GIBBS MEASURES AND FACTOR CODES IN SYMBOLIC DYNAMICS

by

SOPHIE QUERIDA TENNANT MACDONALD

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies

(Mathematics)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

April 2023

© Sophie Querida Tennant MacDonald, 2023

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Gibbs measures and factor codes in symbolic dynamics

submitted by Sophie MacDonald in partial fulfillment of the requirements for

the degree of Doctor of Philosophy

in Mathematics

Examining committee:

Brian Marcus, Professor, Mathematics, UBC
Co-Supervisor

Tom Meyerovitch, Professor, Mathematics, Ben Gurion University of the Negev
Co-Supervisor

Andrew Rechnitzer, Professor, Mathematics, UBC
University Examiner

Bruce Shepherd, Professor, Computer Science, UBC
University Examiner

Additional Supervisory Committee Members:

Lior Silberman, Professor, Mathematics, UBC
Supervisory Committee Member

Omer Angel, Professor, Mathematics, UBC
Supervisory Committee Member

Abstract

This dissertation consists principally of four of the author's research articles, included as Chapters 2 through 5, all within or related to the area of symbolic dynamics, especially to shifts of finite type (SFTs) and sofic shifts. More specifically, each of the articles involves either Gibbs measures or factor codes, with Chapter 3 involving both in an essential way.

The research in Chapter 2, in the thermodynamic formalism of dynamical systems, shows that two commonly used definitions of a Gibbs measure coincide for an arbitrary subshift over an arbitrary countable group, and that two different forms of one of these definitions are equivalent under certain regularity hypotheses. The main innovation is a more careful decomposition of holonomies of the Gibbs relation than had previously appeared.

Chapter 3 is concerned with the Dobrushin and Lanford-Ruelle theorems, which relate Gibbs and equilibrium measures. This chapter establishes these theorems for irreducible sofic shifts, by lifting to an irreducible SFT and using cyclic structure to reduce to the classical mixing SFT case. The result was already known, by work of Haydn-Ruelle and Baladi, but the proof presented in this chapter is more self-contained. In particular, the argument makes extensive use of the properties of doubly transitive points, and relates them in a novel way to the Gibbs relation studied in Chapter 2.

The work in Chapter 4 resolves two new cases of a problem in automata theory that originated in symbolic dynamics. The methods combine the notion of stability, introduced by Culik-Karhumaki-Kari and used by Trahtman in the solution to the road colouring problem, with the framework of graph homomorphisms introduced by Ashley-Marcus-Tuncel. The chapter also includes several new algorithms relevant to the problem.

Chapter 5 solves a symbolic dynamics problem related to zero-error coding. The main result, which generalizes Krieger's embedding theorem, characterizes the subshifts of a given mixing SFT on which a given sliding block code is injective. The proof follows the overall strategy of Krieger's proof, but with significant technical innovations required to make the embedding compatible with the code.

Lay Summary

In certain areas of physics and computer science, a system of interest is represented by a large collection of simple pieces arranged in a line or a grid. Each of the pieces can be in one of a few states, and each piece is influenced by other pieces, especially the ones nearby. In physics, the classic example is a magnet, made up of many pieces each pointing north or south, with each piece trying to align itself with nearby pieces. In computing, the main examples are methods of storing data in patterns of 0's and 1's in a way that is resilient against errors, by only allowing certain patterns. There may be several ways of writing down models like this for the same system. This thesis investigates several questions about when one system is essentially the same as another, even though they may be written down in different-looking ways.

Preface

Chapter 2. A version of this chapter has been published [15], having previously appeared as an arXiv preprint [17]. An account of the research based closely on the published version appeared in Borsato's dissertation at the University of Sao Paulo, defended in October 2022 and not yet publicly available as of the date of submission of this dissertation.

We began work on the research that is presented in this chapter after a meeting in the summer of 2019 in which I raised the question of whether several definitions of a Gibbs measure extant in the literature were in fact equivalent. I formulated the paper's new definitions, namely Gibbs measures with respect to general measurable cocycles, groups with bounded sphere ratios, shell-regular and volume-regular potentials, and full-dimensional interactions. I devised the proofs of the main results, building on Borsato's detailed knowledge of the works of Kimura and Muir. I wrote the majority of the manuscript, the quality of which Borsato brought to a substantially higher standard by identifying errors and gaps in my reasoning and presentation while the work was in progress and while we were preparing the manuscript for submission.

Chapter 3. A version of this chapter appeared as an arXiv preprint [16] but was not submitted for publication. As with Chapter 2, an account of the research based closely on the preprint appeared in Borsato's dissertation at the University of Sao Paulo, defended in October 2022 and not yet publicly available as of the date of submission of this dissertation. As with Chapter 2, I devised the overall structure of the argument and wrote the majority of the manuscript, while Borsato contributed crucially to the review of the literature and corrected many errors throughout the work and the writing.

Chapter 4. A version of this chapter appeared as an arXiv preprint [51]. A manuscript based closely on that preprint has been submitted for publication and is in review as of the date of submission of this dissertation. I carried out the work and the writing with frequent supervisory feedback from Brian Marcus and Tom Meyerovitch.

Chapter 5. A version of this chapter appeared as an arXiv preprint [50]. A manuscript based closely on that preprint has been submitted for publication and is in review as of the date of submission of this dissertation. I carried out the work and the writing with frequent supervisory feedback from Brian Marcus and Tom Meyerovitch.

Contents

Acknowledgments

I offer my thanks to my supervisory committee, especially my supervisors Brian Marcus and Tom Meyerovitch, who have been extraordinarily generous with their time, attention, and advice over the last several years. The research presented in this dissertation would not have happened without their extensive, careful feedback and stimulating questions. Similar thanks are owed to Doug Lind.

I thank the many mentors at UBC who have enriched and supported my scholarly development beyond research, especially Malabika Pramanik, Matt Coles, Joseph Topornycky, and Elyse Yeager.

I thank Luísa Borsato for being a tremendously skilled, persistent, and supportive collaborator and a dear friend. In a similar spirit, I thank my close friends in or near this department—Caleb Marshall, Maricela Best McKay, Katie Faulkner, Mia Kramer, and Anshula Gandhi—for our years of studying, teaching, strategizing, and celebrating. It has been a joy to be confused and frustrated together.

Beyond the Math Department, I thank the many friends who have accompanied, supported, and entertained me these last few years, especially Matthew Jordan, Kayla Besse, Joseph Burkhart, Walker Williams, Alison Stevens, Yotam Ronen (and Noga Vieman), and Alexis Kellum-Creer.

Finally, I thank my family for supporting me in our tradition of vigorous curiosity from the beginning, especially my mother, without whom I would have been entirely sunk countless times along the way.

Chapter 1

Introduction

1.1 Symbolic dynamics and entropy

1.1.1 Symbolic dynamics

This dissertation treats several problems in or related to symbolic dynamics. Symbolic dynamics is a branch of the theory of abstract dynamical systems with a strong computational and combinatorial flavour, drawing heavily on automata theory, formal language theory, computability theory, and information theory as well as on ergodic theory, topological dynamics, and statistical thermodynamics.

Much of the distinctive character of symbolic dynamics (see [49] for a standard introduction) comes from the fact that, in contrast to other branches of ergodic theory and topological dynamics, which study a diverse collection of spaces and often a wide array of transformations of any given space (for instance, the many manifolds on which one can consider smooth flows, or the many automorphisms of the two-dimensional torus), the dynamical system under consideration in symbolic dynamics is always a subsystem of one of a small number of systems, i.e. spaces and transformations of them, with very similar topological and measure-theoretic properties. Once the "ambient" space and the transformation (or set of transformations) are fixed, one studies the restrictions of the transformation to various invariant subsystems, which exhibit remarkable variety in their dynamical properties, as well as maps between these subsystems that commute with the overall transformation.

The ambient system in question is a *full shift* space. A full shift is the set of all functions from some fixed countably infinite group or semigroup, often the integers or a higher-dimensional integer lattice, to a fixed finite (or sometimes, but never in this dissertation, countably infinite) set of symbols, called the alphabet. There is therefore a full shift for each countable (semi)group and each positive natural number (the cardinality of the alphabet). The (semi)group acts on this space of functions by translation (hence the term *shift space*), with the transformation or set of transformations that define the system generating the group. In the most classical theory, the alphabet consists of the binary symbols 0 and 1 and the group is isomorphic to the group of integers, with a single generator, namely the shift transformation.

When one equips the alphabet with the discrete topology and takes the product topology over the group, the full shift is a zero-dimensional compact metric space on which the shift map is a homeomorphism. The aforementioned functions from the group to the alphabet are the points in the space. Symbolic dynamics studies the dynamical properties of the restrictions of the shift map to the closed, shift-invariant subspaces of the full shift, known as *subshifts*, and the continuous, shift-commuting (i.e. equivariant) maps between these systems. Any subshift is characterized by the set of finite patterns of symbols that appear in its points; in particular, a one-dimensional subshift (i.e. a subshift over the integers) is characterized by the language consisting of the finite words it allows. This is the key idea of symbolic dynamics: a subshift is a topological dynamical system defined by constraints on finite configurations of symbols.

1.1.2 Entropy

Many of the stationary finite-valued stochastic processes or random fields studied in statistical physics and information theory can be understood as shift-invariant Borel probability measures on subshifts, which leads to fruitful interactions of these areas with symbolic dynamics via the application of ergodic theory, the study of measure-preserving transformations in general.

In particular, a central concept in symbolic dynamics, imported from ergodic theory, is that of *entropy*, which refers to several different, closely related quantities. The (topological) entropy of a subshift is the asymptotic exponential growth rate of the number of finite patterns appearing in points in the subshift, with respect to the size of the patterns. Although elementary, this definition does not explain the significance of entropy in the subject. We need some history in order to convey the context and meaning of the present dissertation to a reader not trained in ergodic theory.

The definition of the entropy of a measure-preserving system was introduced in its modern generality by Kolmogorov [46], who introduced it in order to resolve isomorphism problems in ergodic theory. In particular, he and Sinai [71] showed that entropy is an invariant of measure-theoretic isomorphism and was able to distinguish systems which were not previously known to be non-isomorphic, the best-known pair being the uniform Bernoulli processes on the full 2- and 3-shifts. Previously, the main invariants known had been qualitative, such as ergodicity and other mixing properties, and spectral, referring to the spectrum of the Koopman operator, which translates functions along orbits. The main result on spectral invariants was due to von Neumann, who was motivated to work on the isomorphism problem because he considered the essential properties of a physical system with an invariant measure to be those that were preserved by measure-theoretic isomorphism [60].

When Kolmogorov introduced entropy into ergodic theory, he was drawing on the work of Shannon, who had defined entropy in the special case of finite Markov chains and shown that it was the essential quantity for determining whether a given communication channel could transmit a given Markov source of messages with acceptably low probability of error [69]. Shannon had observed that entropy, as defined and used in statistical physics by Boltzmann, Gibbs, and later von Neumann, is a generalization of a quantity introduced to information theory by Hartley [35] to calculate, among other things, the relationships between the redundancy of the code used for a telegraph and the electrical properties of the telegraph lines required for a receiver to reliably decode the sender's messages.

In these thermodynamic and information-theoretic contexts for entropy, although the probability measures involved can often be formulated as measures on subshifts, the dynamical and topological aspects are secondary. However, Kolmogorov also introduced another notion of entropy, namely a kind of covering number for a metric space–specifically, the ε-entropy of a metric space is the logarithm of the largest number of points that can be packed into the space such that any two are at distance at least ε from each other [47]. A point in the space can be approximated to within a distance of $\varepsilon/2$ by the nearest of these points, which is in a precise sense analogous both to the decoding of continuously varying voltages over an analog channel back into symbols, and to the approximation of the continuously varying thermodynamic variables in a small region of a system by a finite set of possible values. Moreover, by considering a continuous transformation of a metric space (i.e. a topological dynamical system), one can ask more generally about the largest number of points such that after n iterations of the transformation, any two of the points have been separated by a distance of at least ε at least once in their orbits. This number grows exponentially in n, and the limit, as ε tends to 0, of its asymptotic exponential growth rate is the topological entropy of the system [76].

The concept of topological entropy allows us, both by analogy and by the essential theorems relating it to measure-theoretic entropy, to ask information-theoretic and thermodynamic questions about dynamical systems with no obvious connection to telecommunications or thermodynamics. Both the information-theoretic and the statistical thermodynamic influences on symbolic dynamics are present in this dissertation, with the thermodynamics emphasized in Chapters 2 and 3 and the information theory in Chapter 5, as we now elaborate. In what follows, see the respective chapters for precise definition and statements.

1.2 Chapter 2

This chapter is concerned with two notions of a Gibbs measure on a subshift over a countable group. The first of these is defined by the Dobrushin-Lanford-Ruelle (DLR) equations, or equivalently a Gibbsian specification. This notion of a Gibbs measure appears for instance in the classical theorems of Dobrushin [28] and Lanford-Ruelle [38]. The second is the notion of a conformal measure, introduced in [63] and [27] and used for instance by Meyerovitch in [56] as the setting for a stronger Lanford-Ruelle theorem. There are other definitions in the literature, such as a Gibbs measure in the sense of Bowen, but we do not consider these here.

The purpose of this chapter is to show that the two notions of Gibbs measure recalled above coincide in some generality. Our results build on those of Kimura, obtained in his Master thesis [44], who proves two results relevant here. The first is that every conformal measure, with respect to an appropriately regular potential, satisfies the DLR equations for that potential. The second is a partial converse, namely that every measure satisfying the DLR equations for such a potential is a topological Gibbs measure (see e.g [44], Definition 5.14). This is a weaker property than being conformal, although equivalent on certain subshifts, such as shifts of finite type (SFTs) [56]. Sarig ([68], Proposition 2.1) shows that the two notions of Gibbsianness are equivalent in the case of a one-sided topological Markov shift, using martingale and Ruelle transfer operator methods (note that he uses the word "conformal" for a notion that is related to but different from the one we consider). Cioletti-Lopes-Stadlbauer ([24]) prove the equivalence of the DLR and conformal notions of Gibbs measure for one-sided, one-dimensional shifts with a compact metric alphabet and a continuous potential. Muir [58] also obtains the full equivalence for the full shift on $\mathbb{Z}^d$ over a countable alphabet.

Our main result in this chapter, Theorem 2.3.4, strengthens one of Kimura's results in a more general setting. Specifically, we show that any measure satisfying certain equations with respect to a measurable cocycle on the Gibbs relation must also be conformal with respect to that cocycle. When the cocycle is induced either by an interaction or by a potential in the standard way, these equations reduce to the classical DLR equations. We prove this result for arbitrary subshifts with finite alphabet on an arbitrary countable group. The results of Kimura and Sarig in the forward direction (conformal implies DLR) can also be generalized to our setting; in §2.3, we mention the idea for the proof but refer readers to [44] for the details in Kimura's setting, as the proof strategy changes very little.

1.3 Chapter 3

The Dobrushin theorem establishes sufficient conditions on shift spaces X and potentials $f \in C(X)$ such that every Gibbs measure for f is an equilibrium measure for f. This theorem holds in any shift space, not necessarily of finite type, with a certain mixing condition known in the literature as condition (D) [66]. This condition is implied, for instance, by strong irreducibility, and in this paper we only use the strongly irreducible case of the classical theorem.

The classical converse to the Dobrushin theorem is known as the Lanford-Ruelle theorem. To our knowledge, the most general natural hypothesis known for the Lanford-Ruelle theorem is the topological Markov property [9], which is satisfied in particular by SFTs. Examples are also known, however, of shift spaces which lack the topological Markov property, but for which the conclusion of the Lanford-Ruelle theorem nevertheless holds, at various levels of generality [56]. It is therefore desirable to extend the Lanford-Ruelle theorem beyond the class of shifts with the topological Markov property, in the hope of explaining such examples.

In this chapter, we do not treat the examples in [56], but we do prove a Lanford-Ruelle theorem for irreducible sofic shifts in one dimension (Theorem 3.4.7), which generally lack the topological Markov property. This is related to a question of Kitchens-Tuncel ([45], Remark 7.10(iii)). The proof relies on a preservation of Gibbsianness result for almost invertible factor codes on irreducible SFTs (Proposition 3.3.7), which we generalize to finite-to-one factor codes in Corollary 3.4.10. We prove Theorem 3.4.7 by lifting an equilibrium measure on a sofic shift to an equilibrium measure on a covering SFT, which is

Gibbs by the classical Lanford-Ruelle theorem, then concluding by Proposition 3.3.7 that the original equilibrium measure is Gibbs. Irreducibility of the sofic shift is essential: the Lanford-Ruelle theorem holds for SFTs with no irreducibility assumption, but it is false in general for reducible sofic shifts. The simplest counterexample is the sunny-side-up shift (the set of sequences in $\{0,1\}^{\mathbb{Z}}$ with at most a single 1) with its unique shift-invariant measure.

We also extend the Dobrushin theorem to irreducible sofic shifts (Theorem 3.4.11), which generally lack the mixing properties hypothesized in the classical version. Here, our approach is based on the cyclic structure of an irreducible SFT, combined with our other results.

After posting a version of this chapter on the arXiv, we became aware of the work of Viviane Baladi [8], which obtains the same Dobrushin-Lanford-Ruelle result for Hölder potentials on finitely presented systems, of which irreducible sofic shifts are a particular case. We also became aware of the work of Haydn-Ruelle [36], which does the same for expansive homeomorphisms with specification, of which mixing sofic shifts are a particular case. We hope that the self-contained symbolic approach of this paper, which leverages the classical Dobrushin and Lanford-Ruelle theorems essentially as black boxes, may be accessible to a wider audience and may draw attention to the earlier work by Baladi and Haydn-Ruelle.

1.4 Chapter 4

1.4.1 Background: from toral automorphisms to finite automata

In this chapter, we are concerned with a generalization of the road (colouring) problem. That problem, posed by Adler-Goodwyn-Weiss in [2] and solved by Trahtman in [74], asked whether every strongly connected, aperiodic directed graph of constant out-degree is the underlying graph of a synchronizing deterministic finite automaton (DFA). A *finite automaton* is a very simple model of a computer, consisting of a directed graph with edges labeled with symbols from a finite alphabet. When there is a distinguished initial state (vertex) and a distinguished set of *accepting* states, one considers the question of which words over the alphabet label a route from the initial state to one of the accepting states. More generally, one can ask about the mappings induced on the set of states by various words. An automaton, or graph labeling, is *synchronizing* if for some word it routes all states to a common state–that is, if the image of the mapping induced by some word is a singleton. Trahtman's road colouring theorem (Theorem 4.4.1 below) gives an affirmative answer to the road problem, showing that every graph with the aforementioned properties indeed admits a synchronizing colouring, and [21, 10] give a generalization to periodic graphs (Theorem 4.4.2).

The motivation for the road problem comes from ergodic theory. Specifically, a weak form of the road colouring theorem was used to prove the main theorem of [2], which gives a criterion for measure-theoretic isomorphism of certain Markov chains. This was the first time the road problem was explicitly posed, although the real origin of the problem is the earlier paper [3], which concerns the analogous isomorphism problem for hyperbolic automorphisms of the two-dimensional torus.

This chapter concerns a generalization of the road problem motivated by graph-theoretic invariants for a different, but related, isomorphism relation in ergodic theory. Specifically, Ashley-Marcus-Tuncel [6] identified a graph-theoretic criterion for isomorphism of one-sided stationary Markov chains, implicit in [19], and gave a complete, effectively computable set of isomorphism invariants. They observed that a certain conjectural uniqueness result (the $O(G)$ conjecture, below in §4.3.1) would, if proven, simplify the set of invariants, and proved the conjecture in a family of special cases. The conjecture reduces to a generalization of the road problem involving certain right-resolving graph homomorphisms (which we call *right-resolvers*; see §4.2.1). For graphs of constant out-degree, these homomorphisms coincide with road colourings.

1.4.2 Main ideas and contributions

The main purpose of this chapter is to present new results toward the $O(G)$ conjecture. There are two main ideas in the paper. The first idea concerns the *stability relation* of a right-resolver, which was introduced in [37] for DFAs or road colourings. For context, Kari [39] solved the road problem in the Eulerian case by finding, for a given graph, a road colouring with nontrivial (i.e. not merely diagonal) stability relation, then recursively finding a synchronizing road colouring of the strictly smaller quotient graph (in which the states are stability classes), and lifting it to the original graph. Trahtman's solution of the full road problem uses the same inductive strategy, paired with a more sophisticated technique for obtaining a colouring with a nontrivial stability relation.

By determining how the stability relation behaves with respect to composition of right-resolvers, we are able to study the aforementioned recursive lifting constructions more systematically, allowing us to apply them toward the $O(G)$ conjecture. In particular, we adapt Trahtman's proof of the road colouring theorem to cover a larger family of cases (Theorem 4.4.3), in which the out-degrees of states are allowed to vary cyclically. To do so, we generalize, to the setting of right-resolvers, a sufficient condition for a road colouring to have a nontrivial stable pair, based on the idea of a function graph with a unique tallest tree. This condition is at the heart of all proofs of the road colouring theorem to date, and its importance has motivated detailed analysis [12].

The second main idea in this chapter is a graph property that we call *bunchiness*, along with a weaker property called *almost bunchiness*. Bunchy and almost bunchy graphs are characterized by the property that the right-resolvers they admit are unique up to automorphisms in a certain sense (Proposition 4.5.4). We highlight the implicit role of bunchiness both in [6] and in the road colouring literature, and prove the $O(G)$ conjecture for bunchy and almost bunchy graphs (Theorem 4.5.6). Furthermore, we show that the fiber product of right-resolving homomorphisms satisfies a universal property (Theorem 4.5.12) that further highlights the essential role of bunchy graphs.

Motivated by these results, we introduce a new conjecture, which we call the bunchy factor conjecture (see §4.6), asserting essentially that the $O(G)$ conjecture can be proved using the stability approach that Trahtman used to prove the road colouring theorem, with bunchy graphs as the base of the recursion. Another way of articulating this conjecture is that the barrier to proving the $O(G)$ conjecture is our lack of a sufficiently general method of producing homomorphisms with nontrivial stability relation. The fact that the bunchy factor conjecture implies the $O(G)$ conjecture is made explicit in Proposition 4.6.2, which relies primarily on Proposition 4.5.15, a uniqueness result that uses the universal property of the fiber product in an essential way.

1.5 Chapter 5

As mentioned above, Shannon introduced a model of a noisy communication channel [69], in which the input and output are modeled by stationary probability measures on a space of sequences of symbols. Shannon gave conditions under which the input can be recovered from the output, at least with an acceptable rate of error or ambiguity, in the case of a Bernoulli source, and this work has since been extended to more general sources [43].

This chapter is motivated by the particular question of when one can ensure zero error, not just almost surely as in information theory but in fact deterministically. A deterministic channel can be modeled by a sliding block code, i.e. a continuous, shift-commuting map on a subshift, on which a stationary process could be supported. In this model, we can use symbolic dynamics to investigate the effects of *deterministic noise* [54], also called *distortion* [69], which we can interpret as a failure of injectivity of the sliding block code representing the channel, even in the absence of random errors.

The main result of this chapter, Theorem 5.1.1, determines the extent to which the non-injectivity of a sliding block code on a mixing SFT can be avoided by restricting to a subshift of the domain. Interpreting the sliding block code as a channel with deterministic noise, Theorem 5.1.1 characterizes the sources with entropy strictly lower than that of the output which can be transmitted without error or ambiguity. That is, let X be a mixing SFT, Y a mixing sofic shift, and $\pi : X \to Y$ a factor code.

Let Z be a subshift. For there to exist an embedding $\psi : Z \to X$ such that $\pi \circ \psi$ is injective, the entropy of Z must be at most that of Y (with equality possible only if Z is conjugate to Y), and Z must have few enough periodic points that $\pi \circ \psi$ can be injective at least on the periodic points in Z. The content of Theorem 5.1.1 is that these two sufficient conditions are enough.

In that spirit, Theorem 5.1.1 is a generalization of a theorem of Krieger (Theorem 2 in [48], quoted as Theorem 5.1.2 in this dissertation), which shows that two necessary conditions for the existence of an embedding of a subshift Z into a mixing SFT X are also sufficient. These conditions are that the entropy of Z must be at most that of X, with equality possible if and only if Z and X are conjugate, and that there must a shift-commuting injection at least from the periodic points of Z to those of X. Krieger's theorem is false if X is taken to be merely mixing sofic; indeed, the embedding problem for mixing sofic shifts remains open, and seems to be delicate [73].

Because the asymptotic exponential growth rate, with respect to least period, of the number of periodic points in a mixing sofic shift is equal to the entropy of the shift, the periodic point condition, in both Krieger's theorem and Theorem 5.1.1, presents only a finite number of possible obstructions, coming from low-order periodic points in the source shift Z. In the proof of Theorem 5.1.1, which is closely modeled on the proof of Krieger's theorem as presented in [49], points in Z are broken into blocks of moderate length and long blocks coming from lower-order periodic points. The entropy and periodic point conditions are used to construct injections on the moderate and long segments respectively, first to blocks in Y and then to blocks in X. The moderate and long blocks are then stitched together using blocks which we call "stamps", in such a way that the locations of the original blocks in Z can be unambiguously identified once the factor code π has been applied. Much of the chapter is concerned with establishing the existence of properties of these stamps.

Chapter 2

Conformal measures and the Dobrushin-Lanford-Ruelle equations

2.1 Introduction

The plan is as follows. In §2.2, we review the definitions and basic facts required to prove our main result in §2.3. In §2.4 and §2.5, we recall well-known material on interactions and potentials, respectively, in order to show that the equations involved in our main theorem do in fact reduce to the classical DLR equations. In §2.6, we recall results of Muir and Kimura, elaborating on Ruelle, by which a potential can be constructed from a sufficiently regular interaction, and vice versa, with "physical" data (Gibbs and equilibrium measures) preserved.

In §2.5 and §2.6, we require that the underlying group admits a finite generating set that yields a certain spherical growth condition, defined in §2.5. This condition is satisfied, for any generating set, by any group of polynomial growth of nilpotency class at most 2, such as $\mathbb{Z}^d$, the case of greatest physical interest. It is also satisfied by any free group F_n, with the usual generating set of cardinality n, and, conditional on a folklore conjecture, is satisfied by any nilpotent group.

2.2 Cocycles and the Gibbs relation

Throughout, let G be a countable group with identity e. Let $\mathcal{A}$ be a finite alphabet equipped with the discrete topology, and $X \subseteq \mathcal{A}^G$ a subshift, i.e., a closed set in the product topology, invariant under the shift action of G via $(g \cdot x)_h = x_{g^{-1}h}$. The topology on X is generated by cylinders, i.e., sets of the form $[\omega] = \{x \mid x_\Lambda = \omega\}$ for finite sets $\Lambda \Subset G$. We use the notation $\Lambda \Subset G$ to indicate that Λ is a finite subset of G. This topology can be induced by a metric such that the resulting metric space is complete and separable; that is, $\mathcal{A}^G$ is a Polish space. We equip X with the Borel σ-algebra $\mathcal{F}$. For a set $A \subseteq G$, we write $\mathcal{F}_A$ for the sub-σ-algebra of $\mathcal{F}$ determined by A—that is, for any $E \in \mathcal{F}_A$ and any $x, x' \in X$, if $x \in E$ and $x_A = x'_A$, then $x' \in E$.

The *Gibbs relation*, also called the asymptotic relation, is the equivalence relation $\mathfrak{T}_X \subset X \times X$ such that $(x, y) \in \mathfrak{T}_X$ if and only if $x_{\Lambda^c} = y_{\Lambda^c}$ for some finite set $\Lambda \Subset G$. Let $(\Lambda_N)_{N=1}^{\infty}$ be a sequence of finite sets exhausting G, i.e., $(\Lambda_N)_{N=1}^{\infty}$ is an increasing sequence and $G = \cup_{N=1}^{+\infty} \Lambda_N$. Define the subrelation $\mathfrak{T}_{X,N} = \{(x,y) : x_{\Lambda_N^c} = y_{\Lambda_N^c}\} \subseteq \mathfrak{T}_X$. Observe that, for each subrelation $\mathfrak{T}_{X,N}$, each equivalence class is a finite set, and that $\mathfrak{T}_X = \cup_{N=0}^{\infty} \mathfrak{T}_{X,N}$. (In the language of Borel equivalence relations, this means that $\mathfrak{T}_X$ is *hyperfinite* [41], which we mention for context, although we do not use any theorems about hyperfiniteness in this paper.) In particular, every equivalence class in $\mathfrak{T}_X$ is at most countable. Note that we can write each subrelation as $\mathfrak{T}_{X,N} = \cap_{n=N}^{\infty} \cup_{\omega \in \mathcal{A}^{\Lambda_n \setminus \Lambda_N}} [\omega] \times [\omega]$, which shows that $\mathfrak{T}_{X,N}$ is a measurable subset of $X \times X$ in the product σ-algebra $\mathcal{F} \otimes \mathcal{F}$, as is $\mathfrak{T}_X$.

For Borel sets $A, B \subseteq X$, a *holonomy* of $\mathfrak{T}_X$ is a Borel isomorphism $\psi : A \to B$ such that $(x, \psi(x)) \in \mathfrak{T}_X$ for all $x \in A$. We say that a holonomy ψ is *global* if $A = B = X$. The definitions for $\mathfrak{T}_{X,N}$ are analogous, with a holonomy of $\mathfrak{T}_{X,N}$ also a holonomy of $\mathfrak{T}_X$, for every N.

For a Borel set $A \subseteq X$, we denote $\mathfrak{T}_X(A) = \bigcup_{x \in A}\{y \in X | (x,y) \in \mathfrak{T}_X\}$, and the same for the subrelations. The saturations $\mathfrak{T}_X(A)$ and $\mathfrak{T}_{X,N}(A)$ are easily shown to be Borel using the fact that the diagonal in $X \times X$ is measurable in the product σ-algebra, which follows as an easy exercise from the fact that X is Polish.

Lemma 2.2.1. *There exists a countable group Γ of global holonomies of X such that*

$$\mathfrak{T}_X = \{(x, \gamma(x)) : x \in X, \gamma \in \Gamma\}.$$

In other words, Γ generates $\mathfrak{T}_X$.

Proof. The group Γ can be described explicitly as a countable increasing union of finite groups Γ_N. For each N, the group Γ_N generates $\mathfrak{T}_{X,N}$ and is isomorphic to the symmetric group of order $|\mathcal{A}^{\Lambda_N}|$. Take Γ_N to be generated by holonomies ψ of the following form: given $\omega, \eta \in \mathcal{A}^{\Lambda_N}$, define $\psi_{\omega,\eta} : X \to X$ by

$$\psi_{\omega,\eta}(x) = \begin{cases} \eta x_{\Lambda_N^c} & x_{\Lambda_N} = \omega,\ \eta x_{\Lambda_N^c} \in X \\ \omega x_{\Lambda_N^c} & x_{\Lambda_N} = \eta,\ \omega x_{\Lambda_N^c} \in X \\ x & \text{otherwise} \end{cases}$$

That is, $\psi_{\omega,\eta}$ exchanges ω and η, wherever possible, and otherwise does nothing. These involutions were considered in [56] and [44], for slightly different purposes.

Observe that $(x,y) \in \mathfrak{T}_{X,N}$ if and only if there exists $\psi \in \Gamma_N$ with $\psi(x) = y$, so $\mathfrak{T}_{X,N}$ is precisely the orbit relation of Γ_N. The result for Γ is immediate. $\square$

Remark 2.2.2. We mention for context that Lemma 2.2.1 is a special case of the main theorem of [30], which in fact asserts the same for any Borel equivalence relation on a Polish space in which every equivalence class is countable. This result was adapted to the symbolic setting in [56], with the countability of the equivalence classes established via the expansivity of the shift action. The proof is presented for subshifts over $\mathbb{Z}^d$, but the same proof goes through for arbitrary countable groups without modification. However, since we establish Lemma 2.2.1 directly, we do not need to appeal to the theorem of [30] (nor the symbolic corollary in [56]).

We say that a measure μ on X (by which we always mean a Borel probability measure) is $\mathfrak{T}_X$-nonsingular if for every Borel $A \subset X$ with $\mu(A) = 0$, we have $\mu(\mathfrak{T}_X(A)) = 0$. Note that if μ is $\mathfrak{T}_X$-nonsingular and $\psi : A \to B$ is a holonomy of $\mathfrak{T}_X$, then whenever $E \subset A$ has $\mu(E) = 0$, we have $\mu(\psi(E)) \leq \mu(\mathfrak{T}_X(E)) = 0$. In particular, the Radon-Nikodym derivative $\frac{d(\mu\circ\psi)}{d\mu}$ is well-defined. The same holds with $\mathfrak{T}_X$ replaced by $\mathfrak{T}_{X,N}$.

A (real-valued) cocycle on $\mathfrak{T}_X$ is a Borel measurable function $\phi : \mathfrak{T}_X \to \mathbb{R}$ such that $\phi(x,y) + \phi(y,z) = \phi(x,z)$ for all $x,y,z \in X$ with $(x,y),(y,z) \in \mathfrak{T}_X$ (so that $(x,z) \in \mathfrak{T}_X$ as well). Any cocycle on $\mathfrak{T}_X$ clearly restricts to a cocycle on $\mathfrak{T}_{X,N}$, for any given N. Given a $\mathfrak{T}_X$-nonsingular measure μ on X, we say that a Borel function $D : \mathfrak{T}_X \to \mathbb{R}$ is a *Radon-Nikodym cocycle* on $\mathfrak{T}_X$ with respect to μ if the pushforward of μ by any holonomy $\psi : A \to B$ of $\mathfrak{T}_X$ satisfies $\frac{d(\mu\circ\psi)}{d\mu}(x) = D(x,\psi(x))$ for μ-a.e. $x \in A$. It is routine to show, using Lemma 2.2.1, that any $\mathfrak{T}_X$-nonsingular measure μ on X has a μ-a.e. unique Radon-Nikodym cocycle.

Definition 2.2.1 (conformal measure)**.** Let μ be a $\mathfrak{T}_X$-nonsingular Borel probability measure on X, and let $\phi : \mathfrak{T}_X \to \mathbb{R}$ be a cocycle. We say that μ is $(\phi, \mathfrak{T}_X)$-*conformal* if for any holonomy $\psi : A \to B$ of $\mathfrak{T}_X$, with A and B Borel sets, we have

$$\mu(B) = \int_A \exp(\phi(x,\psi(x)))\, d\mu(x)$$

Note that this is equivalent to the condition that

$$D_{\mu,\mathfrak{T}_X}(x,\psi(x)) = \exp(\phi(x,\psi(x)))$$

for μ-a.e. $x \in A$. Note also that a $\mathfrak{T}_X$-nonsingular measure is conformal precisely with respect to the logarithm of its Radon-Nikodym cocycle.

Remark 2.2.3. The name "conformal measure" was given to a related kind of measure in [27] with respect to a $\mathbb{N}$-action. It can be shown that these are precisely those measures which are conformal in the present sense and are non-singular with respect to the $\mathbb{N}$-action. Note that [27] uses a multiplicative rather than an additive cocycle, and since Denker-Urbanski use only a single transformation, their cocycle is a function of one variable rather than of two. The name "conformal" was motivated, in [27], by analogy with Patterson's study [62] of measures on the limit sets of Fuchsian groups of conformal mappings of the unit disc in the complex plane.

Conformal measures in our sense were introduced in [22], where they were simply called Gibbs measures. For links between similar notions on one- and two-sided shifts in one dimension, see [11].

Definition 2.2.2 (DLR equations for a cocycle)**.** Let $X \subseteq \mathcal{A}^G$ be a subshift, ϕ a cocycle on $\mathfrak{T}_X$, and μ a measure on X. For a Borel set $A \subseteq X$ and a finite set $\Lambda \Subset G$, the DLR equation for $x \in X$ is as follows:

$$\mu(A \,|\, \mathcal{F}_{\Lambda^c})(x) = \sum_{\eta \in \mathcal{A}^\Lambda} \left[\sum_{\zeta \in \mathcal{A}^\Lambda} \exp(\phi(\eta x_{\Lambda^c}, \zeta x_{\Lambda^c})) \mathbf{1}_X(\zeta x_{\Lambda^c}) \right]^{-1} \mathbf{1}_A(\eta x_{\Lambda^c}) \tag{2.1}$$

We say that μ is DLR with respect to ϕ if, for any Borel $A \subseteq X$ and any $\Lambda \Subset G$, (2.1) holds for μ-a.e. $x \in X$.

2.3 Equivalence of the conformal and DLR properties

For us, the main value of Lemma 2.2.1 is the following lemma, which reveals in particular that to show that a given measure is conformal (such as in Theorem 2.3.4), it is sufficient to consider only global holonomies.

Lemma 2.3.1. *Let μ be a Borel probability measure on X, let ϕ be a cocycle on $\mathfrak{T}_X$, and let Γ be a countable group generating $\mathfrak{T}_X$. Then μ is $(\phi, \mathfrak{T}_X)$-conformal if and only if, for each $\gamma \in \Gamma$, the pushforward $\mu \circ \gamma$ is absolutely continuous with respect to μ, with $\frac{d(\mu\circ\gamma)}{d\mu}(x) = \exp(\phi(x, \gamma(x)))$ for μ-a.e. $x \in X$.*

Proof. The "only if" direction is immediate from the definition of conformal measure. To confirm the "if" direction, we first check nonsingularity. Let $A \subset X$ be Borel with $\mu(A) = 0$. Then $\mathfrak{T}_X(A) = \bigcup_{\gamma\in\Gamma} \gamma(A)$, which is a countable union and thus has measure zero by the explicit expression for $\frac{d(\mu\circ\gamma)}{d\mu}$.

Now let $\psi : A \to B$ be a holonomy of $\mathfrak{T}_X$ and let $E \subseteq A$ be Borel. Let $\Gamma = (\gamma_n)_{n\in\mathbb{N}}$ be an enumeration of Γ. For each $n \in \mathbb{N}$, let $E_n = \{x \in E : \psi(x) = \gamma_n(x)\}$. To see that each E_n is Borel, define the map $\tau_n : X \to X \times X$ by $\tau_n(x) = (\psi(x), \gamma_n(x))$, which is clearly measurable in the product σ-algebra. Then $E_n = \tau_n^{-1}(D)$ where $D \subset X \times X$ is the diagonal, which, as discussed above, is also Borel in the product σ-algebra, because X is Polish.

Now let $E'_0 = E_0$, and for $n \geq 1$, let $E'_n = E_n \setminus \cup_{k=0}^{n-1} E_k$. The Borel sets E'_n partition E, so

$$\mu(\psi(E)) = \sum_{n=0}^{\infty} \mu(\gamma_n(E'_n)) = \int_E \exp(\phi(x, \psi(x)))\, d\mu(x)$$

Thus $\frac{d(\mu\circ\psi)}{d\mu}(x) = \exp(\phi(x, \psi(x))$ for μ-a.e. $x \in A$, as required. □

We will use Lemma 2.3.1 in concert with the following lemma, which reduces the question of $(\phi, \mathfrak{T}_X)$-conformality to that of conformality with respect to the finite-order subrelations.

