

# SHIFTS OF FINITE TYPE OBTAINED BY FORBIDDING A SINGLE PATTERN

NISHANT CHANDGOTIA<sup>1</sup>, BRIAN MARCUS<sup>2</sup>, JACOB RICHEY<sup>3</sup>, AND CHENGYU WU<sup>4</sup>

ABSTRACT. Given a finite word  $w$ , Guibas and Odlyzko (*J. Combin. Theory Ser. A*, 30, 1981, 183-208) showed that the autocorrelation polynomial  $\phi_w(t)$  of  $w$ , which records the set of self-overlaps of  $w$ , explicitly determines for each  $n$ , the number  $|\mathcal{B}_n(w)|$  of words of length  $n$  that avoid  $w$ . We consider this and related problems from the viewpoint of symbolic dynamics, focusing on the setting of  $X_{\{w\}}$ , the space of all bi-infinite sequences that avoid  $w$ . We first summarize and elaborate upon (*J. Combin. Theory Ser. A*, 30, 1981, 183-208) and other work to show that the sequence  $|\mathcal{B}_n(w)|$  is equivalent to several invariants of  $X_{\{w\}}$ . We then give a finite-state labeled graphical representation  $L_w$  of  $X_{\{w\}}$  and show that  $w$  can be recovered from the graph isomorphism class of the unlabeled version of  $L_w$ . Using  $L_w$ , we apply ideas from probability and Perron-Frobenius theory to obtain results comparing features of  $X_{\{w\}}$  for different  $w$ . Next, we give partial results on the problem of classifying the spaces  $X_{\{w\}}$  up to conjugacy. Finally, we extend some of our results to spaces of multi-dimensional arrays that avoid a given finite pattern.

## 1. INTRODUCTION

For a positive integer  $q$  and a finite word  $w = w_1 \cdots w_k$  over the alphabet  $[q] := \{0, \dots, q-1\}$ , consider the set  $\mathcal{B}_n(w)$  of words over  $[q]$  of length  $n$  which do *not* contain  $w$ , i.e., in which the word  $w$  is forbidden as a contiguous subword. A natural problem is to determine how the sequence of numbers  $|\mathcal{B}_n(w)|$  depends on the word  $w$ . A complete answer to this problem was given over forty years ago in seminal work by Guibas and Odlyzko [GO81a]: the starting point is an explicit formula for the generating function  $G_w(t) = \sum_{n=0}^{\infty} |\mathcal{B}_n(w)| t^n$ . Their formula is a simple function of the (*auto-*) *correlation polynomial*,  $\phi_u(t)$ , whose coefficients record the lengths of self-overlaps of  $u$ . The formula shows that

$$(1) \quad G_u(t) = G_v(t) \text{ iff } \phi_u(t) = \phi_v(t).$$

Since the sequence  $|\mathcal{B}_n(w)|$  grows at most exponentially in  $n$ , the correlation polynomial  $\phi_w(t)$  completely determines the sequence  $|\mathcal{B}_n(w)|$ .

A natural setting for problems of this type is symbolic dynamics. For a positive integer  $q$ , the *full  $q$ -shift* is the set  $[q]^{\mathbb{Z}}$  of bi-infinite sequences over the alphabet  $[q]$ . A *shift space* over  $[q]$  is a subset  $X \subset [q]^{\mathbb{Z}}$  defined by a list  $\mathcal{F}$  (finite or infinite) of finite forbidden words. A prominent example is  $X = X_{\{w\}}$ , the set of all bi-infinite sequences over  $[q]$  that do not contain the (finite) word  $w$ . This combines all the  $\mathcal{B}_n(w)$  above neatly into a single object. A simple example is to take  $q = 2$  and  $w = 11$ , and then  $X_{\{w\}}$  is called the *golden mean shift*.

---

2020 *Mathematics Subject Classification*. Primary 37B10; Secondary 60J10.

*Key words and phrases*. Correlation polynomial, entropy, shifts of finite type, Markov chains, conjugacy of SFTs.

<sup>1</sup>Tata Institute of Fundamental Research-Centre for Applicable Mathematics, <sup>2</sup> The University of British Columbia, <sup>3</sup>The Alfréd Rényi Institute of Mathematics, <sup>4</sup> The University of British Columbia

More generally one can consider a *shift of finite type (SFT)* which is a shift space defined by a finite list of finite forbidden words. Of course,  $X_{\{w\}}$  is an SFT defined by a single forbidden word  $w$ . Another prominent class of SFTs is defined by finite directed graphs  $G$ : the *edge shift*  $X_G$  is the set of bi-infinite edge walks on  $G$ .

