [14]. We can prove stronger results in the latter case.

3.3 The Morphic Case

Theorem 3.7. *Let $\mathcal{A} = (Q, q_{\text{init}}, \Sigma, \delta, \Gamma, \lambda)$ be a transducer, and let $x \in \Sigma^\omega$ be a primitive morphic word given as the image under τ of a particular fixed point of a primitive substitution σ . When $\mathcal{A}(x)$ is an infinite word:*

1. *The word $\mathcal{A}(x)$ has a primitive morphic suffix y . We can compute the index at which y begins, and compute substitutions $\hat{\sigma}, \hat{\tau}$ such that $\hat{\sigma}$ is primitive and y is the image under $\hat{\tau}$ of a fixed point of $\hat{\sigma}$.*
2. *The word $\mathcal{A}(x)$ admits computable factor frequencies with closed-form expressions.*

3.4 Preservation of Boshernitzan's Condition

The following is the technical highlight of Chapter 3 of the thesis.

Theorem 3.8. *Let $\mathcal{A}$ be a transducer, and let $x \in \Sigma^\omega$ be a uniformly recurrent word that satisfies Boshernitzan's condition. We have that if $\mathcal{A}(x)$ is an infinite word, then its uniformly recurrent suffix also satisfies Boshernitzan's condition. Furthermore, if x is effectively uniformly recurrent, then we can compute the factor frequencies of $\mathcal{A}(x)$ too.*

Recall our strategy to decompose $\mathcal{A}$ as a cascade of permutation-reset automata, and prove the theorem for these simpler automata. The most technically involved case arises when we have a purely permutation automaton. It is here that we use Boshernitzan's condition in the proof, which involves cohomological arguments.

Sturmian words in particular enjoy elegant algebraic properties [19] that simplify the cohomology above. We get the following corollary, which we acknowledge was known to Olivier Carton and Vincent Delecroix.

Corollary 3.9. *Let $\mathcal{A}$ be a permutation automaton, and let $x \in \Sigma^\omega$ be a Sturmian word whose factor frequencies have effective closed-form expressions. The factor frequencies of $\mathcal{A}(x)$ also have effective closed-form expressions.*

4 Automata on S -adic Words

In this chapter of the thesis, we will show that the directive sequence of a word is an “automatic presentation” of it in the following sense: given a set of substitutions S (throughout this chapter, we assume all substitutions are non-erasing) and an ω -regular language L , we can compute an ω -regular language L' such that an S -adic word is in L if and only if its directive sequence is in L' . This result allows us, for instance, to explicitly characterise the set of pairs of slopes and intercepts that define Sturmian words included in a given ω -regular language L ; in particular we can determine whether this set is empty.

This application can be regarded as the MSO satisfiability problem for Sturmian words, and thus a counterpart in spirit to the Presburger satisfiability

problem for Sturmian words studied in [21]. The results in [21] also rely on an automatic presentation of Sturmian words, which is also based on the Ostrowski numeration system, and technically only slightly different from the directive sequence. We will describe a reconciliation at the end of the thesis chapter.

This chapter can also be viewed as a generalisation of the analogous results of Carton and Thomas [11] for morphic words.

4.1 Structure Lemmas for Directive Sequences

The techniques we develop hinge on the notion of a *congenial* directive sequence. A directive sequence $(\sigma_n)_{n \in \mathbb{N}}$ may be augmented with letters into the sequence $((\sigma_n, a_n))_{n \in \mathbb{N}}$. This augmentation is said to be congenial if for every n , the word $\sigma_{n+1}(a_{n+1})$ begins with the letter a_n . By a straightforward induction, we have that for all $n \geq 1$,

$$\sigma_0 \cdots \sigma_n(a_n) = \sigma_0(a_0) \cdot \sigma_0(v_1) \cdot \sigma_0\sigma_1(v_2) \cdots (\sigma_0 \cdots \sigma_{n-1}(v_n)), \quad (6)$$

where $\sigma_n(a_n) = a_{n-1}v_n$. Observe that for a congenial augmentation of a sequence, $\lim_n \sigma_0 \cdots \sigma_n(a_n)$ exists; we denote the set of such limits for a directive sequence σ by $\text{congenials}(\sigma)$.