Lemma 2.3.2. *Let μ be a measure on X and ϕ a cocycle on $\mathfrak{T}_X$. Suppose that μ is $(\phi, \mathfrak{T}_{X,N})$-conformal for each $N \geq 0$. Then, μ is $(\phi, \mathfrak{T}_X)$-conformal.*

Proof. By Lemma 2.3.1, it is enough to consider only global holonomies. Let $\psi : X \to X$ be a global holonomy of the Gibbs relation $\mathfrak{T}_X$ and let $A \subseteq X$ be a Borel set. We begin by writing A as the increasing union $A = \cup_{N=0}^{\infty} A_N$, where $A_N = \{x \in A : (x, \psi(x)) \in \mathfrak{T}_{X,N}\}$. Since $\psi|_{A_N}$ is a holonomy of $\mathfrak{T}_{X,N}$ and μ is $(\phi, \mathfrak{T}_{X,N})$-conformal, we have

$$\begin{aligned}\mu(\psi(A)) &= \lim_{N\to\infty} \mu(\psi(A_N)) \\ &= \lim_{N\to\infty} \int_{A_N} \exp(\phi(x, \psi(x)))\, d\mu(x) \\ &= \int_A \exp(\phi(x, \psi(x)))\, d\mu(x),\end{aligned}$$

by monotone convergence. Thus, μ is indeed $(\phi, \mathfrak{T}_X)$-conformal by Lemma 2.3.1. □

To echo the comment above on hyperfiniteness, we remark here that both of these results apply, with the same proofs, to any hyperfinite Borel equivalence relation on any Polish space. The following lemma, by contrast, seems to rely more specifically on the structure of X as a subshift.

Lemma 2.3.3. *Let $X \subseteq \mathcal{A}^G$ be a subshift, let ϕ be a cocycle on X, and let μ be a DLR measure on X with respect to ϕ. Let $N \geq 1$. Then μ is $(\phi, \mathfrak{T}_{X,N})$-conformal.*

Proof. It is enough to show that $\mu(\psi([\omega])) = \int_{[\omega]} \exp(\phi(x, \psi(x)))\, d\mu(x)$ for any cylinder $[\omega]$ and (by Lemma 2.3.1) any global holonomy ψ of $\mathfrak{T}_{X,N}$. Fix a holonomy $\psi : X \to X$ of $\mathfrak{T}_{X,N}$. Since the equivalence classes of $\mathfrak{T}_{X,N}$ are finite, and in fact have bounded cardinality, there exists some $r \geq 0$ such that $\psi^r(x) = x$, for all $x \in X$. Let $m \geq N$ and fix $\omega \in \mathcal{A}^{\Lambda_m}$. We now partition X according to the orbits of points under ψ, in such a way that $[\omega]$ is partitioned into sets that are easy to control. Specifically, for each $\overline{\eta} = (\eta_0, \ldots, \eta_{r-1}) \in (\mathcal{A}^{\Lambda_m})^r$, let

$$T_{\overline{\eta}} = \{x \in X : \psi^j(x)_{\Lambda_m} = \eta_j, 0 \leq j \leq r-1\}$$

Note that $T_{\overline{\eta}}$ can be empty. We have $[\omega] = \sqcup_{\overline{\eta}:\eta_0=\omega} T_{\overline{\eta}}$, and $\psi(T_{\overline{\eta}}) = T_{\overline{\sigma\eta}}$, where $\overline{\sigma\eta} = (\eta_1, \ldots, \eta_{r-1}, \eta_0)$ is a cyclic permutation of $\overline{\eta}$. It is enough to show that, for all $\overline{\eta} \in (\mathcal{A}^{\Lambda_m})^r$, we have

$$\mu(\psi(T_{\overline{\eta}})) = \int_{T_{\overline{\eta}}} \exp\left(\phi(x, \psi(x))\right) d\mu(x).$$

By the equality $\psi(T_{\overline{\eta}}) = T_{\overline{\sigma\eta}}$, we have

$$\mu(\psi(T_{\overline{\eta}})) = \int_X \mu(T_{\overline{\sigma\eta}} \,|\, \mathcal{F}_{\Lambda_m^c})\, d\mu(x)$$

For any $x \in X$, we know that

$$\mathbf{1}_{T_{\overline{\sigma\eta}}}(\eta_1 x_{\Lambda_m^c}) = \mathbf{1}_{T_{\overline{\eta}}}(\eta_0 x_{\Lambda_m^c})$$

By this identity, as well as the DLR hypothesis and the defining property of a cocycle, we have the following manipulations:

$$\begin{aligned}\mu(T_{\overline{\sigma\eta}} \,|\, \mathcal{F}_{\Lambda_m^c})(x) &= \left[\sum_{\zeta \in \mathcal{A}^{\Lambda_m}} \exp(\phi(\eta_1 x_{\Lambda_m^c}, \zeta x_{\Lambda_m^c}))\mathbf{1}_X(\zeta x_{\Lambda_m^c})\right]^{-1} \mathbf{1}_{T_{\overline{\sigma\eta}}}(\eta_1 x_{\Lambda_m^c}) \\ &= \left[\sum_{\zeta \in \mathcal{A}^{\Lambda_m}} \exp(\phi(\eta_0 x_{\Lambda_m^c}, \zeta x_{\Lambda_m^c}))\mathbf{1}_X(\zeta x_{\Lambda_m^c})\right]^{-1} \\ &\qquad \times \mathbf{1}_{T_{\overline{\eta}}}(\eta_0 x_{\Lambda_m^c})\, \exp(\phi(\eta_0 x_{\Lambda_m^c}, \eta_1 x_{\Lambda_m^c})) \\ &= \mu(T_{\overline{\eta}} \,|\, \mathcal{F}_{\Lambda_m^c})(x)\, \exp(\phi(\eta_0 x_{\Lambda_m^c}, \eta_1 x_{\Lambda_m^c}))\end{aligned}$$

Integrating this equation yields the result. □

We have therefore done all the work required to prove the following:

Theorem 2.3.4. *Let $X \subseteq \mathcal{A}^G$ be a subshift, ϕ a cocycle on X, and μ a DLR measure on X with respect to ϕ. Then μ is $(\phi, \mathfrak{T}_X)$-conformal.*

Proof. By Lemma 2.3.3, μ is $(\phi, \mathfrak{T}_{X,N})$-conformal for each N. The result is then immediate from Lemma 2.3.2. $\square$

Theorem 2.3.4 was proven by Kimura ([44], Theorem 5.30) in the special case that $G = \mathbb{Z}^d$, X is a shift of finite type, and the cocycle ϕ is induced by a potential, in the manner that we discuss in Proposition 2.5.3 below. Furthermore, Kimura proved the following converse ([44], Corollary 5.33), again in the case of $G = \mathbb{Z}^d$ and ϕ induced by a potential, but with no finite type assumption on X.

Theorem 2.3.5. *Let $X \subseteq \mathcal{A}^G$ be a subshift, ϕ a cocycle on X, and μ a $(\phi, \mathfrak{T}_X)$-conformal measure on X. Then μ is DLR with respect to ϕ.*

The proof of Theorem 2.3.5 is a straightforward adaptation of the methods that Kimura used for the case that he treated. The rough idea is to show that two cylinder sets have conditional measures with the appropriate ratio by considering the holonomy that exchanges them, as in the proof of Lemma 2.2.1 above, then applying the conformal hypothesis. The main difference required to adapt the proof is that the version stated here concerns the DLR equations for an arbitrary measurable cocycle, not necessarily one induced by a potential.

2.4 Interactions

In this section, we show that, when a cocycle is induced by an interaction, the DLR equations for the cocycle reduce to those for the interaction.

Definition 2.4.1 (interaction)**.** An interaction is a family $\Phi = (\Phi_\Lambda)_{\Lambda \Subset G}$ of functions $\Phi_\Lambda : X \to \mathbb{R}$ such that for each $\Lambda \Subset G$, Φ_Λ is $\mathcal{F}_\Lambda$-measurable, and for all $\Lambda \Subset G$, $x \in X$, the *Hamiltonian series*

$$H_\Lambda^\Phi(x) = \sum_{\substack{\Delta \Subset G \\ \Delta \cap \Lambda \neq \emptyset}} \Phi_\Delta(x)$$

converges in the sense that there exists a real number $H_\Lambda^\Phi(x)$ and, for every $\varepsilon > 0$, there exists some $F \Subset G$ such that, for all $F' \supseteq F$,

$$\left| H_\Lambda^\Phi(x) - \sum_{\substack{\Delta \subseteq F' \\ \Delta \cap \Lambda \neq \emptyset}} \Phi_\Delta(x) \right| < \varepsilon$$

Proposition 2.4.1. *Let Φ be an interaction. For each $(x, y) \in \mathfrak{T}_X$, the series*

$$\sum_{\Lambda \Subset G} [\Phi_\Lambda(x) - \Phi_\Lambda(y)]$$

converges in the same sense as the Hamiltonian series. Moreover, the function $\phi_\Phi : \mathfrak{T}_X \to \mathbb{R}$ defined by

$$\phi_\Phi(x, y) = \sum_{\Lambda \Subset G} [\Phi_\Lambda(x) - \Phi_\Lambda(y)]$$

is a cocycle on $\mathfrak{T}_X$.

Proof. Let $(x, y) \in \mathfrak{T}_X$ be such that $x_{\Delta^c} = y_{\Delta^c}$. We claim that

$$\sum_{\Lambda \Subset G} [\Phi_\Lambda(x) - \Phi_\Lambda(y)] = H_\Delta^\Phi(x) - H_\Delta^\Phi(y)$$

with the equality understood in the sense of convergence discussed in the statement of the proposition. Indeed, choose $\varepsilon > 0$. By the definition of an interaction, there exists some $F \Subset G$ sufficiently large that whenever $F \subseteq F' \Subset G$, we have (noting that $\Phi_E(x) - \Phi_E(y) = 0$ when $E \cap \Delta = \emptyset$),

$$\begin{aligned}
&\left| [H_\Delta^\Phi(x) - H_\Delta^\Phi(y)] - \sum_{E \subseteq F'} [\Phi_E(x) - \Phi_E(y)] \right| \\
&\leq \left| H_\Delta^\Phi(x) - \sum_{\substack{E \subseteq F' \\ E \cap \Delta \neq \emptyset}} \Phi_E(x) \right| + \left| H_\Delta^\Phi(y) - \sum_{\substack{E \subseteq F' \\ E \cap \Delta \neq \emptyset}} \Phi_E(y) \right| \\
&< \varepsilon
\end{aligned}$$

This establishes that the series converges, in the sense claimed, to a real number $\phi_\Phi(x, y) = H_\Delta^\Phi(x) - H_\Delta^\Phi(y)$. Moreover, this energy difference expression makes it obvious that ϕ_Φ is a cocycle, concluding the proof. □

We now observe that the DLR equations for the cocycle ϕ_Φ, in the sense of Definition 2.1, are equivalent to the classical DLR equations for the interaction Φ. Indeed, if μ is a DLR measure with respect to ϕ_Φ, then for any $\Lambda \Subset G$, any Borel $A \subseteq X$, and μ-a.e. $x \in X$, we have

$$\begin{aligned}
\mu(A \,|\, \mathcal{F}_{\Lambda^c})(x) &= \sum_{\zeta \in \mathcal{A}^\Lambda} \left[\sum_{\eta \in \mathcal{A}^\Lambda} \exp(\phi_\Phi(\zeta x_{\Lambda^c}, \eta x_{\Lambda^c})) \mathbf{1}_X(\zeta x_{\Lambda^c}) \right]^{-1} \mathbf{1}_A(\zeta x_{\Lambda^c}) \\
&= \sum_{\zeta \in \mathcal{A}^\Lambda} \left[\sum_{\eta \in \mathcal{A}^\Lambda} \exp\left(H_\Lambda^\Phi(\zeta x_{\Lambda^c}) - H_\Lambda^\Phi(\eta x_{\Lambda^c}) \right) \mathbf{1}_X(\eta x_{\Lambda^c}) \right]^{-1} \mathbf{1}_A(\zeta x_{\Lambda^c}) \\
&= \frac{1}{Z_\Lambda^\Phi(x)} \sum_{\zeta \in \mathcal{A}^\Lambda} \exp\left(-H_\Lambda^\Phi(\zeta x_{\Lambda^c}) \right) \mathbf{1}_A(\zeta x_{\Lambda^c})
\end{aligned}$$

where

$$Z_\Lambda^\Phi(x) = \sum_{\eta \in \mathcal{A}^\Lambda} \exp\left(-H_\Lambda^\Phi(\eta x_{\Lambda^c}) \right) \mathbf{1}_X(\eta x_{\Lambda^c})$$

By Theorem 2.3.4, if μ satisfies these (classical) DLR equations for Φ, then μ is $(\phi_\Phi, \mathfrak{T}_X)$-conformal.

2.5 Potentials

In this section and the next, we restrict to finitely generated groups G satisfying a certain growth condition. We need this condition in order to construct a cocycle from a potential in a way that is compatible with interactions, in a sense to be made precise in §2.6. The condition is as follows. It concerns the *spherical growth function* $|B_k \setminus B_{k-1}|$, which is a basic quantity studied in geometric group theory, discussed for instance in ([34], §VI.A).

Definition 2.5.1 (bounded sphere ratios)**.** Let G be a finitely generated group. With respect to a finite generating set $S \Subset G$, we can consider the open balls $B_k = \{g \in G \,:\, d(g, e) < n\}$ of radius

k centered at the identity in the Cayley graph of G with respect to S. We say that a group G has *bounded sphere ratios* if there exists a finite generating set S such that

$$\sup_{m\geq 1}\frac{|B_{m+1}\setminus B_m|}{|B_m\setminus B_{m-1}|}<+\infty.$$

In this section and the next, when we refer to balls in a group G with bounded sphere ratios, we always mean balls with respect to a generating set that witnesses the bounded sphere ratios. Note also that if G has bounded sphere ratios, then (for some generating set S) we have

$$\sup_{m\geq 1}\frac{|B_{m+n}\setminus B_{m+n-1}|}{|B_m\setminus B_{m-1}|}<+\infty.$$

for any n.

Remark 2.5.1. A finitely generated group G has *polynomial growth* if $|B_n|\leq cn^d$ for some $c>0, d\in\mathbb{N}$ and all n; *exponential growth* if $|B_n|\geq c\alpha^n$ for some $\alpha>1, c>0$ and all n; and *intermediate growth* otherwise. Here we outline certain types of polynomial and exponential growth known to imply bounded sphere ratios.

In the polynomial case, recall that a group has polynomial growth if and only if it is virtually nilpotent, i.e. has a finite-index nilpotent group [32]. It is conjectured ([20], Conjecture 10) that for any nilpotent group, we have $|B_n|=cn^d+O(n^{d-1})$, where $c>0$ is a constant depending only on the group, with the coefficients of the lower-order terms depending on the generating set. This would imply ([20], Corollary 11) positive constant upper and lower bounds on the ratio $|B_k\setminus B_{k-1}|/k^{d-1}$, and thus that the group has bounded sphere ratios. What is known is more restricted. Associated to any nilpotent group is its *nilpotency class*, a number measuring how far the group is from being abelian (abelian groups, like $\mathbb{Z}^d$, have class 1). By a result of Stoll [72], the conjectured asymptotics for $|B_n|$ hold at least for groups of nilpotency class at most 2.

In the exponential case, we say that a group (with a given generating set) has *exact exponential growth* if there exist $\alpha>1$ and $0<c<C<c\alpha$ with $c\leq|B_n|/\alpha^n\leq C$ for all $n\geq 1$. (This is not a standard definition.) This condition is satisfied, for example, by a free group with the usual generating set. To see that exact exponential growth implies bounded sphere ratios, note that

$$\frac{|B_{k+1}\setminus B_k|}{|B_k\setminus B_{k-1}|}\leq\frac{C\alpha^{k+1}-c\alpha^k}{c\alpha^k-C\alpha^{k-1}}=\left(\frac{C\alpha-c}{c\alpha-C}\right)\alpha<\infty$$

We now turn our attention to potentials. For a function $f:X\to\mathbb{R}$ and $k\geq 1$, define the kth variation of f as

$$v_k(f):=\sup\left\{|f(y)-f(x)|\,\middle|\,x,y\in X,\,x_{B_k}=y_{B_k}\right\}.$$

We separately define $v_0(f)=\|f\|_\infty$. It is also convenient to define $B_0=\emptyset$. We define the *shell norm* $\|\cdot\|_{\mathrm{ShVar}}$ by

$$\|f\|_{\mathrm{ShVar}}:=\sum_{k=0}^{\infty}|B_{k+1}\setminus B_k|v_k(f).$$

We define the space $\mathrm{ShReg}(X)$ as the space of *shell-regular potentials*, i.e., functions $f:X\to\mathbb{R}$ with $\|f\|_{\mathrm{ShVar}}<\infty$. It is elementary to show that shell-regularity implies continuity, and that $\mathrm{ShReg}(X)$, with the shell norm, is a Banach space. Note that this space depends, in general, on the generating set chosen.

Remark 2.5.2. In earlier work on subshifts over $\mathbb{Z}^d$ [56], the relevant space of potentials is known as $\mathrm{SV}_d(X)$, the space of potentials with d-summable variation, defined by the norm $\|f\|_{SV_d}=\sum_{k=1}^{\infty}k^{d-1}v_{k-1}(f)$. This space is also known as $\mathrm{Reg}_{d-1}(X)$ [59]. With $B_n=\mathbb{Z}^d\cap[-n,n]^d$, we have $|B_{k+1}\setminus B_k|=2^d d(1+o(1))k^{d-1}$. Thus, on $\mathbb{Z}^d$, we have $\mathrm{ShReg}(X)=\mathrm{SV}_d(X)$, with the identity a continuous linear map.

Proposition 2.5.3. *Let G be a group with bounded sphere ratios and X a subshift over G. For any $f \in \mathrm{ShReg}(X)$ and any $(x,y) \in \mathfrak{T}_X$, the series*

$$\sum_{g\in G}[f(g\cdot y) - f(g\cdot x)]$$

converges absolutely and defines a cocycle ϕ_f on $\mathfrak{T}_X$.

Proof. Fix $(x,y) \in \mathfrak{T}_X$, and let $n \geq 1$ be such that $x_{B_n^c} = y_{B_n^c}$. If $g \in G$ and $m \geq 1$ are such that $B_{m-1} \subseteq g^{-1}B_n^c$, then $(g\cdot x)|_{B_{m-1}} = (g\cdot y)|_{B_{m-1}}$ so $|f(g\cdot y) - f(g\cdot x)| \leq v_{m-1}(f)$. For $m \geq 1$ and $g \in B_k \setminus B_{k-1}$, the triangle inequality guarantees that $gB_{m-1} \subseteq B_n^c$ if $k - n \geq m$. Since the shells $B_{k+1} \setminus B_k$ partition G, we have

$$\begin{aligned}\sum_{g\in G}|f(g\cdot y) - f(g\cdot x)| &\leq 2|B_n|\|f\|_\infty + \sum_{k=n+1}^{\infty}|B_k \setminus B_{k-1}|v_{k-n-1}(f)\\ &\leq 2|B_n|\|f\|_\infty + \left(\sup_{k\geq 1}\frac{|B_{k+n}\setminus B_{k+n-1}|}{|B_k\setminus B_{k-1}|}\right)\|f\|_{\mathrm{ShVar}}\end{aligned}$$

so indeed the cocycle is well-defined by an absolutely convergent series. □

Just as in the case of an interaction, this expression for the cocycle ϕ_f allows us to rewrite the DLR equations in a more classical form. Let $f \in \mathrm{ShReg}(X)$. It follows from a simple manipulation that for any $(x,y) \in \mathfrak{T}_X$, we have

$$\exp(\phi_f(x,y)) = \lim_{m\to+\infty}\exp\left(\sum_{g\in B_m}[f(g\cdot y) - f(g\cdot x)]\right) = \lim_{m\to+\infty}\frac{\exp f_m(y)}{\exp f_m(x)}.$$

where $f_m(z) = \sum_{g\in B_m} f(g\cdot z)$. Now, let $A \subseteq X$ be a Borel set. If μ is a DLR measure with respect to ϕ_f, then for μ-a.e. $x \in X$, we have

$$\begin{aligned}\mu(A\,|\,\mathcal{F}_{\Lambda^c})(x) &= \sum_{\eta\in\mathcal{A}^\Lambda}\left[\sum_{\zeta\in\mathcal{A}^\Lambda}\exp(\phi_f(\eta x_{\Lambda^c},\zeta x_{\Lambda^c}))\mathbf{1}_X(\zeta x_{\Lambda^c})\right]^{-1}\mathbf{1}_A(\eta x_{\Lambda^c})\\ &= \sum_{\eta\in\mathcal{A}^\Lambda}\left[\sum_{\zeta\in\mathcal{A}^\Lambda}\lim_{m\to+\infty}\frac{\exp f_m(\zeta x_{\Lambda^c})}{\exp f_m(\eta x_{\Lambda^c})}\mathbf{1}_X(\zeta x_{\Lambda^c})\right]^{-1}\mathbf{1}_A(\eta x_{\Lambda^c})\\ &= \lim_{m\to\infty}\frac{\sum_{\eta\in\mathcal{A}^\Lambda}\exp\left(f_m(\eta x_{\Lambda^c})\right)\mathbf{1}_A(\eta x_{\Lambda^c})}{\sum_{\zeta\in\mathcal{A}^\Lambda}\exp\left(f_m(\zeta x_{\Lambda^c})\right)\mathbf{1}_X(\zeta x_{\Lambda^c})}\end{aligned}$$

These are the DLR equations as found in Kimura [44]. Applying Theorem 2.3.4 therefore shows that any DLR measure with respect to a potential $f \in \mathrm{ShReg}(X)$ is necessarily $(\phi_f, \mathfrak{T}_X)$-conformal, providing the full converse for Kimura's result described in the introduction.

2.6 Potentials induced by interactions, and vice versa

We have seen that the DLR property implies the conformal property for an arbitrary cocycle on the Gibbs relation, with Gibbs measures for interactions and for potentials as two special cases. These cases are not independent. In this section, we adapt the methods and results of Muir [59] and Ruelle [66] to construct potentials from interactions and vice versa. In this section, all interactions are translation-invariant, i.e., for any $\Lambda \Subset G$ and any $x \in X$, we require that $\Phi_{g\Lambda}(g\cdot x) = \Phi_\Lambda(x)$. We recall a classical space of particularly well-behaved interactions:

Definition 2.6.1. For an interaction Φ, let

$$\|\Phi\|_B = \sum_{\substack{\Lambda \Subset G \\ e \in \Lambda}} \|\Phi_\Lambda\|_\infty$$

We define $\mathcal{B}$ as the space of *absolutely summable* Φ, i.e., those for which $\|\Phi\|_B < \infty$.

It is routine to check that $(\mathcal{B}, \|\cdot\|_B)$ is a Banach space. Moreover, for$\Phi \in \mathcal{B}$, we in fact have absolute convergence of the series defining the cocycle ϕ_Φ, since for any $(x, y) \in \mathfrak{T}_X$ with $x_{\Delta^c} = y_{\Delta^c}$ for some $\Delta \Subset G$, we have

$$\begin{aligned}
\sum_{\Lambda \Subset G} |\Phi_\Lambda(x) - \Phi_\Lambda(y)| &\leq 2 \sum_{\substack{\Lambda \Subset G \\ \Lambda \cap \Delta \neq \emptyset}} \|\Phi_\Lambda\|_\infty \\
&\leq 2|\Delta| \sum_{\substack{\Lambda \Subset G \\ e \in \Lambda}} \|\Phi_\Lambda\|_\infty \\
&= 2|\Delta| \|\Phi\|_B < \infty
\end{aligned}$$

We introduce a family of linear maps that convert interactions into potentials.

Definition 2.6.2 (translate-weighting maps)**.** Let $(a_\Lambda)_{\Lambda \Subset G,\, e \in \Lambda}$ be a collection of nonnegative real coefficients such that, for each $\Lambda \Subset G$ with $e \in \Lambda$, we have $\sum_{g \in \Lambda} a_{g^{-1}\Lambda} = 1$. Then, for an interaction Φ, define the potential A_Φ via

$$A_\Phi(x) = -\sum_{\substack{\Lambda \Subset G \\ e \in \Lambda}} a_\Lambda \Phi_\Lambda(x)$$

The map $\Phi \mapsto A_\Phi$ is clearly linear. We refer to this map as the translate-weighting map determined by the weights $(a_\Lambda)_{\Lambda \Subset G, e \in \Lambda}$.

Remark 2.6.1. Two important examples are the following.

- The *uniform map*, where $a_\Lambda \in \{0, \frac{1}{|\Lambda|}\}$ for every nonempty $\Lambda \Subset G$. Muir uses the letter A to denote this specific operator, i.e., $A(\Phi) = A_\Phi$.
- The class of *dictator maps*, where $a_\Lambda \in \{0, 1\}$ for every $\Lambda \Subset G$. For instance, on $\mathbb{Z}^d$, Ruelle studies the operator for which $a_\Lambda = 1$ if and only if 0 is the middle element, or more precisely the $\lfloor(|\Lambda| + 1)/2\rfloor$-th element, of Λ in lexicographic order. In [59], Muir refers to this operator as $\hat{A}$.

In Fact 7.8 in [59], it is claimed that $A_\Phi \in \mathrm{ShReg}(X)$ for every translate-weighting map and every $\Phi \in \mathcal{B}$. This claim is incorrect, as we demonstrate with an example below. However, the argument presented for this claim is correct in the case of what Muir calls "cubic-type" interactions. Here we reproduce a version of this proof for a larger class of interactions.

Definition 2.6.3. An interaction Φ is *full-dimensional* if there exists some $C > 0$ such that, for all $\Lambda \Subset G$ with $e \in \Lambda$ and $\Phi_\Lambda \not\equiv 0$, we have the bound

$$\sup\{|B_n| : n \in \mathbb{N}, \Lambda \cap B_{n-1}^c \neq \emptyset\} \leq C|\Lambda|$$

Proposition 2.6.2. *Let G be a group with bounded sphere ratios and let X be a subshift over G. If $\Phi \in \mathcal{B}$ is full-dimensional, then $A_\Phi \in \mathrm{ShReg}(X)$, where A_Φ is the image of Φ under an arbitrary translate-weighting map.*

Proof. We first estimate $v_{k-1}(A_\Phi)$:

$$v_{k-1}(A_\Phi) = \sup\left\{\left|\sum_{\substack{\Lambda\Subset G\\ e\in\Lambda}} a_\Lambda[\Phi_\Lambda(x)-\Phi_\Lambda(y)]\right| : x,y\in X,\, x_{B_{k-1}}=y_{B_{k-1}}\right\}$$
$$\leq 2\sum_{\substack{\Lambda\Subset G\\ e\in\Lambda\\ \Lambda\cap B_{k-1}^c\neq\emptyset}} a_\Lambda\|\Phi_\Lambda\|_\infty$$

We can now estimate the shell norm by an exchange of summations:

$$\|A_\Phi\|_{\mathrm{ShVar}} \leq 2\sum_{k=0}^{\infty}|B_{k+1}\setminus B_k| \sum_{\substack{\Lambda\Subset G\\ e\in\Lambda\\ \Lambda\cap B_k^c\neq\emptyset}} a_\Lambda\|\Phi_\Lambda\|_\infty$$
$$= 2\sum_{\substack{\Lambda\Subset G\\ e\in\Lambda}} a_\Lambda\|\Phi_\Lambda\|_\infty \sum_{\substack{k\geq 0\\ \Lambda\cap B_k^c\neq\emptyset}} |B_{k+1}\setminus B_k|$$

Observe that

$$\sum_{\substack{k\geq 0\\ \Lambda\cap B_{k-1}^c\neq\emptyset}} |B_{k+1}\setminus B_k| = \sup\{|B_n| : n\in\mathbb{N},\, \Lambda\cap B_n^c\neq\emptyset\} \leq C|\Lambda|$$

so in fact

$$\|A_\Phi\|_{\mathrm{ShVar}} \leq 2C\sum_{\substack{\Lambda\Subset G\\ e\in\Lambda}} a_\Lambda|\Lambda|\|\Phi_\Lambda\|_\infty$$

We need to rearrange this sum. For a given $\Lambda\Subset G$, consider the set of translates of Λ containing the identity, denoted $T(\Lambda)=\{g^{-1}\Lambda,\, g\in\Lambda\}$. For instance, in $\mathbb{Z}$, if $\Lambda=\{0,1\}$, then $T(\Lambda)=\{\{-1,0\},\{0,1\}\}$. Let $\mathcal{T}$ denote the set of such sets of translates, i.e., $\mathcal{T}=\{T(\Lambda) : \Lambda\Subset G,\, e\in\Lambda\}$. Note that $\mathcal{T}$ is a partition of the set $\{\Lambda\Subset G,\, e\in\Lambda\}$. Observe furthermore that $|T|=|\Lambda|$ for any $\Lambda\in T$.

For any given $T\in\mathcal{T}$, the value $|\Lambda|\|\Phi_\Lambda\|_\infty$ is the same for any $\Lambda\in T$, i.e., any Λ such that $T=T(\Lambda)$. so we denote it by c_T. We can then express the bound on $\|A_\Phi\|_{\mathrm{ShVar}}$ by summing over $T\in\mathcal{T}$, as follows:

$$\begin{aligned}
\sum_{\substack{\Lambda\Subset G\\ e\in\Lambda}} a_\Lambda|\Lambda|\|\Phi_\Lambda\|_\infty &= \sum_{T\in\mathcal{T}}\sum_{\Lambda\in T} a_\Lambda c_T\\
&= \sum_{T\in\mathcal{T}} c_T\sum_{\Lambda\in T} a_\Lambda\\
&= \sum_{T\in\mathcal{T}} c_T\\
&= \sum_{T\in\mathcal{T}} |\Lambda|\|\Phi_\Lambda\|_\infty\\
&= \sum_{T}\sum_{\Lambda\in T} \|\Phi_\Lambda\|_\infty\\
&= \|\Phi\|_B
\end{aligned}$$

Thus $\|A_\Phi\|_{\mathrm{ShVar}} \leq 2C\|\Phi\|_B < \infty$. □

In the next example, $\Phi \in \mathcal{B}$ is not full-dimensional, and A_Φ is not shell-regular.

Example 1. Let $X = \{0,1\}^{\mathbb{Z}}$, with $B_k = (-k,k) \cap \mathbb{Z}$. Define $\Phi = (\Phi_\Lambda)_{\Lambda \Subset \mathbb{Z}}$ as follows: for any $i, j \in \mathbb{Z}$, $\Phi_{\{i,j\}}(x) = \frac{1}{(j-i)^2}$ if $x_i = x_j = 1$ and 0 otherwise; and $\Phi_\Lambda \equiv 0$ for all other $\Lambda \Subset G$. Clearly Φ is translation-invariant. We claim that $\Phi \in \mathcal{B}$ but $A_\Phi \notin \mathrm{ShReg}(X)$, where A_Φ is the image of Φ under the dictator map that ignores $\Lambda \Subset \mathbb{Z}$ unless $0 = \inf \Lambda$. Indeed, $\|\Phi\|_{\mathcal{B}} = 2\sum_{j=1}^{\infty} \frac{1}{j^2} < \infty$, but

$$v_k(A_\Phi) = \sum_{l=k}^{\infty} \frac{1}{l^2} \geq \frac{1}{k}$$

which implies that

$$\|A_\Phi\|_{\mathrm{ShVar}} \geq 2\sum_{k=1}^{+\infty} \frac{1}{k} = +\infty$$

We now show that for full-dimensional $\Phi \in \mathcal{B}$, the images A_Φ and A'_Φ of Φ under any two translate-weighting maps have the same Gibbs and equilibrium measures.

Proposition 2.6.3. *Let G be a group with bounded sphere ratios, let X be a subshift on G, and let Φ be an absolutely summable, full-dimensional interaction on X. Then Φ and A_Φ induce the same cocycle, i.e., $\phi_{A_\Phi} = \phi_\Phi$, where A_Φ is the image of Φ under an arbitrary translate-weighting map.*

Proof. Suppose that $(x,y) \in \mathfrak{T}_X$ with $x_{\Delta^c} = y_{\Delta^c}$. Observe that

$$\phi_\Phi(x,y) = \sum_{\substack{\Lambda \Subset G \\ \Lambda \cap \Delta \neq \emptyset}} [\Phi_\Lambda(x) - \Phi_\Lambda(y)]$$

Consider a translate-weighting map with weights a_Λ. To compute ϕ_{A_Φ}, we first obtain a convenient expression for $A_\Phi(g \cdot x) - A_\Phi(g \cdot y)$:

$$\begin{aligned} A_\Phi(g \cdot x) - A_\Phi(g \cdot y) &= -\sum_{\substack{\Lambda \Subset G \\ e \in \Lambda \\ \Lambda \cap g\Delta \neq \emptyset}} a_\Lambda \Phi_{g^{-1}\Lambda}(x) + \sum_{\substack{\Lambda \Subset G \\ e \in \Lambda \\ \Lambda \cap g\Delta \neq \emptyset}} a_\Lambda \Phi_{g^{-1}\Lambda}(y) \\ &= -\sum_{\substack{\Lambda' \Subset G \\ g \in \Lambda' \\ \Lambda' \cap \Delta \neq \emptyset}} a_{g^{-1}\Lambda'}[\Phi_{\Lambda'}(x) - \Phi_{\Lambda'}(y)] \end{aligned}$$

We then compute:

$$\begin{aligned} \phi_{A_\Phi}(x,y) &= \sum_{g \in G} [A_\Phi(g \cdot x) - A_\Phi(g \cdot y)] \\ &= \sum_{g \in G} \sum_{\substack{\Lambda \Subset G \\ g \in \Lambda \\ \Lambda \cap \Delta \neq \emptyset}} a_{g^{-1}\Lambda}[\Phi_\Lambda(x) - \Phi_\Lambda(y)] \\ &= \sum_{\substack{\Lambda \Subset G \\ \Lambda \cap \Delta \neq \emptyset}} \left(\sum_{g \in \Lambda} a_{g^{-1}\Lambda} \right) [\Phi_\Lambda(x) - \Phi_\Lambda(y)] \\ &= \phi_\Phi(x,y) \end{aligned}$$

The interchange of summations is justified by the absolute convergence of the series defining the cocycles ϕ_{A_Φ} and ϕ_Φ, implied by the regularity of Φ and A_Φ. $\square$

Proposition 2.6.3 is similar to Theorem 5.42 in [44], which is stated for Ruelle's operator A, using specifications rather than cocycles.

Proposition 2.6.4. *Let G be a group with bounded sphere ratios and let X be a subshift on G. Let μ be a G-invariant measure on X, let $\Phi \in \mathcal{B}$ be full-dimensional. Let A_Φ be the image of Φ under a translate-weighting map with weights $(a_\Lambda)_{\Lambda \Subset G,\, e \in \Lambda}$. Then the integral $\int_X A_\Phi \, d\mu$ depends only on Φ and μ, and not on the weights a_Λ.*

Proof. As in the proof of Proposition 2.6.2, for each finite $\Lambda \Subset G$ with $e \in \Lambda$, let $T(\Lambda) = \{g^{-1}\Lambda \mid g \in \Lambda\}$. For any given T, the quantity $\int_X \Phi_\Lambda \, d\mu$ is constant as Λ ranges over T, so we denote it by b_T. We now compute:

$$\begin{aligned}\int_X A_\Phi \, d\mu &= -\int_X \sum_{T \in \mathcal{T}} \sum_{\Lambda \in T} a_\Lambda \Phi_\Lambda \, d\mu \\ &= -\sum_{T \in \mathcal{T}} b_T \sum_{\Lambda \in T} a_\Lambda \\ &= -\sum_{\substack{\Lambda \Subset G \\ e \in \Lambda}} \frac{1}{|\Lambda|} \int_X \Phi_\Lambda \, d\mu\end{aligned}$$

which does not depend on the weights a_Λ, and in addition clearly expresses the integral $\int_X A_\Phi \, d\mu$ as the average energy at the identity due to the interaction Φ.

To justify exchanging the integral and the sum above, let $|\Phi|$ be the interaction given by $|\Phi|_\Lambda = |\Phi_\Lambda|$. Then $|\Phi|$ is still full-dimensional, with $\||\Phi|\|_B = \|\Phi\|_B$, so

$$-\sum_{T \in \mathcal{T}} \sum_{\Lambda \in T} a_\Lambda |\Phi_\Lambda| = A_{|\Phi|} \in \mathrm{ShReg}(X)$$

by Proposition 2.6.2. Thus the sum converges absolutely to a continuous function. $\square$

Finally, we introduce a smaller Banach space $\mathrm{VolReg}(X)$ of *volume-regular functions*, defined analogously to $\mathrm{ShReg}(X)$ by a volume norm rather than a shell norm. That is, $\mathrm{VolReg}(X) = \{f : X \to \mathbb{R} : \|f\|_{\mathrm{VolVar}} < \infty\}$ where we define

$$\|f\|_{\mathrm{VolVar}} := \sum_{k=0}^{\infty} |B_k| v_k(f)$$

Volume-regularity clearly implies shell-regularity. The following result of Muir ([59], proof of Fact 7.6) is stated for $\mathbb{Z}^d$, with the name $\mathrm{Reg}_d(X)$ for $\mathrm{VolReg}(X)$, but is valid, with the same proof, on any finitely generated group.

Theorem 2.6.5. *Let G be a finitely generated group and let $f \in \mathrm{VolReg}(X)$ be a volume-regular potential. Then there exists an absolutely summable $\Phi \in \mathcal{B}$ with $A_\Phi = f$ where A_Φ is the image of Φ under some dictator map.*

In particular, any Gibbs measure for $f \in \mathrm{VolReg}(X)$ is also a Gibbs measure for any potential $\Phi \in \mathcal{B}$ with $A_\Phi = f$, and vice versa.

Chapter 3

A Dobrushin-Lanford-Ruelle theorem for irreducible sofic shifts

3.1 Introduction

As advertised in the introduction to the dissertation, this chapter proves the equivalence of Gibbs and equilibrium measures for irreducible sofic shifts in one dimension, extending the classical Dobrushin and Lanford-Ruelle theorems, using a preservation of Gibbsianness approach. We repeat some definitions from the previous chapter because we are now restricting to the group $\mathbb{Z}$ rather than more general countable groups, and in order to be clear about the present notation.

3.2 Definitions, notations, and conventions

3.2.1 Symbolic dynamics

Let $\mathcal{A}$ be a finite set with the discrete topology, to be thought of as an alphabet, and $\mathcal{A}^{\mathbb{Z}}$ be the *full shift* with the product topology, with respect to which $\mathcal{A}^{\mathbb{Z}}$ is compact and metrizable. The group $\mathbb{Z}$ acts naturally on $\mathcal{A}^{\mathbb{Z}}$ by the shift action σ, given by $(\sigma^n x)_0 = x_n$. A *shift space* is any closed, σ-invariant subset $X \subseteq \mathcal{A}^{\mathbb{Z}}$.

For each $n \geq 1$, we write $\mathcal{B}_n(X)$ to denote the set of words of length n in the language $\mathcal{B}(X)$ of X—that is, the set of patterns $w \in \mathcal{A}^n$ such that $x_{[0,n-1]} = w$ for some $x \in X$. For $w \in \mathcal{A}^n$ we denote by $[w]_i$ the set of $x \in X$ with $x_{[i,i+n-1]} = w$.

We will make extensive use of continuous, shift-equivariant factor codes $\pi : X \to Y$ between shift spaces X and Y. By the Curtis-Hedlund-Lyndon theorem, any such map π is a sliding block code, induced by a map $\Pi : \mathcal{B}_m(X) \to \mathcal{B}_1(Y)$ for some $m \geq 1$. Up to a conjugacy of X, we can in fact assume that Π maps symbols to symbols, i.e., $m = 1$ ([49], Proposition 1.5.12). When $\pi : X \to Y$ is surjective, it is known as a *factor* code, and Y is a *factor* of X.

Our main results in this paper concern shifts of finite type and sofic shifts, which we now define.

Definition 3.2.1. [shift of finite type] A *shift of finite type* with alphabet $\mathcal{A}$ is any shift space of $\mathcal{A}^{\mathbb{Z}}$ defined by excluding a finite number of finite words. In other words, $X \subseteq \mathcal{A}^{\mathbb{Z}}$ is a shift of finite type if for some $n \geq 1$ there exists a finite set $\mathcal{F} \subseteq \mathcal{A}^n$ of forbidden words such that

$$X = \{x \in \mathcal{A}^{\mathbb{Z}} : \sigma^m(x)_{[0,n-1]} \notin \mathcal{F}, \text{ for all } m \in \mathbb{Z}\}$$

Definition 3.2.2. [sofic shift] A *sofic shift* is any shift space that is a continuous factor of a shift of finite type.