Of particular interest in symbolic dynamics is the *entropy*  $h(X)$  which is the asymptotic growth rate of the number of allowed finite sequences in elements of the shift space  $X$ . Another useful object is the *zeta function*  $\zeta(X)$ , defined in Section 3, which is a function of a variable  $t$  that encapsulates the number of periodic sequences in the shift space  $X$  of all periods. We often write  $\zeta(X_{\{w\}})$  as  $\zeta_w(t)$ . See [LM21] for more background on symbolic dynamics.

In Section 2, we piece together results from Guibas and Odlyzko [GO81a], Lind [Lin89], and Eriksson [Eri97], to show in Proposition 2.1, that  $\phi_w(t)$ ,  $G_w(t)$ ,  $h(X_{\{w\}})$ ,  $\zeta(X_{\{w\}})$ , as well as a probabilistic quantity  $E(\tau_w)$ , all determine essentially the same information (here  $E(\tau_w)$  is the expected time to generate the word  $w$  in a process that generates symbols in  $[q]$  uniformly and independently). In Proposition 2.2 we summarize related results involving inequalities in these quantities. Our aim in both propositions is to clarify the various implications with simple proofs: to do so we mostly rely on statements and methods of previous authors, plus a few new ideas.

One of the main results of our paper (Theorem 5.4 of Section 5) is a far-reaching generalization of the sufficiency (“if” part) of (1) above, using an entirely different method. Consider a finite subset  $S$  of the  $d$ -dimensional lattice  $\mathbb{Z}^d$ . A *pattern* on  $S$  is an element  $w \in [q]^S$ . We consider an analogue of the self-overlap set, called the *self-agreement set* of  $w$  on  $S$ .

Given such a  $w$  and another finite subset  $T \subset \mathbb{Z}^d$ , one can consider the set  $\mathcal{B}_T(w)$  of patterns  $x \in [q]^T$  which do not contain any translate of  $w$ . We show that for any given  $S$  and  $u, v \in [q]^S$ , if the self-agreement sets of  $u$  and  $v$  on  $S$  are the same, then for all  $T \subset \mathbb{Z}^d$ ,  $|\mathcal{B}_T(u)| = |\mathcal{B}_T(v)|$ . The special case  $d = 1$  is the setting described above in terms of words where  $S = [0, k - 1]$ ,  $T = [0, n - 1]$ . A much harder question is to compare  $|\mathcal{B}_T(u)|$  and  $|\mathcal{B}_T(v)|$  when the patterns  $u, v \in [q]^S$  don’t necessarily have the same set of overlaps. While we cannot address this completely, we mention some preliminary ideas in Section 6.5.

The proof of this generalization is obtained using a simple inclusion-exclusion argument by replacing occurrences of  $u$  by  $v$  in patterns on finite subsets. This seems to be related to ideas in Jacquet and Szpankowski [JS15, Section 4.3]; see also [CHR24]. We also consider periodic configurations and show that whenever  $u, v \in [q]^S$  have the same “self-overlap” sets, then for any finite fundamental domain  $D \subset \mathbb{Z}^2$ , the number of periodic configurations with fundamental domain  $D$  that avoid  $u$  is the same as the number that avoid  $v$ . These results can be generalized to arbitrary dimension and possibly to the setting of shift spaces over arbitrary countable groups.

Returning to the case  $d = 1$ , given a word  $w$  we consider finite-state directed graphical representations of  $X_{\{w\}}$  in Section 3. This point of view allows us to make use of Perron-Frobenius theory. We give a simple explicit description of an edge-labelled graph  $L_w$  whose edge sequences are in one-to-one correspondence with elements of  $X_{\{w\}}$ . Here  $L_w$  is a simple variation of the well-known follower set graph [LM21, Section 3.2] from symbolic dynamics.

We show that the graph isomorphism class of the unlabelled version of  $L_w$  completely determines the word  $w$ , up to a possible permutation of the alphabet (Corollary 3.9). We also give certain graphical conditions on  $L_u$  and  $L_v$  which imply  $h(X_{\{u\}}) \geq h(X_{\{v\}})$ . This shows

how comparisons between different  $X_{\{w\}}$  can be viewed through a graphical lens, involving Perron-Frobenius theory. We show that the same graphical conditions imply  $E(\tau_u) \geq E(\tau_v)$ ; in fact, the graphical conditions imply that  $\tau_u$  stochastically dominates  $\tau_v$ .