The following is proven using König's lemma and a pigeonhole argument.

Lemma 4.1. *Let Σ be an alphabet, and let σ be a directive sequence. The set of congenial sequences that augment σ is nonempty, and has at most $|\Sigma|$ elements.*

Words generated by congenial expansions form the building blocks of words that admit directive sequences.

Lemma 4.2. *Let σ be a directive sequence. We have that $x \in \Sigma^+ \cup \Sigma^\omega$ is directed by σ if and only if x is a (possibly infinite) concatenation of words in $\text{congenials}(\sigma)$.*

4.2 Congeniality and Congruence

We seek to recognise when a word x , generated by a congenial sequence as $\lim_n \sigma_0 \cdots \sigma_n(a_n)$, is included in a given ω -regular language L . To this end, we use a morphism h into an ω -semigroup (M_f, M_ω) with accepting elements F to recognise L . We thus need to recognise whether

$$h \circ \sigma_0(a_0) \cdot \prod_{n=1}^{\infty} h \circ \sigma_0 \cdots \sigma_{n-1}(v_n) \in F. \quad (7)$$

We show that each term of the product can be computed by an automaton, and thus by Lem. 1.2, we can also use an automaton to determine whether the overall product is an element of F .

We define the following three functions, each with finite image. In what follows, morphisms_M denotes the finite set of morphisms from Σ^+ into M_f .

- $\text{head} : S(\Sigma) \rightarrow (\Sigma \rightarrow \Sigma)$ maps (σ, a) to $\sigma(a)(0)$
- $\text{compose} : S(\Sigma) \rightarrow (\text{morphisms}_M \rightarrow \text{morphisms}_M)$ maps (σ, g) to $g \circ \sigma$
- $\text{tail} : S(\Sigma) \rightarrow (\text{morphisms}_M \times \Sigma \rightarrow M_f \cup \{1_M\})$ maps (σ, g, a) to $g(v)$, where $\sigma(a) = \text{head}(\sigma, a) \cdot v$

The product in eq. (7) is then expressed as

$$h \circ \sigma_0(a_0) \cdot \prod_{n=1}^{\infty} \text{tail}(\sigma_n, \text{compose}(\sigma_0 \cdots \sigma_{n-1}, h), a_n).$$

Equivalence of substitutions modulo these functions is, as the title of the discussion suggests, a congruence with a finite set Ξ_M of classes. This allows us to prove several technical results culminating in the following.

Theorem 4.3. *Fix an alphabet Σ . Let $L \subseteq \Sigma^+ \cup \Sigma^\omega$ be the union of a regular language and an ω -regular language (recognised by ω -semigroup M).*

1. *We can construct an automaton $\mathcal{A}$ over the alphabet Ξ_M such that the automaton $\mathcal{A}$ accepts $\text{trace}(\sigma)$ if and only if σ directs a word $x \in L$.*
2. *As a special case, when we have a finite set $S \subseteq S(\Sigma)$, we can construct an ω -regular language $L' \subseteq S^\omega$ such that $\sigma \in L'$ if and only if σ directs a word $x \in L$.*

4.3 Partial Quotients

In this technical discussion, we consider the special case of primitive directive sequences that generate words with a priori known factor complexity. Sturmian words are an immediate application.

Theorem 4.4. *Let $L \subseteq \{0, 1\}^\omega$ be an ω -regular language recognised by an ω -semigroup M . We can compute N such that for any Sturmian word x with slope η and intercept χ , the first N terms of the continued fraction expansion of η and the first N terms of the base- η Ostrowski expansion of χ determine whether $x \in L$.*

4.4 Periodic Generating Sequences

We also study the special case of morphic words, with the aim to understand the extent to which we can decrease our reliance on congeniality.