Sofic shifts have an alternative characterization in terms of bi-infinite walks on finite edge-labeled graphs (see [49]), but we will not need this here. Every shift of finite type is sofic, since the identity code is continuous (more generally, conjugacies preserve the class of shifts of finite type), but not every sofic shift has finite type. An example of a shift that is sofic but not of finite type is the even shift, which is the shift $X \subset \{0,1\}^{\mathbb{Z}}$ consisting of sequences in which 10^n1 may appear only if n is even.

We will, in particular, consider measures on shift spaces, which will always be Borel probability measures. We will refer to these as σ-invariant measures to avoid any possible ambiguity, since in §3.4 we also consider measures that are σ^p-invariant for some positive power p, but in general are not σ-invariant. In particular, we will often refer to ergodic measures, and these will always be ergodic with respect to σ.

3.2.2 The Gibbs relation, cocycles, and Gibbs measures

The *Gibbs relation* on a shift space X, also called the tail, asymptotic, or homoclinic relation, is the equivalence relation $\mathfrak{T}_X \subset X \times X$ such that $(x,y) \in \mathfrak{T}_X$ if and only if $x_{[-N,N]^c} = y_{[-N,N]^c}$ for some $N \geq 1$. For Borel sets $A, B \subseteq X$, a *holonomy* of $\mathfrak{T}_X$ is a Borel isomorphism $\psi : A \to B$ such that $(x, \psi(x)) \in \mathfrak{T}_X$ for all $x \in A$. We say that a measure μ on X is $\mathfrak{T}_X$-nonsingular if for every Borel $A \subset X$ with $\mu(A) = 0$, we have $\mu(\mathfrak{T}_X(A)) = 0$, where the saturation $\mathfrak{T}_X(A)$ is defined as

$$\mathfrak{T}_X(A) = \{x' \in X \,:\, \exists\, x \in A \text{ such that } (x,x') \in \mathfrak{T}_X\}$$

Note that if μ is $\mathfrak{T}_X$-nonsingular and $\psi : A \to B$ is a holonomy of $\mathfrak{T}_X$, then whenever $E \subset A$ has $\mu(E) = 0$, we have $\mu(\psi(E)) \leq \mu(\mathfrak{T}_X(E)) = 0$. In particular, the Radon-Nikodym derivative $\frac{d(\mu \circ \psi)}{d\mu}$ is well-defined.

We note that $\mathfrak{T}_X$ is generated by a countable group Γ of holonomies, in the sense that $(x,x') \in \mathfrak{T}_X$ if and only if there exists $\gamma \in \Gamma$ with $\gamma(x) = x'$. This is a special case of the main theorem of [30]. One could choose Γ consisting of holonomies of the form $\psi_{u,v,a,b}$, where $u, v \in \mathcal{B}_{b-a+1}(X)$, for some $a, b \in \mathbb{Z}$ with $a \leq b$, and

$$\psi_{u,v,a,b}(x) = \begin{cases} x_{(-\infty,a)} u x_{(b,\infty)}, & \text{if } x_{[a,b]} = v \text{ and } x_{(-\infty,a)} u x_{(b,\infty)} \in X \\ x_{(-\infty,a)} v x_{(b,\infty)}, & \text{if } x_{[a,b]} = u \text{ and } x_{(-\infty,a)} v x_{(b,\infty)} \in X \\ x, & \text{otherwise} \end{cases}$$

That is, $\psi_{u,v,a,b}$ replaces u with v, or vice versa, whenever possible, and otherwise does nothing.

A (real, additive) *cocycle* on $\mathfrak{T}_X$ is a Borel measurable function $\phi : \mathfrak{T}_X \to \mathbb{R}$ such that $\phi(x,y) + \phi(y,z) = \phi(x,z)$ for all $x,y,z \in X$ with $(x,y),(y,z) \in \mathfrak{T}_X$ (so that $(x,z) \in \mathfrak{T}_X$ as well). By exponentiating or taking logarithms, we can easily convert between additive and multiplicative notation for cocycles. Additive notation is more natural when the cocycle is intended to represent an energy-difference function, and multiplicative notation is more natural when the cocycle serves as a Jacobian for a change of variables.

Given a $\mathfrak{T}_X$-nonsingular measure μ on X, we say that a Borel function $D_{\mu,\mathfrak{T}_X} : \mathfrak{T}_X \to \mathbb{R}^+$ is a (multiplicative) *Radon-Nikodym cocycle* on $\mathfrak{T}_X$ with respect to μ if the pushforward of μ by any holonomy $\psi : A \to B$ of $\mathfrak{T}_X$ satisfies $\frac{d(\mu\circ\psi)}{d\mu}(x) = D_{\mu,\mathfrak{T}_X}(x, \psi(x))$ for μ-a.e. $x \in A$. It is routine to show that any $\mathfrak{T}_X$-nonsingular measure μ on X has a μ-a.e. unique Radon-Nikodym cocycle.

Definition 3.2.3. [Gibbs measure] Let μ be a $\mathfrak{T}_X$-nonsingular Borel measure on a shift space X, and let $\phi : \mathfrak{T}_X \to \mathbb{R}$ be a cocycle. We say that μ is a *Gibbs measure* if for any holonomy $\psi : A \to B$ of $\mathfrak{T}_X$, and μ-a.e. $x \in A$, we have $D_{\mu,\mathfrak{T}_X}(x, \psi(x)) = \exp(\phi(x, \psi(x)))$.

To put it another way, a measure is by definition Gibbs if and only if it is nonsingular, and a nonsingular measure is Gibbs precisely with respect to the logarithm of its own Radon-Nikodym cocycle. These measures are also known as conformal measures in the literature [15, 56].

In [15] it is shown, building on results of Kimura [44], Keller [42], and others, that the definition of a Gibbs measure that we have given is equivalent to another well-known one involving the Dobrushin-Lanford-Ruelle equations, in terms of which the theorems of Dobrushin (Theorem 3.4.2) and Lanford-Ruelle (Theorem 3.4.3) were originally stated. These Gibbs measures do not coincide, in general, with Gibbs measures in the sense of Bowen; see the second remark after Proposition 3.3.7 for further discussion.

3.3 Preservation of Gibbsianness

In this section, we prove a pair of preservation of Gibbsianness results, namely Propositions 3.3.7 and 3.3.10, which are essential to our main theorems, Theorem 3.4.7 and 3.4.11.

Definition 3.3.1. [irreducibility] A shift space X is irreducible if for every ordered pair of blocks $u, v \in \mathcal{B}(X)$ there is $w \in \mathcal{B}(X)$ so that $uwv \in \mathcal{B}(X)$.

Definition 3.3.2. [period] The *period* of an irreducible sofic shift is the greatest common divisor of the least periods of its periodic points.

Definition 3.3.3. [strong irreducibility] A shift space X is *strongly irreducible* if there exists some $r \geq 1$ such that for any $u, v \in \mathcal{B}(X)$ and any $s \geq r$, there exists $w \in \mathcal{B}_s(X)$ with $uwv \in \mathcal{B}(X)$.

Remark 3.3.1. A shift of finite type is strongly irreducible if and only if it is topologically mixing, if and only if it is irreducible and has period 1.

The following proposition generalizes part of the proof of Lemma 4.1 in [55]. (They treat the case of a uniform Gibbs measure on a strongly irreducible shift of finite type over $\mathbb{Z}^d$.) The proof is also very similar to that of Proposition 5.2 in [53].

Proposition 3.3.2. *Let X be a strongly irreducible shift space. Then any $\mathfrak{T}_X$-invariant nonsingular measure on X has full support.*

Proof. Let μ be a $\mathfrak{T}_X$-invariant nonsingular measure on X. Let $r \geq 1$ be a witness for the strong irreducibility of X. Fix $a, b \in \mathbb{Z}$ with $a < b$, and fix $w \in \mathcal{B}_{b-a}(X)$. By strong irreducibility, for each $n \geq 1$ and each $p, s \in \mathcal{B}_n(X)$, there exists $w' \in \mathcal{B}_{(b-a)+2r}(X)$ such that $w'_{[a,b]} = w$ and $pw's \in \mathcal{B}(X)$. By compactness, it follows that for every $x \in X$, there exists some $u \in \mathcal{B}(X)$ with $u_{[a,b]} = w$ and $x_{(-\infty,a-r-1]}ux_{[b+r+1,\infty)} \in X$.

For each pair $u, v \in \mathcal{B}_{(b-a)+2r+1}(X)$ with $u_{[a,b]} = w$, let

$$E_{u,v} = [v]_{a-r} \cap \{x \in X \,:\, x_{(-\infty,a-r-1]}ux_{[b+r+1,\infty)} \in X\}$$

Then $\bigcup_{u,v} E_{u,v} = X$. In particular, $\mu(E_{u,v}) > 0$ for at least one pair u, v. Let $\psi_{u,v,a,b}$ be the holonomy of $\mathfrak{T}_X$ that exchanges u and v on $[a, b]$ when possible and does nothing else, as in §3.2.2. By the definition of $E_{u,v}$, we have that for every $x \in E_{u,v}$, $\psi_{u,v,a,b}(x) \in [u]_{a-r}$. Then we have

$$\begin{aligned}
\mu([w]_a) &\geq \mu([u]_{a-r}) \\
&\geq \mu(\psi_{u,v,a,b}(E_{u,v})) \\
&= \int_{E_{u,v}} D_{\mu,\mathfrak{T}_X}(x, \psi_{u,v,a,b}(x))\, d\mu(x) \\
&> 0
\end{aligned}$$

Since w was an arbitrary word at an arbitrary position, μ has full support. □

Remark 3.3.3. The statement and proof of Proposition 3.3.2 generalize essentially without modification when $\mathbb{Z}$ is replaced by an arbitrary countable group, with a suitable generalization of strong irreducibility [23, 9]. We also mention that strong irreducibility is used in the proof of Proposition 3.3.2 to show that all $\mathfrak{T}_X$ equivalence classes are dense in X. These equivalence classes are the orbits of the action of the countable group Γ generating $\mathfrak{T}_X$. In the special case of a shift of finite type, Γ can be taken to be generated by homeomorphisms [56]. We can thus interpret Proposition 3.3.2 as the statement that a nonsingular measure for a minimal continuous action must have full support.

Definition 3.3.4. [doubly transitive point] Let X be a shift space. A point $x \in X$ is *doubly transitive* if every word $w \in \mathcal{B}(X)$ appears in x infinitely often to the left and to the right.

It is easy to check that a shift space X contains a doubly transitive point if and only if X is irreducible (see §9.1, [49] for more details). The following lemma will also be useful.

Lemma 3.3.4. *Let X be a shift space. Then the set $D_X \subset X$ of doubly transitive points is $\mathfrak{T}_X$-invariant.*

Proof. If X is not irreducible then $D_X = \emptyset$, which is trivially $\mathfrak{T}_X$-invariant, so assume that X is irreducible. Let $x \in D_X$ and suppose that $x' \in X$ with $(x, x') \in \mathfrak{T}_X$. Then there exists some $\Delta = [a, b] \cap \mathbb{Z}$ such that $x_{\Delta^c} = x'_{\Delta^c}$. Since $x \in D_X$, every word $w \in \mathcal{B}(X)$ appears infinitely often both in $x_{(-\infty,a-1]} = x'_{(-\infty,a-1]}$ and in $x_{[b+1,\infty)} = x'_{[b+1,\infty)}$. Therefore $x' \in D_X$. □

It is then immediate that $X \setminus D_X$ is also $\mathfrak{T}_X$-invariant. (In general, if $\mathcal{R}$ is an equivalence relation on X and A is an $\mathcal{R}$-invariant subset, then A^c is also $\mathcal{R}$-invariant.)

The following is a generalization of Theorem 9.4.9 in [49], which appears to be well-known ([78], proof of Lemma 4.5). We thank Tom Meyerovitch for this short proof.

Proposition 3.3.5. *Let X be an irreducible shift space and let μ be a fully supported σ-invariant ergodic measure on X. Let D_X denote the set of doubly transitive points. Then $\mu(D_X) = 1$.*

Proof. Recall that a *generic point*, for a continuous transformation of a compact space with an σ-invariant measure, is a point that satisfies the conclusion of the pointwise ergodic theorem (Theorem 1.14, [76]) for every continuous function. Since μ is ergodic with respect to σ, the sets of generic points with respect to σ and σ^{-1} have full measure ([31], Proposition 3.7), so their intersection has full measure as well. Since μ has full support, every point that is generic for both σ and σ^{-1} (in particular, almost every point in X) is doubly transitive. In greater detail: we prove the contrapositive. Suppose x is not doubly transitive. Then there is a word w which appears at most finitely often, without loss of generality to the right, in x. Then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_{[w]_0}(\sigma^k x) = 0 \neq \mu([w]_0)$$

so x does not satisfy the conclusion of the ergodic theorem for the continuous function $\mathbf{1}_{[w]_0}$. Therefore x is not generic for σ. □

Let X be a shift of finite type, let Y be a sofic shift, and let $\pi : X \to Y$ be a finite-to-one factor code. Then there is an integer $d \geq 1$, known as the *degree* of π, such that each doubly transitive point $y \in Y$ has exactly d π-preimages [49]. An important special case is when the degree is one, which is in fact equivalent to the following condition, which a priori might seem more general.

Definition 3.3.5. [almost invertibility] Let X be an irreducible shift of finite type, let Y be an irreducible sofic shift, and let $\pi : X \to Y$ be a factor code. We say that π is *almost invertible* if every doubly transitive point $y \in D_Y$ has a unique preimage.

Remark 3.3.6. A factor code on an irreducible shift of finite type is almost invertible if and only if it is finite-to-one with degree one ([49], Proposition 9.2.2).

The following is our first preservation of Gibbsianness result.

Proposition 3.3.7 (preservation of Gibbsianness for almost invertible factor codes)**.** *Let X be an irreducible shift of finite type, let Y be a sofic shift, and let $\pi : X \to Y$ be an almost invertible factor code. Let μ be a measure on X which is fully supported, σ-invariant, ergodic, and $\mathfrak{T}_X$-nonsingular. Let $\nu = \pi_*\mu$. Then ν is $\mathfrak{T}_Y$-nonsingular. Moreover, if μ is Gibbs with respect to an additive cocycle ϕ, then ν is Gibbs with respect to $\phi \circ (\pi^{-1} \times \pi^{-1})$, where π^{-1} is well-defined ν-almost everywhere.*

Proof. First, for a technical reason described below, we assume that π is a one-block code, in the sense that it is induced by a map $\Pi\colon \mathcal{B}_1(X) \to \mathcal{B}_1(Y)$. We assume further that it has a *magic symbol*—that is, a symbol $b \in \mathcal{B}_1(Y)$ with a unique Π-preimage $a \in \mathcal{B}_1(X)$, such that if $\pi(x) = y$ and $y_0 = b$, then $x_0 = a$. These assumptions incur no loss of generality ([49], §9.1).

These assumptions made, we see how to lift holonomies of $\mathfrak{T}_Y$ to holonomies of $\mathfrak{T}_X$. Since π is a Borel map between complete separable metric spaces X, Y which restricts to an injection on the Borel set D_X, the inverse $\pi|_{D_X}^{-1} : D_Y \to D_X$ is Borel ([40], Theorem 15.1). Now let $\psi : A \to B$ be a holonomy of $\mathfrak{T}_Y$. Define the map $\tilde{\psi} : \pi^{-1}(A) \cap D_X \to \pi^{-1}(B) \cap D_X$ by $\tilde{\psi} = \pi^{-1} \circ \psi \circ \pi$; this is clearly a measurable bijection.

Moreover, to see that $(x, \tilde{\psi}(x)) \in \mathfrak{T}_X$ for each $x \in \pi^{-1}(A) \cap D_X$, observe that $(\pi(x), \pi \circ \tilde{\psi}(x)) = (\pi(x), \psi \circ \pi(x)) \in \mathfrak{T}_Y$, so there exist $m < n$ with $\pi(x)_{[m,n]^c} = \psi \circ \pi(x)_{[m,n]^c}$. Since $\pi(x)$ and $\psi \circ \pi(x)$ are doubly transitive, the magic symbol b for π appears infinitely often to the left and right. We may therefore assume, by taking m and n larger if necessary, that

$$\pi(x)_{m-1} = \psi \circ \pi(x)_{m-1} = b = \pi(x)_{n+1} = \psi \circ \pi(x)_{n+1}$$

Moreover, since the magic symbol occurs infinitely often to the left and right, any word outside $[m, n]$ appears within a word beginning and ending with b. Proposition 9.1.9 in [49] asserts, in the almost invertible case, that such a word has a unique π-preimage. This shows that $(x, \tilde{\psi}(x))$ agree outside of $[m, n]$, so $\tilde{\psi}$ is indeed a holonomy of $\mathfrak{T}_X$.

Now, let Γ_X, Γ_Y be countable groups of holonomies generating $\mathfrak{T}_X, \mathfrak{T}_Y$ respectively, and let $A \subset Y$ be Borel with $\nu(A) = 0$. Observe that $\mathfrak{T}_Y(A) = \bigcup_{\gamma \in \Gamma_Y} \gamma(A)$. Observe further that, for each $\gamma \in \Gamma_Y$,

$$\begin{aligned}
\nu(\gamma(A)) &= \nu(\gamma(A) \setminus D_Y) + \nu(\gamma(A) \cap D_Y) \\
&= \mu\left(\pi^{-1}(\gamma(A) \cap D_Y)\right) \\
&= \mu(\tilde{\gamma}(\pi^{-1}(A \cap D_Y))) \\
&= 0
\end{aligned}$$

since $\pi^{-1}(A \cap D_Y)$ is μ-null and μ is $\mathfrak{T}_X$-nonsingular. Therefore ν is indeed $\mathfrak{T}_Y$-nonsingular.

Again, let $\psi : A \to B$ be a holonomy of $\mathfrak{T}_Y$ and let $\tilde{\psi}$ be as above. Then

$$\begin{aligned}
\nu(B) &= \mu(\pi^{-1}(\psi(A \cap D_Y))) \\
&= \mu(\tilde{\psi}(\pi^{-1}(A) \cap D_X)) \\
&= \int_{\pi^{-1}(A) \cap D_X} D_{\mu, \mathfrak{T}_X}(x, \tilde{\psi}(x))\, d\mu(x) \\
&= \int_{A \cap D_Y} D_{\mu, \mathfrak{T}_X}(\pi^{-1}(y), \pi^{-1}(\psi(y)))\, d\nu(y),
\end{aligned}$$

where the last equality follows from the change of variables formula.

On the other hand, since ν is $\mathfrak{T}_Y$-nonsingular, we know that

$$\nu(B) = \int_A D_{\mu, \mathfrak{T}_Y}(y, \psi(y))\, d\nu(y)$$

By the uniqueness of the Radon-Nikodym derivative, we then have, for almost all $(y, y') \in D_Y^2 \cap \mathfrak{T}_Y$, that $D_{\nu, \mathfrak{T}_Y}(y, y') = D_{\mu, \mathfrak{T}_X}(\pi^{-1}(y), \pi^{-1}(y'))$. $\square$

Remark 3.3.8. In Corollary 3.4.10, we generalize Proposition 3.3.7 from almost invertible to finite-to-one codes, in the case that the cocycle on the range is induced by a sufficiently regular potential. It is therefore natural to ask whether the proof of Proposition 3.3.7 can be adapted to the finite-to-one setting. However, such an adaptation would not be straightforward, because at higher degrees we lose a key condition on which the proof of Proposition 3.3.7 relied: namely, that preimages of asymptotic doubly transitive points must themselves be asymptotic. Indeed, if π has degree $d > 1$, then every doubly transitive point $y \in Y$ has d preimages, no two of which ever exhibit the same symbol in the same position (see Exercise 9.1.3, [49]), and are therefore about as far from asymptotic as one could imagine.

Remark 3.3.9. The hypotheses of Proposition 3.3.7 are likely more restrictive than would be needed simply to show that the pushforward of a $\mathfrak{T}_X$-nonsingular measure is $\mathfrak{T}_Y$-nonsingular. The reason is that, for the application in Theorem 3.4.7, we need pointwise control over the potential inducing the Radon-Nikodym cocycle of the pushforward measure, whereas for the purposes of [64], for instance, it is sufficient to determine the potential's regularity. For instance, consider a very simple symbol amalgamation code from the full 3-shift $X = \{0, 1, 2\}^{\mathbb{Z}}$ to the full 2-shift $Y = \{0, 1\}^{\mathbb{Z}}$, given by amalgamating the symbols $1, 2$ into the symbol 1. This code takes the uniform Bernoulli measure on X, which is Gibbs for the zero cocycle, to the $(1/3, 2/3)$ Bernoulli measure on Y, which is Gibbs with respect to a cocycle obtained from a locally constant potential. This would count as preservation of Gibbsianness in the sense of [64] but not in ours.

We now prove Proposition 3.3.10, which is a converse result to Proposition 3.3.7, showing that every nonsingular measure on an irreducible sofic shift is the pushforward, through an almost invertible code, of a nonsingular measure on an irreducible shift of finite type. We observe that an almost invertible factor code $\pi : X \to Y$ yields a Borel isomorphism $\pi_{D_X} : D_X \to D_Y$, so a fully supported σ-invariant measure ν on Y lifts to a unique measure μ on X, giving an isomorphism of measure-preserving systems. In particular, if ν is ergodic, then so is μ.

Proposition 3.3.10 (lifting Gibbs measures through almost invertible factor codes)**.** *Let X be an irreducible shift of finite type, let Y be a sofic shift, and let $\pi : X \to Y$ be an almost invertible factor code. Let ν be a measure on Y which is fully supported, σ-invariant, ergodic, and $\mathfrak{T}_Y$-nonsingular. If ν is Gibbs for an additive cocycle ϕ, then the unique σ-invariant measure μ on X with $\pi_*\mu = \nu$ is $\mathfrak{T}_X$-nonsingular and, in particular, is a σ-invariant Gibbs measure for $\phi \circ (\pi \times \pi)$.*

Proof. As in the proof of Proposition 3.3.7, we assume without loss of generality (that is, up to a conjugacy of X) that π is a one-block code, induced by a block map $\Pi : \mathcal{B}_1(X) \to \mathcal{B}_1(Y)$. We assume further that π has a magic symbol $b \in \mathcal{B}_1(Y)$, which has a unique Π-preimage $a \in \mathcal{B}_1(X)$.

Since $\pi : D_X \to D_Y$ is a Borel isomorphism and $\nu(D_Y) = 1$ by Proposition 3.3.5, there is a unique measure μ on X with $\pi_*\mu = \nu$, and μ is σ-invariant and ergodic. We now show that μ has full support. Let $w \in \mathcal{B}(X)$ be arbitrary. By irreducibility, there exist $u, v \in \mathcal{B}(X)$ such that $auwva \in \mathcal{B}(X)$. Let $s = \Pi(uwv)$, so that, by the same argument as in the proof of Proposition 3.3.7 (using [49], §9.1), $[auwva]_0 = \pi^{-1}([bsb]_0)$. Since ν has full support by hypothesis, $\nu([bsb]_0) > 0$, so $\mu([w]_0) > 0$, since $[auwva] \subseteq [w]$, with coordinates lined up appropriately. Since w was arbitrary and μ is σ-invariant, μ has full support.

Let $\psi \colon X \to X$ be a holonomy of $\mathfrak{T}_X$. As in the proof of Proposition 3.3.7 (but with the holonomies being pushed in the opposite direction), π forms a Borel isomorphism between D_X and D_Y, and if $x, x' \in D_X$ then $(x, x') \in \mathfrak{T}_X$ if and only if $(\pi(x), \pi(x')) \in \mathfrak{T}_Y$. Therefore, we have a holonomy $\tilde{\psi} = \pi \circ \psi \circ (\pi|_{D_X})^{-1} : D_Y \to D_Y$ of $\mathfrak{T}_Y$.

Let $N \subset X$ be a Borel set with μ-measure zero. Observe that

$$\psi(N) = \psi(N \cap D_X) \cup \psi(N \setminus D_X)$$

This yields that $\mu(\psi(N)) = \mu(\psi(N) \cap D_X)$, since $\psi(N \setminus D_X) = \psi(N) \setminus D_X$, and D_X has full measure by ergodicity and Proposition 3.3.5. Moreover,

$$\psi(N \cap D_X) = \pi^{-1} \circ \tilde{\psi}(\pi(N) \cap D_Y)$$

Now, $\nu(\pi(N) \cap D_Y) = \mu(N \cap D_X) = 0$, so $\nu(\tilde{\psi}(\pi(N) \cap D_Y)) = 0$, since ν is $\mathfrak{T}_Y$-nonsingular. Therefore $\mu(\psi(N \cap D_X)) = 0$ as well, so in fact $\mu(\psi(N)) = 0$, which shows that μ is indeed $\mathfrak{T}_X$-nonsingular. A calculation in the same spirit, very similar to the calculation that concludes the proof of Proposition 3.3.7, shows that μ is Gibbs for $\phi \circ (\pi \times \pi)$, as claimed. □

3.4 Equilibrium measures

Definition 3.4.1. [topological pressure and equilibrium states] Let X be a shift space and let $f\colon X \to \mathbb{R}$ be a continuous function. The *topological pressure* of f is the value

$$P_X(\sigma, f) = \sup\left\{h(\mu) + \int_X f\, d\mu\right\}$$

where the supremum is over all σ-invariant measures μ on X. Any measure μ attaining the supremum is known as an *equilibrium measure* for f.

Remark 3.4.1. The pressure is sometimes defined as above (see [42]), but is often defined differently (see e.g. [76]), with the variational property by which we defined it stated as a theorem. However, this variational property is the only one we will need, apart from the two classical theorems below.

Definition 3.4.2. [the function space $\mathrm{SV}(X)$] Let X be a shift space and let $f\colon X \to \mathbb{R}$ be a continuous function. We define the k^{th} *variation* of f as

$$v_k(f) = \sup\{|f(x) - f(y)| \,:\, x_{[-k,k]} = y_{[-k,k]}\}$$

for $k \geq 0$; it is convenient to define $\mathrm{var}_{-1}(f) = \|f\|_\infty$. We then define

$$\|f\|_{\mathrm{SV}(X)} = \sum_{k=0}^{\infty} v_{k-1}(f)$$

and define the class of potentials $\mathrm{SV}(X) = \{f \in C(X) \,:\, \|f\|_{\mathrm{SV}(X)} < \infty\}$.

It is easy to verify that $\left(\mathrm{SV}(X), \|\cdot\|_{\mathrm{SV}(X)}\right)$ is a Banach space, and that ([56], [15]) a potential $f \in \mathrm{SV}(X)$ defines a cocycle ϕ_f on $\mathfrak{T}_X$ via the following absolutely convergent series:

$$\phi_f(x, y) = \sum_{n\in\mathbb{Z}} [f(\sigma^n y) - f(\sigma^n x)]$$

We refer to a Gibbs measure for ϕ_f simply as a Gibbs measure for f. With this definition, we recall the following classical theorems, which we state only for the special cases that we require. These theorems were originally proved in a somewhat different form, using the formalism of interactions rather than potentials as we have used, but the methods adapt easily.

Theorem 3.4.2 (Dobrushin; see [66])**.** *Let X be a strongly irreducible shift space and let $f \in \mathrm{SV}(X)$. Every σ-invariant Gibbs measure for f is an equilibrium measure for f.*

Theorem 3.4.3 (Lanford-Ruelle; see [56])**.** *Let X be a shift of finite type and let $f \in \mathrm{SV}(X)$. Every equilibrium measure for f is a Gibbs measure for f.*

We also require the following result, closely following Lemma 4.5 in [78].

Proposition 3.4.4. *Let X be an irreducible shift of finite type. Any equilibrium measure for $f \in \mathrm{SV}(X)$ has full support.*

Proof. When X is a mixing shift of finite type, the result follows from Theorem 3.4.3 and Proposition 3.3.2. Now, let X be an irreducible shift of finite type such that X has period p. We decompose X into p cyclically moving classes (see [49], §4.5), i.e., $X = \sqcup_{i=0}^{p-1} X_i$ where $\sigma(X_i) = X_{i+1 \mod p}$ and each X_i is mixing with respect to σ^p. In particular, σ^p is a homeomorphism of each clopen set X_i; the system (X, σ^p) is known as the p^{th} higher power shift of X ([49], §1.4).

There is a bijection between σ-invariant measures μ on X and σ^p-invariant measures μ' on X_0, which is given by normalized restriction in one direction and averaging in the other. That is, we take

$\mu' = p\mu|_{X_0}$ and $\mu = p^{-1}\sum_{j=0}^{p-1}\sigma_*^j\mu'$. Observe that μ has full support if and only if μ' does. Moreover, we have

$$h_X(\mu) + \int_X f\,d\mu = \frac{1}{p}\left(h_{X_0}(\mu') + \int_{X_0} R_p f\,d\mu'\right)$$

where $R_p f = \sum_{j=0}^{p-1} f\circ\sigma^{-j}$. Therefore μ is an equilibrium measure for f on X precisely when μ' is an equilibrium measure for $R_p f$ on X_0. So far, our discussion is essentially identical to Yoo's.

We now need to show that $R_p f \in \mathrm{SV}(X_0)$; this is where we part from Yoo, who considers a different class of potentials. Observe that, for each $j \in \{0,\ldots,p-1\}$, the k-th variation of a function on X_0 behaves like the kp-th variation of that function on X, so we have

$$\left\|f\circ\sigma^{-j}\right\|_{\mathrm{SV}(X_0)} \le \sum_{k=0}^{\infty} v_{kp}(f\circ\sigma^{-j}) \le \left\|f\circ\sigma^{-j}\right\|_{\mathrm{SV}(X)}$$

It is also easy to check that $\mathrm{SV}(X)$ is closed under translation, showing that indeed $R_p f \in \mathrm{SV}(X_0)$.

Let μ be an equilibrium measure for f. Then the unique μ' on X_0 such that $\mu = p^{-1}\sum_{j=0}^{p-1}\sigma_*^j\mu'$ is an equilibrium measure for $R_p f$; since X_0 is mixing, μ' is fully supported. Thus μ is fully supported as well. □

We also require a result showing that equilibrium measures lift through almost invertible codes, indeed through any finite-to-one codes. We first prove the following lemma, which is well-known, and can be regarded as a relative version of the Krylov-Bogliubov theorem. One standard proof uses the Hahn-Banach theorem. For completeness, and possibly independent interest, we include a different proof, based on the standard proof of the (non-relative) Krylov-Bogliubov theorem [76]. The Hahn-Banach argument requires no dynamical assumptions at all, whereas our argument relies on the dynamical setting to make the argument somewhat more constructive. We state it only for ergodic measures on shift spaces; the argument goes through for any continuous transformation of a compact metric space, and it easily generalizes to any σ-invariant measure by convexity.

Lemma 3.4.5. *Let X and Y be shift spaces, let $\pi : X \to Y$ be a factor code, and let ν be a σ-invariant measure on Y. Then there exists a σ-invariant measure μ on X with $\pi_*\mu = \nu$. If ν is ergodic then μ can be chosen to be ergodic as well.*

Proof. Let $y \in Y$ be a generic point with a preimage $x \in X$, and let $\nu_n = \frac{1}{n}\sum_{k=0}^{n-1}\delta_y\circ\sigma^{-k}$ be the n^{th} empirical measure for y. By the ergodic theorem, ν_n converges to ν in the weak-$*$ topology. Similarly, let $\mu_n = \frac{1}{n}\sum_{k=0}^{n-1}\delta_x\circ\sigma^{-k}$, and appeal to compactness to extract a weak-$*$ convergent subsequence with limit μ. It is not hard to see that μ must be σ-invariant (since $\|\mu_n - \mu_n\circ\sigma\|_{\mathrm{TV}} \le 2/n$). Moreover, since $\nu_n = \pi_*\mu_n$ and the pushforward operation is continuous, we conclude that $\pi_*\mu = \nu$. The fact that μ can be chosen to be ergodic follows by a standard convexity argument (see [70], Chapter 8). □

Lemma 3.4.6. *Let X be an irreducible shift of finite type, Y a sofic shift, and $\pi : X \to Y$ a finite-to-one factor code. Let $f\colon Y \to \mathbb{R}$ be a function with summable variation, that is, $f \in \mathrm{SV}(Y)$, and let ν be an equilibrium measure for f. Then any σ-invariant measure μ on X with $\pi_*\mu = \nu$ is an equilibrium measure is an equilibrium measure μ for $f\circ\pi$; in particular, there exists an equilibrium measure μ for $f\circ\pi \in \mathrm{SV}(X)$ with $\pi_*\mu = \nu$. If ν is ergodic, then μ can be chosen to be ergodic as well.*

Proof. By Lemma 3.4.5, there exists a σ-invariant measure μ on X with $\pi_*\mu = \nu$, and we can choose μ to be ergodic whenever ν is ergodic. We now show that any such μ is an equilibrium measure for $f\circ\pi$. Since π is finite-to-one, the Abramov-Rokhlin formula [1] shows that $h(\mu) = h(\nu)$. Then

$$h(\mu) + \int_X f\circ\pi\,d\mu = h(\nu) + \int_Y f\,d\nu = P_Y(\sigma, f)$$

We need to show that $P_X(\sigma, f \circ \pi) = P_Y(\sigma, f)$. Clearly $P_X(\sigma, f \circ \pi) \geq P_Y(\sigma, f)$, so we only need the reverse inequality. Let λ be any σ-invariant measure on X. Again $h(\pi_*\lambda) = h(\lambda)$ by the Abramov-Rokhlin formula, so we have

$$h(\lambda) + \int_X f \circ \pi \, d\lambda = h(\pi_*\lambda) + \int_Y f \, d\pi_*\lambda \leq P_Y(f)$$

Therefore $P_X(\sigma, f \circ \pi) = P_Y(\sigma, f)$, so μ is indeed an equilibrium measure for $f \circ \pi$. $\square$

We can now state and prove the first main result of this section.

Theorem 3.4.7 (Lanford-Ruelle theorem for irreducible sofic shifts)**.** *Let Y be an irreducible sofic shift and let $f\colon Y \to \mathbb{R}$ be a potential with summable variation, that is, $f \in SV(Y)$. Let ν be an equilibrium measure for f. Then ν is a Gibbs measure for f.*

Proof. Since Y is an irreducible sofic shift, there exist an irreducible shift of finite type X and an almost invertible factor code $\pi\colon X \to Y$, for instance the minimal right-resolving presentation ([49], §3.3, §9.2).

Suppose first that ν is ergodic. By Lemma 3.4.6, there exists an ergodic equilibrium measure μ for $f \circ \pi$ such that $\nu = \pi_*\mu$. Then, by Theorem 3.4.3, μ is a Gibbs measure for $f \circ \pi$, and by Proposition 3.4.4, μ has full support. Then, by Proposition 3.3.7, which requires ergodicity and full support, ν is a Gibbs measure for f.

The general result follows from the ergodic case via the Krein-Milman theorem [70], together with the compactness and convexity of the sets of Gibbs and equilibrium measures [42] and the fact that the extreme points of these sets are precisely their ergodic elements. $\square$

We now change course and proceed towards a Dobrushin type theorem for irreducible sofic shifts, which is the second main result of this section.

Lemma 3.4.8. *Let X be an irreducible shift of finite type of period p, partitioned into p cyclically moving classes X_i, $0 \leq i \leq p-1$. Then any $\mathfrak{T}_X$-equivalence class is contained in a single cyclically moving class. That is, if $x \in X_i$ and $x' \in X_j$ with $(x, x') \in \mathfrak{T}_X$, then $i = j$.*

Proof. Suppose without loss of generality that X is an edge shift with alphabet $\mathcal{A} = \mathcal{B}_1(X)$. Then $\mathcal{A}$ can be partitioned into p subsets $\mathcal{A}_i$, $0 \leq i \leq p-1$, corresponding to the cyclically moving classes X_i, such that for any $x \in X$, $x \in X_i$ if and only if $x \in \mathcal{A}_i$. Thus for $x, x' \in X$, if $x_n = x'_n$ for some n, then x, x' are in the same cyclically moving class. But if $(x, x') \in \mathfrak{T}_X$ then $x_n = x'_n$ for all but finitely many n, so in particular x and x' are in the same cyclically moving class. $\square$

Lemma 3.4.9 (Dobrushin theorem for irreducible shifts of finite type)**.** *Let X be an irreducible shift of finite type and $f\colon X \to \mathbb{R}$ be a potential with summable variation, that is $f \in SV(X)$. Then every σ-invariant Gibbs measure for f is an equilibrium measure for f.*

Proof. Let X have period p and let $X_0, \ldots, X_{p-1}$ be the cyclically moving classes of X as in the proof of Proposition 3.4.4. Note that each X_i, $0 \leq i \leq p-1$ is $\mathfrak{T}_X$-invariant and that we can regard each one as a mixing shift of finite type with alphabet a subset of $\mathcal{B}_p(X)$, so that it is meaningful to speak of the Gibbs relations $\mathfrak{T}_{X_i}$. Moreover, since each class X_i is $\mathfrak{T}_X$-invariant, we have for each i that $\mathfrak{T}_{X_i}$ is simply the restriction of $\mathfrak{T}_X$ to $X_i \times X_i$.

Let μ be a Gibbs measure for f and μ' be the normalized restriction of μ of X to X_0, i.e., $\mu'(E) = p\mu(E)$ for any Borel set $E \subseteq X_0$. Lemma 3.4.8 shows that every holonomy $\psi : A \to B$ of $\mathfrak{T}_X$ restricts to a holonomy $\psi_0 : A \cap X_0 \to B \cap X_0$ of the relation $\mathfrak{T}_{X_0}$. Moreover, every holonomy of $\mathfrak{T}_{X_0}$ clearly arises as such a restriction: if ψ_0 is a holonomy of $\mathfrak{T}_{X_0}$, $\psi_0 = \psi|_{X_0}$ where $\psi|_{X_i} = \sigma^i \circ \psi_0 \circ \sigma^{-i}$. Therefore the measure μ' inherits the Gibbsianness of μ, with respect to $R_p f = \sum_{j=0}^{p-1} f \circ \sigma^j$. Since each X_i,

$0 \le i \le p-1$ is a mixing shift of finite type, therefore strongly irreducible, Theorem 3.4.2 shows that μ' is an equilibrium measure for $R_p f$. As in the proof of Proposition 3.4.4, we have

$$h_X(\mu) + \int_X f\, d\mu = \frac{1}{p}\left(h_{X_0}(\mu') + \int_{X_0} R_p f\, d\mu' \right),$$

which concludes that μ is an equilibrium measure for f.

□

The next corollary generalizes Proposition 3.3.7 from almost invertible codes to finite-to-one codes, in the case of a cocycle induced by a sufficiently regular potential. As discussed in §3.3, the proof technique that we used for Proposition 3.3.7 does not generalize to higher degrees. To obtain Corollary 3.4.10, we apply Theorem 3.4.7, even though the statement of Corollary 3.4.10 does not mention equilibrium measures explicitly.

Corollary 3.4.10 (preservation of Gibbsianness for finite-to-one factor codes)**.** *Let X be an irreducible shift of finite type and Y an irreducible sofic shift. Let $\pi : X \to Y$ be a finite-to-one factor code, let $f\colon Y \to \mathbb{R}$ be a potential with summable variation, that is, $f \in SV(Y)$, and let μ be a σ-invariant Gibbs measure for $f \circ \pi$. Then $\nu = \pi_* \mu$ is a Gibbs measure for f.*

Proof. First note that μ is an equilibrium measure for $f \circ \pi$ by Lemma 3.4.9. Since π is finite-to-one, ν is an equilibrium measure for f ([78], Lemma 4.4), so by Theorem 3.4.7, ν is Gibbs for f. □

While Corollary 3.4.10 may look like it generalizes Proposition 3.3.7 in that μ need no longer be ergodic and π need no longer have degree one. However, the cocycle in Proposition 3.3.7 need not a priori be induced by the pullback of a potential with summable variation. Corollary 3.4.10 therefore generalizes Proposition 3.3.7 only in this special case.

Theorem 3.4.11 (Dobrushin theorem for irreducible sofic shifts)**.** *Let Y be an irreducible sofic shift and let $f\colon Y \to \mathbb{R}$ be a potential with summable variation, that is, $f \in SV(Y)$. Let ν be a σ-invariant Gibbs measure for f. Then ν is an equilibrium measure for f.*

Proof. First suppose that ν is ergodic. Let $\pi : X \to Y$ be the minimal right-resolving presentation of Y. By Lemma 3.3.10, there exists an ergodic σ-invariant Gibbs measure μ for $f \circ \pi$ with $\pi_*\mu = \nu$. By Lemma 3.4.9, μ is an equilibrium measure for $f \circ \pi$. Finally, again by Lemma 4.4 in [78], ν is an equilibrium measure for f. The result for general (not necessarily ergodic) ν follows from the ergodic case by compactness and convexity as in the proof of Theorem 3.4.7. □

Taken together, Theorems 3.4.7 and 3.4.11 show that for a potential with summable variations on an irreducible sofic shift, the equilibrium measures are precisely the σ-invariant Gibbs measures.

Finally, we can use Theorem 3.4.11 to generalize Proposition 3.3.10 from almost invertible to finite-to-one codes, in the same special case for which Corollary 3.4.10 generalizes Proposition 3.3.7.

Corollary 3.4.12 (lifting Gibbs measures through almost invertible factor codes)**.** *Let X be an irreducible shift of finite type and Y an irreducible sofic shift. Let $\pi : X \to Y$ be a finite-to-one factor code, let $f\colon Y \to \mathbb{R}$ be a potential with summable variation, that is, $f \in SV(Y)$, and let ν be a σ-invariant Gibbs measure for f. Then there exists a σ-invariant Gibbs measure μ for $f \circ \pi$ with $\pi_*\mu = \nu$.*

Proof. By Theorem 3.4.11, ν is an equilibrium measure for f. By Lemma 3.4.6, there is an equilibrium measure μ for $f \circ \pi$ with $\pi_*\mu = \nu$. By Theorem 3.4.3, μ is a Gibbs measure for $f \circ \pi$, and is certainly σ-invariant. □

In closing, we note that it is an open problem to determine the existence of a finite-to-one factor code from a given shift of finite type X onto a given sofic shift Y, as in the hypotheses of Corollaries 3.4.10 and 3.4.12. Equal entropy is necessary, but there are additional necessary conditions; see [49], §12.2.

Chapter 4

The road problem and homomorphisms of directed graphs

4.1 Introduction

In §4.2, we recall and adapt standard material on graphs and homomorphisms. In §4.3, we define the stability relation for a right-resolver, give its main structural properties, and relate it to synchronization. In §4.4, we recall from [6] the connection between the road problem and the $O(G)$ conjecture, and state our generalization of the road colouring theorem.

In §4.5, we introduce the concepts of bunchiness and almost bunchiness and present results involving them, including the $O(G)$ conjecture for bunchy and almost bunchy graphs and the universal property of the fiber product. In §4.6, we pose the bunchy factor conjecture, which has several equivalent formulations, and discuss its relation to the $O(G)$ conjecture and the road problem. In §4.7, we give polynomial-time algorithms for construction and decision problems involving right-resolvers, and discuss the algorithmic implications of the $O(G)$ and bunchy factor conjectures.

The proofs of many results in §§4.2–4.7, comprising the structural properties of right-resolvers, are deferred to §4.8. The proof of our generalization of the road colouring theorem is deferred to §4.9.

4.2 Graphs and graph homomorphisms

4.2.1 Basic definitions

We take all graphs to be finite and directed. A graph G consists of a set $V(G)$ of states, or vertices, and a set $E(G)$ of edges, together with a pair of maps $s, t : E(G) \to V(G)$ giving the source and target of each edge. Loops (edges e with $s(e) = t(e)$) and parallel edges (distinct edges e, e' with $s(e) = s(e')$, $t(e) = t(e')$) are allowed. For $I \in V(G)$, we write $E_I(G) = s^{-1}(I)$ for the set of outgoing edges from I. We write $F(I) = t(E_I(G))$ for the set of follower states of I, and we write $E_{IJ}(G) = s^{-1}(I) \cap t^{-1}(J)$ for the set of edges from I to J. We write $L(G)$ for the language of G, i.e. the set of finite edge paths in G, i.e. $e_1 e_2 \dots e_n$ where $t(e_i) = s(e_{i+1})$, $1 \leq i \leq n-1$. We also refer to elements of $L(G)$ as *words*. The maps s, t extend to $s, t : L(G) \to V(G)$ by $s(e_1 \dots e_n) = s(e_1)$, $t(e_1 \dots e_n) = t(e_n)$. We define $L_I(G) = \{u \in L(G) \,|\, s(u) = I\}$, and $L_{IJ}(G) = \{u \in L_I(G) \,|\, t(u) = J\}$. A *cycle* in G is a path $u \in L(G)$ with $s(u) = t(u)$, i.e. an element of $L_{II}(G)$ for some $I \in V(G)$.

A *graph homomorphism* $\Phi : G \to H$ is a pair of maps $\Phi : E(G) \to E(H)$, $\partial\Phi : V(G) \to V(H)$ such that $s \circ \Phi = \partial\Phi \circ s$ and $t \circ \Phi = \partial\Phi \circ t$. If there is a surjective homomorphism from G to H, then we say that H is a *factor* of G, and that G is an *extension* of H. Observe that every factor of a strongly connected graph is strongly connected. A graph homomorphism $\Phi : G \to H$ induces a map $L(G) \to L(H)$ (also written Φ) in the obvious way.

A graph isomorphism is a homomorphism that is injective and surjective (i.e. on both edges and states), and an automorphism of a graph G is an isomorphism from G to itself. We denote the group of automorphisms of G by $\mathrm{Aut}(G)$, and by $P(G)$ the (normal) subgroup of $\mathrm{Aut}(G)$ which acts trivially on states and permutes parallel edges. We generally identify isomorphic graphs, and use the symbol

$\cong$ to denote isomorphism, except when discussing algorithms for deciding isomorphism, or confirming that the automorphism group of a given graph is trivial.

We now introduce the class of homomorphisms with which we are concerned.

Definition 4.2.1 (right-resolver)**.** Let G, H be graphs. Let $\Phi : G \to H$ be a surjective graph homomorphism. We say that Φ is *right-resolving*, or is a *right-resolver*, if, for each $I \in V(G)$, the restriction $\Phi|_{E_I(G)} : E_I(G) \to E_{\partial\Phi(I)}(H)$ is a bijection. We denote the set of right-resolvers $G \to H$ by $\hom_R(G, H)$, and we write $H \leq_R G$ if $\hom_R(G, H) \neq \emptyset$.

Remark 4.2.1. The term "right-resolving" comes from symbolic dynamics, where the words are of primary importance and the actual graph is secondary. A graph homomorphism $\Phi : G \to H$ is right-resolving if and only if the associated map $\Phi : L(G) \to L(H)$ satisfies a certain condition on the symbols (edges) appearing to the right of a given symbol in a word. See [6], §8.2 for details if interested.

Remark 4.2.2. The class of right-resolving graph homomorphisms is closed under composition. This reduces to the fact, applied to the outgoing edges from each state, that a composition of bijections is a bijection. This means that the relation $\leq_R$ is transitive, and since the graphs are finite, it is clearly antisymmetric, so it is indeed a partial order on the set of all graphs (really on the set of equivalence classes of graphs up to isomorphism).

The following lemma is evident but we state it explicitly for future reference.

Lemma 4.2.3. *The image of a right-resolver is determined up to graph isomorphism by the partition of the domain into fibers. That is, if $H_1, H_2 \leq_R G$ via $\Phi_i \in \hom_R(G, H_i)$, and for any $I_1, I_2 \in V(G)$ we have $\partial\Phi_1(I_1) = \partial\Phi_1(I_2)$ if and only if $\partial\Phi_2(I_1) = \partial\Phi_2(I_2)$, then in fact the H_i are isomorphic.*

Note that the converse is not true: for a given G, H with $H \leq_R G$, there may exist $\Phi, \Phi' \in \hom_R(G, H)$ with distinct partitions $\{(\partial\Phi)^{-1}(I) \mid I \in V(H)\}$, $\{(\partial\Phi')^{-1}(I) \mid I \in V(H)\}$ of $V(G)$. However, this cannot occur when H is $\leq_R$-minimal:

Theorem 4.2.4 ([6], Theorem 3.2 and Corollary 3.3(a))**.** *For any graph G, there exist a unique $\leq_R$-minimal graph $M(G) \leq_R G$ and a unique map $\Sigma_G : V(G) \to V(M(G))$ such that $\partial\Phi = \Sigma_G$ for any $\Phi \in \hom_R(G, M(G))$.*

The construction of $M(G)$ was first given in [25], though not in this notation. We discuss the proof of Theorem 4.2.4 in §4.8.1. The notion of $M(G)$, for a graph G, provides context for road colourings:

Definition 4.2.2 (M_D and road colourings)**.** For $D \geq 1$, let M_D be the graph with a single state and D self-loops. For a graph G of constant out-degree D, a *road colouring* of G is a right-resolver $G \to M_D$.

Note that each M_D is $\leq_R$-minimal, and that $\hom_R(G, M_D)$ is nonempty if and only if G has constant out-degree D, in which case $M(G) = M_D$.

4.2.2 Subgraphs and connectedness

A *sink* in a graph G is a state $I \in V(G)$ is a state with no outgoing edges, i.e. $F(I) = \emptyset$. We assume throughout that all graphs are sink-free; this is purely for convenience, as all of the results that do not require strong connectedness can be proved for graphs with sinks, with routine but tedious modifications to the proofs. We say that a graph G is *strongly connected*, or irreducible, if for any ordered pair $I, J \in V(G)$, there is a (directed) edge path in G from I to J, i.e. $L_{IJ}(G) \neq \emptyset$. Note that strongly connected graphs are sink-free. The *period* $\mathrm{per}(G)$ of a strongly connected graph G is the gcd of its cycle lengths.

A graph H is a *subgraph* of a graph G if $E(H) \subseteq E(G)$, $V(H) \subseteq V(G)$, and the maps s, t with respect to H agree with their counterparts on G, restricted to H. An *induced subgraph* of a graph G

is a subgraph H such that $E_{IJ}(H) = E_{IJ}(G)$ for every $I, J \in V(H)$. A *strong component* of a graph is a maximal strongly connected subgraph, i.e. a strongly connected subgraph that is not a proper subgraph of another strongly connected graph.

A *principal subgraph* of a graph G is a subgraph H such that $E_I(H) = E_I(G)$ for every $I \in V(H)$. Note that every principal subgraph is induced. Note also that if H is a principal subgraph of G, and K is a principal subgraph of H, then K is a principal subgraph of G. A *principal component* is a strongly connected principal subgraph; note that the principal components are precisely the minimal principal subgraphs. In particular, any two principal components of a given graph have disjoint sets of states. The principal components of G correspond to the sink states in the *condensation* of G, which is the directed acyclic graph in which the states are the strong components, or maximal strongly connected subgraphs, of G, and with an edge $C_1 \to C_2$ in the condensation if there is an edge $I_1 \to I_2$ for $I_i \in V(C_i)$ in G.

Let G, H be graphs with $H \leq_R G$, and let $\Phi \in \hom_R(G, H)$. Let K be a subgraph of G. Note, by the right-resolving property, that in order for $\Phi|_K : K \to H$ to be surjective, it is necessary and sufficient that $\partial\Phi|_{V(K)} : V(K) \to V(H)$ be surjective and that K be a principal subgraph of G.

Remark 4.2.5. The road problem and the $O(G)$ problem were both originally raised for strongly connected graphs, which is a natural restriction given the origins of both problems in the ergodic theory of stationary Markov chains. Moreover, strong connectedness is used in an important way in a lemma used to prove both the road colouring theorem and the almost bunchy case of the bunchy factor conjecture. This is why the $O(G)$ conjecture and the bunchy factor conjecture are stated only for strongly connected graphs.

However, it is quite natural from an automata-theoretic perspective, especially concerning computational complexity, to consider graphs that are not strongly connected. For instance, Eppstein [29] shows that it is NP-complete to determine whether the minimum length of a synchronizing word for a given synchronizing DFA is at most some given value. Eppstein's examples are not strongly connected.

4.3 Stability and synchronization

4.3.1 Transitions, stability, and synchronization

A right-resolver on a graph G induces transition maps on $V(G)$ in the standard way:

Definition 4.3.1 (transition map)**.** Let G, H be graphs with $H \leq_R G$. Let $\Phi \in \hom_R(G, H)$. For $I \in V(G)$ and $u \in L_{\partial\Phi(I)}(H)$, we write $I \cdot u$ for the terminal state $t(\gamma)$ of the unique $\gamma \in L_I(G)$ with $\Phi(\gamma) = u$. That is, $I \cdot u = t((\Phi|_{L_I(G)})^{-1}(u))$. We denote by S_Φ the set of maps of the form $I \mapsto I \cdot u$ with respect to Φ.

We now introduce the notion of a *congruence* (see [14], Chapter 1, or [39], §3), of which we will see two important examples. The main example will be the stability relation, but we will also use a congruence in Proposition 4.8.7 to construct the maximal bunchy factor $B(G)$ of a given graph G.

Definition 4.3.2 (congruences and quotients)**.** Let G, H be graphs with $H \leq_R G$, let $\Phi \in \hom_R(G, H)$, and let $\sim$ be an equivalence relation on $V(G)$. We say that $\sim$ is a *congruence* with respect to Φ if it is invariant under transitions, i.e. for all $I \in V(H)$, all $u \in L_I(H)$, and all $I_1, I_2 \in (\partial\Phi)^{-1}(I)$ with $I_1 \sim I_2$, we have $I_1 \cdot u \sim I_2 \cdot u$. We "overload" a congruence $\sim$ by defining it also on paths (in particular, edges), by saying that $\gamma_1 \sim \gamma_2$, for $\gamma_1, \gamma_2 \in L(G)$, if $\Phi(\gamma_1) = \Phi(\gamma_2)$ and $s(\gamma_1) \sim s(\gamma_2)$. Define the *quotient graph* $G/\sim$ by $V(G/\sim) = V(G)/\sim$, $E(G/\sim) = E(G)/\sim$, $s([e]_\sim) = [s(e)]_\sim$, and $t([e]_\sim) = [t(e)]_\sim$.

Remark 4.3.1. Let G, H be graphs with $H \leq_R G$, let $\Phi \in \hom_R(G, H)$, and let $\sim$ be a congruence on G with respect to Φ. Observe that there are right-resolvers $G \to G/\sim$, $G/\sim \to H$ which compose to Φ, where the right-resolver $G \to G/\sim$ is the quotient map, and the right-resolver $G/\sim \to H$ takes a $\sim$ class to the image in H of any of its representatives.

Remark 4.3.2. The coarsest congruence, with respect to a right-resolver $\Phi \in \hom_R(G, H)$, is the total relation on the fibers, i.e. the relation $\bigsqcup_{I \in V(H)}((\partial\Phi)^{-1}(I))^2$. The quotient of G by this relation is simply H, with Φ as the quotient map.

Definition 4.3.3 (stability relation for a right-resolver)**.** Let G, H be graphs with $H \leq_R G$ and let $\Phi \in \hom_R(G, H)$. The *stability relation* for Φ, written $\sim_\Phi$, is the equivalence relation on $V(G)$ defined as follows: for $I \in V(H)$ and $I_1, I_2 \in (\partial\Phi)^{-1}(I)$, $I_1 \sim_\Phi I_2$ if and only if for all $u \in L_I(H)$, there exists $v \in L_{t(u)}(H)$ such that $I_1 \cdot uv = I_2 \cdot uv$.

Lemma 4.3.3. *Let G, H be graphs with $H \leq_R G$ and let $\Phi \in \hom_R(G, H)$. The stability relation $\sim_\Phi$ is a congruence with respect to Φ.*

Proof. Let $I \in V(H)$ and $I_1, I_2 \in (\partial\Phi)^{-1}(I)$ with $I_1 \sim_\Phi I_2$. Let $u \in L_I(H)$ and let $v \in L_{t(u)}(H)$. Since $I_1 \sim_\Phi I_2$, there exists $w \in L_{t(v)}(H)$ such that $I_1 \cdot uvw = I_2 \cdot uvw$. Therefore $I_1 \cdot u \sim_\Phi I_2 \cdot u$, so $\sim_\Phi$ is indeed a congruence. □

We now define *synchronizing* right-resolvers in terms of the stability relation, then show in Proposition 4.3.4 that, at least in the strongly connected case, this definition is equivalent to a more obvious notion of synchronization for a right-resolver.

Definition 4.3.4 (synchronizer)**.** Let G, H be graphs with $H \leq_R G$ and let $\Phi \in \hom_R(G, H)$. We say that Φ is *synchronizing*, or is a *synchronizer*, if each fiber $(\partial\Phi)^{-1}(I)$, $I \in V(H)$, is a $\sim_\Phi$ class. We denote the set of synchronizers $G \to H$ by $\hom_S(G, H)$, and we write $H \leq_S G$ if $\hom_S(G, H) \neq \emptyset$.

Note that $\sim_\Phi$ depends on Φ only through S_Φ. That is, if Φ, Φ' are such that $S_\Phi = S_{\Phi'}$, then $\sim_\Phi = \sim_{\Phi'}$.

Proposition 4.3.4. *Let G, H be graphs with $H \leq_R G$ and let $\Phi \in \hom_R(G, H)$. Then Φ is synchronizing if and only if for every $I \in V(H)$, there exists $u \in L_I(H)$ with $|(\partial\Phi)^{-1}(I) \cdot u| = 1$.*

Proof. First, suppose that Φ is synchronizing and let $I \in V(H)$. If $|(\partial\Phi)^{-1}(I)| = 1$, then we are done. Otherwise, there exist at least two distinct states $I_1, I_2 \in (\partial\Phi)^{-1}(I)$, and $I_1 \sim_\Phi I_2$ since $(\partial\Phi)^{-1}(I)$ is a $\sim_\Phi$ class by assumption. Therefore, there exists $u_1 \in L_I(H)$ with $I_1 \cdot u_1 = I_2 \cdot u_1$. In particular, $|(\partial\Phi)^{-1}(I) \cdot u_1| < |(\partial\Phi)^{-1}(I)|$. Continuing inductively, we can produce a sequence of words $u_1, \dots, u_n$ such that $t(u_i) = s(u_{i+1})$ and $|(\partial\Phi)^{-1}(I) \cdot u_1 \cdots u_n| = 1$. This proves the first claim.

For the converse, let $I \in V(H)$ and $I_1, I_2 \in (\partial\Phi)^{-1}(I)$. We will show that $I_1 \sim_\Phi I_2$. Let $v \in L_I(H)$ be arbitrary. By the assumption that each fiber can be collapsed to a single state, let $w \in L_{t(v)}(H)$ be such that $|(\partial\Phi)^{-1}(t(v)) \cdot w| = 1$. Then $|(\partial\Phi)^{-1}(I) \cdot vw| = 1$; in particular, $I_1 \cdot vw \sim_\Phi I_2 \cdot vw$. Therefore $(\partial\Phi)^{-1}(I)$ is a $\sim_\Phi$ class, and Φ is synchronizing. □

We now summarize the structure of stability, in the sense of its behaviour with respect to composition of right-resolvers. For the proof of Theorem 4.3.5, see §4.8.2.

Theorem 4.3.5. *Let G, K, H be graphs with $H \leq_R K \leq_R G$. Let $\Psi \in \hom_R(G, K)$, $\Delta \in \hom_R(K, H)$, and let $\Phi = \Delta \circ \Psi$.*

1. *The $\sim_\Psi$ classes in $V(G)$ are the intersections of $\sim_\Phi$ classes with $\partial\Psi$ fibers. In particular, Ψ is synchronizing if and only if every $\partial\Psi$ fiber is contained in a $\sim_\Phi$ class.*
2. *If $K = G/\sim_\Phi$ and Ψ is the quotient map for $\sim_\Phi$, then Ψ is synchronizing and $\sim_\Delta$ is trivial.*
3. *If $\sim_\Delta$ is trivial, then $\sim_\Phi = \sim_\Psi$.*
4. *Φ is synchronizing if and only if both Ψ and Δ are synchronizing.*

The following observation follows immediately from Theorem 4.3.5(4).

Corollary 4.3.6. *The relation $\leq_S$ is transitive, and is thus a partial order on the class of graphs (again, really isomorphism classes of graphs), refining the partial order $\leq_R$.*

Conjecture ($O(G)$ conjecture, Question 4.6 in [6]). *Let G be a strongly connected graph. Then the set of graphs H with $H \leq_S G$ has a unique $\leq_S$-minimal element $O(G)$.*

Remark 4.3.7. This remark is intended for readers interested in algebraic or categorical perspectives on automata theory. Recall that for a complete DFA, or road colouring $\Phi \in \hom_R(G, M_D)$, where G is a graph of constant out-degree D, the set S_Φ of transition maps forms a transformation semigroup under composition. Indeed, a complete DFA is essentially a finite transformation semigroup together with a choice of generators; this perspective is taken explicitly in [5, 21] and mentioned in [61], the first paper on the road problem after [2].

For a general right-resolver $\Phi \in \hom_R(G, H)$, one could see S_Φ as the semigroup of transitions of a partial finite automaton (PFA), where a given transition is defined only on a single fiber. However, it is more helpful to see S_Φ as a *semigroupoid* (equivalently, if the empty word is included, a small category). One reason is that, as we show in §4.7, it can be decided in polynomial time whether Φ is synchronizing, and the length of a word synchronizing a given fiber is bounded by a polynomial in $|V(G)|$. This is in contrast to the high level of complexity typical of related problems in subset synchronization and synchronization of PFAs [13, 75].

The reader may verify as an exercise, generalizing Cayley's theorem or specializing the Yoneda lemma, that every finite semigroupoid is isomorphic to S_Φ for some graphs G, H (although possibly with sinks) and some $\Phi \in \hom_R(G, H)$, with appropriate generalizations to the infinite case. Moreover, just as every group is a quotient of a free group, S_Φ is a quotient of the free semigroupoid $L(H)$.

4.3.2 Sufficient conditions for stability

We now give a pair of sufficient conditions for nontrivial stability, both of which are used in the proof of the road colouring theorem and one of which is also used in §4.5 to obtain a right-resolver with nontrivial stability on an almost bunchy graph.

The first condition involves a special case of the operation known as in-amalgamation ([49], §2.4):

Lemma 4.3.8. *Let G, H be graphs with $H \leq_R G$ and let $\Phi \in \hom_R(G, H)$ be a right-resolver. Let $I \in V(H)$. Suppose that there exist $I_1, I_2 \in (\partial\Phi)^{-1}(I)$ such that $|E_{I_1 J}(G)| = |E_{I_2 J}(G)|$ for all $J \in V(G)$. Then there exists $\Phi' \in \hom_R(G, H)$ such that $I_1 \sim_{\Phi'} I_2$.*

Proof. We first claim that $F(I_1) = F(I_2)$, where we recall from §4.2 the notation $F(\cdot)$ for the set of follower states of a given state. Indeed, the assumption that $|E_{I_1 J}(G)| = |E_{I_2 J}(G)|$ for all $J \in V(G)$ implies in particular that, for any $V(G)$, we have $|E_{I_1 J}(G)| > 0$, equivalently $J \in F(I_1)$, if and only $|E_{I_2 J}(G)| > 0$, equivalently $J \in F(I_2)$.

Let $F = F(I_1) = F(I_2)$. For each $J \in F$, choose a bijection $\Theta_J : E_{I_2 J}(G) \to E_{I_1 J}(G)$. Define Φ' as follows: $\partial\Phi' = \partial\Phi$, $\Phi'|_{E(G)\setminus E_{I_2}(G)} = \Phi|_{E(G)\setminus E_{I_2}(G)}$, and, for each $J \in F$, $\Phi'|_{E_{I_2 J}(G)} = \Phi|_{E_{I_1 J}(G)} \circ \Theta_J$. That is, Φ' agrees with Φ on states and on all edges with initial state other than I_2, but may disagree with Φ on the outgoing edges from I_2, specifically by permutations of parallel edges. Since $\Phi'|_{E_{I_2}(G)} : E_{I_2}(G) \to E_I(G)$ is a bijection (being a composition of bijections), Φ' is indeed right-resolving.

By the construction of Φ', for any $a \in E_I(G)$, we have $I_1 \cdot a = I_2 \cdot a$. Any $w \in L_I(H)$ is of the form $w = au$ with $a \in E_I(G)$, so $I_1 \cdot w = I_2 \cdot w$. Therefore $I_1 \sim_{\Phi'} I_2$. □

In Lemma 4.3.8, the states I_1, I_2 are said to be *in-amalgamated* by the operation $G \to G/\sim_{\Phi'}$; the inverse operation is known as *in-splitting*. The lemma shows in particular that no fiber of a $\leq_S$-minimal graph G over $M(G)$ has two states that can be in-amalgamated. Trahtman applies a special case of this fact to graphs of constant out-degree, and we follow his line of application; see the first paragraph of the proof of Theorem 4.4.3, found in §4.9.2.

The second sufficient condition, given in Proposition 4.3.12, concerns *minimal images*:

Definition 4.3.5 (minimal image)**.** Let G, H be graphs with $H \leq_R G$, and let $\Phi \in \hom_R(G, H)$. A *minimal image* is a set of the form $U = (\partial\Phi)^{-1}(I) \cdot u$ for some $I \in V(H)$ and $u \in L_I(H)$, such that $|U \cdot v| = |U|$ for any $v \in L_{t(u)}(H)$.

Remark 4.3.9. Let G, H be graphs with $H \leq_R G$, and let $\Phi \in \hom_R(G, H)$. For any $I \in V(H)$, any $u \in L_I(H)$, and any $v \in L_{t(u)}(H)$, if $U = (\partial\Phi)^{-1}(I) \cdot u$, we clearly have $|U \cdot v| \leq |U|$. If there exists $v \in L_{t(u)}(H)$ such that this inequality is strict, then $|U|$ is not minimal, i.e. U is not a minimal image. This is the reason for the term.

For a right-resolver Φ on a strongly connected graph, all minimal images have the same size, which is called the *degree* of Φ, and a word with minimal image is called a *magic word*. See [6], §9.1 for a treatment of degrees, using symbolic dynamics. In this chapter, we only need a small fragment of the theory of degree, which we establish in a self-contained way with no connectedness assumptions, using the properties of stability.

We use minimal images to give a criterion for stability that can be seen as a pairwise version of Proposition 4.3.4.

Lemma 4.3.10. *Let G, H be graphs with $H \leq_R G$, and let $\Phi \in \hom_R(G, H)$. For $I \in V(H)$ and $I_1, I_2 \in (\partial\Phi)^{-1}(I)$, we have $I_1 \sim_\Phi I_2$ if and only if $I_1 \cdot u = I_2 \cdot u$ for every word $u \in L_I(H)$ such that $(\partial\Phi)^{-1}(I) \cdot u$ is a minimal image.*

Proof. Let $I \in V(H)$ and let $I_1, I_2 \in (\partial\Phi)^{-1}(I)$. First suppose that $I_1 \sim_\Phi I_2$, and let $u \in L_I(H)$. If $I_1 \cdot u \neq I_2 \cdot u$, then let $v \in L_{t(u)}(H)$ be such that $I_1 \cdot uv = I_2 \cdot uv$. Then $|(\partial\Phi)^{-1}(I) \cdot uv| < |(\partial\Phi)^{-1}(I) \cdot u|$, so $(\partial\Phi)^{-1}(I) \cdot u$ is not a minimal image. Contrapositively, if $(\partial\Phi)^{-1}(I) \cdot u$ is a minimal image, then $I_1 \cdot u = I_2 \cdot u$.

Conversely, suppose that $I_1 \cdot u = I_2 \cdot u$ for every word $u \in L_I(H)$ such that $(\partial\Phi)^{-1}(I) \cdot u$ is a minimal image. Let $r = \min_{u \in L_I(H)} |(\partial\Phi)^{-1}(I) \cdot u|$ and let $u \in L_I(H)$. We claim that there exists $w \in L_{t(u)}(H)$ with $I_1 \cdot uw = I_2 \cdot uw$. Indeed, if $I_1 \cdot u \neq I_2 \cdot u$, then $(\partial\Phi)^{-1}(I) \cdot u$ is not a minimal image, so there exists $v \in L_{t(u)}(H)$ such that $|(\partial\Phi)^{-1}(I) \cdot uv| < |(\partial\Phi)^{-1}(I) \cdot u|$. We can thus inductively extend v to obtain the desired w, so indeed $I_1 \sim_\Phi I_2$. □

The following easy observation about minimal images is the main reason that our results toward the bunchy factor conjecture (the generalized road colouring theorem and the related result for almost bunchy graphs) require strong connectedness.

Lemma 4.3.11. *Let G, H be strongly connected graphs with $H \leq_R G$, and let $\Phi \in \hom_R(G, H)$. Every minimal image for Φ has the same cardinality, and for every $I' \in V(G)$, there exists a minimal image U with $I' \in U$.*

In the proof of the generalized road colouring theorem, we apply Lemma 4.3.11 both directly and via Proposition 4.3.12. The proof of Proposition 4.3.12 is adapted from the proof of Lemma 10.4.4 in [14].

Proposition 4.3.12. *Let G, H be strongly connected graphs with $H \leq_R G$ and let $\Phi \in \hom_R(G, H)$. Let $I, J \in V(H)$ and let $u_1, u_2 \in L_{IJ}(H)$ be such that $U_i = (\partial\Phi)^{-1}(I) \cdot u_i$ are minimal images. If $|U_1 \Delta U_2| = 2$, say $U_1 \Delta U_2 = \{J_1, J_2\}$ (where Δ denotes the symmetric difference), then $J_1 \sim_\Phi J_2$.*

Proof. Let $r = |U_1| = |U_2|$. Suppose without loss of generality that $J_i \in U_i$, and let $U_0 = U_1 \cap U_2$, so that $U_i = U_0 \cup \{J_i\}$. For any $v \in L_J(H)$, we have $(U_1 \cup U_2) \cdot v = (U_0 \cdot v) \cup (\{J_1, J_2\} \cdot v)$. We must have $|U_0 \cdot v| = |U_0| = r - 1$ and $J_i \cdot v \notin U_0 \cdot v$, since otherwise we would have $|(\partial\Phi)^{-1}(I) \cdot u_i v| = |U_i \cdot v| < r$, contradicting the minimality assumption. Therefore $r - 1 + |\{J_1, J_2\} \cdot v| = |(U_1 \cup U_2) \cdot v|$. Note that $|(U_1 \cup U_2) \cdot v| \in \{r, r+1\}$.

Let $v \in L_J(H)$ be such that $(\partial\Phi)^{-1}(J) \cdot v$ is a minimal image. By Lemma 4.3.10, to show that $J_1 \sim_\Phi J_2$, we need to show that $|\{J_1, J_2\} \cdot v| = 1$, or equivalently $|(U_1 \cup U_2) \cdot v| \leq r$. By strong connectedness and Lemma 4.3.11, we have $|(\partial\Phi)^{-1}(J) \cdot v| = r$. But $(U_1 \cup U_2) \cdot v \subseteq (\partial\Phi)^{-1}(J) \cdot v$, so indeed $|(U_1 \cup U_2) \cdot v| \leq r$. □

4.4 The $O(G)$ conjecture and the road problem

4.4.1 Generalization of the road colouring theorem

We first introduce the class of graphs involved in the theorem. A *bunch* in a graph G is a state $I \in V(G)$ with $|F(I)| = 1$. (This terminology, introduced in [74] and used also in [14], is the origin of our term *bunchy*, introduced in §4.5.) A strongly connected graph in which every state is a bunch is a *cycle of bunches*. Let M be a cycle of bunches with $V(M) = \{I_0, \ldots, I_{p-1}\}$, where $F(I_i) = \{I_{i+1}\}$ (subscripts indexing states in a cycle of length p should be read modulo p throughout), and let $D_i = |E_{I_i}(M)|$. Note that M is $\leq_R$-minimal if and only if the sequence of out-degrees $D_0, \ldots, D_{p-1}$ is not a cyclic shift of a sequence obtained by concatenating a strictly shorter sequence with itself more than once.

Let M be a $\leq_R$-minimal cycle of bunches. Let $O_{M,q}$ be the cycle of bunches in which the sequence of out-degrees consists of q cyclic repetitions of $D_0, \ldots, D_{p-1}$. Note that $O_{M,1} = M$. Observe that, for a strongly connected graph G with $M = M(G)$ a cycle of bunches, if $q = \mathrm{per}(G)/\mathrm{per}(M)$ and H is a cycle of bunches with $H \leq_S G$, then $H = O_{M,q}$. Let $O_{D,p} = O_{M_D,p}$ be the cycle of bunches of period p and constant out-degree D. Note that $O_{D,1} = M_D$. For a strongly connected, aperiodic graph G of constant out-degree D, a synchronizer $G \to M_D$ is precisely a synchronizing road colouring of G (recall Definition 4.2.2).

The *road problem*, posed in [2], was the problem of showing, in the notation and conceptual framework of this chapter, that $M_D \leq_S G$ for every strongly connected, aperiodic graph of constant out-degree D. The problem was solved by Trahtman, and the statement of the solution is known as the *road colouring theorem*:

Theorem 4.4.1 (Trahtman, [74]). *Let G be a strongly connected, aperiodic graph of constant out-degree D. Then $M_D \leq_S G$.*

Theorem 4.4.2 (Béal-Perrin [10], Budzban-Feinsilver [21]). *Let G be a strongly connected graph of constant out-degree D and period p. Then $O_{D,p} \leq_S G$.*

We prove the following generalization:

Theorem 4.4.3. *Let G be a strongly connected graph such that $M(G)$ is a cycle of bunches. Let $q = \mathrm{per}(G)/\mathrm{per}(M(G))$. Then $O_{M(G),q} \leq_S G$.*

The proof (see §4.9) follows that of Theorems 4.4.1 and 4.4.2. The strategy is to show that if G is not itself a cycle of bunches, then there exists $\Phi \in \hom_R(G, M(G))$ with $\sim_\Phi$ nontrivial, by constructing two minimal images that differ by a pair and applying Proposition 4.3.12. A very similar strategy is used to prove the bunchy factor conjecture for almost bunchy graphs (Corollary 4.5.8), the (substantial) difference being the different techniques used to obtain the requisite pair of minimal images.

4.4.2 The $O(G)$ conjecture implies the road colouring theorem

We now recall the sense in which the $O(G)$ conjecture was first understood to relate to the road problem. Although the road colouring theorem clearly implies the $O(G)$ conjecture for strongly connected, aperiodic graphs of constant out-degree, the converse implication may not be apparent. Indeed, the $O(G)$ conjecture asserts that $O(G)$ is well-defined for every strongly connected graph, but does not immediately say how to compute $O(G)$, whereas the road colouring theorem explicitly specifies the form of $O(G)$ for the graphs G to which it applies. However, the $O(G)$ conjecture does imply the road colouring theorem, via a key result from [2], for which we require a definition.

Definition 4.4.1 (higher edge graph). Let G be a graph. For $k \geq 2$, the k-th *higher edge graph* of G is the graph $G^{[k]}$ with edge set consisting of edge paths $e_1 e_2 \cdots e_{k-1} e_k$ of length k in G, and states given by $s(e_1 e_2 \cdots e_{k-1} e_k) = e_1 e_2 \cdots e_{k-1}$, $t(e_1 e_2 \cdots e_{k-1} e_k) = e_2 \cdots e_{k-1} e_k$. We define $G^{[1]} = G$.

It is a standard result ([49], Chapter 2) that $G \leq_S G^{[k]}$ for any strongly connected graph G and any $k \geq 1$. In our terminology, Adler-Goodwyn-Weiss showed the following:

Lemma 4.4.4 ([2], Lemma 4)**.** *Let G be a strongly connected, aperiodic graph of constant out-degree D. Then for all sufficiently large k, we have $M_D \leq_S G^{[k]}$.*

Together with an easy observation about partially ordered sets, the Adler-Goodwyn-Weiss result shows that the $O(G)$ conjecture implies the road colouring theorem.

Lemma 4.4.5. *Let $(\mathcal{P}, \preceq)$ be a partially ordered set such that, for any $y \in \mathcal{P}$, there exists a unique $\preceq$-minimal element $O(y) \preceq y$. If $x \preceq y$, then $O(x) = O(y)$.*

Proposition 4.4.6 ([6])**.** *Suppose that the $O(G)$ conjecture is true. Let G be a strongly connected, aperiodic graph of constant out-degree D. Then $M_D \leq_S G$.*

Proof. Let k be large enough that $M_D \leq_S G^{[k]}$, by Lemma 4.4.4. Then in fact $M_D = O(G^{[k]})$. Since $G \leq_S G^{[k]}$ as well, the result follows by Lemma 4.4.5. □

4.5 Bunchiness

In this section, we define and characterize the classes of bunchy and almost bunchy graphs, and demonstrate the importance of bunchy graphs to the structural properties of right-resolvers.

4.5.1 Bunchy and almost bunchy graphs

We recall from Theorem 4.2.4, for a graph G, the notation $\Sigma_G : V(G) \to V(M(G))$ for the unique state map among right-resolvers $G \to M(G)$.

Definition 4.5.1 (bunchy states and graphs)**.** Let G be a graph. We say that a state $I \in V(G)$ is *bunchy* if $\Sigma_G|_{F(I)} : F(I) \to F(\Sigma_G(I)) \subseteq V(M(G))$ is a bijection. We say that G is bunchy if every $I \in V(G)$ is bunchy. We say that G is *almost bunchy* if for each $I, J \in V(M)$, there exists at most one $I' \in \Sigma_G^{-1}(I)$ such that $|F(I') \cap \Sigma_G^{-1}(J)| \geq 2$.

Remark 4.5.1. The definition of almost bunchiness means that for every ordered pair of Σ_G fibers in G, say the fibers of states $I, J \in V(M)$, there is at most one state in the fiber of I that does not "look bunchy", in the sense that it has non-parallel outgoing edges into the fiber of J. In other words, an almost bunchy graph almost satisfies the conditions for bunchiness, but an exception is allowed for each ordered pair of fibers.

The following is evident but we state it explicitly for reference:

Lemma 4.5.2. *The classes of bunchy and almost bunchy graphs are closed under right-resolvers. Moreover, if G is a bunchy graph and C is a principal subgraph of G with $M(C) = M(G)$, then C is also bunchy.*

Remark 4.5.3. We briefly discuss examples of bunchy and almost bunchy graphs. The only strongly connected bunchy graphs of constant out-degree are the cycles of bunches. For a given $\leq_R$-minimal cycle of bunches M with sequence of out-degrees $D_0, \ldots, D_{p-1}$, the only strongly connected bunchy graphs G with $M(G) = M$ are the graphs $O_{M,q}$ introduced in the previous section.

A strongly connected almost bunchy (but not bunchy) graph of constant out-degree is a graph with a unique non-bunchy state, i.e. a state I with $|F(I)| \geq 2$, together with a path from each element of $F(I)$ back to I. One example that has been considered in the literature is the graph W_n studied in [4], first discussed in [77], and of interest due to its slow synchronization.

An almost bunchy graph G can have at most $|V(M(G))|^2$ non-bunchy states, one for each ordered pair of Σ_G fibers. One way to obtain an almost bunchy graph is to start with a bunchy graph and perform a sequence of in-splittings (recall Lemma 4.3.8, and see also [49], §2.4), but not all in-splittings will preserve almost bunchiness, and not all almost bunchy graphs arise this way.