Theorem 4.5. *Let $L \subseteq \Sigma^+ \cup \Sigma^\omega$ be recognised by a morphism h into a finite ω -semigroup M with accepting elements F . For any word $u \in \Sigma^+$, we can compute $G_u \subseteq \Xi_M$ such that the word $\sigma^\omega(u)$ is well defined and in L if and only if the congruence class of σ is in G_u .*

5 The Monadic Theories of Arithmetic Predicates

In this chapter of the thesis, we apply the results and techniques we have developed to prove the decidability of Monadic Second-Order (MSO) theories of expanded structures $\langle \mathbb{N}; <, (P_i)_{i \in I} \rangle$. The unary predicates $(P_i)_{i \in I}$ (where I is some indexing set) are of “arithmetic” nature, e.g., the powers of a for a natural number a , the set of perfect k -th powers, the set of Fibonacci numbers. The key technical properties of these predicates are that they are ultimately periodic modulo any divisor, and can be proven “effectively sparse” by Baker’s theory of linear forms in logarithms.

It is equivalent to consider the *characteristic word* $x \in \{0, 1, \dots, 2^d - 1\}$ (the bits of the n -th letter, when written in binary, indicate which of the predicates hold for n) of the structure, and decide whether a given (deterministic parity) automaton $\mathcal{A}$ accepts x . We outline our methods with the example of $P_0 = \{2^n \mid n \in \mathbb{N}\}$ and $P_1 = \{3^n \mid n \in \mathbb{N}\}$. For the sake of intuitive convenience, we record position 1 with the letter 1, positions of other powers of 2 with the letter 2, and positions of other powers of 3 with the letter 3, and remaining positions with the letter 0. With this encoding, the characteristic word is $x = 01232000230000002 \dots 3 \dots 2 \dots$. Our plan is as follows.

1. Consider the order word $y = 12322323223 \dots$, which is obtained by deleting intervening occurrences of the letter 0.
2. Show, using periodicity and effective sparsity, that for any automaton $\mathcal{A}$, we can construct a transducer $\mathcal{B}$ such that x and $\mathcal{B}(y)$ are indistinguishable to $\mathcal{A}$. This is reminiscent of the contraction method of [18]: the problem Acc_x reduces to Acc_y .
3. Identify number-theoretic dynamical systems that generate the word y . In our running example, the suffix $y(1, \infty)$ is a cutting sequence (Fig. 2) and hence an effective Sturmian word of slope $\log_6 2$ and intercept 0. More generally, higher-dimensional cutting sequences z are uniformly recurrent, but in three or more dimensions, we often need the conditional decidability of the first-order theory of $\langle \mathbb{R}; <, 0, 1, +, \cdot, \exp \rangle$ (which is subsumed by Schanuel’s conjecture) in order to demonstrate that the factor language of z is recursive before we can apply the results of Chapter 2 to get decidability.

Step (2) can be iterated, yielding some interesting consequences. Its compositional nature gives a direct proof that the MSO theory with P_0 being the set of squares and P_1 being the set of powers of 4 is decidable; on the other hand we can prove that when P_1 is the set of powers of 2 instead, the theory reduces to the acceptance problem for the word z which gives the binary representation of $\sqrt{2}$. If we assume the conjecture that z is disjunctive, the theory is decidable by Prop. 2.6.

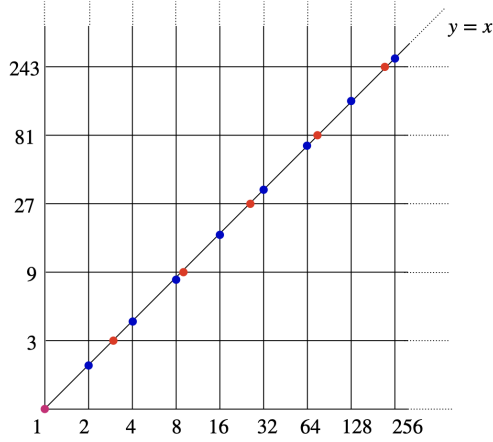


Figure 2: Order word as a two-dimensional cutting sequence. Cutting sequences may be generalised to higher dimensions, e.g., the word recording the order in which powers of 2, 3, 5 occur.