Bunchy and almost bunchy graphs are characterized in terms of automorphisms, with an important uniqueness consequence for the sets of transition maps induced on them by right-resolvers:

Proposition 4.5.4. *A graph G is almost bunchy if and only if there is a unique right-resolver $G \to M(G)$ up to permutations of parallel edges: that is, if and only if, for any $\Phi_1, \Phi_2 \in \hom_R(G, M(G))$, there exist $\sigma \in P(G), \tau \in P(M(G))$ such that $\Phi_1 = \tau \circ \Phi_2 \circ \sigma$. Moreover, G is bunchy if and only if we can take $\tau = \mathrm{id}$ regardless of Φ_1, Φ_2.*

Proposition 4.5.5. *Let G be an almost bunchy graph. Let $\Phi_1, \Phi_2 \in \hom_R(G, M(G))$. Then $S_{\Phi_1} = S_{\Phi_2}$. In particular, $\sim_{\Phi_1} = \sim_{\Phi_2}$.*

For the proofs of Propositions 4.5.4 and 4.5.5, see §4.8.3. The following definition is now justified.

Definition 4.5.2. Let G be an almost bunchy graph. We denote by $\sim_G$ the unique relation on $V(G)$ with $\sim_G = \sim_\Phi$ for any $\Phi \in \hom_R(G, M(G))$.

4.5.2 Proof of the $O(G)$ conjecture in the bunchy case

We now resolve the $O(G)$ conjecture in the almost bunchy case (which includes the bunchy case). This extends Corollary 4.3 in [6], which resolves the conjecture for graphs G such that $M(G)$ has no parallel edges (so G is trivially bunchy). The proof here is quite different from the proof in the no-parallel-edges case, and yields a polynomial-time algorithm (Algorithm 4.7.2) for constructing $O(G)$.

Theorem 4.5.6. *Let G be an almost bunchy graph and let $H \leq_S G$. If H is $\leq_S$-minimal, then $H = G/\sim_G$. In particular, the set $\{K \mid K \leq_S G\}$ has a unique $\leq_S$-minimal element $O = G/\sim_G$.*

Proof. Let $\Psi \in \hom_S(G, H)$ and $\Delta \in \hom_R(H, M)$. We can naturally identify $V(H) = V(G)/\sim_\Psi$ by Lemma 4.2.3. The hypothesis that H is $\leq_S$-minimal implies that $\sim_\Delta$ is trivial, so $\sim_\Psi = \sim_{\Delta \circ \Psi}$ by Theorem 4.3.5(3). By Proposition 4.5.5, we have $\sim_{\Delta\circ\Psi} = \sim_G$. We can thus identify $V(H)$ with $V(G)/\sim_G = V(G/\sim_G)$. Thus $H = G/\sim_G$ by a second application of Lemma 4.2.3. □

In the strongly connected case, we can apply Proposition 4.3.12, which is also used in the proof of the road colouring theorem, to say more.

Proposition 4.5.7. *Let G be a strongly connected almost bunchy graph. If G is not bunchy, then $\sim_G$ is nontrivial.*

Proof. Let $I, J \in V(M(G))$ and $I' \in \Sigma_G^{-1}(I)$ with $|F(I') \cap \Sigma_G^{-1}(J)| \geq 2$. Let $J_1, J_2 \in F(I') \cap \Sigma_G^{-1}(J)$, $J_1 \neq J_2$, and let $e_i \in E_{I'J_i}(G)$. Let $\Phi \in \hom_R(G, M(G))$ and let $a_i = \Phi(e_i)$. By strong connectedness and Lemma 4.3.11, there exists a minimal image $U \subseteq \Sigma_G^{-1}(I)$ with $I' \in U$. Let $U_0 = U \setminus \{I'\}$. Then $U_0 \cdot a_1 = U_0 \cdot a_2$ since G is almost bunchy. Moreover, $J_i = I' \cdot a_i \notin U_0 \cdot a_i$ (otherwise, minimality would be contradicted). Thus $(U \cdot a_1)\Delta(U \cdot a_2) = \{J_1, J_2\}$. By Proposition 4.3.12, $J_1 \sim_\Phi J_2$. □

Corollary 4.5.8. *Let G be a strongly connected almost bunchy graph. Then $O(G)$ is bunchy.*

Proof. If $|V(G)| = 1$ then the claim is clearly true. Suppose that the conclusion is true for all almost bunchy G with $|V(G)| \leq_R N$, and let G be almost bunchy with $|V(G)| = N + 1$. If G is bunchy, then $O(G)$ is clearly bunchy by Lemma 4.5.2. If G is not bunchy, then $\sim_G$ is nontrivial by Proposition 4.5.7, so $|V(G/\sim_G)| \leq_R N$. Moreover, since $G/\sim_G \leq_S G$, it follows that $O(G) = O(G/\sim_G)$ is bunchy by the inductive hypothesis and Lemma 4.5.2. □

4.5.3 Universal property of the fiber product

We recall a standard construction known as the fiber product, and derive several new properties. Chief among these is the one exhibited in Theorem 4.5.12, which is an analogue of the universal property often enjoyed by the fiber product, or pullback, in other categories (see e.g. [65], Definition 3.1.15 and subsequent discussion).

Definition 4.5.3 (fiber product)**.** Let H_1, H_2, K be graphs and let $\Psi_i : H_i \to K$ be graph homomorphisms. The *fiber product* of Ψ_1, Ψ_2 is the graph $P = H_1 \times_{\Psi_1,\Psi_2} H_2$ where

$$V(P) = \bigsqcup_{I \in V(K)} (\partial\Psi_1)^{-1}(I) \times (\partial\Psi_2)^{-1}(I)$$
$$E(P) = \{(e_1, e_2) \,|\, e_i \in E(H_i),\ \Psi_1(e_1) = \Psi_2(e_2)\}$$

together with the coordinate projections $\hat{\Psi}_i : P \to H_i$. We write $\Psi_P = \Psi_i \circ \hat{\Psi}_i : P \to K$.

Remark 4.5.9. To see that $\Psi_1 \circ \hat{\Psi}_1 = \Psi_2 \circ \hat{\Psi}_2$ and thus that Φ_P is well-defined, note that for every $(I_1, I_2) \in V(P)$ and every $(e_1, e_2) \in E_{(I_1,I_2)}(P)$, we have, by the definition of P,

$$\Psi_1 \circ \hat{\Psi}_1(e_1, e_2) = \Psi_1(e_1) = \Psi_2(e_2) = \Psi_2 \circ \hat{\Psi}_2(e_1, e_2)$$

Remark 4.5.10. Observe that the $\hat{\Psi}_i$ are surjective (respectively, right-resolving) when the Ψ_i are surjective (respectively, right-resolving). Moreover, if C is a principal subgraph of P such that the restricted state maps $\partial\hat{\Psi}_i|_{V(C)} : V(C) \to V(H_i)$ are surjective, then $H_i \leq_R C$, indeed $\hat{\Psi}_i|_C \in \hom_R(C, H_i)$. In particular, this condition is satisfied if the H_i are strongly connected and C is a principal component of P.

Remark 4.5.11. Often the convention is taken that $V(P) = V(H_1) \times V(H_2)$. However, all of the elements of the full Cartesian product that are not elements of $V(P)$, as we have defined it, would be isolated states, and in particular would be sinks. Our definition has the feature that the fiber product of two sink-free graphs (or rather, of two right-resolvers defined on such graphs) is also sink-free.

We now state the universal property of the fiber product. Compare with a similar diagram in [6] (p. 289). See §4.8.4 for the proof.

Theorem 4.5.12. *Let H_1, H_2 be bunchy graphs with $M(H_1) = M(H_2) = M$. Let $\Psi_i \in \hom_R(H_i, M)$ be right-resolvers, and let $P = H_1 \times_{\Psi_1,\Psi_2} H_2$. Let G be a common right-resolving extension of H_1, H_2 via $\Phi_i \in \hom_R(G, H_i)$. Then there exist a principal subgraph C of P and right-resolvers $\Delta_i \in \hom_R(G, C)$ such that $\Phi_i = \hat{\Psi}_i \circ \Delta_i$ and $\partial\Delta_1 = \partial\Delta_2$. In particular, $H_i \leq_R C \leq_R G$, with $\hat{\Psi}_i|_C \in \hom_R(C, H_i)$.*

Remark 4.5.13. The bunchiness hypothesis on the H_i cannot be dropped, as the following construction illustrates. Let G be a graph and let $\Phi_1, \Phi_2 \in \mathrm{Aut}(G)$. In the notation of the the theorem, we will take $H_1 = H_2 = G$. (Recall that any automorphism is right-resolving.) Let $M = M(G)$. Let $\Psi_i \in \hom_R(G, M)$ and $P = G \times_{\Psi_1,\Psi_2} G$. Let C be a principal subgraph of P with $\hat{\Psi}_i|_V(C)$ surjective, and let $\Delta_i \in \hom_R(G, C)$ with $\Phi_i = \hat{\Psi}_i|_C \circ \Delta_i$. Then $\hat{\Psi}_i|_C$ and Δ_i are isomorphisms, since they compose to an isomorphism. In particular, since $\Psi_1 \circ \hat{\Psi}_1|_C = \Psi_2 \circ \hat{\Psi}_2|_C$, we have $\Psi_1 = \Psi_2 \circ \tau$ where $\tau = \hat{\Psi}_2|_C \circ \left(\hat{\Psi}_1|_C\right)^{-1}$ is an isomorphism. In other words, any two elements of $\hom_R(G, M)$ agree up to an automorphism of G. That condition always holds when G is bunchy (see Proposition 4.5.4), but fails in general.

We now give two applications of the universal property. The first, Proposition 4.5.15, is applied in Proposition 4.6.2, which is the main motivation for the bunchy factor conjecture. See §4.8.5 for the proof of Lemma 4.5.14.

Lemma 4.5.14. *Let H_1, H_2 be bunchy graphs with $M(H_1) = M(H_2) = M$. Let $\Psi_i \in \hom_R(H_i, M)$, and let $P = H_1 \times_{\Psi_1,\Psi_2} H_2$. Then P is bunchy. In particular, if C is a principal subgraph of P such that the restrictions $\partial\hat{\Psi}_i|_{V(C)} : V(C) \to V(H_i)$ are surjective, then C is bunchy.*

Proposition 4.5.15. *Let G be a graph. Let $H_1, H_2 \leq_S G$ be bunchy. Then $O(H_1) = O(H_2)$, i.e. G has at most one $\leq_S$-minimal bunchy synchronizing factor.*

Proof. Let $M = M(G)$, let $\Phi_i \in \hom_S(G, H_i)$, and let $\Psi_i \in \hom_R(H_i, M)$. Let $P = H_1 \times_{\Psi_1,\Psi_2} H_2$. By Theorem 4.5.12, there is a principal subgraph C of P admitting $\Delta_i \in \hom_R(G, C)$ such that $\Phi_i = \hat{\Psi}_i|_C \circ \Delta_i$. Since each Φ_i is synchronizing, each restriction $\hat{\Psi}_i|_C$ is synchronizing as well, so $H_i \leq_S C$. Since C is bunchy by Lemma 4.5.14, we know that $O(C)$ is well-defined, and thus, by Lemma 4.4.5, we have $O(H_1) = O(C) = O(H_2)$ as claimed. □

For the second application of the universal property, recall that the only strongly connected bunchy graphs of constant out-degree are the cycles of bunches. In particular, by the periodic road colouring theorem, for any strongly connected graph G of constant out-degree D and period p, the unique maximal bunchy right-resolving factor of G, namely $O_{D,p}$, is a synchronizing factor of G (and is indeed equal to $O(G)$). We now show that every graph G has a unique maximal bunchy right-resolving factor $B(G)$. The construction is similar to that of the auxiliary graph $\tilde{G}$ in [6], §5. See §4.8.5 for the proof, as well as an explicit construction of $B(G)$ yielding a polynomial-time algorithm (Algorithm 4.7.3).

Proposition 4.5.16. *Let G be a graph.*

(1) The set $\{H \leq_R G \,|\, H \text{ is bunchy}\}$ has a unique $\leq_R$-maximal element $B = B(G)$.

(2) Let $H \leq_R G$ be bunchy and $\Phi \in \hom_R(G, H)$. Then Φ factors through B, i.e. there exist $\Delta \in \hom_R(G, B)$, $\Theta \in \hom_R(B, H)$ such that $\Phi = \Theta \circ \Delta$.

4.6 The $O(G)$ conjecture and bunchy synchronizing factors

As we have seen, the $O(G)$ conjecture holds for strongly connected graphs G such that $M(G)$ is a cycle of bunches, and for almost bunchy graphs (including the bunchy graphs) by Theorem 4.5.6. Moreover, for strongly connected almost bunchy graphs, and strongly connected graphs that factor onto cycles of bunches, we know that there is a bunchy synchronizing factor, which we show inductively by assuming non-bunchiness and obtaining a right-resolver with a nontrivial stability relation. It seems plausible that, if the $O(G)$ conjecture is true, then it can be proven by a similar approach: assume non-bunchiness, find a right-resolver with nontrivial stability relation, recursively find a bunchy synchronizing factor, and apply Proposition 4.5.15. The next proposition gives several equivalent formulations of the hypothesis that this approach can be made to work. See §4.8.6 for the proof.

Proposition 4.6.1. *The following statements are equivalent.*

(1) Any strongly connected $\leq_S$-minimal graph is bunchy.

(2) For any strongly connected graph G, there exists some bunchy $H \leq_S G$.

(3) For any non-bunchy strongly connected graph G, there exists some $\Phi \in \hom_R(G, M(G))$ with $\sim_\Phi$ nontrivial.

(4) For any strongly connected graph G, $B(G) \leq_S G$.

Conjecture (bunchy factor conjecture)**.** *The assertions in Proposition 4.6.1 are true.*

Proposition 4.6.2. *The bunchy factor conjecture implies the $O(G)$ conjecture.*

Proof. Let G be a strongly connected graph and let $A = \{H \,|\, H \leq_S G,\, H \text{ is } \leq_S\text{-minimal}\}$. Clearly $|A| \geq 1$. By hypothesis, every element of A is bunchy. By Proposition 4.5.15, $|A| \leq 1$, so A has a single element, namely $O(G)$. □

Observe that the bunchy factor conjecture is a straightforward generalization of the road problem. As discussed above, the $O(G)$ conjecture was already known to imply the road colouring theorem, via the higher-edge result from [2]. By contrast, the bunchy factor conjecture implies the road colouring theorem more directly, without reference to [2].

4.7 Computing with right-resolvers

We now discuss the computational problems of constructing $O(G_1), O(G_2)$, for input graphs G_1, G_2 such that the $O(G_i)$ are known to exist, and deciding whether the $O(G_i)$ are isomorphic. Although one could apply generic graph isomorphism algorithms, which are efficient in practice (Theorem 1 in [57] gives a polynomial-time reduction from directed to undirected graph isomorphism, and see [7] for a survey of the state of the art), it is desirable to have a polynomial-time algorithm, in particular one that only uses constructions involved in the theory of right-resolvers and synchronization. We do not attempt detailed complexity analyses, noting only that all of the procedures we give can be easily seen to run in polynomial time.

4.7.1 Basic routines

In [6], a polynomial-time algorithm is given for computing $M(G)$, along with Σ_G and a total ordering of $V(M(G))$ such that any graph isomorphism $M(G) \to M(H)$ must preserve the order of the states. Deciding whether $M(G)$, $M(H)$ are isomorphic is therefore no harder than constructing them. Moreover, we can use Σ_G to construct right-resolvers, as follows.

Algorithm 4.7.1. Construct a right-resolver from a graph to its minimal right-resolving factor.

1. Input: a graph G.
2. Construct $M(G)$ and Σ_G.
3. For each $I, J \in V(M(G))$:
    1. Choose a total ordering on $E_{IJ}(M)$.
    2. For each $I' \in \Sigma_G^{-1}(I)$:
        1. Choose a total ordering on $\bigcup_{J' \in \Sigma_G^{-1}(J)} E_{I'J'}(G)$, i.e. the edges $e \in E_{I'}(G)$ with $\Sigma_G(t(e)) = J$.
        2. For each $e \in E_{I'}(G)$ with $\Sigma_G(t(e)) = J$, record as $\Phi(e)$ the edge in $E_{IJ}(M)$ with the same position in the total ordering of $E_{IJ}(M)$ as e has in $\bigcup_{J' \in \Sigma_G^{-1}(J)} E_{I'J'}(G)$.
4. Return: Φ.

There are also obvious polynomial-time procedures for constructing the fiber product of two right-resolvers, and for determining whether there is a path from one given state to another (in a graph that may not be strongly connected). With these basic routines, we can construct the stability relation of a right-resolver in polynomial time, as follows.

Algorithm 4.7.2. Construct the stability relation of a right-resolver.

1. Input: graphs G, H, and $\Phi \in \hom_R(G, H)$.
2. Construct $P = G \times_{\Phi,\Phi} G$.
3. Populate the set U of states $(I_1, I_2) \in V(P)$ with no outgoing path to the diagonal in $V(P)$.
4. Populate and output: the set $\sim_\Phi$ of states $(I_1, I_2) \in V(P)$ with no outgoing path to U.

Recall that, by definition, $\Phi \in \hom_S(G, H)$ if and only if $H = G/\sim_\Phi$, i.e. the $\partial\Phi$-fibers are precisely the $\sim_\Phi$ classes. Since this is easy to check, Algorithm 4.7.2 can be used to decide whether Φ is synchronizing.

A similar procedure can be used to construct the maximum bunchy factor $B(G)$, following the construction in Proposition 4.8.7:

Algorithm 4.7.3. Construct the maximum bunchy right-resolving factor of a graph.

1. Input: a graph G.
2. Construct $M(G)$, along with the quotient map $\Sigma_G : V(G) \to V(M)$.
3. Construct a graph H with the following data:

$$V(H) = V(G) \times V(G)$$
$$|E_{(I_1,I_2)(J_1,J_2)}(H)| = \begin{cases} 1, & \Sigma_G(I_1) = \Sigma_G(I_2), \Sigma_G(J_1) = \Sigma_G(J_2), \text{ and } J_i \in F(I_i) \\ 0, & \text{otherwise} \end{cases}$$

4. Populate the set $\approx_0$ of pairs (I_1, I_2) with a path in H from the diagonal to (I_1, I_2).
5. Construct and output: the transitive closure $\approx$ of the relation $\approx_0$.

Regarding step 4: referring to the description of $\approx_0$ in Proposition 4.8.7, a path in H from the diagonal to (I_1, I_2) corresponds to a pair of paths $\gamma, \delta \in L(G)$ witnessing $I_1 \approx_0 I_2$.

4.7.2 Decision procedures for common synchronizing factors and extensions

See §4.8.6 for the proof of the following Proposition, which collects several similar statements relating common synchronizing factors and common synchronizing extensions. These results allow us to use fiber products of two graphs to decide questions about common factors of the graphs.

Proposition 4.7.1. *Let G_1, G_2 be strongly connected graphs with $M(G_1) = M(G_2) = M$.*

1. *If G_1, G_2 have a common synchronizing factor, then there exist $\Phi_i \in \hom_R(G_i, M)$ and a principal component C of $Q = G_1 \times_{\Phi_1,\Phi_2} G_2$ such that $\hat{\Phi}_i|_C \in \hom_S(C, G_i)$.*
2. *Assume the $O(G)$ conjecture. Then the converse holds in (1). That is, suppose that there exist $\Phi_i \in \hom_R(G_i, M)$ and a principal component C of $Q = G_1 \times_{\Phi_1,\Phi_2} G_2$, such that $\hat{\Phi}_i|_C \in \hom_S(C, G_i)$. Then G_1, G_2 have a common synchronizing factor, specifically $O(G_1) = O(C) = O(G_2)$.*
3. *If the G_i are bunchy, then the equivalence described in (1) and (2) holds unconditionally (i.e. without assuming any unproven conjectures).*
4. *Assume the bunchy factor conjecture. Then we have $O(G_1) = O(G_2)$ if and only if, for the essentially unique $\Phi_i \in \hom_R(B(G_i), M)$, there is a principal component C of $P = B(G_1) \times_{\Phi_1,\Phi_2} B(G_2)$ such that $\hat{\Phi}_i|_C \in \hom_S(C, B(G_i))$.*

Ashley-Marcus-Tuncel give a polynomial-time algorithm for deciding whether two strongly connected graphs have a common strongly connected synchronizing extension. Their algorithm relies on the construction of a graph that they call $\tilde{G}$, from an input graph G; see Theorem 5.2 and Remark 5.10 in [6]. If the $O(G)$ conjecture is true, then by Proposition 4.7.1(2), the Ashley-Marcus-Tuncel algorithm also decides whether $O(G_1) = O(G_2)$ for G_1, G_2 strongly connected. Without assuming any unproven conjectures, a negative result from the algorithm shows, by Proposition 4.7.1(1), that G_1, G_2 have no common synchronizing factor, while a positive result is inconclusive.

For bunchy G_1, G_2, however, Theorem 4.5.6 and Proposition 4.7.1(3) show that it can be decided in polynomial time whether $O(G_1), O(G_2)$ are isomorphic without the Ashley-Marcus-Tuncel algorithm:

Algorithm 4.7.4. Decide whether $O(G_1), O(G_2)$ are isomorphic for G_1, G_2 strongly connected and bunchy.

1. Input: strongly connected bunchy graphs G_1, G_2 such that $M(G_1) = M(G_2) = M$.
2. Choose $\Phi_i \in \hom_R(G_i, M)$, using Algorithm 4.7.1.
3. Construct $Q = G_1 \times_{\Phi_1,\Phi_2} G_2$.
4. For each principal component C, decide whether $\hat{\Phi}_i|_C \in \hom_S(C, G_i)$, using Algorithm 4.7.2. If so, halt, and return the result that $O(G_1) = O(G_2)$.
5. If no affirmative result is returned in step 4, then halt and return the result that $O(G_1) \neq O(G_2)$.

Regarding steps 4–5: Proposition 4.7.1(3) shows that we have $O(G_1) = O(G_2)$ if and only if $\hat{\Phi}_i|_C \in \hom_S(C, G_i)$ for some principal component C of Q. The same algorithm would work even if G_1, G_2 are not strongly connected but only bunchy, with "principal component C" replaced by "principal subgraph C such that each $\partial\hat{\Phi}_i|_{V(C)}$ is surjective", but it is not clear that the number of such subgraphs is bounded by a polynomial in $|V(G)|$.

Furthermore, if the bunchy factor conjecture is true, then isomorphism of $O(G_1), O(G_2)$ is equivalent, by Proposition 4.6.1, to isomorphism of $O(B(G_1)), O(B(G_2))$, which can be decided without the use of the Ashley-Marcus-Tuncel $\tilde{G}$ algorithm, using the following procedure. Note that this works even if we do not have an efficient way to find an element of $\hom_S(G_i, B(G_i))$.

Algorithm 4.7.5. Decide isomorphism of $O(G_1), O(G_2)$ (assuming the bunchy factor conjecture).

1. Input: strongly connected graphs G_1, G_2.
2. Construct $B(G_1)$ and $B(G_2)$, using Algorithm 4.7.3.
3. Decide whether $O(B(G_1)) = O(B(G_2))$, using Algorithm 4.7.4.
4. Output: the Boolean value of "$O(B(G_1)) = O(B(G_2))$".

Observe that the procedure of Algorithm 4.7.5 is essentially what is described in Proposition 4.7.1(4).

4.8 Proofs of structural results and additional details

4.8.1 Remarks on the proof of Theorem 4.2.4

In this subsection, we revisit the first proof of Theorem 4.2.4 given in [6]. The proof of the uniqueness of $M(G)$, for a given graph G, is without issue, but the proof of the uniqueness of Σ_G is not quite complete. That proof seems to assume that, for graphs G, H with $H \leq_R G$, two right-resolvers $G \to H$ with distinct state maps must partition $V(G)$ differently. In general, this is false: per Lemma 4.8.3 below, counterexamples arise precisely when $\mathrm{Aut}(H)$ acts nontrivially on $V(H)$.

The second proof given in [6], constructing $V(M(G))$ by successive refinements of an initial state partition, can be made complete by observing that all of the state maps corresponding to the successive refinements are invariant under $\mathrm{Aut}(G)$.

Proposition 4.8.1 (Lemma 3.1 in [6])**.** *Let G, H_1, H_2 be graphs with $H_i \leq_R G$ via $\Psi_i \in \hom_R(G, H_i)$, with partitions $\alpha_i = V(H_i)$ of $V(G)$. Then there is a graph $K \leq_R H_1, H_2$, with $V(K)$ equal to the finest common coarsening of α_1, α_2.*

Corollary 4.8.2 (Theorem 3.2 in [6])**.** *Let G be a graph. Then there exists a unique $\leq_R$-minimal graph $M(G) \leq_R G$, with $V(M(G))$ given by the finest common coarsening of the partitions $\alpha = V(H)$ of $V(G)$, where $H \leq_R G$.*

We will show further that for any $\Phi_1, \Phi_2 \in \hom_R(G, M(G))$, we have $\partial\Phi_1 = \partial\Phi_2$. Note that $\partial\Phi_1, \partial\Phi_2$ at least have the same sets of fibers, since, if they did not, then by Proposition 4.8.1, we could take their finest common coarsening, contradicting the minimality of $M(G)$. It is therefore enough to show that $\mathrm{Aut}(M(G))$ acts trivially on $M(G)$. For this, we need a lemma.

Lemma 4.8.3. *Let G, H be graphs with $H \leq_R G$, and let $\Phi_1, \Phi_2 \in \hom_R(G, H)$. Suppose that the $\partial\Phi_i$ have the same fibers, i.e. for any $I_1, I_2 \in V(G)$, we have $\partial\Phi_1(I_1) = \partial\Phi_1(I_2)$ if and only if $\partial\Phi_2(I_1) = \partial\Phi_2(I_2)$. Then there exists an automorphism $\tau \in \mathrm{Aut}(H)$ such that $\partial\Phi_2 = \partial(\tau \circ \Phi_1)$.*

Proof. By the assumption of equal fibers, there exists a (unique) bijection $T : V(H) \to V(H)$ with $\partial\Phi_2 = T \circ \partial\Phi_1$. We need to find $\tau \in \mathrm{Aut}(H)$ with $\partial\tau = T$. Let $I \in V(H)$. For $J \in F(I)$, note, by the right-resolving hypothesis on the Φ_i and the choice of T, that $|E_{IJ}(H)| = |E_{T(I)T(J)}(H)|$. Letting $\tau|_{E_{IJ}(H)} : E_{IJ}(H) \to E_{T(I)T(J)}(H)$ be any bijection, we have $\tau \in \mathrm{Aut}(H)$ with $T = \partial\tau$ and $\partial\Phi_2 = \partial(\tau \circ \Phi_1)$. □

We now discuss the quotient of a graph by its automorphism group. Let G be a graph. We will construct a graph $K = G/\mathrm{Aut}(G) \leq_R G$ as follows. Let $V(K)$ consist of the orbits in $V(G)$ under $\mathrm{Aut}(G)$. Let $I, J \in V(K)$ and $I_1, I_2 \in I$. We need to specify $|E_{IJ}(K)|$. We claim that $\sum_{J' \in J} |E_{I_1 J'}(G)| = \sum_{J' \in J} |E_{I_2 J'}(G)|$. Indeed, let $\tau \in \mathrm{Aut}(G)$ such that $I_2 = \partial\tau(I_1)$. Then $\partial\tau(F(I_1)) = \partial\tau(F(I_2))$, and $\tau(J) = J$, so for any $J' \in J \cap F(I_1)$, we have $\partial\tau(J') \in J \cap F(I_2)$. Therefore $\sum_{J' \in J} |E_{I_1 J'}(G)| \leq_R \sum_{J' \in J} |E_{I_2 J'}(G)|$. Replacing τ with τ^{-1}, we obtain equality. Define K by specifying that $|E_{IJ}(K)| = \sum_{J' \in J} |E_{I_i J'}(G)|$. This edge count ensures that $K \leq_R G$.

In particular, if $\mathrm{Aut}(G)$ acts nontrivially on $V(G)$, then $K \neq G$, so G is not $\leq_R$-minimal. It follows that, for a $\leq_R$-minimal graph M, $\mathrm{Aut}(M)$ acts trivially on $V(M)$. This observation, together with Lemma 4.8.3, shows that for any $\Phi_1, \Phi_2 \in \hom_R(G, M(G))$, we have $\partial\Phi_1 = \partial\Phi_2$.

4.8.2 Proof of Theorem 4.3.5

We prove Theorem 4.3.5 in several stages. First, in Lemma 4.8.4, we determine how stability classes intersect with fibers in compositions of right-resolvers.

Lemma 4.8.4. *Let G, K, H be graphs with $H \leq_R K \leq_R G$. Let $\Psi \in \hom_R(G, K)$, $\Delta \in \hom_R(K, H)$, and let $\Phi = \Delta \circ \Psi$. Then $\sim_\Psi$-classes are intersections of $\sim_\Phi$-classes with $\partial\Psi$-fibers. That is, for $I_1, I_2 \in V(G)$, if $I_1 \sim_\Psi I_2$, then $I_1 \sim_\Phi I_2$; conversely, if $I_1 \sim_\Phi I_2$ and moreover $\partial\Psi(I_1) = \partial\Psi(I_2)$, then $I_1 \sim_\Psi I_2$.*

Proof. First suppose that $I_1 \sim_\Psi I_2$. Then in particular $\partial\Psi(I_1) = \partial\Psi(I_2) = \hat{I}$. Let $I = \partial\Delta(\hat{I})$ and $u \in L_I(H)$, and consider the unique $\lambda \in L_{\hat{I}}(K)$ such that $\Delta(\lambda) = u$. Since $I_1 \sim_\Psi I_2$, there exists $\mu \in L_{t(\lambda)}(K)$ such that $I_1 \cdot \lambda\mu = I_2 \cdot \lambda\mu$. Letting $v = \Delta(\mu)$, we have $I_1 \cdot uv = I_2 \cdot uv$, so $I_1 \sim_\Phi I_2$.

For the converse, suppose that $I_1 \sim_\Phi I_2$ and $\partial\Psi(I_1) = \partial\Psi(I_2) = \hat{I}$. Let $I = \Delta(\hat{I})$, let $\lambda \in L_{\hat{I}}(K)$, let $u = \Delta(\lambda)$, and let $v \in L_{t(u)}(H)$ be such that $I_1 \cdot uv = I_2 \cdot uv$. Consider the unique $\mu \in L_{t(\lambda)}(K)$ such that $\Delta(\mu) = v$. Then $I_1 \cdot \lambda\mu = I_2 \cdot \lambda\mu$, so indeed $I_1 \sim_\Psi I_2$. □

Corollary 4.8.5 follows immediately from Lemma 4.8.4, and together they comprise Theorem 4.3.5(1).

Corollary 4.8.5. *If $\Phi = \Delta \circ \Psi$ is a composition of right-resolvers, then Ψ is synchronizing if and only if every $\partial\Psi$-fiber is contained in a $\sim_\Phi$-class.*

Proof of Theorem 4.3.5(2). By assumption, the $\sim_\Phi$ classes are precisely the $\partial\Psi$-fibers, so by Lemma 4.8.4, they are also the $\sim_\Psi$ classes. Thus indeed $\Psi \in \hom_S(G, K)$.

Moreover, let $I \in V(H)$, let $I_1', I_2' \in (\partial\Delta)^{-1}(I)$, and suppose that $I_1' \sim_\Delta I_2'$. We claim that $I_1' = I_2'$. Toward this end, we claim that $(\partial\Psi)^{-1}(\{I_1', I_2'\})$ is a subset of a $\sim_\Phi$ class. Indeed, let $u \in L_I(H)$, and let $I_i \in (\partial\Psi)^{-1}(I_i')$. Let $v \in L_{t(u)}(H)$ be such that $I_1' \cdot uv = I_2' \cdot uv = J'$. Let $J_i = I_i \cdot uv$. Then $J_1, J_2 \in (\partial\Psi)^{-1}(J')$. Since Ψ is synchronizing, there exists $\gamma \in L_{J'}(K)$ with $J_1 \cdot \gamma = J_2 \cdot \gamma$. Let $w = \Delta(\gamma)$. Then $I_1 \cdot uvw = I_2 \cdot uvw$. Since $u \in L_I(H)$ was arbitrary, it follows that $I_1 \sim_\Phi I_2$ as claimed. Since, by assumption, a $\sim_\Phi$ class is precisely a $\partial\Psi$-fiber of a single state of K, we must have $I_1' = I_2'$ as claimed. Therefore $\sim_\Delta$ is indeed trivial. $\square$

Lemma 4.8.6. *Let G, K, H be graphs with $H \leq_R K \leq_R G$. Let $\Psi \in \hom_R(G, K)$, $\Delta \in \hom_R(K, H)$, and let $\Phi = \Delta \circ \Psi$. Let $I \in V(H)$, let $I_1', I_1' \in (\partial\Delta)^{-1}(I)$, and let $I_i \in (\partial\Psi)^{-1}(I_i')$. If $I_1 \sim_\Phi I_2$, then $I_1' \sim_\Delta I_2'$.*

Proof. Suppose that $I_1 \sim_\Phi I_2$. Let $u \in L_I(H)$. Let $v \in L_{t(u)}(H)$ be such that $I_1 \cdot uv = I_2 \cdot uv$. Then, since $I_i' \cdot uv = \partial\Phi(I_i \cdot uv)$, we have $I_1' \cdot uv = I_2' \cdot uv$. Therefore $I_1' \sim_\Delta I_2'$. $\square$

Proof of Theorem 4.3.5(3). Suppose that $\sim_\Delta$ is trivial. Let $I \in V(H)$, let $I_1', I_2' \in (\partial\Delta)^{-1}(I)$ with $I_1' \neq I_2'$, and let $I_i \in (\partial\Psi)^{-1}(I_i')$. Since $\sim_\Delta$ is trivial, there exists $u \in L_I(H)$ such that, for every $v \in L_{t(u)}(H)$, we have $I_1' \cdot uv \neq I_2' \cdot uv$. This implies that $I_1 \cdot uv \neq I_2 \cdot uv$, so $I_1 \not\sim_\Phi I_2$. It follows that each $\sim_\Phi$ class is contained inside a $\partial\Psi$ fiber. By Lemma 4.8.4, this shows that $\sim_\Phi = \sim_\Psi$. $\square$

The final part of Theorem 4.3.5 can be proved in the strongly connected case using symbolic dynamics, via the multiplicativity of degree under composition of right-resolvers; see [49], §9.1. Using the theory developed so far, we give a self-contained proof without the assumption of strong connectedness.

Proof of Theorem 4.3.5(4). Suppose that Ψ, Δ are synchronizing. Let $I \in V(H)$. Let $I_1, I_2 \in (\partial\Phi)^{-1}(I)$. We need to show that $I_1 \sim_\Phi I_2$. Let $u \in L_I(K)$. We need to find $v \in L_{t(u)}(H)$ such that $I_1 \cdot uv = I_2 \cdot uv$. Let $I_i' = \partial\Psi(I_i)$. Then $\partial\Delta(I_1') = \partial\Delta(I_2')$, so $I_1' \sim_\Delta I_2'$ since Δ is synchronizing. Let $v_1 \in L_{t(u)}(H)$ be such that $I_1' \cdot uv_1 = I_2' \cdot uv_1$. Note that $I_i' \cdot uv_1 = \partial\Psi(I_i \cdot uv_1)$. Let $J' = I_i' \cdot uv_1$ and $J_i = I_i \cdot uv_1$. Then $\partial\Psi(J_1) = J' = \partial\Psi(J_2)$, so $J_1 \sim_\Psi J_2$ since Ψ is synchronizing. Let $\gamma \in L_{J'}(K)$ be such that $J_1 \cdot \gamma = J_2 \cdot \gamma$. Let $v_2 = \Delta(\gamma)$. Then $I_i \cdot uv_1v_2 = J_i \cdot \gamma$, so taking $v = v_1v_2$, we have $I_1 \cdot uv = I_2 \cdot uv$. Thus $I_1 \sim_\Phi I_2$, so Φ is indeed synchronizing.

We prove the converse by the contrapositive. Suppose that Δ is not synchronizing. Then there exist $I \in V(H)$ and $I_1', I_2' \in (\partial\Delta)^{-1}(I)$ such that $I_1' \not\sim_\Delta I_2'$. By Lemma 4.8.6, there exist $I_i \in (\partial\Psi)^{-1}(I_i')$ such that $I_1 \not\sim_\Phi I_2$. Thus Φ is not synchronizing. Similarly, suppose that Ψ is not synchronizing. Then there exist $I' \in V(K)$ and $I_1, I_2 \in (\partial\Psi)^{-1}(I')$ such that $I_1 \not\sim_\Psi I_2$. By Lemma 4.8.4, $I_1 \not\sim_\Phi I_2$, so Φ is not synchronizing. $\square$

4.8.3 Proofs of Propositions 4.5.4 and 4.5.5

Proof of Proposition 4.5.4. Let $M = M(G)$. First, suppose that G is almost bunchy. Let $\Phi_1, \Phi_2 \in \hom_R(G, M)$. For each $I, J \in V(M)$, and each $I' \in \Sigma_G^{-1}(I)$, let $A_{I',J} = \{e \in E_{I'}(G) \mid \Sigma_G(t(e)) = J\}$. If there exists $I' \in \Sigma_G^{-1}(I)$ with $|F(I') \cap \Sigma_G^{-1}(J)| \geq 2$ (by almost bunchiness, there is at most one such I' for any given I, J), then let $\tau_{IJ} \in P(M)$ be the permutation of parallel edges in M given by

$$\begin{aligned}\tau_{IJ}|_{E(M)\setminus E_{IJ}(M)} &= \mathrm{id}|_{E(M)\setminus E_{IJ}(M)}\\ \tau_{IJ}|_{E_{IJ}(M)} &= \Phi_1|_{A_{I',J}} \circ (\Phi_2|_{A_{I',J}})^{-1}\end{aligned}$$

If there is no $I' \in \Sigma_G^{-1}(I)$ with $|F(I') \cap \Sigma_G^{-1}(J)| \geq 2$, then let $\tau_{IJ} = \mathrm{id}$. Distinct τ_{IJ} are permutations of disjoint sets and therefore commute. Let $\tau = \prod_{I,J \in V(M)} \tau_{IJ}$. Note that if G is bunchy, then $\tau_{IJ} = \mathrm{id}$ for all I, J, so $\tau = \mathrm{id}$.

Now, for each $I, J \in V(M)$ and each $I' \in \Sigma_G^{-1}(I)$, let $\sigma_{IJ,I'} \in P(G)$ be given by

$$\begin{aligned}\sigma_{IJ,I'}|_{E(G)\setminus A_{I',J}} &= \mathrm{id}|_{E(G)\setminus A_{I',J}}\\ \sigma_{IJ,I'}|_{A_{I',J}} &= \Phi_2|_{A_{I',J}}^{-1} \circ \tau^{-1} \circ \Phi_1|_{A_{I',J}}\end{aligned}$$

All of the $\sigma_{IJ,I'}$ commute. Let $\sigma = \prod_{I,J \in V(M),\, I' \in \Sigma_G^{-1}(I)} \sigma_{IJ,I'}$. Then for $I \in V(M)$, $J \in F(I)$, and $I' \in \Sigma_G^{-1}(I)$,

$$\tau \circ \Phi_2 \circ \sigma|_{A_{I',J}} = \tau \circ \Phi_2 \circ (\Phi_2|_{A_{I',J}}^{-1} \circ \tau^{-1} \circ \Phi_1|_{A_{I',J}}) = \Phi_1|_{A_{I',J}}$$

This concludes the proof in the “only if” direction.