Administrative Matters

My advisor, Joël Ouaknine (MPI-SWS, Universität des Saarlandes), and I intend to assemble the following thesis committee:

- Joël Ouaknine
- Georg Zetsche
- Thomas Colcombet
- Reem Yassawi

The rationale behind choosing the examiners other than the advisor is as follows. We regard Georg Zetsche (MPI-SWS, RPTU Kaiserslautern) as an expert in Verification, Thomas Colcombet (IRIF Paris) as an expert in Automata Theory, and Reem Yassawi (Queen Mary University London) as an expert in Symbolic Dynamics.

In keeping with MPI-SWS requirements, we shall schedule the thesis proposal talk on some date within the first two weeks of October 2026.

As of writing this proposal, most of the thesis has already been drafted, with the exception of the parts in Chapter 5 which describe the acceptance problem for order words. I therefore target a date of submission between November 15, 2026 and December 15, 2026. This gives sufficient time to attend to ongoing projects, complete the draft, and proofread rigorously.

We expect the reviews within two months of submission, i.e., around January 30, 2027. The university regulations stipulate a waiting period of two weeks; this gives us a defence date tentatively in the third week of February 2027. Of course, there may be delays, and my working assumption for all professional purposes is that I will have defended by the end of March 2027.

References

- [1] Shaul Almagor, Toghrul Karimov, Edon Kelmendi, Joël Ouaknine, and James Worrell. Deciding omega-regular properties on linear recurrence sequences. *POPL*, 5, January 2021. doi:10.1145/3434329.
- [2] Christel Baier, Florian Funke, Simon Jantsch, Toghrul Karimov, Engel Lefauchaux, Joël Ouaknine, David Purser, Markus A. Whiteland, and James Worrell. Parameter Synthesis for Parametric Probabilistic Dynamical Systems and Prefix-Independent Specifications. In *CONCUR 2022*, volume 243 of *LIPIcs*, 2022. doi:10.4230/LIPIcs.CONCUR.2022.10.
- [3] Heinz Bauer. *Measure and Integration Theory*. De Gruyter, Berlin, New York, 2001. URL: <https://doi.org/10.1515/9783110866209> [cited 2025-12-08], doi:doi:10.1515/9783110866209.
- [4] Valérie Berthé and Vincent Delecroix. Beyond substitutive dynamical systems: S -adic expansions. *RIMS Kôkyûroku Bessatsu*, B46:81–123, 2014.
- [5] Valérie Berthé, Charles Holton, and Luca Q. Zamboni. Initial powers of Sturmian sequences. *Acta Arithmetica*, 122(4):315–347, 2006.
- [6] Valérie Berthé, Toghrul Karimov, Joris Nieuwveld, Joël Ouaknine, Mihir Vahanwala, and James Worrell. On the decidability of monadic second-order logic with arithmetic predicates. In *Proceedings of the 39th Annual ACM/IEEE Symposium on Logic in Computer Science*, pages 1–14, 2024.
- [7] Valérie Berthé, Toghrul Karimov, and Mihir Vahanwala. Automata on S -Adic Words. In *ICALP*, 2026. doi:10.4230/LIPIcs.ICALP.2026.165.
- [8] Valérie Berthé, Herman Goulet-Ouellet, Toghrul Karimov, Dominique Perrin, and Mihir Vahanwala. Preservation theorems for transducer outputs, 2026. URL: <https://arxiv.org/abs/2606.30013>, arXiv:2606.30013.
- [9] Michael D. Boshernitzan. A condition for unique ergodicity of minimal symbolic flows. *Ergodic Theory and Dynamical Systems*, 12(3):425–428, 1992. doi:10.1017/S0143385700006866.
- [10] J. Richard Büchi. On a decision method in restricted second order arithmetic. In Ernest Nagel, Patrick Suppes, and Alfred Tarski, editors, *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress*, pages 1–11, Stanford, California, 1962. Stanford University Press.