For the “if” direction, which we prove in the contrapositive, suppose that G is not almost bunchy. Let $I, J \in V(M)$ and $I_1, I_2 \in \Sigma_G^{-1}(I)$ such that $|F(I_i) \cap \Sigma_G^{-1}(J)| \geq 2$ for $i = 1, 2$. Let $e_{i,1}, e_{i,2} \in A_{I_i,J}$ be such that $t(e_{i,1}) \neq t(e_{i,2})$. Let $a_1, a_2 \in E_{IJ}(M)$, and let $\Phi_1, \Phi_2 \in \hom_R(G, M)$ be such that $\Phi_1(e_{1,j}) = a_j$ but $\Phi_2(e_{2,1}) = a_2$ and $\Phi_2(e_{2,2}) = a_1$. (The behaviour of Φ_i on $E(G) \setminus \{e_{i,j}\}_{i,j=1,2}$ is irrelevant.) Then there do not exist $\sigma \in P(G)$, $\tau \in P(M)$ such that $\Phi_1 = \tau \circ \Phi_2 \circ \sigma$.

Finally, suppose that G is almost bunchy, but not bunchy. Let $I, J \in V(M)$ and $I' \in \Sigma_G^{-1}(I)$ such that $|F(I') \cap \Sigma_G^{-1}(J)| \geq 2$. Let $e_1, e_2 \in A_{I',J}$ be such that $t(e_1) \neq t(e_2)$. Let $a_1, a_2 \in E_{IJ}(M)$, and let $\Phi_1, \Phi_2 \in \hom_R(G, M)$ be such that $\Phi_1(e_i) = a_i$, $\Phi_2(e_1) = a_2$, and $\Phi_2(e_2) = a_1$. Then there does not exist $\sigma \in P(G)$ such that $\Phi_1 = \Phi_2 \circ \sigma$. □

Proof of Proposition 4.5.5. Let $M = M(G)$. It is enough to show that S_{Φ_1}, S_{Φ_2} share a generating set T, which we now construct and examine. For $I, J \in V(M)$, if $J \notin F(I)$, let $T_{I,J,1} = T_{I,J,2} = \emptyset$. If $J \in F(I)$, then let $T_{I,J,i}$ be the set of maps $f_{a,i} : \Sigma_G^{-1}(I) \to \Sigma_G^{-1}(J)$ of the form $I' \mapsto t \circ (\Phi_i|_{E_{I'}})^{-1}(a)$, i.e. $I' \cdot f_{a,i} = I' \cdot a$ with respect to Φ_i, where $a \in E_{IJ}(M)$. Clearly $T_{I,J,i}$ generates S_{Φ_i}, in the sense that S_{Φ_i} is the smallest collection of maps closed under composition and containing the $T_{I,J,i}$ as I, J range over $V(M)$.

We claim that $T_{I,J,1} = T_{I,J,2}$. Indeed, let $a \in E_{IJ}(M)$ and let $I' \in \Sigma_G^{-1}(I)$ such that $|F(I') \cap \Sigma_G^{-1}(J)| \geq 2$. It is enough to show that $f_{a,1} \in T_{I,J,2}$, as this will show that $T_{I,J,1} \subseteq T_{I,J,2}$, from which equality follows by symmetry. By almost bunchiness, there is at most one state $I' \in \Sigma_G^{-1}(I)$ with $|F(I') \cap \Sigma_G^{-1}(J)| \geq 2$. If there is no such state, then clearly $f_{a,1} = f_{a,2}$, so assume that such a state I' exists. Observe that $I' \cdot f_{b,2} = I' \cdot f_{a,1}$ where $b = \Phi_2 \circ (\Phi_1|_{E_{I'(G)}})^{-1}(a)$. Moreover, for every $I'' \in \Sigma_G^{-1}(I)$ with $I'' \neq I'$, we also have, by almost bunchiness, that $I'' \cdot f_{b,2} = I'' \cdot f_{a,1}$. Therefore $f_{a,1} = f_{b,2} \in T_{I,J,2}$, so indeed $T_{I,J,1} = T_{I,J,2}$ as claimed. Let $T_{I,J} = T_{I,J,i}$ and let $T = \bigcup_{I,J} T_{I,J}$. Then T generates both S_{Φ_1}, S_{Φ_2}, so indeed $S_{\Phi_1} = S_{\Phi_2}$. □

4.8.4 Proof of Theorem 4.5.12

Proof of Theorem 4.5.12. Define $T : V(G) \to V(P)$ as follows: for $I' \in V(G)$, let $T(I') = (\partial\Phi_1(I'), \partial\Phi_2(I'))$. Note that $\partial\Psi_P \circ T = \partial(\Psi_i \circ \Phi_i) = \Sigma_G$ for each i. Let C be the subgraph of P induced by $T(V(G))$. Let $I' \in V(G)$ and $J' \in F(I')$; $I = \Sigma_G(I')$ and $J = \Sigma_G(J')$; and $I_i = \partial\Phi_i(I')$ and $J_i = \partial\Phi_i(J')$. Observe that $(I_1, I_2) = T(I')$ and $(J_1, J_2) = T(J')$.

As in the proof of Proposition 4.5.4, let $A_{I',J} = \{e \in E_{I'}(G) \,|\, \Sigma_G(t(e)) = J\}$. Since H_i is bunchy, we have $F(I_i) \cap \Sigma_{H_i}^{-1}(J) = \{J_i\}$, so $\Phi_i(A_{I',J}) = E_{I_iJ_i}(H_i)$. It therefore makes sense to define

$$\Delta_i|_{A_{I',J}} = (\hat{\Psi}_i|_{E_{(I_1,I_2)(J_1,J_2)}(P)})^{-1} \circ \Phi_i|_{A_{I',J}} : A_{I',J} \to E_{(I_1,I_2)(J_1,J_2)}(P)$$

Gluing these together, we obtain maps $\Delta_i : E(G) \to E(P)$. For $e \in A_{I',J}$, we have $s(\Delta_i(e)) = T(s(e))$ and $t(\Delta_i(e)) = T(t(e))$, so the $\Delta_i : G \to P$ are graph homomorphisms with $\partial\Delta_i = T$.

We now claim that $\Delta_i(E(G)) = E(C)$. Enumerate $E_{IJ}(M) = \{a^{(1)}, \ldots, a^{(k)}\}$, and let $e_i^{(j)} = (\Psi_i|_{E_{I_iJ_i}(H_i)})^{-1}(a^{(j)})$. Then

$$(e_1^{(j)}, e_2^{(j)}) \in E_{(I_1,I_2)(J_1,J_2)}(P) = E_{(I_1,I_2)(J_1,J_2)}(C)$$

where equality holds because $(I_1, I_2), (J_1, J_2) \in V(C)$ and C is an induced subgraph of P. Fixing (I_1, I_2) and varying J', thus varying (J_1, J_2), the sets $E_{(I_1,I_2)(J_1,J_2)}(P)$ exhaust $E_{(I_1,I_2)}(P)$, so $E_{(I_1,I_2)}(C) = E_{(I_1,I_2)}(P)$. Since $\Delta_i|_{E_{I'}(G)} : E_{I'}(G) \to E_{(I_1,I_2)}(P)$ is surjective, we have $E_{(I_1,I_2)}(C) = \Delta_i(E_{I'}(G))$. Thus indeed $E(C) = \Delta_i(E(G))$.

Since we already know that $E_{(I_1,I_2)}(C) = E_{(I_1,I_2)}(P)$ for each $(I_1, I_2) \in V(C)$, it follows that C is indeed a principal subgraph of P. Finally, we have

$$\hat{\Psi}_i \circ \Delta_i|_{A_{I',J}} = \hat{\Psi}_i \circ (\hat{\Psi}_i|_{E_{(I_1,I_2)(J_1,J_2)}(P)})^{-1} \circ \Phi_i|_{A_{I',J}} = \Phi_i|_{A_{I',J}}$$

as claimed. This shows that $\hat{\Psi}_i|_C : C \to H_i$ is surjective, so $H_i \leq_R C$. □

4.8.5 Proofs of Lemma 4.5.14 and Proposition 4.5.16, and construction of $B(G)$

Proof of Lemma 4.5.14. Let $(I_1, I_2) \in V(P)$ with $I_i \in V(H_i)$. Let $e = (e_1, e_2), e' = (e'_1, e'_2) \in E_{(I_1,I_2)}(P)$ such that $t(\partial\Psi_P(e)) = t(\partial\Psi_P(e'))$. We must show that $t(e) = t(e')$. Toward that goal, note that for each $i = 1, 2$, we have

$$t(\partial\Psi_P(e)) = \partial\Psi_P(t(e)) = \partial\Psi_i(t(e_i))$$
$$t(\partial\Psi_P(e')) = \partial\Psi_P(t(e')) = \partial\Psi_i(t(e'_i))$$

Therefore $\partial\Psi_i(t(e_i)) = \partial\Psi_i(t(e'_i))$ for each i, since $t(\partial\Psi_P(e)) = t(\partial\Psi_P(e'))$ by hypothesis. But since the H_i are bunchy, we in fact have $t(e_i) = t(e'_i)$ for each i, so indeed $t(e) = t(e')$ as required. This shows that $\partial\Psi_P|_{F((I_1,I_2))}$ is a bijection, so P is bunchy.

Let C be a principal subgraph of P such that $\partial\hat{\Psi}_i|_{V(C)} : V(C) \to V(H_i)$ are surjective. Then $M(C) = M$. For each $J \in V(M)$, we have $\Sigma_C^{-1}(J) = \Sigma_P^{-1}(J) \cap V(C)$. Thus C is bunchy. □

Proof of Proposition 4.5.16. To prove (1), let $H_1, H_2 \leq_R G$ be $\leq_R$-maximal among the bunchy right-resolving factors of G. Let $\Phi_i \in \hom_R(G, H_i)$ and $\Psi_i \in \hom_R(H_i, M)$. Let $P = H_1 \times_{\Psi_1,\Psi_2} H_2$. By Theorem 4.5.12, there exist a principal subgraph C of P, and $\Delta_i \in \hom_R(G, C)$, such that $\Phi_i = \hat{\Psi}_i|_C \circ \Delta_i$, so $H_i \leq_R C$. By Lemma 4.5.14, C is bunchy, so by the maximality of the H_i, we have $H_1 = C = H_2$. This proves uniqueness, so we can take $B = H_i = C$.

To prove (2), let $H \leq_R G$ be bunchy and let $\Phi \in \hom_R(G, H)$. Let $\Phi' \in \hom(G, B)$, $\Psi \in \hom_R(H, M)$, $\Psi' \in \hom(B, M)$. Let $P = H \times_{\Phi,\Phi'} B$. By the universal property (Theorem 4.5.12), there exist a principal subgraph C of P and $\Delta, \Delta' \in \hom_R(G, C)$ with $\Phi = \hat{\Psi} \circ \Delta$ and $\Phi' = \hat{\Psi}' \circ \Delta'$. Again by Lemma 4.5.14, C is bunchy, so by (1) and the fact that $B \leq_R C$, we have $B = C$. This proves (2) with $\Theta = \hat{\Psi}'|_C$. □

We now present the construction of $B(G)$ described in Algorithm 4.7.3.

Proposition 4.8.7. *Let G be a graph. For $J_1, J_2 \in V(G)$ with $\Sigma_G(J_1) = \Sigma_G(J_2)$, write $J_1 \approx_0 J_2$ if there exist paths $\gamma = \gamma_1 \cdots \gamma_n, \delta = \delta_1 \cdots \delta_n \in L(G)$ where $\gamma_i, \delta_i \in E(G)$, such that $s(\gamma_1) = s(\delta_1)$, $t(\gamma_n) = J_1$, $t(\delta_n) = J_2$, and $\Sigma_G(t(\gamma_i)) = \Sigma_G(t(\delta_i))$ for each i. Let $\approx$ denote the transitive closure of $\approx_0$ and let $\Phi \in \hom_R(G, M(G))$. Then $\approx$ is a congruence with respect to Φ, and $B(G) = G/\approx$.*

Proof. Let $M = M(G)$. Let $I \in V(M)$, $I_1, I_2 \in \Sigma_G^{-1}(I)$, and $J \in F(I)$. Suppose that $I_1 \approx I_2$. Let $J_i \in F(I_i) \cap \Sigma_G^{-1}(J)$. We need to show that $J_1 \approx J_2$. Let $I_1 = I^{(0)}, I^{(1)}, \ldots, I^{(n)} = I_2 \in \Sigma_G^{-1}(I)$ with $I^{(j)} \approx_0 I^{(j+1)}$. Let $\gamma^{(j)}, \delta^{(j)} \in L(G)$ witness the relation $I^{(j)} \approx_0 I^{(j+1)}$. Choose $J^{(j)} \in F(I^{(j)}) \cap \Sigma_G^{-1}(J)$, with $J^{(0)} = J_1$ and $J^{(n)} = J_2$. Let $e^{(j)} \in E_{I^{(j)}J^{(j)}}(G)$. Then $\gamma^{(j)}e^{(j)}, \delta^{(j)}e^{(j+1)}$ witness $J^{(j)} \approx_0 J^{(j+1)}$. This shows that $J_1 \approx J_2$.

Let $\Phi \in \hom_R(G, M(G))$. As in the previous paragraph, let $I \in V(M)$, let $J \in F(I)$, and let $I_1, I_2 \in \Sigma_G^{-1}(I)$. Suppose that $I_1 \approx I_2$. Let $a \in E_{IJ}(M)$, let $J_i = I_i \cdot a$, and let $e_i = (\Phi|_{E_{I_i}(G)})^{-1}(a)$. Then $t(e_i) = J_i$, so $J_1 \approx J_2$. Therefore $\approx$ is indeed a congruence for Φ. It follows that $F([I_i]_\approx) \cap \Sigma_{G/\approx}^{-1}(J) = \{[J_i]_\approx\}$, so indeed $G/\approx$ is bunchy.

To see that $G/\approx$ is $\leq_R$-maximal among the bunchy factors of G, let $H \leq_R G$ be bunchy and $\Psi \in \hom_R(G, H)$. If $I_1 \approx_0 I_2$ and H is bunchy, then we must have $\partial\Psi(I_1) = \partial\Psi(I_2)$. Therefore the partition into $\partial\Psi$-fibers corresponds to an equivalence relation that coarsens the symmetric, reflexive relation $\approx_0$, and thus also coarsens the transitive closure $\approx$. Considering $V(H)$ as a partition of $V(G)$, we must therefore have $V(H) \preceq V(G)/\approx$, so $(G/\approx) = B(G)$ by the maximality of $B(G)$. □

4.8.6 Proofs of Propositions 4.6.1 and 4.7.1

Lemma 4.8.8. *Let G_1, G_2, H, K be graphs with $K \leq_R H \leq_R G_i$. Let $\Delta \in \hom_R(H, K)$. Let $\Psi_i \in \hom_R(G_i, H)$ and $P = G_1 \times_{\Psi_1,\Psi_2} G_2$. Let $\Phi_i = \Delta \circ \Psi_i$ and $Q = G_1 \times_{\Phi_1,\Phi_2} G_2$. Then, noting that*

$V(P) = V(Q)$ and $E(P) \subseteq E(Q)$, we have $\hat{\Phi}_i|_P = \hat{\Psi}_i$, and every principal subgraph of P is a principal subgraph of Q.

Proof. Let C be a principal subgraph of P. By the definition of the fiber product, we have $V(P) = V(Q) = V(G_1) \times V(G_2)$. Let $I \in V(H)$ and let $I_i \in (\partial\Psi_i)^{-1}(I)$. Suppose that $(I_1, I_2) \in V(C)$. We need to show that $E_{(I_1,I_2)}(C) = E_{(I_1,I_2)}(Q)$.

Since $E_{(I_1,I_2)}(C) = E_{(I_1,I_2)}(P)$ by the definition of a principal subgraph, it is enough to show that $E_{(I_1,I_2)}(P) = E_{(I_1,I_2)}(Q)$ for any $(I_1, I_2) \in V(P)$. Clearly $E_{(I_1,I_2)}(P) \subseteq E_{(I_1,I_2)}(Q)$, so it is enough to show that $|E_{(I_1,I_2)}(P)| = |E_{(I_1,I_2)}(Q)|$. To see this equality, note that since $\Delta \circ \Psi_P = \Phi_Q$, we have

$$|E_{(I_1,I_2)}(P)| = |E_{\partial\Psi_P(I_1,I_2)}(H)| = |E_{\partial\Phi_Q(I_1,I_2)}(K)| = |E_{(I_1,I_2)}(Q)|$$

where the equalities follow from the facts that Ψ_P, Δ, and Φ_Q respectively are right-resolving. Therefore $E_{(I_1,I_2)}(P) = E_{(I_1,I_2)}(Q)$. This shows that C is indeed a principal subgraph of Q. Moreover, for $e = (e_1, e_2) \in E(P)$, we have $\hat{\Phi}_i(e) = e_i = \hat{\Psi}_i(e)$, so indeed $\hat{\Phi}_i|_P = \hat{\Psi}_i$. □

Lemma 4.8.9. *Let G_1, G_2, K be graphs with $K \leq_S G_i$ via $\Phi_i \in \hom_S(G_i, K)$. Let C be a principal subgraph of $P = G_1 \times_{\Phi_1,\Phi_2} G_2$ such that the $\partial\hat{\Phi}_i|_{V(C)} : V(C) \to V(G_i)$ are surjective. Then $\hat{\Phi}_i|_C \in \hom_S(C, G_i)$. In particular, C is a common synchronizing extension of the G_i.*

Proof. Let $I \in V(K)$ and let $I_i, I'_i \in V(G_i)$ with $\partial\Phi_i(I_i) = \partial\Phi_i(I'_i) = I$. Since the Φ_i are synchronizing, we have $I_i \sim_{\Phi_i} I'_i$. Suppose that $(I_1, I_2), (I'_1, I'_2) \in V(C)$. In order to show that $\hat{\Phi}_i|_C \in \hom_S(C, G_i)$, we claim that $(I_1, I_2) \sim_{\Phi_P} (I'_1, I'_2)$. This will show that $\Phi_P|_C \in \hom_S(C, K)$. Since $\Phi_P|_C = \Phi_i \circ \hat{\Phi}_i|_C$, it will then follow that $\hat{\Phi}_i|_C \in \hom_S(C, G_i)$ by Theorem 4.3.5(4).

To prove this claim, let $u \in L_I(K)$. Since $I_1 \sim_{\Phi_1} I'_1$, there exists $v_1 \in L_{t(u)}(K)$ such that $I_1 \cdot uv_1 = I'_1 \cdot uv_1$. Similarly, since $I_2 \sim_{\Phi_2} I'_2$, there exists $v_2 \in L_{t(v_1)}(K)$ such that $I_2 \cdot uv_1v_2 = I'_2 \cdot uv_1v_2$. Then in particular $(I_1, I_2) \cdot uv_1v_2 = (I'_1, I'_2) \cdot uv_1v_2$. Since $u \in L_I(K)$ was arbitrary, we have $(I_1, I_2) \sim_{\Phi_P} (I'_1, I'_2)$ as claimed. □

Proof of Proposition 4.7.1(1). Let $K \leq_S G_i$ and $\Psi_i \in \hom_S(G_i, K)$. Let $P = G_1 \times_{\Psi_1,\Psi_2} G_2$ and let C be a principal subgraph of P such that $\partial\hat{\Psi}_i|_{V(C)} : V(C) \to V(G_i)$ are surjective. By Lemma 4.8.9, we have $\hat{\Psi}_i|_C \in \hom_S(C, G_i)$. Let $\Delta \in \hom_R(K, M)$, let $\Phi_i = \Delta \circ \Psi_i$, and let $Q = G_1 \times_{\Phi_1,\Phi_2} G_2$. Then, by Lemma 4.8.8, C is a principal subgraph of Q, with $\hat{\Phi}_i|_C = \hat{\Psi}_i|_C$. In particular, $\hat{\Phi}_i|_C \in \hom_S(C, G_i)$. □

We now give an equivalent form of the $O(G)$ conjecture. Fragments of this result appear in Proposition 4.7.1. The logical structure of Proposition 4.8.10 (one statement is equivalent to the equivalence of four other statements) is unusual.

Proposition 4.8.10. *Let $\mathcal{F}$ be a family of graphs satisfying the following conditions:*

(i) If $G \in \mathcal{F}$ and $H \leq_R G$, then $H \in \mathcal{F}$.

(ii) Let $G_1, G_2, K \in \mathcal{F}$ with $K \leq_R G_i$. Let $\Phi_i \in \hom_R(G_i, K)$, let $P = G_1 \times_{\Phi_1,\Phi_2} G_2$, and let C be a principal subgraph of P such that the $\partial\hat{\Phi}_i|_{V(C)}$ are surjective. Then $C \in \mathcal{F}$.

Then the following assertions are equivalent.

(1) For any $G \in \mathcal{F}$, there exists a unique $\leq_S$-minimal graph $O(G) \leq_S G$.

(2) For any $G_1, G_2 \in \mathcal{F}$, the following assertions are equivalent.

(a) $O(G_1), O(G_2)$ exist and are equal.

(b) G_1, G_2 have a common synchronizing factor.

(c) $M(G_1) = M(G_2) = M$ and there exist $\Phi_i \in \hom_R(G_i, M)$ such that $\hat{\Phi}_i \in \hom_S(C, G_i)$ for some principal subgraph C of $Q = G_1 \times_{\Phi_1,\Phi_2} G_2$.

(d) *G_1, G_2 have a common synchronizing extension $K \in \mathcal{F}$.*

Remark 4.8.11. The proof of Proposition 4.8.10 in fact shows that both (1) and (2) are equivalent to the assertion that (d) implies (b). Note that for $G_1 = G_2 = G$, (a) states that $O(G)$ is well-defined, while (b)–(d) are trivial.

Remark 4.8.12. The $O(G)$ conjecture states that the equivalent statements (1) and (2) in Proposition 4.8.10 hold with $\mathcal{F}$ equal to the class of all strongly connected graphs.

Proof of Proposition 4.8.10. First, assume (2). Let G be a graph and let $H_1, H_2 \leq_S G$ be $\leq_S$-minimal. Since H_1, H_2 have the common synchronizing extension G, they satisfy condition (d), so they also satisfy condition (b), i.e. there exists a common synchronizing factor $K \leq_S H_i$. But since the H_i were assumed minimal, we must have $H_1 = K = H_2 = O(G)$.

Now, assume (1) and deduce (2) as follows. Trivially, (a) implies (b) and (c) implies (d). Moreover, (b) implies (c) by Proposition 4.7.1(1). Finally, assume (d). Suppose that G_1, G_2 have a common synchronizing extension K. Then $O(G_1) = O(K) = O(G_2)$, so (d) implies (a). □

Proof of Proposition 4.7.1(2). This is immediate from Proposition 4.8.10, specifically, the equivalence of 2(b) and 2(c), with $\mathcal{F}$ taken to be the class of all strongly connected graphs. □

Proof of Proposition 4.7.1(3). By Theorem 4.5.6, this is immediate from Proposition 4.8.10, again via the equivalence of 2(b) and 2(c), but with $\mathcal{F}$ taken to be the class of strongly connected almost bunchy graphs. □

Proposition 4.8.13. *Let $\mathcal{F}$ be a family of graphs such that, if $G \in \mathcal{F}$ and $H \leq_R G$, then $H \in \mathcal{F}$. Then the following assertions are equivalent.*

(1) *Any $\leq_S$-minimal graph $H \in \mathcal{F}$ is bunchy.*

(2) *For any $G \in \mathcal{F}$, there exists some bunchy $H \leq_S G$.*

(3) *For any non-bunchy $G \in \mathcal{F}$, there exists some $\Phi \in \hom_R(G, M(G))$ with $\sim_\Phi$ nontrivial.*

(4) *For any $G \in \mathcal{F}$, $B(G) \leq_S G$.*

Remark 4.8.14. Note that Proposition 4.8.13 is a more detailed version of Proposition 4.6.1.

Proof. To see that (1) implies (2), let $G \in \mathcal{F}$ and consider the set $\{H \,|\, H \leq_S G\}$. Being a finite partially ordered set, this set must have at least one minimal element. If (1) holds, then this minimal element is bunchy. Thus (2) holds.

To see that (2) implies (3), let $G \in \mathcal{F}$ be non-bunchy. There exists at least one $\leq_S$-minimal graph $H \leq_S G$. If (2) holds, then G is not $\leq_S$-minimal, so $H \neq G$. Let $\Psi \in \hom_S(G, H)$. The $\partial\Psi$-fibers are precisely the $\sim_\Psi$-classes. Since $H \neq G$, the $\partial\Psi$-fibers are not merely singletons, so $\sim_\Psi$ is nontrivial. Let $M = M(G)$ and $\Delta \in \hom_R(H, M)$, and let $\Phi = \Delta \circ \Psi$. By Lemma 4.8.4, the $\sim_\Phi$-classes are unions of $\sim_\Psi$-classes, so in particular, $\sim_\Phi$ is nontrivial.

To see that (3) implies (1), let $G \in \mathcal{F}$ be non-bunchy. If (3) holds, then there exists $\Phi \in \hom_R(G, M(G))$ with $\sim_\Phi$ nontrivial. Then $G/\sim_\Phi \leq_S G$ and $G/\sim_\Phi \neq G$; in particular, G is not $\leq_S$-minimal. This proves (1) in the contrapositive.

Finally, we show that (2) and (4) are equivalent. If (4) holds, then (2) holds since $B(G)$ is bunchy. Conversely, assume (2) and let $G \in \mathcal{F}$. Then there exists $H \leq_S G$ bunchy. Let $\Phi \in \hom_S(G, H)$ and let $B = B(G)$. By Proposition 4.5.16(2) and Theorem 4.3.5(4), we have $\Phi = \Delta \circ \Theta$ for some $\Theta \in \hom_S(G, B)$, $\Delta \in \hom_S(B, H)$. Therefore (4) holds. □

Corollary 4.8.15. *Let $\mathcal{F}$ be a family of graphs such that, if $G \in \mathcal{F}$ and $H \leq_R G$, then $H \in \mathcal{F}$. Suppose that any $\leq_S$-minimal element of $\mathcal{F}$ is bunchy. Then $O(G)$ exists for any $G \in \mathcal{F}$.*

The proof of Corollary 4.8.15 is a trivial adaptation of the proof of Proposition 4.6.2.

Proof of Proposition 4.7.1(4). Assuming the bunchy factor conjecture, the $O(G_i)$ are well-defined. By the equivalence of conditions (a),(c) in Proposition 4.8.13, the hypothesis on the $B(G_i)$ is equivalent to the equality $O(B(G_1)) = O(B(G_2))$, which in turn is equivalent to the equality $O(G_1) = O(G_2)$ by Proposition 4.8.13. □

4.9 Proof of Theorem 4.4.3, following Trahtman

In this section, we recall Trahtman's proof of the road colouring theorem and reformulate it in a form applicable to Theorem 4.4.3.

4.9.1 Systems of maps with unique tallest trees

The central idea in Trahtman's proof is an idea well known in the literature on the combinatorics of transformations of finite sets—namely, that, the graph of a transformation f of a set V is a directed graph of constant out-degree 1 with $V(G)$‘V, where, for each state $I \in V$, the unique edge with source I has target $I \cdot f$. When V is finite, the graph consists of a set of state-disjoint directed cycles together with trees rooted on the cycles, directed toward their roots. Trahtman's proof is based on the construction of a transformation with a unique tallest tree (the height of a tree being the maximum distance from the root to a leaf). We need a few definitions in order to extend the notion of a unique tallest tree to a system of maps taking one set to another, cyclically, as opposed to a transformation of a single set.

Let $p \in \mathbb{N}$ and let $\{V_i\}_{0 \leq i \leq p-1}$ be disjoint finite sets. Let $a_i : V_i \to V_{i+1}$ be maps, with subscripts read modulo p. For $k \geq 0$, let $b_k = a_k a_{k+1} \cdots a_{p-1} a_0 \cdots a_{k-1} : V_k \to V_k$. As noted above in the general discussion of graphs of transformations, since each b_k is a transformation of a finite set, it is eventually periodic—that is, for each k, there exist $m \geq 0$, $z \geq 1$ such that $b_k^{m+z} = b_k^m$. Moreover, the orbits of individual elements $I \in V_k$ may vary in their eventually periodic behaviour. Specifically, for a given $I \in V_k$, consider the lexicographically minimal (ℓ, m, z) with $0 \leq \ell \leq p-1$, $m \geq 0$, $z \geq 1$, such that $I \cdot a_k \cdots a_{k+\ell-1} b_{k+\ell}^{m+z} = I \cdot a_k \cdots a_{k+\ell-1} b_{k+\ell}^m$. Define the *height* $h(I) = mp + \ell$ and the *root* $\rho(I) = I \cdot a_k \cdots a_{k+\ell-1} b_{k+\ell}^{m+z}$. The idea is that $h(I)$ is the number of steps required until the orbit of I reaches the root $\rho(I)$ and becomes periodic, with $z(I) = z$ being the period as a multiple of p.

To talk about unique tallest trees, let $h_{\max,k} = \max\{h(I) \mid I \in V_k\}$, and let $h_k(J) = \max\{h(I) \mid I \in V_k,\ \rho(I) = J\}$. Let $z_k = \operatorname{lcm}\{z(I) \mid I \in V_k\}$. We say that the system $(V_i, a_i)_{0 \leq i \leq p-1}$ has a *unique tallest tree* at V_k if there is a unique J with $h_k(J) = h_{\max,k}$. Note that the terms we have defined here still make sense even if the V_i are not all pairwise disjoint, as long as any two are either equal or disjoint, since we can make them disjoint by replacing V_i with $V_i \times \{i\}$.

We now present our interpretation of a key step in Trahtman's proof, applying Proposition 4.3.12 and closely following [14].

Lemma 4.9.1. *Let G, H be strongly connected graphs with $H \leq_R G$. Let $\Phi \in \hom_R(G, H)$. Let $I_0, \ldots I_{p-1} \in V(H)$, not necessarily distinct, be such that $I_{i+1} \in F(I_i) \neq \emptyset$, and let $a_i \in E_{I_i I_{i+1}}(H)$ (with subscripts read modulo p). Suppose that the system $((\partial\Phi)^{-1}(I_i), a_i)_{0 \leq i \leq p-1}$ has a unique tallest tree, where we write a_i for the map $I \mapsto I \cdot a_i$, $I \in (\partial\Phi)^{-1}(I_i)$. Then $\sim_\Phi$ is nontrivial.*

Proof. For $0 \leq i \leq p-1$, let $V_i = (\partial\Phi)^{-1}(I_i)$. Suppose without loss of generality that the system has a unique tallest tree at V_0. Let $I \in V_0$ be a state of maximal height $h(I) = h_{\max,0}$. Let $m \geq 0$, $0 \leq p-1$ be such that $h_{\max,0} = mp + \ell$, and let $z = z(I)$ and let $R = \rho(I)$. Note that $R \in V_\ell$. That is, upon cyclic application of the maps a_i, I is eventually mapped to R, then returns to R every z cycles around the graph.

By strong connectedness and Lemma 4.3.11, let $U \subseteq I_0$ be a minimal image such that $I \in U$. We claim that there is no other state $I' \in U$ with $\rho(I') = R$ and $I' \neq I$. Indeed, suppose that there is such a state I'. Then $I \cdot a_0 \cdots a_{\ell-1} b_\ell^m = I' \cdot a_0 \cdots a_{\ell-1} b_\ell^m = R$, so $|U \cdot a_0 \cdots a_{\ell-1} b_\ell^m| < |U|$, contradicting the minimality of U. This proves the claim. Let $U_0 = U \setminus \{I\}$. Then every element of U_0 has height strictly less than $h_{\max,0}$.

Let $u_1 = a_0 \cdots a_{\ell-1} b_\ell^{m-1} a_\ell \cdots a_{\ell-2} : V_0 \to V_{\ell-1}$, where the tail $a_\ell \cdots a_{\ell-2}$ includes each a_i exactly once, other than $a_{\ell-1}$, and the subscripts are read modulo p. The effect of applying u_1 to V_0 is to bring I to one step before its first encounter its root R; since I has maximal height, every other $I' \in V_0$ has already reached its root and is in the periodic part of its orbit after application of u_1.

Let $u_2 = u_1 b_{\ell-1}^{z_\ell} = a_0 \cdots a_{\ell-1} b_{\ell-1}^{m+z_k} a_\ell \cdots a_{\ell-2} : V_0 \to V_{\ell-1}$. Let $J_i = I \cdot u_i$ and $U_i = U \cdot u_i$. As observed in the previous paragraph, after application of u_1, I is not yet in the periodic part of its orbit (i.e. the orbit under cyclic application of the maps a_i), but every other element of

Observe that $J_1 \neq J_2$ by the assumed value of $h(I)$.

However, since I is the unique element of U with this maximal height, we have $U_1 \Delta U_2 = \{J_1, J_2\}$. Since the U_i are minimal images, we have $J_1 \sim_\Phi J_2$ by Proposition 4.3.12. □

4.9.2 Obtaining a right-resolver with a unique tallest tree

Let G be a graph. We define a *total order colouring* to be a total ordering of each edge set $E_I(H)$, i.e. a labeling of the edges of G such that, if $|E_I(G)| = k$, then the edges in $E_I(G)$ are labeled bijectively by $\{0, \ldots, k-1\}$. Suppose that $M = M(G)$ is a cycle of bunches. Then, once a total order colouring of M is fixed, total order colourings of G correspond bijectively with right-resolvers $\Phi \in \hom_R(G, M)$.

Letting $V(M) = \{I_0, \ldots, I_{p-1}\}$, there is exactly one edge $a_i \in E_{I_i}(H)$ labeled 0 for each i. This yields a subgraph W of G consisting of edges labeled 0, which is a spanning subgraph of G of constant out-degree 1. Every graph of constant out-degree 1 consists of a set of state-disjoint cycles, together with trees rooted on the cycles, directed toward their roots. The height of a tree is the maximum path length from a state in the tree to its root. Observe that the system $((\partial\Phi)^{-1}(I_i), a_i)_{i=0}^{p-1}$ has a unique tallest tree, in the sense of mappings, if and only if there is a unique tallest tree in W.

We now present our interpretation of the main technical lemma in the proof of the road colouring theorem (Lemma 10.4.6 in [14]). The following is not how the lemma is stated in [14], but one can follow the proof and observe that it is equivalent.

Lemma 4.9.2 (Trahtman)**.** *Let G be a strongly connected graph such that $M(G)$ is a cycle of bunches. At least one of the following is true:*

1. *G is itself a cycle of bunches.*
2. *G has two distinct bunches whose outgoing edges have the same target.*
3. *G admits a total order colouring with a unique tallest tree.*

With this result, we can prove our generalization of the road colouring theorem.

Proof of Theorem 4.4.3. Let $M = M(G)$. The claim is trivially true if $|V(G)| = |V(M)|$, in which case $G = M$. Suppose that it is true for all H with $|V(H)| \leq_R N$ and $M(H) = M$. Suppose that $|V(G)| = N + 1$. If G is bunchy, then we are done. If G is not bunchy, but has two states that can be in-amalgamated, then they are stable for some $\Phi \in \hom_R(G, M)$ by Lemma 4.3.8. Let $G' = G/\sim_\Phi$. Then $|V(G/\sim_\Phi)| < |V(G)|$, so by the inductive hypothesis, there is some bunchy $H \leq_S G/\sim_\Phi \leq_S G$.

Now, suppose that G does not have two states that can be in-amalgamated—in particular, G does not have two distinct bunches whose outgoing edges have the same target. Since G is not bunchy, it is in particular not a cycle of bunches, so by Lemma 4.9.2, it admits a total order colouring with unique tallest tree. As remarked above, this total order colouring corresponds to some right-resolver $\Phi \in \hom_R(G, M)$, which has $\sim_\Phi$ nontrivial by Lemma 4.9.1. Then once more $G/\sim_\Phi$ is strictly smaller than G. We can then apply Proposition 4.8.13 to conclude that any $\leq_S$-minimal O with $M = M(O)$ is a cycle of bunches. If $O \leq_S G$, then $O = O_{M,p}$ for $p = \mathrm{per}(G)/\mathrm{per}(M)$. □

Chapter 5

Encoding subshifts through sliding block codes

5.1 Introduction

As indicated in the introduction to the dissertation, this chapter proves the following theorem.

Theorem 5.1.1. *Let X be a mixing SFT, Y a mixing sofic shift, and $\pi : X \to Y$ a factor code. Let Z be a subshift with topological entropy strictly less than that of Y. Then there exists a subshift Z' of X conjugate to Z such that $\pi|_{Z'}$ is injective, if and only if for every $n \geq 1$, the number of periodic points of least period n in Z is at most the number of periodic points of least period n in Y with a π-preimage of equal least period.*

Theorem 5.1.1 is a generalization of the following theorem of Krieger in the case of unequal entropy; in particular, Theorem 5.1.1 reduces to Theorem 5.1.2 in the case that $Y = X$ and π is the identity.

Theorem 5.1.2 (Theorem 2 in [48])**.** *Let Y be a mixing shift of finite type and Z a subshift. Then there is a subshift $Z' \subseteq Y$ conjugate to Z if and only if Z and Y are conjugate or the (topological) entropy of Z is less than that of Y and, for every $n \geq 1$, the number of periodic points of least period n in Z is at most the corresponding number in Y.*

We note that with X, Y, Z, π as in the statement of Theorem 5.1.1, clearly there exists a subshift Z' of X conjugate to Z such that $\pi|_{Z'}$ is injective if and only if there exists a sliding block code $\psi : Z \to X$ such that $\pi \circ \psi$ is injective, in which case $Z' = \psi(Z) \subset X$. To verify the "only if" statement in Theorem 5.1.1, suppose that there is a subshift Z' of X conjugate to Z such that $\pi|_{Z'}$ is injective. Let $y \in \pi(Z')$ be periodic. Let $x = \pi|_{Z'}^{-1}(y)$ be the unique preimage of y in Z'. Then the orbit of x is in bijection with the orbit of y; otherwise, π would fail to be injective on the orbit of x, which is contained in Z'. In particular, x has finite orbit, so x is periodic, moreover with $\mathrm{per}(x) = \mathrm{per}(y)$. Thus, every periodic point in $\pi(Z') \subset Y$ has a periodic preimage in $Z' \subset X$ of equal least period, which shows the necessity of the stated condition.

Both Theorem 5.1.1 and Theorem 5.1.2 give conditions for the existence of an embedding in terms of entropy and a periodic point condition. The following corollary, which we prove in Section 5.5, shows that the periodic point condition can be removed in exchange for a small loss of injectivity.

Corollary 5.1.3. *Let X be a mixing SFT, Y a mixing sofic shift, and $\pi : X \to Y$ a factor code. Let Z be a subshift with topological entropy strictly less than that of Y. Then there exist a subshift Z', a finite-to-one factor code $\chi : Z' \to Z$, and a sliding block code $\psi : Z' \to X$ such that $\pi \circ \psi$ is injective. Moreover, if Z is mixing sofic with positive entropy (i.e. not a single fixed point), then Z' can be taken to be a mixing SFT and χ can be taken to be almost invertible.*

The code χ is in fact injective except on points in Z' whose images in Z are backward-asymptotic to one of finitely many periodic points in Z. See Lemma 5.2.2 and Remark 5.2.3. From Corollary 5.1.3, we can immediately conclude the following, with h denoting the topological entropy of a subshift.

Corollary 5.1.4. *Let X be a mixing SFT, Y a mixing sofic shift, and $\pi : X \to Y$ a factor code. For any $\varepsilon > 0$, there exists a mixing SFT $Z \subset X$ with $h(Z) > h(Y) - \varepsilon$ such that $\pi|_Z$ is injective.*

The proof of Theorem 5.1.1 adapts the strategy used to prove Theorem 5.1.2 in [48, 49] and related results in [18]. The outline of the proof is as follows. We use a marker set, as in the proof of Theorem 5.1.2, to break points in Z into moderate blocks and long periodic blocks, separated by marker coordinates. We code these separately using certain "data blocks" in Y, some of moderate length and some long and periodic, where the long periodic data blocks come from periodic points with π-preimages of equal least period in X. A block in Z between marker coordinates is coded to a data block in Y which is shorter by an additive constant, so that there are gaps between the data blocks, filled with repetitions of a "blank" symbol. We then lift the data blocks from Y to data blocks from X, then replace the blanks with a "stamp" block from X to form a valid point in X. The stamp block is chosen to ensure that once the point in X is coded into Y by π, the locations of the stamp, and thus of the marker coordinates, can be recognized. These manipulations of markers, blanks, and stamps are presented in detail in Section 5.3, while the quantitative arguments required to construct the data blocks and stamps are given in Section 5.4.

The statement of Theorem 5.1.2 is false for X merely mixing sofic, and to date there is no known characterization of the subshifts that embed into a given mixing sofic shift, though some sufficient conditions are known [18, 73]. Theorem 5.1.1 sheds some light on this problem, without resolving it. Salo-Törmä have answered [67] the following related question: let Y be a mixing sofic shift and $Z \subset Y$ a mixing SFT. For which such Y, Z do there exist a mixing SFT extension $\pi : X \to Y$ and a (mixing SFT) $Z' \subset X$ such that $\pi|_{Z'} : Z' \to Z$ is a conjugacy? However, it is unclear how the conditions given in that answer compare to those in Theorem 5.1.1, or to the results given in [73]. As a final related question, when Y is an SFT and Z is conjugate to Y, the existence of an SFT $Z' \subset X$ conjugate to Z such that $\pi|_{Z'} : Z' \to Y$ is a conjugacy, i.e. is surjective as well as injective, has been studied in [33], continuing work from [54].

5.2 Conventions, definitions, and background

5.2.1 Subshifts and sliding block codes

We recall the definition, from earlier in the paper, of a subshift over $\mathbb{Z}$ with alphabet $\mathcal{A}$. A subshift $X \subset \mathcal{A}^{\mathbb{Z}}$ is characterized by the set $\mathcal{B}(X)$ of blocks $w \in \mathcal{A}^*$ such that $X \cap [w] \neq \emptyset$, called the *language* of X. When the intended subshift X is clear, we write $[w]_i$ for $X \cap [w]_i$. We write $\mathcal{B}_n(X) = \mathcal{B}(X) \cap \mathcal{A}^n$. We can equivalently characterize a subshift by a set of forbidden words $\mathcal{F} \subset \mathcal{A}^*$, writing $\mathsf{X}_{\mathcal{F}} := \mathcal{A}^{\mathbb{Z}} \setminus \bigcup_{w \in \mathcal{F}} \bigcup_{i \in \mathbb{Z}} [w]_i$. Note that in general $\mathcal{F} \subsetneq \mathcal{A}^* \setminus \mathcal{B}(\mathsf{X}_{\mathcal{F}})$. For a given subshift $X \subset \mathcal{A}^{\mathbb{Z}}$, there may be several different sets of forbidden words $\mathcal{F} \subset \mathcal{A}^*$ such that $X = \mathsf{X}_{\mathcal{F}}$. A *shift of finite type* (SFT) is a subshift X such that $X = \mathsf{X}_{\mathcal{F}}$ for some finite set $\mathcal{F}$. A *k-step* SFT over $\mathcal{A}$ is an SFT of the form $X = \mathsf{X}_{\mathcal{F}}$ for some set $\mathcal{F} \subset \mathcal{A}^{k+1}$.

We recall from Chapter 3 that, for subshifts X, Y, a function $\phi : X \to Y$ is continuous and shift-equivariant if and only if it is a *sliding block code*, which means that there exist $m, n \geq 0$ and $\Phi : \mathcal{B}_{m+n+1}(X) \to \mathcal{B}_1(Y)$ such that for every $x \in X$ and every $i \in \mathbb{Z}$, $\phi(x)_i = \Phi(x_{[i-m,i+n]})$. We say that ϕ is a k-block code if $m + n + 1 = k$. A *factor code* is a surjective sliding block code, and for a sliding block code ϕ defined on a subshift X, we say that the image $\phi(X)$ is a *factor* of X, and that X, or more properly $\phi : X \to \phi(X)$, is an *extension* of $\phi(X)$. We recall also that a *sofic shift* is any factor of a shift of finite type. An injective sliding block code is called an *embedding*, and a bijective sliding block code is called a *conjugacy*. The properties of being sofic and of finite type are both invariant under conjugacy.

As in Chapter 3, a subshift X is said to be *irreducible* if for all $u, w \in \mathcal{B}(X)$, there exists $v \in \mathcal{B}(X)$ such that $uvw \in \mathcal{B}(X)$, and *strongly irreducible* with gap $g \geq 1$ if, for any u, w, we can take always take $v \in \mathcal{B}_g(X)$. Any factor of an irreducible (resp. strongly irreducible) subshift is irreducible (resp. strongly irreducible). A *periodic point* in a subshift X is a point $x \in X$ with $x = \sigma^n x$ for some

$n \geq 1$—we say that x has period n. The *least period* $\mathrm{per}(x)$ of a periodic point x is the least n such that $\sigma^n x = x$. Note that $|\{\sigma^n x \,|\, n \in \mathbb{Z}\}| = \mathrm{per}(x)$. We write $P(X)$ for the set of periodic points in a subshift X, $Q_n(X)$ for the set of periodic points of least period n, and $q_n(X) = |Q_n(X)|$. The number of periodic points of a given least period is a conjugacy invariant.

It is a theorem that periodic points are dense in any irreducible shift of finite type. As in Chapter 3, the *period* $\mathrm{per}(X)$ of an irreducible shift of finite type X is the gcd of the periods of the periodic points of X. An irreducible SFT with period 1 is said to be *aperiodic*. An irreducible SFT is strongly irreducible if and only if it is aperiodic, if and only if has periodic points of all sufficiently high periods. For irreducible sofic shifts, strong irreducibility is equivalent to having periodic points of all sufficiently high periods, which clearly implies that the periods have gcd 1, but the reverse implication fails. For example, consider the odd shift over $\{0, 1\}$, in which the block 10^n1 is permitted only for odd n. This is an irreducible sofic shift which contains the fixed point 0^∞, so the periods of periodic points trivially have gcd 1. However, the odd shift has no other periodic points of odd period. We follow the convention of the literature in referring to strongly irreducible sofic shifts (in particular SFTs) as mixing sofic shifts (mixing SFTs), because they are also characterized by a topological mixing property, but we will not use that property explicitly, so we do not define it here.

The following definition is new, and we use it extensively.

Definition 5.2.1. Let X and Y be subshifts and let $\pi : X \to Y$ be a factor code. We write $R_n(\pi)$ for the set of periodic points $y \in Y$ such that $y = \pi(x)$ for some periodic point $x \in X$ with $\mathrm{per}(x) = \mathrm{per}(y)$. We write $r_n(\pi) = |R_n(\pi)|$.

For a subshift X, the (topological) *entropy* of X is the value $h(X) = \inf_{n \geq 1} \frac{1}{n} \log |\mathcal{B}_n(X)|$; in fact, the limit $\lim_{n\to\infty} \frac{1}{n} \log |\mathcal{B}_n(X)|$ exists and is equal to $h(X)$. For a mixing sofic shift (in particular, a mixing SFT) X, we also have $h(X) = \lim_{n\to\infty} \frac{1}{n} \log q_n(X)$. Entropy is non-increasing under factor codes and is thus a conjugacy invariant, though certainly not a complete invariant. For any irreducible sofic shift X, and any proper subshift $V \subset X$, we have $h(V) < h(X)$. In Section 5.4, we use the following lemma of Marcus, which allows us to approximate a sofic shift from the inside by SFTs in terms of entropy.

Lemma 5.2.1 (Proposition 3 in [52])**.** *Let Y be a sofic shift. For every $\varepsilon > 0$, there exists an irreducible SFT $U \subseteq Y$ with $h(U) > h(Y) - \varepsilon$.*

For any subshift X and any $k \geq 1$, we can form the *kth higher block shift* $X^{[k]}$ with alphabet $\mathcal{B}_k(X)$, where

$$w = (a_{1,1}a_{1,2}\dots a_{1,k})(a_{2,1}a_{2,2}\dots a_{2,k})\dots(a_{\ell,1}a_{\ell,2}\dots a_{\ell,k}) \in \mathcal{B}(X^{[k]})$$

if any only if for each i, j we have $a_{i,j} = a_{i+1,j-1}$, so that

$$w = (a_1a_2\dots a_k)(a_2a_3\dots a_{k+1})\dots(a_{\ell+1}a_{\ell+2}\dots a_{\ell+k}),$$

and $a_1a_2\dots a_{k+\ell} \in \mathcal{B}(X)$. Observe that X and $X^{[k]}$ are conjugate for any subshift X and any $k \geq 1$. Moreover, if X is an m-step SFT and $k \leq m-1$, then $X^{[k]}$ is an $(m-k)$-step SFT. In particular, every SFT is conjugate to a 1-step SFT, and every sliding block code on an SFT can be written as a composition of a conjugacy and a 1-block code. We will therefore frequently assume WLOG that a given sliding block code on an SFT is a 1-block code on a 1-step SFT.

For a sliding block code on an irreducible shift of finite type, either every fiber is a finite set (indeed, of bounded cardinality), in which case the code is said to be *finite-to-one* and the entropy of the image is equal to that of the domain, or almost every fiber is uncountable, and the entropy of the image is strictly less than that of the domain. In the finite-to-one case, the minimum fiber cardinality is generic and is known as the *degree*. In particular, a code (on an irreducible SFT) with degree 1 is said to be *almost invertible*. It is a theorem that every irreducible (resp. mixing) sofic shift is an almost invertible factor of an irreducible (resp. mixing) SFT. We use the following construction of almost invertible codes, known as the "blowing-up lemma", in the proof of Corollary 5.1.3 in Section 5.5.

Lemma 5.2.2 (Lemma 10.3.2, [49])**.** *Let Z be a mixing SFT and let $z \in Z$ be a periodic point with least period p. Let $M \geq 1$. Then there exist a mixing SFT Z' and an almost invertible factor code $\chi : Z' \to Z$ such that the preimage of the orbit of z under χ is a single orbit of length Mp. Moreover, every periodic point in Z not in the orbit of z has unique preimage under χ.*

Remark 5.2.3. Note that in [49], the extension χ in Lemma 5.2.2 is only stated to be finite-to-one, but the existence of periodic points having unique preimage already implies almost invertibility. Indeed, the construction in [49], based on work in [18], in fact shows that χ is injective except on the points that are backward-asymptotic to points in the preimage of the orbit of z, where we say that two points z, z' are backward-asymptotic if $d(\sigma^n z, \sigma^n z') \to 0$ as $n \to -\infty$.

5.2.2 Markers and Markov approximations

We now recall the constructions with markers and long periodic blocks that are at the heart of the proof of Theorem 5.1.1. For an alphabet $\mathcal{A}$, we say that a block $w = w_1 \dots w_n \in \mathcal{A}^n$ is k-periodic, or has self-overlap of $n-k$, if for $1 \leq i \leq n-k$ we have $w_i = w_{i+k}$. A given block may be k-periodic for several different k.

Lemma 5.2.4 (Lemma 2.3 in [18])**.** *Let Z be a subshift, let $N \geq 1$, and $a, b \in \mathbb{Z}$ with $b - a \geq 2N$. Let $z \in Z$. If for every $i \in [a+N, b-N]$ there exists $p \leq N-1$ such that $z_{[i-N,i+N]}$ is p-periodic, then there is at most one periodic point $\zeta \in Z$ with $\operatorname{per}(\zeta) \leq N-1$ and $\zeta_{[a,b]} = z_{[a,b]}$. If Z is a 1-step SFT, then such a ζ exists.*

Lemma 5.2.5 (Lemma 2 in [48])**.** *Let Z be a subshift. For any $N \geq 1$, there exists a subset $F \subset Z$, which can be taken to be a finite union of cylinders, such that:*

1. *the sets $\sigma^i F$, $0 \leq i \leq N-1$, are all disjoint, and*
2. *if $z \notin \sigma^i F$ for all $-(N-1) \leq i \leq (N-1)$, then $z_{[-N,N]}$ is p-periodic for some $p \leq N-1$.*

Definition 5.2.2 (marker set)**.** With Z, N, and F as in Lemma 5.2.5, we refer to F as a *marker set* for Z with parameter N.

For any subshift X and any $n \geq 1$, we can form the nth Markov approximation X_n, which is the SFT defined by allowing precisely the blocks of length n which appear in X. Clearly $X_{n+1} \subset X_n$. It is an exercise to show that for any $\varepsilon > 0$ and any $N \geq 1$, there exists $N' \geq N$ such that $h(X_{N'}) < h(X) + \varepsilon$ and $q_n(X_{N'}) = q_n(X)$ for all $n \leq N$. In Lemma 5.2.6, we use the Markov approximation, together with higher block shifts, to show that in the proof of Theorem 5.1.1, we can assume WLOG that Z is a 1-step SFT, which allows us to apply Lemma 5.2.4.

We remark that there are versions of Lemma 5.2.5 which obviate the need for Lemma 5.2.4. However, for our purposes in this paper, embedding Z into an SFT has the additional benefit that the rate of convergence of $\frac{1}{n} \log q_n(Z)$ to $h(Z)$ can be easily estimated when Z is an SFT (see e.g. [49], pp. 349-351), which gives a procedure for deciding whether a given X, Y, π, Z satisfy the periodic point condition in Theorem 5.1.1, assuming that $h(Z) < h(Y)$ (namely, compute $N \geq 1$ such that for all $n \geq N$, $q_n(Z) < r_n(\pi)$, then check all $n \leq N$ to determine whether $q_n(Z) \leq r_n(\pi)$).

Lemma 5.2.6. *Let X be a mixing SFT, Y a mixing sofic shift, and $\pi : X \to Y$ a factor code. Let Z be a subshift with $h(Z) < h(Y)$ and $q_n(Z) \leq r_n(\pi)$ for all $n \geq 1$. Then there exists a 1-step SFT Z' such that Z embeds into Z', $h(Z') < h(Y)$, and $q_n(Z') \leq r_n(\pi)$ for all $n \geq 1$.*

We defer the proof of Lemma 5.2.6 to Section 5.5.

5.3 Coding

In Section 5.3.1, we introduce two coding constructions, namely subshifts with blanks adjoined, Definition 5.3.1, and stamps, Definition 5.3.2, then use them to create one side of an interface between Z on the one hand and $\pi : X \to Y$ on the other. In Section 5.3.2, we use markers in Z to construct the other side of this interface. In Section 5.3.3, we use stamps to give a construction of SFTs analogous to S-gap shifts. We use this construction in Section 5.4.2 to construct the shifts that are used in Section 5.3.1 and Section 5.3.2.

5.3.1 Blanks and stamps

As outlined in Section 5.1, the proof of Theorem 5.1.1 involves coding Z into X via certain intermediate subshifts which consist of long "data" blocks separated by blanks. We now define this construction precisely.

Definition 5.3.1 (subshift with blanks adjoined)**.** Let W be a subshift and let $N, \ell \geq 1$ with $\ell < N$. Let $*$ be a symbol not appearing in the alphabet of W. Let $M \subset \bigcup_{n=1}^{2N} \mathcal{B}_n(W)$ be a set of blocks and let $Q \subset \cup_{n=1}^{2N-1} Q_n(W)$ be a set of periodic points. Denote by Blanks$(M, Q, N, *, \ell)$ the subshift in which each point is of the form $\ldots w_{-1} *^{\ell} w_0 *^{\ell} w_1 \ldots$ where either $w_i \in M$ or $w_i = y_T$ where $y \in Q$ and $T = (-\infty, 0], [0, +\infty), (-\infty, \infty)$, or $[0, m]$ with $m \geq 2N$.

The purpose of the Blanks construction is to provide an interface between the channel $\pi : X \to Y$ and the subshift Z to be embedded. One side of this interface, namely the embedding of a Blanks subshift into X, is specified in Proposition 5.3.4. The construction involves particular blocks, which we call *stamps*, that can be unambiguously recognized in the following sense:

Definition 5.3.2 (stamp)**.** Let Y be a subshift, $W \subset Y$ a proper subshift, and $k \geq 1$. We say that $\mu \in \mathcal{B}(Y) \setminus \mathcal{B}(W)$ is a (Y, W, k) *stamp* if for all $u_1, u_2 \in \mathcal{B}(W)$ and $v_1, v_2 \in \mathcal{B}_k(Y)$, μ appears exactly once in $u_1 v_1 \mu v_2 u_2$.

Remark 5.3.1. In Definition 5.3.2, continuing with the notation there, we do not explicitly require $u_1 v_1 \mu v_2 u_2$ to be legal in Y. Doing so would neither affect the results nor simplify the proofs. In all of the examples we consider, such blocks will in fact be legal in Y.

Proposition 5.3.2. *Let Y be a strongly irreducible subshift with gap g and $W \subset Y$ a proper subshift. For every $k \geq g$ and every sufficiently large n, there exists a (Y, W, k) stamp of length n.*

We defer the proof of Proposition 5.3.2 to Section 5.4.1, but before applying stamps in Proposition 5.3.4, we prove a lemma that expresses how stamps are actually used in our constructions.

Lemma 5.3.3. *Let Y be a subshift, $W \subset Y$ a proper subshift, $k \geq 1$, and $\mu \in \mathcal{B}(Y) \setminus \mathcal{B}(W)$ a (Y, W, k) stamp. Let $N \geq |\mu|$. Then for any $\gamma^{\pm} \in \mathcal{B}_k(Y)$, and any $w \in \mathcal{B}(W)$ with $|w| \geq N$, the stamp μ appears exactly twice in the block $\mu \gamma^- w \gamma^+ \mu$.*

Proof. By the hypotheses on μ, $\gamma^{\pm}$, and w, and Definition 5.3.2, μ appears exactly once in each subblock $\mu \gamma^- w$, $w \gamma^+ \mu$. An appearance of μ other than at the positions explicitly indicated must therefore overlap both of these subblocks. Since $|w| \geq |\mu|$, μ must therefore be a subblock of w, contradicting the hypothesis that $w \in \mathcal{B}(W)$ and $\mu \in \mathcal{B}(Y) \setminus \mathcal{B}(W)$. □

We now give one of the main coding constructions (Proposition 5.3.4), embedding a subshift with blanks adjoined, and with blocks from a subshift $V \subset X$, into X via a sliding block code γ, such that $\pi \circ \gamma$ is injective. The large amount of data in the statement is representative of the complexity of the construction and the modular nature of the proof.

Proposition 5.3.4. *Let X be a mixing SFT with gap g, let Y be a mixing sofic shift, and let $\pi : X \to Y$ be a 1-block factor code.*

Let $V \subset X$, $W = \pi(V) \subset Y$ be proper subshifts. Let $*$ be a symbol not appearing in the alphabets of X, Y. Let $N \geq 1$. Let $M \subset \bigcup_{n=1}^{2N-1} \mathcal{B}_n(W)$ be a collection of blocks, and let $R \subset \bigcup_{n=1}^{N-1} R_n(\pi|_V)$ be a union of finite (i.e. periodic) orbits in W with π-preimages of equal cardinality in V. Let $\kappa : M \to \mathcal{B}(V)$ be an injection such that $\pi \circ \kappa(w) = w$ for each $w \in M$, and let $\hat{M} = \kappa(M)$. Similarly, let $\lambda : R \to P(V)$ be a shift-commuting injection such that $\pi \circ \lambda(y) = y$ for each $y \in R$, and let $\hat{R} = \lambda(R)$. Then for any $\ell \geq 1$, $\mathrm{Blanks}(M, R, N, *, \ell)$ and $\mathrm{Blanks}(\hat{M}, \hat{R}, N, *, \ell)$ are conjugate.

Moreover, let $\mu \in \mathcal{B}(Y) \setminus \mathcal{B}(W)$ be a (Y, W, g) stamp such that $|\mu| \leq N$, and suppose that $M \subset \bigcup_{n=N}^{2N-1} \mathcal{B}_n(W)$, i.e. M contains no blocks of length less than N. Then there exists a sliding block code $\gamma : \mathrm{Blanks}(\hat{M}, \hat{R}, N, *, |\mu| + 2g) \to X$ such that $\pi \circ \gamma$ is injective.

Proof. First, the conjugacy. Let $W[*] = \mathrm{Blanks}(M, R, N, *, \ell)$ and $V[*] = \mathrm{Blanks}(\hat{M}, \hat{R}, N, *, \ell)$. Consider the 1-block code $\pi[*]$ defined on $V[*]$ by the block map $\pi[*](a) = \pi(a)$ for a in the alphabet of V and $\pi[*](*) = *$. We claim that $W[*] = \pi[*](V[*])$ and that $\pi[*] : V[*] \to W[*]$ is a conjugacy. To see that $W[*] \subseteq \pi[*](V[*])$, note that any $\xi \in V[*]$ is of the form

$$\xi = \dots w_{-1} *^{\ell} w_0 *^{\ell} w_1 \dots$$

where either $w_i \in \hat{M}$ or $w_i = x_T$ for some $x \in \hat{R}$ and T an interval with $2N + 1 \leq |T|$. If $w_i \in \hat{M}$, then $\pi(w_i) \in M$; if $w_i = x_T$ for some $x \in \hat{R}$, then $\pi(w_i) = \pi(x)_T$, and $\pi(x) \in R$. Therefore

$$\pi[*](\xi) = \dots \pi(w_{-1}) *^{\ell} \pi(w_0) *^{\ell} \pi(w_1) \dots \in W[*]$$

This shows that indeed $W[*] \subset \pi[*](V[*])$. Similarly, any $\eta \in W[*]$ is of the form

$$\eta = \dots w_{-1} *^{\ell} w_0 *^{\ell} w_1 \dots$$

where either $w_i \in M$ or $w_i = y_T$ for some $y \in R$ and T an interval with $2N + 1 \leq |T| \leq \infty$. For η of this form, using Lemma 5.2.4, we can use the injections κ, λ to reconstruct a unique $\xi \in V[*]$ such that $\pi[*](\xi) = \eta$.

We now suppose that each block in M has length at least N and that we have a (Y, W, g) stamp $\mu \in \mathcal{B}(Y) \setminus \mathcal{B}(W)$ such that $|\mu| \leq N$. Under these assumptions, we construct a sliding block code $\gamma : V[*] \to X$ and show that $\pi \circ \gamma$ is injective. Fix a π-preimage $\hat{\mu}$ of μ, and let $\ell = |\mu| + 2g$. Using the hypothesis that X is a mixing 1-step SFT, define maps $\gamma^{\pm} : \mathcal{B}_1(V) \to \mathcal{B}_g(X)$ such that, for $a, b \in \mathcal{B}_1(V)$, we have $\hat{\mu}\gamma^-(a)a, b\gamma^+(b)\hat{\mu} \in \mathcal{B}(X)$. We then have a sliding block code $\gamma : V[*] \to X$, given by replacing each block $b *^{\ell} a$ by $b\gamma^+(b)\hat{\mu}\gamma^-(a)a$, and leaving the non-blank symbols unchanged.

Let

$$\xi = \dots *^{\ell} v_{-1} *^{\ell} v_0 *^{\ell} v_1 *^{\ell} \dots \in V[*]$$

Then

$$\gamma(\xi) = \dots \hat{\mu}\gamma^-(a_0) v_0 \gamma^+(b_0)\hat{\mu} \dots$$

where a_i, b_i are, respectively, the initial and terminal symbols of v_i. In turn, we have

$$\pi \circ \gamma(\xi) = \dots \mu(\pi \circ \gamma^-(a_0))\pi(v_0)(\pi \circ \gamma^+(b_0))\mu \dots$$

Moreover, by Lemma 5.3.3 and the lower bound on lengths of blocks in M, it follows that μ appears in $\pi \circ \gamma(\xi)$ only where $\hat{\mu}$ appears at the same position in $\gamma(\xi)$. By replacing, in $\pi \circ \gamma(\xi)$, each appearance of μ, and the blocks of length k to the left and right of μ, with $*^{\ell}$, we obtain the point $\dots *^{\ell} \pi(v_0) *^{\ell} \dots = \pi[*](\xi) \in \mathrm{Blanks}(M, R, N, *, \ell)$, from which ξ can be recovered since $\pi[*]$ is a conjugacy. $\square$

5.3.2 Blanks and markers

We now prove a lemma that encapsulates the use of marker constructions in our proof of Theorem 5.1.1.

Lemma 5.3.5. *Let Z, W be subshifts with Z a 1-step SFT. Let $N, \ell \geq 1$ be such that $q_n(Z) \leq q_n(W)$ for $n \leq N-1$ and $|\mathcal{B}_n(Z)| \leq |\mathcal{B}_{n-\ell}(W)|$ for $N+\ell \leq n \leq 2N+\ell-1$. Let $M \subset \bigcup_{n=N}^{2N-1} \mathcal{B}_n(W)$ and $Q \subset \bigcup_{n=1}^{N-1} Q_n(W)$ be a union of finite (i.e. periodic) orbits such that $|\mathcal{B}_n(Z)| \leq |M \cap \mathcal{B}_{n-\ell}(W)|$ for $N+\ell \leq n \leq 2N+\ell-1$, and $q_n(Z) \leq |Q \cap Q_n(W)|$ for $n \leq N-1$. Then Z embeds into* $\mathrm{Blanks}(M, Q, N, *, \ell)$.

Remark 5.3.6. The lower bound on the length of blocks in M is not in fact needed for Lemma 5.3.5, but it is needed in order to apply Lemma 5.3.5 in conjunction with Proposition 5.3.4 in the proof of Theorem 5.1.1 below.

Proof. Let F be a marker set for Z with parameter N. For $z \in Z$, let $A(z) = \{i \in \mathbb{Z} \,|\, \sigma^i z \in F\}$. Enumerate each $A(z)$ as $\{a_j(z)\}_{j \in J(z)}$ where the index set $J(z)$ may be the empty set, or a finite set, or the integers, or the positive or negative natural numbers, and where $a_j(z) < a_{j+1}(z)$ for each j. We refer to the elements of $A(z)$ as marker coordinates for z. Say that T is a *marker interval* for z if: $T = [a_j(z), a_{j+1}(z))$ where $a_j(z), a_{j+1}(z)$ are both defined; or $T = [a_0(z), \infty)$ if $a_0(z) = \max A(z) < \infty$; or $T = (-\infty, a_0(z)]$ if $a_0(z) = \min A(z) > -\infty$; or $T = (-\infty, \infty)$ if $A(z) = \emptyset$.

We construct an embedding of Z into $\mathrm{Blanks}(M, Q, N, *, \ell)$ by constructing a function Φ that maps a block occurring between marker coordinates to a data block padded with $*^\ell$. Let $c_n : Q_n(Z) \to Q_n(W)$ be shift-commuting injections for $n \leq N-1$. Let $d_n : \mathcal{B}_n(Z) \to \mathcal{B}_{n-\ell}(W) \cap M$ be injections for $N+\ell \leq n \leq 2N+\ell-1$. (This is despite the fact that the parameter for the marker set F is N. We need the extra space in order to pad data blocks with blanks.) For a block $w \in \mathcal{B}_n(Z)$ with $N+\ell \leq n \leq 2N+\ell-1$, let $\Phi(w) = *^\ell d_n(w)$. For $z \in Z$ periodic with $n = \mathrm{per}(z) \leq N-1$, if $m \geq 2N+\ell$, let $\Phi(z_{[0,m]}) = *^\ell c_n(z)_{[\ell,m]}$. Similarly, let $\Phi(z_{[0,\infty)}) = *^\ell c_n(z)_{[\ell,\infty)}$. Finally, let $\Phi(z_{(-\infty,0]}) = c_n(z)_{(-\infty,0]}$ and let $\Phi(z) = c_n(z)$. Observe that Φ is injective, by Lemma 5.2.4.

Define $\phi : Z \to W$ by declaring that $\phi(z)_T = \Phi(z_T)$ whenever T is a marker interval for z. We need to show that ϕ is an embedding. Certainly ϕ is shift-commuting, since, if T is a marker interval for z, then $T-1$ is a marker interval for σz, so

$$\phi(\sigma z)_{T-1} = \Phi((\sigma z)_{T-1}) = \Phi(z_T) = \phi(z)_T = (\sigma\phi(z))_{T-1}$$

Thus indeed $\phi(\sigma z) = \sigma\phi(z)$. Moreover, ϕ is injective because the appearances of $*^\ell$ in $\phi(z)$ allow us to reconstruct the marker coordinates, and then the injectivity of Φ allows us to reconstruct z_T for each marker interval T for z.

We need to show finally that ϕ is continuous, i.e. that for $z \in Z$, $\phi(z)_0$ depends only on $z_{[-L,L]}$ for some finite L independent of z. To see this, let L' be such that F is a union of cylinders on $[-L', L']$. Let $L = L' + 2N$. By examining $z_{[-L,L]}$, we can determine whether there are marker coordinates for z in $[-2N, 0)$ and/or $[0, 2N]$. If each of these intervals contains a marker coordinate, then $\phi(z)_0$ is determined by z_T where $T \subset [-2N, 2N]$ is the unique marker interval for z containing 0. If at least one of $[-2N, 0), [0, 2N]$ has no marker coordinates, then 0 is in a long marker interval for z. If there is a marker coordinate in $(-\ell, 0]$, then $\phi(z)_0 = *$. Otherwise, by Lemma 5.2.4, $\phi(z)_0$ is determined by any subblock $z_{[a,b]}$ where $a < 0 \leq b$, $b - a \geq 2N$, and $[a, b]$ contains no marker coordinate for z. This concludes the proof that ϕ is continuous. □

The remainder of the proof of Theorem 5.1.1 follows from the following proposition, the proof of which is taken up in Section 5.4.

Proposition 5.3.7. *Let X be a mixing SFT with gap g, Y a mixing sofic shift, and $\pi : X \to Y$ a factor code. Let Z be a subshift with $h(Z) < h(Y)$ and $q_n(Z) \leq r_n(\pi)$ for every $n \geq 1$. Then there exist: $N \geq 1$, subshifts $V \subset X$, $W = \pi(V) \subset Y$, and a (Y, W, g) stamp $\mu \in \mathcal{B}(Y) \setminus \mathcal{B}(W)$, such that $|\mu| \leq N$, $q_n(Z) \leq r_n(\pi|_V)$ for $n \leq N-1$ and $|\mathcal{B}_n(Z)| \leq |\mathcal{B}_{n-\ell}(W)|$ for $N+\ell \leq n \leq 2N+\ell-1$, where $\ell = |\mu| + 2g$.*

Proof of Theorem 5.1.1. By Lemma 5.2.6, assume WLOG that Z is a 1-step SFT. Let N, ℓ, $V \subset X$, $W = \pi(V) \subset Y$, and μ be as in Proposition 5.3.7. Let $M \subset \bigcup_{n=N}^{2N-1} \mathcal{B}_n(W)$ be as in Lemma 5.3.5, and let $R \subset \bigcup_{n=1}^{N-1} R_n(\pi|_V)$ be a union of finite orbits, such that $q_n(Z) \leq |R \cap R_n(\pi|_V)|$ for $n \leq N-1$. Each of which the orbits in R is, by the definition of R_n, necessarily the image of an orbit with equal cardinality in V. Here, R takes the role that Q plays in Lemma 5.3.5, but in Lemma 5.3.5, there was no channel π, and thus no preimage requirement, hence the change in notation. By Lemma 5.3.5, let $\phi : Z \to \mathrm{Blanks}(M, R, N, *, \ell)$ be an embedding.

Let $\hat{M}$, $\hat{R}$ be as in Proposition 5.3.4, let $\pi[*] : \mathrm{Blanks}(\hat{M}, \hat{R}, N, *, \ell) \to \mathrm{Blanks}(M, R, N, *, \ell)$ be a conjugacy, and let $\gamma : \mathrm{Blanks}(\hat{M}, \hat{R}, N, *, \ell) \to X$ be an embedding such that $\pi \circ \gamma$ is injective (by Proposition 5.3.4, using μ). Then $\psi = \gamma \circ (\pi[*])^{-1} \circ \phi : Z \to X$ is a sliding block code such that $\pi \circ \psi$ is injective. $\square$

5.3.3 Stamps and SFTs

In this subsection, we prove Lemma 5.3.10, which, in conjunction with Lemma 5.2.1, allows us, in Proposition 5.4.4, to construct a mixing SFT $V \subset X$ such that the image $\pi(V) \subset Y$ is a proper subshift of Y but has entropy at least $h(Y) - \varepsilon$ for a given $\varepsilon > 0$. It may be possible to give a more efficient construction of such a V, but we have not found one. We first prove Lemma 5.3.8, which is related to the characterization of SFTs among S-gap shifts (Theorem 3.3 in [26]).

Lemma 5.3.8. *Let X be a mixing SFT with gap g and let $V_0 \subset X$ be an SFT. Let $k \geq g$ and let $\mu \in \mathcal{B}(X) \setminus \mathcal{B}(V_0)$ be an (X, V_0, k) stamp. Let $N \geq |\mu|$ and let $V_1 \subset X$ be the closure of the set of points of the form*

$$\ldots v_{-1}\gamma_{-1}^{+}\mu\gamma_0^{-}v_0\gamma_0^{+}\mu\gamma_1^{-}v_1 \cdots \in X$$

where, for each i, $\gamma_i^{\pm} \in \mathcal{B}_k(X)$ and $v_i \in \mathcal{B}(V_0)$ with $|v_i| \geq N$. Then V_1 is a mixing SFT.

Remark 5.3.9. Note that V_1, as defined in the statement of Lemma 5.3.8, clearly contains V_0.

Proof. Assume without loss of generality that X is a 1-step SFT. We first perform a small recoding for convenience, specifically to make it easier to recognize stamps, by replacing X by a conjugate shift $\hat{X}$. For each $x \in X$, define $\hat{x}$ as follows: if $x_{[i,i+|\mu|)} = \mu$, then for each $i \in [-k, |\mu| + k)$, let $a = x_i$ and let $\hat{x}_i = \hat{a}$, where for symbols a, b in the alphabet of X, we have $\hat{a} = \hat{b}$ if and only if $a = b$, and the set of symbols with hats is disjoint from the alphabet of X. If there is no $j \in (i - (|\mu| + k), i + k]$ with $x_{[j,j+|\mu|)} = \mu$, then let $\hat{x}_i = x_i$. Clearly the map $x \mapsto \hat{x}$ is a sliding block code, and it is just as clearly injective, since we recover x from $\hat{x}$ by dropping hats. Therefore $\hat{X} = \{\hat{x} \,|\, x \in X\}$ is a mixing SFT, conjugate to X.

Denote by $\hat{V}_1 \subset \hat{X}$ the image of V_1 under the map $x \mapsto \hat{x}$. Let $\ell = |\mu| + 2k$. Since μ is an (X, V_0, k) stamp, and $N \geq |\mu|$, blocks of the form $\gamma_i^{+}\mu\gamma_{i+1}^{-}$ do not overlap in any point in V by Lemma 5.3.3, so symbols with hats occur in $\hat{V}_1$ in blocks of length exactly ℓ. The blocks of symbols with hats are separated by blocks from V_0. Since $\hat{V}_1$ is the image of V under a conjugacy $X \to \hat{X}$, V_1 is an SFT if and only if $\hat{V}_1$ is an SFT.

Let $m \geq N$ be such that $\hat{X}$ and V_0 are m-step SFTs. We claim that if $\hat{x} \in \hat{X}$ is such that $x_{[i,i+m]} \in \mathcal{B}_{m+1}(\hat{V}_1)$ for all $i \in \mathbb{Z}$, then $\hat{x} \in \hat{V}_1$, which means precisely that V is an m-step SFT. To prove this claim, let $\mathcal{F} \subset \mathcal{B}_{m+1}(X)$ be the set of blocks of length $m+1$ which contain at least one of the following: a block of length greater than ℓ in which all symbols have hats; a block without hats that is not in $\mathcal{B}(V_0)$; or a block of symbols without hats, of length less than N, bounded on both sides by symbols with hats. Note that $\mathcal{F}$ is disjoint from $\mathcal{B}_{m+1}(\hat{V}_1)$. Suppose that $\hat{x}_{[i,i+m]} \notin \mathcal{F}$ for all $i \in \mathbb{Z}$. Then any block of symbols with hats in $\hat{x}$ has length exactly ℓ, and is thus of the form $\gamma^{+}\mu\gamma^{-}$, where $\gamma^{\pm} \in \mathcal{B}_g(X)$ (with hats added). Furthermore, the blocks separating the blocks with hats must have length at least N and must be in $\mathcal{B}(V_0)$, since every subblock of length $m+1$ is in $\mathcal{B}(V_0)$ and V_0 is an m-step SFT. Thus indeed $\hat{x} \in \hat{V}_1$, so $\hat{V}_1$ is indeed an SFT.

To see that V_1 is irreducible, let $u_-, u_+ \in \mathcal{B}(V_1)$. We need to construct $u_0 \in \mathcal{B}(V)$ such that $u_-u_0u_+ \in \mathcal{B}(V_1)$. We do so as follows. Extend u_- on the right to form a block $v_- \in \mathcal{B}(V_1)$, which begins with u_- and ends with $\gamma_{-1}^+\mu\gamma_0^-$ where $\gamma_{-1}^+, \gamma_0^- \in \mathcal{B}_k(V_1)$. (It is possible that u_- overlaps $\gamma_{-1}^+\mu\gamma_0^-$.) Let $v_0 \in \mathcal{B}_N(V_0)$ be such that $v_-v_0 \in \mathcal{B}(X)$. Similarly, extend u_+ on the left to form a block $v_+ \in \mathcal{B}(V_1)$ which ends with u_+ and begins with $\gamma_0^+\mu\gamma_1^-$, where $\gamma_0^+, \gamma_1^- \in \mathcal{B}_k(V_1)$ and $v_0\gamma_0^+ \in \mathcal{B}(X)$. Let $x^\pm \in \mathcal{B}(V_1)$ be such that $x^-_{[0,\infty)}$ begins with v_- and $x^+_{(-\infty,-1]}$ ends with v^+. Let $x = x^-_{(-\infty,-1]}v_-v_0v_+x^+_{[0,\infty)}$. Then $x \in X$ since X is a 1-step SFT. Moreover, $x \in V_1$, since the tails $x^-_{(-\infty,-1]}v_-$ and $v_+x^+_{[0,\infty)}$ appear in V_1 and are joined together in a way that creates no violations of the restrictions defining V_1. Letting u_0 be the block appearing between u_-, u_+, such that $v_-v_0v_+ = u_-u_0u_+ \in \mathcal{B}(V)$, the construction is complete, showing that V_1 is indeed irreducible.

To see that V_1 is mixing, let $u_1, u_2 \in \mathcal{B}(V_0)$ with $|u_1| > m$, where m is as above, and $|u_2| = |u_1|+1$. Let $\gamma_i^\pm \in \mathcal{B}(X)$, $i = 1,2$, be such that $u_i\gamma_i^+\mu\gamma_i^-u_i \in \mathcal{B}(X)$. Then $x_i = (u_i\gamma_i^+\mu\gamma_i^-)^\infty \in V_1$ for both $i = 1,2$. Indeed, certainly $x_i \in X$, since $u_i\gamma_i^+\mu\gamma_i^-u_i \in \mathcal{B}(X)$ and X is a 1-step SFT. Moreover, $\mathrm{per}(x_i)$ divides $\ell+|u_i|$, and $\gcd(\ell+|u_1|, \ell+|u_2|) = \gcd(\ell+|u_1|, \ell+|u_1|+1) = 1$, so $\gcd(\mathrm{per}(x_1), \mathrm{per}(x_2)) = 1$. Since V_1 is an irreducible SFT with periodic points of coprime periods, V_1 is mixing. □

As advertised, we now use Lemma 5.3.8 to prove the following lemma, which is applied in the proof of Proposition 5.4.4, which in turn is the main input to the proof of Proposition 5.3.7.