- [11] Olivier Carton and Wolfgang Thomas. The monadic theory of morphic infinite words and generalizations. *Information and Computation*, 176(1):51–65, 2002.
- [12] Jon Chaika. Skew products over rotations with exotic properties. *Geometriae Dedicata*, 168(1):369–385, 2014. doi:10.1007/s10711-013-9835-4.
- [13] Thomas Colcombet. Green’s relations and their use in automata theory. In Adrian-Horia Dediu, Shunsuke Inenaga, and Carlos Martín-Vide, editors, *Language and Automata Theory and Applications*, pages 1–21, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.
- [14] David Damanik and Daniel Lenz. A condition of Boshernitzan and uniform convergence in the multiplicative ergodic theorem. *Duke Mathematical Journal*, 133(1):95 – 123, 2006. doi:10.1215/S0012-7094-06-13314-8.
- [15] Jean-Charles Delvenne, Petr Kůrka, and Vincent Blondel. Decidability and universality in symbolic dynamical systems. *Fundam. Inf.*, 74(4):463–490, December 2006.
- [16] Tomasz Downarowicz and Benjamin Weiss. Lifting generic points. *Ergodic Theory and Dynamical Systems*, 44(9):2565–2580, 2024. doi:10.1017/etds.2023.119.
- [17] Fabien Durand and Dominique Perrin. *Dimension Groups and Dynamical Systems: Substitutions, Bratteli Diagrams and Cantor Systems*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2022.
- [18] Calvin C. Elgot and Michael O. Rabin. Decidability and undecidability of extensions of second (first) order theory of (generalized) successor. *The Journal of Symbolic Logic*, 31(2):169–181, 1966.
- [19] Herman Goulet-Ouellet. Suffix-connected languages. *Theoretical Computer Science*, 923:126–143, 2022. URL: <https://www.sciencedirect.com/science/article/pii/S0304397522002857>, doi:10.1016/j.tcs.2022.05.001.
- [20] Jacques Hadamard. Les surfaces à courbures opposées et leurs lignes géodésiques. *Journal de Mathématiques Pures et Appliquées*, 4:27–73, 1898.
- [21] Philipp Hieronymi, Dun Ma, Reed Oei, Luke Schaeffer, Christian Schulz, and Jeffrey Shallit. Decidability for sturmian words. *Logical Methods in Computer Science*, Volume 20, Issue 3, Aug 2024. URL: <https://lmcs.episciences.org/9980>, doi:10.46298/lmcs-20(3:12)2024.
- [22] Toghrul Karimov. *Algorithmic verification of linear dynamical systems*. PhD thesis, Universität des Saarlandes, 2023. doi:<http://dx.doi.org/10.22028/D291-41630>.

- [23] Kenneth Krohn and John Rhodes. Algebraic theory of machines. i. prime decomposition theorem for finite semigroups and machines. *Transactions of The American Mathematical Society - TRANS AMER MATH SOC*, 116, 04 1965. doi:10.2307/1994127.
- [24] Douglas Lind and Brian Marcus. *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press, 1995.
- [25] A.R. Meyer and C. Thompson. Remarks on algebraic decomposition of automata. *Mathematical Systems Theory*, 3:110–118, 1969. doi:10.1007/BF01746516.
- [26] A. Muchnik, A. Semenov, and M. Ushakov. Almost periodic sequences. *Theoretical Computer Science*, 304(1):1–33, 2003. doi:10.1016/S0304-3975(02)00847-2.
- [27] Dominique Perrin and Jean-Éric Pin. Semigroups and automata on infinite words. In *Semigroups, Formal Languages and Groups*, pages 49–72. Kluwer Academic Publishers, 1995.
- [28] Nir Piterman. From nondeterministic büchi and streett automata to deterministic parity automata. *Logical Methods in Computer Science*, Volume 3, Issue 3, Aug 2007. URL: <https://lmcs.episciences.org/1199>, doi:10.2168/LMCS-3(3:5)2007.
- [29] Yu. L. Pritykin. Finite-automaton transformations of strictly almost-periodic sequences. *Mathematical Notes*, 80(5):710–714, 2006. URL: <https://doi.org/10.1007/s11006-006-0191-7>.
- [30] Martine Queffélec. *Substitution Dynamical Systems - Spectral Analysis*, volume 1294 of *Lecture Notes in Mathematics*. Springer, Berlin, Heidelberg, 2nd edition, 2010.