Lemma 5.3.10. *Let X be a mixing SFT, Y a mixing sofic shift, and $\pi : X \to Y$ a factor code. Let $W_0 \subsetneq Y$ be an SFT. Then there exists a mixing SFT $V_1 \subset X$ with $W_0 \subset \pi(V_1) \subsetneq Y$.*

Proof. Let $V_0 = \pi^{-1}(W_0) \subset X$. Note that V_0 is an SFT since W_0 is an SFT. Let g be the mixing gap of X. Let $y \in Y \setminus W_0$ be a periodic point with least period $k \geq g$. Such a y certainly exists because periodic points are dense in Y and W_0 is a proper subshift. Let k' be such that $y_{[0,k')} \notin \mathcal{B}_{k'}(W_0)$. Let $\ell = k + k'$. Then every ℓ-block in y is forbidden in W_0. In particular, for any $x \in \pi^{-1}(\{y\})$ and any $i \in \mathbb{Z}$, we have $x_{[i,i+\ell)} \notin \mathcal{B}(V_0)$.

By Proposition 5.3.2, let μ be an (X, V_0, g) stamp. Let V_1 consist of the closure of the set of points of the form $\dots v_{-1}\gamma_{-1}^+\mu\gamma_0^-v_0\gamma_0^+\mu\gamma_1^-v_1 \dots \in X$ where each $v_i \in \mathcal{B}(V_0)$ with $|v_i| \geq \ell$ and each $\gamma_i^\pm \in \mathcal{B}_g(X)$. By Lemma 5.3.8, V_1 is indeed a mixing SFT. Note that every point in V_1 contains ℓ-blocks permitted in V_0, so V_1 is disjoint from $\pi^{-1}(\{y\})$, and therefore $\pi(V_1) \subsetneq Y$. □

5.4 Counting

In this section, we prove Proposition 5.3.2 and Proposition 5.3.7, which state the existence and properties respectively of the stamps and the shifts $V \subset X, W \subset Y$ used in Section 5.3. Section 5.4.1 contains two results required for the proof of Proposition 5.3.2, one (Lemma 5.4.1) showing that most blocks in a subshift with positive entropy have little self-overlap, and the other (Lemma 5.4.2) showing that one can assume, at the cost of a small loss of entropy, that a given sufficiently long block appears syndetically in a mixing sofic shift. Section 5.4.2 then gives a crucial asymptotic result on the number of periodic points in Y with a preimage of equal least period in X, and applies the results from Section 5.3.3 to construct the shifts V and W.

5.4.1 Self-overlap and stamps

We begin by showing that most blocks have very little self-overlap, which we use both to construct stamps and to determine the asymptotic number of periodic points in Y with a π-preimage of equal least period.

Lemma 5.4.1. *Let Y be a subshift with $h(Y) > 0$. For every $\alpha \in (0,1)$, there exist $N \geq 1$ and $b > 0$ such that for every $n \geq N$, there are at least $(1 - \exp(-bn))\exp(nh(Y))$ blocks $w \in \mathcal{B}_n(Y)$ with no self-overlap of more than αn.*

Proof. Let $\varepsilon = \frac{1}{2}(\alpha^{-1}-1)h(Y)$, so that $\alpha(h(Y)+\varepsilon) < h(Y)$. Let $r = \exp(h(Y))$ and $s = \exp(h(Y)+\varepsilon)$. Note that $s^\alpha < r < s$ and that $r^n \leq |\mathcal{B}_n(Y)|$ for every n. Let N_0 be large enough that for all $n \geq N_0$, we have $|\mathcal{B}_n(Y)| \leq s^n$. Let $C_1 = \sum_{k=1}^{N_0-1} |\mathcal{B}_k(Y)|$. Then the number of blocks in X of length n with self-overlap of more than αn is at most

$$\begin{aligned}
\sum_{k=1}^{\lceil \alpha n \rceil} |\mathcal{B}_k(Y)| &\leq \sum_{k=1}^{N_0-1} |\mathcal{B}_k(Y)| + \sum_{k=N_0}^{\lceil \alpha n \rceil} |\mathcal{B}_k(Y)| \\
&\leq C_1 + \sum_{k=N_0}^{\lceil \alpha n \rceil} s^k \\
&\leq C_1 + \frac{s^{\alpha n+2} - s^{N_0}}{s-1} \\
&\leq C_2 s^{\alpha n}
\end{aligned}$$

where

$$C_2 = C_1 + \frac{s^2}{s-1}$$

Let $N > \frac{(1-\alpha)h(Y)-\alpha\varepsilon}{\log C_2}$. Then, for $n \geq N$, the number of blocks in Y of length n with no self-overlap by more than αn is at least

$$\begin{aligned}
|\mathcal{B}_n(Y)| - \sum_{k=1}^{\lceil \alpha n \rceil} |\mathcal{B}_k(Y)| &\geq r^n - C_2 s^{\alpha n} \\
&= r^n \left(1 - C_2 \left(\frac{s^\alpha}{r}\right)^n\right) \\
&> (1-\exp(-bn))\exp(nh(Y))
\end{aligned}$$

where we can take

$$\begin{aligned}
b &= \frac{1}{2}\log\left(C_2\left(\frac{r}{s^\alpha}\right)^N\right) \\
&= (1-\alpha)h(Y) - \alpha\varepsilon - \frac{1}{N}\log C_2
\end{aligned}$$

which is positive by the choice of N. □

We now control the entropy loss incurred by requiring a given long block to appear syndetically.

Lemma 5.4.2. *Let Y be a strongly irreducible subshift with $h(Y) > 0$. For every $\varepsilon > 0$, there exist $\beta \in (0,1)$ and $N \geq 1$ such that for every $n \geq N$ and every $\theta \in \mathcal{B}_{\lfloor \beta n \rfloor}(Y)$, the subshift $S \subset Y$ consisting of points $y \in Y$ in which θ appears at least once in $y_{[i,i+n)}$ for every $i \in \mathbb{Z}$ has entropy at least $h(Y)-\varepsilon$.*

Proof. Let g be the gap for Y. Let $\beta = \min\{\varepsilon/(4h(Y)), 1/2\}$ and let $N = \lceil 4(2g-1)h(Y)/\varepsilon \rceil$. Let $n \geq N$ and fix $\theta \in \mathcal{B}_{\lfloor \beta n \rfloor}(Y)$. For $m \geq n$, and for all $u_1, \ldots, u_k \in \mathcal{B}_{n-\lfloor \beta n \rfloor - 2g}(Y)$, where $k = \lfloor m/n \rfloor$, there exist $v_1^\pm, \ldots, v_k^\pm \in \mathcal{B}_g(Y)$ and $v_0 \in \mathcal{B}_{m-kn}(Y)$ such that

$$v_0 \theta v_1^- u_1 v_1^+ \theta v_2^- u_2 v_2^+ \ldots \theta v_k^- u_k v_k^+ \in \mathcal{B}_m(Y)$$

Therefore, by manipulation of logarithms and the fact that $h(Y) = \inf_{\ell \geq 1} \frac{1}{\ell} \log |\mathcal{B}_\ell(Y)|$ by definition,

$$\begin{aligned}
|\mathcal{B}_m(S)| &\geq |\mathcal{B}_{n-\lfloor \beta n \rfloor - 2g}(Y)|^{\lfloor m/n \rfloor} \\
\frac{1}{m} \log |\mathcal{B}_m(S)| &\geq \frac{1}{m} \log \left(|\mathcal{B}_{n-\lfloor \beta n \rfloor - 2g}(Y)|^{(m-1)/n} \right) \\
&= \left(1 - \frac{1}{m}\right) \frac{1}{n} \log |\mathcal{B}_{\lfloor (1-2\beta) n \rfloor - 2g}(Y)| \\
&\geq \left(1 - \frac{1}{m}\right) \frac{1}{n} (n - \lfloor \beta n \rfloor - 2g) h(Y) \\
&> h(Y) - \varepsilon/2
\end{aligned}$$

for large enough m, where the final inequality follows from the choices of β and N. We conclude that $h(S) = \liminf_{m \to \infty} \frac{1}{m} |\mathcal{B}_n(S)| > h(Y) - \varepsilon$. □

Proof of Proposition 5.3.2. It is clearly enough to prove the result for u_1, u_2 sufficiently long, since we can then pass to subwords of u_1, u_2. By Lemma 5.4.2, let $\beta \in (0,1)$, m sufficiently large, and $\theta \in \mathcal{B}_{\beta m}(Y) \setminus \mathcal{B}(W)$ be such that the subshift $S \subset Y$ defined by requiring at least one appearance of θ in any block of length m has $h(S) > 0$. Let $\alpha \in (0,1)$ be arbitrary, and let $n > (m+k)/(1-\alpha)$ be large enough that, by Lemma 5.4.1, there exists $\mu \in \mathcal{B}_n(S)$ such that μ has no self-overlap by more than αn, in particular by more than $n - (m+k)$.

Let $u_1 \in \mathcal{B}_{k_1}(S)$, $u_2 \in \mathcal{B}_{k_2}(S)$ with $k_1, k_2 \geq m$ and let $v_1, v_2 \in \mathcal{B}_k(Y)$. Then μ cannot appear in $u_1 v_2 \mu v_2 u_2$ except at the position explicitly indicated. Indeed, μ cannot appear at a position shifted by at most $m+k$—otherwise, μ would overlap itself by too much—and it cannot appear at a position shifted by more than $m+k$, as it would then overlap with u_1 or u_2 in a block of length at least m, contradicting the fact that $\mu \in \mathcal{B}(S)$, and thus has θ as a subword. □

5.4.2 Entropy and periodic points

We first show that at least a positive fraction of periodic points in Y of sufficient least period have a preimage of equal least period, and in particular that their growth is exponential with rate $h(Y)$.

Proposition 5.4.3. *Let X be a mixing SFT, Y a mixing sofic shift, and $\pi : X \to Y$ a factor code. Then $\lim_{n \to \infty} \frac{1}{n} \log r_n(\pi) = h(Y)$.*

Proof. Let g be the mixing gap of X. By Lemma 5.4.1, let $b > 0$ and $N > 3g$ be such that, for all $n \geq N$, the number of blocks in Y of length $n - g$ with no self-overlap by more than $n/3$ is at least $c \exp(nh(Y))$, where we may take $c = \frac{1}{2} \exp(-gh(X))$. For each block $v \in \mathcal{B}_{n-g}(Y)$, there exists a periodic point $x \in X$ with $\pi(x)_{[0,n-g)} = v$ such that $\operatorname{per}(x)$ divides n. Thus $\pi(x)$ is also periodic with least period dividing n. Moreover, if v has no self-overlap by more than $n/3$, then in fact $\operatorname{per}(\pi(x)) = n$. Therefore $r_n(\pi) \geq c \exp(nh(Y))$, so $\liminf_{n \to \infty} \frac{1}{n} \log r_n(\pi) \geq h(Y)$, matching $\limsup_{n \to \infty} \frac{1}{n} \log r_n(\pi) \leq \lim_{n \to \infty} \frac{1}{n} \log q_n(Y) = h(Y)$. □

We now assemble the quantitative results proven so far.

Proposition 5.4.4. *Let X be a mixing SFT, Y a mixing sofic shift, and $\pi : X \to Y$ a factor code. Let $\varepsilon > 0$ and $N_0 \geq 1$. Then there exist $N_1 \geq N_0$ and proper subshifts $W \subsetneq Y$, $V = \pi^{-1}(W) \subset X$, such that: $h(W) > h(Y) - \varepsilon$; for $n \leq N_1$, $r_n(\pi|_V) = r_n(\pi)$; and for $n \geq N_1$, $r_n(\pi|_V) > \exp(n(h(Y) - \varepsilon))$.*

Proof. By Lemma 5.2.1 and Lemma 5.3.10, let $V_1 \subset X$ be a mixing SFT such that $h(Y) - \varepsilon/2 < h(\pi(V_1)) < h(Y)$. Let $W_1 = \pi(V_1)$. By Proposition 5.4.3, let $N_1 \geq N_0$ be such that for any $n \geq N_1$, we have $\frac{1}{n} \log r_n(\pi|_{V_1}) > h(W_1) - \varepsilon/2 > h(Y) - \varepsilon$. Let $W = W_1 \cup \bigcup_{n=1}^{N_1} R_n(\pi)$ and $V = \pi^{-1}(W)$. Then $r_n(\pi|_V) = r_n(\pi)$ for all $n \leq N_1$.

To see that $W \neq Y$, observe that the only n-blocks in W that may not be in W_1 are those in the low-order periodic points that have been adjoined, which are bounded in number by a constant.

That is, $|\mathcal{B}_n(W)| \leq |\mathcal{B}_n(W_1)| + C$ for all $n \geq N_1$, where we can take $C = \sum_{k=1}^{N_1} k|R_k(\pi)|$. Thus $h(W) = h(W_1) < h(Y)$. $\square$

Proposition 5.4.4 is the final input to the proof of Proposition 5.3.7, and thus of Theorem 5.1.1.

Proof of Proposition 5.3.7. Let $\varepsilon = h(Y) - h(Z)$. Let $N_0 \geq 1$ be large enough that for all $n \geq N_0$,

$$\frac{1}{n} \log \max\{q_n(Z), |\mathcal{B}_n(Z)|\} < h(Z) + \frac{\varepsilon}{4}$$

By Proposition 5.4.4, let $W \subset Y$, $V = \pi^{-1}(W) \subset X$, and $N_1 \geq N_0$ be such that $h(W) > h(Y) - \frac{\varepsilon}{4}$, $r_n(\pi|_V) = r_n(\pi)$ for all $n \leq N_1$, and $\frac{1}{n}\log r_n(\pi|_V) > h(Y) - \frac{\varepsilon}{4}$ for all $n \geq N_1$. Note that $h(W) > h(Z) + \frac{\varepsilon}{2}$ and that $q_n(Z) \leq r_n(\pi|_V)$ for all $n \geq 1$.

Let g be the mixing gap of X. By Proposition 5.3.2, let $\mu \in \mathcal{B}(Y) \setminus \mathcal{B}(W)$ be a (Y, W, g) stamp. Let $\ell = |\mu| + 2g$. Then since $h(Z) < h(W)$, there exists N sufficiently large so that for all $n \geq N$, in particular for $N + \ell \leq n \leq 2N + \ell - 1$, we have $|\mathcal{B}_n(Z)| < |\mathcal{B}_{n-\ell}(W)|$. $\square$

5.5 Proofs of Lemma 5.2.6 and Corollary 5.1.3

We first use Proposition 5.4.3, along with facts about Markov approximations in Section 5.2.2, to prove Lemma 5.2.6, which reduces Theorem 5.1.1 to the case where Z is a 1-step SFT.

Proof of Lemma 5.2.6. We use the properties of Markov approximations mentioned in Section 5.2. Let $\varepsilon = h(Y) - h(Z)$. Let m_0 be such that $h(Z_{m_0}) < h(Z) + \varepsilon/3$, where Z_{m_0} is the m_0th Markov approximation to Z, and such that, by Proposition 5.4.3, for all $n \geq m_0$, $\frac{1}{n}\log r_n(\pi) > h(Y) - \varepsilon/3$. Let $m_1 \geq m_0$ be such that $\frac{1}{n}\log q_n(Z_{m_0}) < h(Z_{m_0}) + \varepsilon/3$ for all $n \geq m_1$. Let $m_2 \geq m_1$ be such that for all periodic points $z \in P(Z_{m_1}) \setminus Z$ (under the natural embedding $Z \hookrightarrow Z_{m_1}$) with $\mathrm{per}(z) \leq m_1$, we have $z_{[0,m_2)} \notin \mathcal{B}_{m_2}(Z)$. Then Z_{m_2} satisfies $q_n(Z_{m_2}) = q_n(Z) \leq r_n(\pi)$ for all $n \leq m_1$. Moreover, since $Z_{m_2} \subset Z_{m_0}$, $\frac{1}{n}\log q_n(Z_{m_2}) \leq \frac{1}{n}\log q_n(Z_{m_0})$ for all n; in particular,

$$\begin{aligned}\frac{1}{n}\log q_n(Z_{m_2}) &< h(Z_{m_0}) + \varepsilon/3 \\ &< h(Z) + 2\varepsilon/3 \\ &< h(Y) - \varepsilon/3 \\ &< \frac{1}{n}\log r_n(\pi)\end{aligned}$$

for all $n \geq m_1$. Taking $Z' = Z_{m_2}^{[m_2]}$ to be the m_2th higher block shift, the lemma is proved. $\square$

To prove Corollary 5.1.3, in the mixing sofic case, we use Lemma 5.2.2 to handle low-order periodic point obstructions, with periodic points of sufficiently high order controlled by Proposition 5.4.3. To handle the arbitrary case, we first give an improved Markov approximation (Lemma 5.5.2), embedding an arbitrary subshift into a mixing SFT with only slightly greater entropy. The construction uses Lemma 5.3.8; in Lemma 5.5.1 we estimate the entropy of the mixing SFT constructed in Lemma 5.3.8.

Lemma 5.5.1. *Let X be a mixing SFT with gap g and let $V_0 \subset X$ be an SFT. Let $k \geq g$ and let $\mu \in \mathcal{B}(X) \setminus \mathcal{B}(V_0)$ be an (X, V_0, k) stamp. For any $\varepsilon > 0$, there exists $N \geq |\mu|$ such that $h(V_1) < h(V_0) + \varepsilon$, where V_1 (depending on N) is as in Lemma 5.3.8.*

Proof. Let $N_0 \geq 1$ be such that for all $n \geq N_0$ we have $\frac{1}{n}\log|\mathcal{B}_n(V_0)| + \varepsilon/4$. Let $N > 2N_0$ be such that

$$\frac{1}{N}\max\{\log N, \log|\mathcal{B}_k(X)|^2, \log|\mathcal{B}_{N_0}(V_0)|\} < \frac{\varepsilon}{4}$$

We will show that $\frac{1}{N}\log|\mathcal{B}_N(V_1)| < h(V_0)+\varepsilon$. Consider a block of length N in V_1. Such a block can contain at most one full or partial block of the form $\gamma^+\mu\gamma^-$ where $\gamma^\pm \in \mathcal{B}_k(X)$. The μ, if present, can begin at any of the N positions. The rest of the block of length N, outside the block $\gamma^+\mu\gamma^-$, consists of one or two blocks from V_0, with length totalling at most N. We thus have

$$|\mathcal{B}_N(V_1)| \le N|\mathcal{B}_k(X)|^2 \max_{0\le \ell\le N/2} |\mathcal{B}_\ell(V_0)||\mathcal{B}_{N-\ell}(V_0)|$$

If $0 \le \ell \le N_0$, then $|\mathcal{B}_\ell(V_0)||\mathcal{B}_{N-\ell}(V_0)| \le |\mathcal{B}_{N_0}(V_0)||\mathcal{B}_N(V_0)|$, so

$$\begin{aligned}\frac{1}{N}\log\left(|\mathcal{B}_\ell(V_0)||\mathcal{B}_{N-\ell}(V_0)|\right) &\le \frac{1}{N}\log|\mathcal{B}_{N_0}(V_0)| + \frac{1}{N}\log|\mathcal{B}_N(V_0)| \\ &< h(V_0)+\frac{\varepsilon}{2}\end{aligned}$$

If $N_0 \le \ell \le N/2$, then

$$\begin{aligned}\frac{1}{N}\log\left(|\mathcal{B}_\ell(V_0)||\mathcal{B}_{N-\ell}(V_0)|\right) &= \frac{\ell}{N}\frac{1}{\ell}\log|\mathcal{B}_\ell(V_0)| + \frac{N-\ell}{N}\frac{1}{N-\ell}\log|\mathcal{B}_{N-\ell}(V_0)| \\ &< \frac{\ell}{N}\left(h(V_0)+\frac{\varepsilon}{4}\right) + \frac{N-\ell}{N}\left(h(V_0)+\frac{\varepsilon}{4}\right) \\ &= h(V_0)+\frac{\varepsilon}{4}\end{aligned}$$

Therefore

$$\begin{aligned}\frac{1}{N}\log|\mathcal{B}_N(V_1)| &< \frac{1}{N}\log N + \frac{1}{N}\log|\mathcal{B}_k(X)|^2 + h(V_0) + \frac{\varepsilon}{2} \\ &< h(V_0)+\varepsilon\end{aligned}$$

by the above choice of N. □

We also use the following standard lemma improving the construction of the Markov approximation. For completeness, we include a proof using Lemma 5.3.8.

Lemma 5.5.2. *Let Z be a subshift and let $\varepsilon > 0$. Then there exists a mixing SFT V containing Z with $h(V) < h(Z)+\varepsilon$.*

Proof. Let m be large enough that $h(Z_m) < h(Z)+\varepsilon/2$, where Z_m is the mth Markov approximation to Z. Let X be the full shift on the alphabet of Z. Certainly $Z_m \subseteq X$. If $Z_m = X$, then we can take $V = X$. If $Z_m \ne X$, then by Proposition 5.3.2, let μ be an (X, Z_m, k) stamp for some $k \ge 0$. Let $V_0 = Z_m$ and $V = V_1$ as in Lemma 5.3.8 where N is large enough that, by Lemma 5.5.1, we have $h(V) < \varepsilon/2$. Thus V is indeed a mixing SFT containing Z with $h(V) < h(Z)+\varepsilon$. □

Proof of Corollary 5.1.3. We first consider the case in which Z is mixing sofic. Let $\tilde{Z}$ be a mixing SFT and $\chi_0 : \tilde{Z} \to Z$ an almost invertible factor code. If already $q_n(\tilde{Z}) \le r_n(\pi)$ for all $n \ge 1$, then we can take $Z' = \tilde{Z}$ and apply Theorem 5.1.1 immediately to construct the claimed embedding $\psi : Z' \to X$. However, if $q_n(\tilde{Z}) > r_n(\pi)$ for some n, so that $X, Y, \pi, \tilde{Z}$ violate the hypotheses of Theorem 5.1.1, then we need to construct a further extension of $\tilde{Z}$ which satisfies the hypotheses of Theorem 5.1.1. The construction, consisting a tower of extensions via Lemma 5.2.2, is as follows.

By Proposition 5.4.3, since $h(\tilde{Z}) = h(Z) < h(Y)$, there are at most finitely many n such that $q_n(\tilde{Z}) > r_n(\pi)$. Let N denote the greatest such n. Let $C = \sum_{k=1}^{N} \max\{0, q_k(\tilde{Z}) - r_k(\pi)\}$. That is, C is the number of periodic points by which $X, Y, \pi, \tilde{Z}$ violate the hypotheses of Theorem 5.1.1. For $1 \le k \le N$ and $1 \le \ell \le k^{-1}\max\{0, q_k(\tilde{Z}) - r_k(\pi)\}$, let $z_{k,\ell}$ be periodic points with pairwise disjoint orbits, such that $\mathrm{per}(z_{k,\ell}) = k$. For a given k, the union of the orbits of the points $z_{k,\ell}$ has cardinality $\max\{0, q_k(\tilde{Z}) - r_k(\pi)\}$. Let $C' = \sum_{k=1}^{N} k^{-1}\max\{0, q_k(\tilde{Z}) - r_k(\pi)\}$ (counting orbits, rather than points), and let $\{z^{(j)}\}_{j=1}^{C'} = \{z_{k,\ell}\}_{k,\ell}$ be an enumeration of the points $z_{k,\ell}$.

Again by Proposition 5.4.3, let $M > N$ be large enough that for all $n \geq M$, we have $q_n(\tilde{Z}) + Cn \leq r_n(\pi)$. We now repeatedly apply Lemma 5.2.2. Let $Z^{(0)} = \tilde{Z}$. For $1 \leq j \leq C'$, let $Z^{(j)}$ be a mixing SFT and $\chi^{(j)} : Z^{(j)} \to Z^{(j-1)}$ an almost invertible factor code such that the preimage of the orbit of z_j under $\chi^{(j)}$ is a single orbit of length $M\mathrm{per}(z_j)$, and such that every periodic point in $Z^{(j-1)}$ not in the orbit of z_j has a unique preimage under $\chi^{(j)}$. Let $\eta^{(1)} = \chi^{(1)}$ and $\eta^{(j+1)} = \eta^{(j)} \circ \chi^{(j+1)}$. Let $Z' = Z^{(C')}$ and $\eta = \eta^{(C')} : Z' \to \tilde{Z}$. Certainly η is almost invertible, so $h(Z') = h(\tilde{Z}) < h(Y)$. We claim that $q_n(Z') \leq r_n(\pi)$ for all $n \geq 1$. Indeed, for each j, if $\mathrm{per}(z_j) = k$, then we have $q_k(Z^{(j)}) = q_k(Z^{(j-1)}) - k$, $q_{Mk}(Z^{(j)}) = q_{Mk}(Z^{(j-1)}) + Mk$, and $q_n(Z^{(j)}) = q_n(Z^{(j-1)})$ for all $n \notin \{k, Mk\}$. Therefore $q_k(Z') = r_k(\pi)$, and

$$\begin{aligned} q_{Mk}(Z') &= q_{Mk}(\tilde{Z}) + M \max\{0, q_k(\tilde{Z}) - r_k(\pi)\} \\ &\leq q_{Mk}(\tilde{Z}) + CM \\ &\leq r_{Mk}(\pi) \end{aligned}$$

where the last inequality follows from the choice of M. Therefore X, Y, π, Z' satisfy the hypotheses of Theorem 5.1.1, so there exists a sliding block code $\psi : Z' \to X$ such that $\pi \circ \psi$ is injective. This concludes the proof in the case that Z is mixing sofic.

We now handle the general case, where Z is an arbitrary subshift with $h(Z) < h(Y)$. By Lemma 5.5.2, let V be a mixing SFT containing Z with $h(V) < h(Y)$. By the mixing sofic case, let V' be a mixing SFT such that X, Y, π, V' satisfy the hypotheses of Theorem 5.1.1, and let $\chi : V' \to V$ be an almost invertible factor code. Let $Z' = \chi^{-1}(Z)$. Then $\chi|_{Z'}$ is still finite-to-one, which concludes the proof. $\square$

Bibliography

[1] R.L. Adler. "A note on the entropy of skew product transformations". In: *Proc. Am. Math. Soc.* 14.4 (1963).

[2] R.L. Adler, L.W. Goodwyn, and B. Weiss. "Equivalence of topological Markov shifts". In: *Isr. J. Math.* 27.1 (1977), pp. 49–63.

[3] R.L. Adler and B. Weiss. "Entropy, a complete metric invariant for automorphisms of the torus". In: *Proc. Nat. Acad. Sci.* 57.6 (1967), pp. 1573–1576.

[4] D. Ananichev, V. Gusev, and M. Volkov. "Slowly synchronizing automata and digraphs". In: *MFCS 2010*. Ed. by Petr Hlinený and Antonín Kucera. Vol. 6281. Lect. Notes Comput. Sci. Springer-Verlag Berlin Heidelberg, 2010, pp. 55–65.

[5] J. Araújo, P. Cameron, and B. Steinberg. "Between primitive and 2-transitive: synchronization and its friends". In: *EMS Surv. Math. Sci.* 40.2 (2017), pp. 101–184.

[6] J. Ashley, B. Marcus, and S. Tuncel. "The classification of one-sided Markov chains". In: *Ergod. Th. Dynam. Syst.* 17.2 (1997), pp. 269–295.

[7] L. Babai. "Groups, graphs, algorithms: the graph isomorphism problem". In: *ICM 2018*. Ed. by B. Sirakov, P. Ney de Souza, and M. Viana. Vol. 3. World Scientific, 2018, pp. 3303–3320.

[8] V. Baladi. "Gibbs states and equilibrium states for finitely presented dynamical systems". In: *J. Stat. Phys.* 62.1/2 (1991), pp. 239–256.

[9] S. Barbieri et al. "Equivalence of relative Gibbs and relative equilibrium measures for actions of countable amenable groups". In: *Nonlinearity* 33.5 (2020), pp. 2409–2454.

[10] M.-P. Béal and D. Perrin. "A quadratic algorithm for road coloring". In: *Discrete Appl. Math.* 169 (2014), pp. 15–29.

[11] S. Berghout, S. Fernández, and E. Verbitskiy. "On the relation between Gibbs and g-measures". In: *Ergod. Th. Dynam. Syst.* 39.12 (2018), pp. 3224–3249.

[12] M.V. Berlinkov. "On the probability of being synchronizable". In: *CALDAM 2016*. Ed. by S. Govindarajan and A. Maheshwari. Vol. 9602. Lect. Notes Comput. Sci. Springer International Publishing Switzerland, 2016, pp. 73–84.

[13] M.V. Berlinkov. "On two algorithmic problems about synchronizing automata". In: *DLT 2014*. Ed. by A.M. Shur and M.V. Volkov. Vol. 8633. LNCS. 2014, pp. 61–67.

[14] J. Berstel, D. Perrin, and C. Reutenauer. *Codes and Automata*. Vol. 129. Encycl. Math. Appl. Cambridge University Press, 2009.

[15] L. Borsato and S. MacDonald. "Conformal measures and the Dobrushin-Lanford-Ruelle equations". In: *Proc. Amer. Math. Soc.* 149.10 (2021), pp. 4355–4369.

[16] Luísa Borsato and Sophie MacDonald. *A Dobrushin-Lanford-Ruelle theorem for irreducible sofic shifts*. Preprint, arXiv:2007.05862. 2020.

[17] Luísa Borsato and Sophie MacDonald. *Conformal measures and the Dobrushin-Lanford-Ruelle equations*. Preprint, arXiv:2003.05532. 2020.

[18] M. Boyle. "Lower entropy factors of sofic systems". In: *Ergod. Th. Dynam. Syst.* 4 (1984), pp. 541–557.

[19] M. Boyle and S. Tuncel. "Regular isomorphism of Markov chains is almost topological". In: *Ergod. Th. Dynam. Syst.* 10.1 (1990), pp. 89–100.

[20] E. Breuillard and E. Le Donne. "On the rate of convergence to the asymptotic cone for nilpotent groups and subFinsler geometry". In: *Proc. Natl. Acad. Sci. USA* 110.48 (2013), pp. 19220–19226.

[21] G. Budzban and P. Feinsilver. "The generalized road coloring problem and periodic digraphs". In: *Appl. Algebra Engrg. Comm. Comput.* 22.1 (2011), pp. 21–35.

[22] D. Capocaccia. "A definition of Gibbs state for a compact set with $\mathbb{Z}^\nu$ action". In: *Comm. Math. Phys.* 48 (1976), pp. 85–88.

[23] T. Ceccherini-Silberstein and M. Coornaert. "On the density of periodic configurations in strongly irreducible subshifts". In: *Nonlinearity* 25 (2012), pp. 2119–2131.

[24] L. Cioletti, A.O. Lopes, and M. Stadlbauer. "Ruelle operator for continuous potentials and DLR-Gibbs measures". In: *Discrete Contin. Dyn. Syst.* 40.8 (2020), p. 4625.

[25] D.G. Corneil and C.C. Gotlieb. "An efficient algorithm for graph isomorphism". In: *J. Assoc. Comput. Mach.* 17.1 (1970), pp. 51–64.

[26] D.A. Dastjerdi and S. Jangjoo. "Dynamics and topology of S-gap shifts". In: *Topology Appl.* 159 (2012), pp. 26554–2661.

[27] M. Denker and M. Urbański. "On the existence of conformal measures". In: *Trans. Amer. Math. Soc.* 328.2 (1991).

[28] R.L. Dobrushin. "Gibbsian random fields for lattice systems with pairwise interactions". In: *Funct. Anal. Appl.* 2.4 (1969), pp. 292–301.

[29] D. Eppstein. "Reset sequences for monotonic automata". In: *SIAM J. Comput.* 19.3 (1990), pp. 500–510.

[30] J. Feldman and C.C. Moore. "Ergodic equivalence relations, cohomology, and von Neumann algebras. I". In: *Trans. Amer. Math. Soc.* 234.2 (1977), pp. 289–324.

[31] H. Furstenberg. *Recurrence in ergodic theory and combinatorial number theory*. M.B. Porter Lectures. Princeton University Press, 1981.

[32] M. Gromov. "Groups of polynomial growth and expanding maps (with an appendix by Jacques Tits)". In: *Publications Mathématiques de l'IHÉS* 53 (1981), pp. 53–78.

[33] G. Han, B. Marcus, and C. Wu. *Markov capacity for codes with an unambiguous symbol*. arXiv preprint. 2022. URL: https://arxiv.org/abs/2210.12251.

[34] P. de la Harpe. *Topics in geometric group theory*. Chicago Lectures in Mathematics. Chicago, 2000.

[35] R.V.L. Hartley. "Transmission of information". In: *Bell Syst. Tech. J.* 7.3 (1928), pp. 535–563.

[36] N.T.A. Haydn and D. Ruelle. "Equivalence of Gibbs and equilibrium states for homeomorphisms satisfying expansiveness and specification". In: *Comm. Math. Phys.* 148 (1992), pp. 155–167.

[37] K. Culik II, J. Karhumäki, and J. Kari. "A note on synchronized automata and road coloring problem". In: *DLT 2001*. Ed. by W. Kuich, G. Rozenberg, and A. Salomaa. Vol. 2295. Lect. Notes Comput. Sci. Springer-Verlag Berlin Heidelberg, 2002, pp. 175–185.

[38] O.E. Lanford III and D. Ruelle. "Observables at infinity and states with short range correlations in statistical mechanics". In: *Comm. Math. Phys.* 13.3 (1969), pp. 194–215.

[39] J. Kari. “Synchronizing finite automata on Eulerian digraphs”. In: *Theor. Comput. Sci.* 295 (2003), pp. 223–232.

[40] A.S. Kechris. *Classical descriptive set theory*. Springer, 1995.

[41] A.S. Kechris. *The theory of countable Borel equivalence relations*. Preprint. http://www.math.caltech.edu/~kechris/. 2019.

[42] G. Keller. *Equilibrium states in ergodic theory*. London Mathematical Society Student Texts 42. Cambridge University Press, 1998.

[43] J.C. Kieffer. “Zero-error stationary coding over stationary channels”. In: *Prob. Th. Rel. Fields* 56 (1981), pp. 113–126.

[44] B. Kimura. “Gibbs measures on subshifts”. MA thesis. University of São Paulo, 2015.

[45] B. Kitchens and S. Tuncel. *Finitary measures for subshifts of finite type and sofic systems*. Vol. 58. Mem. Amer. Math. Soc. 338. 1985.

[46] A.N. Kolmogorov. “Entropy per unit time as a metric invariant of automorphisms”. In: *Doklady Akad. Nauk SSSR* 124 (1959). (Russian), pp. 754–755.

[47] A.N. Kolmogorov. “On certain asymptotic characteristics of completely bounded metric spaces”. In: *Doklady Akad. Nauk SSSR* 108 (1956). (Russian), pp. 385–388.

[48] W. Krieger. “On the subsystems of topological Markov chains”. In: *Ergod. Th. Dynam. Syst.* 2.2 (1982), pp. 195–203.

[49] D. Lind and B. Marcus. *An introduction to symbolic dynamics and coding*. Cambridge University Press, 1995.

[50] Sophie MacDonald. *Encoding subshifts through sliding block codes*. Preprint, arXiv:2210.08150. 2022.

[51] Sophie MacDonald. *The road problem and homomorphisms of directed graphs*. Preprint, arXiv:2201.12942. 2022.

[52] B. Marcus. “Sofic systems and encoding data”. In: *IEEE Trans. Info. Th.* IT-31.3 (1985), pp. 366–377.

[53] B. Marcus and R. Pavlov. “Approximating entropy for a class of $\mathbb{Z}^2$ Markov random fields and pressure for a class of functions on $\mathbb{Z}^2$ shifts of finite type”. In: *Ergod. Th. Dynam. Syst.* 33.1 (2013), pp. 186–220.

[54] B. Marcus, K. Petersen, and S. Williams. “Transmission rates and factors of Markov chains”. In: *Conference in modern analysis and probability: Yale University, New Haven CT, June 8–11, 1982*. Ed. by R. Beals et al. Vol. 26. Contemp. Math. Amer. Math. Soc., 1984, pp. 279–293.

[55] R. Meester and J.E. Steif. “Higher-dimensional subshifts of finite type, factor maps and measures of maximal entropy”. In: *Pacific J. Math.* 200.2 (2001), pp. 497–510.

[56] T. Meyerovitch. “Gibbs and equilibrium measures for some families of subshifts”. In: *Ergod. Th. Dynam. Syst.* 33.3 (2013), pp. 934–953.

[57] G.L. Miller. “Graph isomorphism, general remarks”. In: *J. Comput. Syst. Sci.* 18 (1979), pp. 128–142.

[58] S.R. Muir. “A new characterization of Gibbs measures on $\mathbb{N}^{\mathbb{Z}^d}$”. In: *Nonlinearity* 24 (2011), pp. 2933–2952.

[59] S.R. Muir. “Gibbs/equilibrium measures for functions of multidimensional shifts with countable alphabets”. PhD thesis. University of North Texas, 2011.

[60] J. von Neumann. “Zur Operatorenmethode in der klassiche Mechanik”. In: *Ann. of Math.* 33.3 (1932). German, pp. 587–642.

[61] G.L. O’Brien. “The road-colouring problem”. In: *Isr. J. Math.* 39.1–2 (1981), pp. 145–154.

[62] S.J. Patterson. "The limit set of a Fuchsian group". In: *Acta Math.* 136 (1976), pp. 241–273.

[63] K. Petersen and K. Schmidt. "Symmetric Gibbs measures". In: *Trans. Amer. Math. Soc.* 349.7 (1997), pp. 2775–2811.

[64] M. Pollicott and T.M.W. Kempton. "Factors of Gibbs measures for full shifts". In: *Entropy of hidden Markov processes and connections to dynamical systems.* Ed. by B. Marcus, K. Petersen, and T. Weissman. Cambridge University Press, 2011. Chap. 8, pp. 246–257.

[65] E. Riehl. *Category Theory in Context.* Aurora. Dover, 2017.

[66] D. Ruelle. *Thermodynamic formalism: the mathematical structures of equilibrium statistical mechanics.* 2nd. Cambridge, 2004.

[67] V. Salo. *Cohomology for extension problems in symbolic/topological dynamics?* MathOverflow. URL: https://mathoverflow.net/q/424145.

[68] O.M. Sarig. "Lecture notes on thermodynamic formalism for countable Markov shifts". Unpublished lecture notes available from the author's webpage. 2009.

[69] C. Shannon. "A mathematical theory of communication". In: *Bell Syst. Tech. J.* 27 (1948), pp. 379–423, 623–656.

[70] B. Simon. *Convexity: an analytic viewpoint.* Vol. 187. Cambridge Tracts in Mathematics. Cambridge University Press, 2011.

[71] Ya. Sinai. "On the concept of entropy for a dynamic system. (Russian)". In: *Dolk. Akad. Nauk SSSR* 124 (1959), pp. 768–771.

[72] M. Stoll. "On the asymptotics of 2-step nilpotent groups". In: *J. London. Math. Soc.* 58.1 (1998), pp. 38–48.

[73] K. Thomsen. "On the structure of a sofic shift space". In: *Trans. Amer. Math. Soc.* 356.9 (2004), pp. 3557–3619.

[74] A.N. Trahtman. "The road coloring problem". In: *Isr. J. Math.* 172.1 (2009), pp. 51–60.

[75] V. Vorel. "Subset synchronization and careful synchronization of binary finite automata". In: *Intern. J. Found. Comput. Sci.* 27.5 (2016), pp. 557–577.

[76] P. Walters. *An introduction to ergodic theory.* Springer, 1982.

[77] H.W. Wielandt. "Unzerlegbare, nicht negative Matrizen". In: *Math. Z.* 52 (1950). (German), pp. 642–648.

[78] J. Yoo. "Decomposition of infinite-to-one factor codes and uniqueness of relative equilibrium states." In: *J. Mod. Dyn.* 13 (2018), pp. 271–284.