For shift spaces  $X$  and  $Y$ , a *sliding block code*  $\phi : X \rightarrow Y$  is a stationary coding  $\phi(x) = y$ , with each symbol  $y_n$  determined by a “local” rule  $y_n = \Phi(x_{[n-m, n+a]})$ , for some fixed  $m, a$ . A *conjugacy* from one shift space  $X$  to another  $Y$  is a bijective sliding block code from  $X$  onto  $Y$ , in which case one says that  $X$  and  $Y$  are *conjugate*. The main classification problem in symbolic dynamics, still unsolved after many years, is the problem of deciding, given two SFTs  $X$  and  $Y$ , whether  $X$  and  $Y$  are conjugate.

It is not hard to see that given finite directed graphs  $G$  and  $H$ , the conjugacies  $\phi$  from  $X_G$  to  $X_H$  with  $m = a = 0$  for both  $\phi$  and  $\phi^{-1}$  are precisely those given by graph isomorphisms from  $G$  to  $H$ . But these are rather trivial conjugacies. It is also not hard to see that for any finite word  $w$  over  $[q]$ ,  $X_{\{w\}}$  is conjugate to the edge shift  $X_{L_w}$ , in a natural way. Thus, any graph isomorphism from  $L_u$  to  $L_v$  defines a conjugacy from  $X_{\{u\}}$  to  $X_{\{v\}}$ . Since the graph isomorphism class of  $L_u$  determines  $u$  (modulo a permutation of the letters) there doesn’t exist a conjugacy of this type from  $X_{\{u\}}$  to  $X_{\{v\}}$  (if  $u$  and  $v$  are essentially different).

Are there other conjugacies? Since the entropy  $h(X_{\{u\}})$  determines the correlation polynomial  $\phi_u(t)$  and entropy is preserved under conjugacy, a necessary condition for  $X_{\{u\}}$  to be conjugate to  $X_{\{v\}}$  is that  $\phi_u(t) = \phi_v(t)$ . In the case where  $\phi_u(t) = \phi_v(t) = t^{k-1}$ ,  $u$  and  $v$  have no non-trivial self-overlaps, and in Theorem 4.3 of Section 4, we show that  $X_{\{u\}}$  and  $X_{\{v\}}$  are conjugate (under an irreducibility condition). So, from the viewpoint of conjugacy of  $X_{\{u\}}$ , the words with no non-trivial self-overlaps are all the “same,” but from the viewpoint of graph isomorphism of  $L_u$ , these words are all “different.”

As a simple example,  $u = 110100$  and  $v = 111010$  have no non-trivial self-overlaps, and so  $X_{\{u\}}$  and  $X_{\{v\}}$  are conjugate. If  $u = 110110$  and  $v = 011011$  (the reverse of  $u$ ), then  $u$  and  $v$  have the same non-trivial set of self-overlaps, but we do not know if  $X_{\{u\}}$  and  $X_{\{v\}}$  are conjugate.

Instead of starting with a full shift and forbidding a single word, one could start with another simple (*ambient*) SFT  $X$  and then forbid a single word from  $X$ . For instance,  $X$  could be the SFT defined by forbidding a single word  $w$  of length two,  $X = X_{\{w\}}$ , in which case  $(X_{\{w\}})_{\{u\}}$  would be the SFT obtained by forbidding  $w$  and  $u$ . A few results in this framework, with  $w = 11$ , are sprinkled throughout the paper (see Sections 3.2 and 4.2). In fact, this is the framework considered by Lind [Lin89], where  $w$  is an arbitrary word of length 2.

There is a connection between this work and hitting times of stationary processes, or synonymously, the subject of escape rates in dynamical systems with holes. For a given stationary process, one is often concerned with the time taken to reach a particular state. A well-known instance of this is Penney’s game [Pen69], concerning the hitting time of certain sequences upon tossing of a fair coin, and the related ‘ABRACADABRA problem’. In the language of discrete dynamical systems, one has a phase space  $X$  and a self-map  $T : X \rightarrow X$ . For each  $x \in X$ , one forms an orbit of iterates  $x, T(x), T(T(x)), T(T(T(x))), \dots$ . The system has a hole  $H \subset X$  if  $T$  is defined only on  $X \setminus H$ , so if the orbit ever falls into  $H$  it cannot continue any further. This hole can be thought of as a state of a certain stationary process where the starting point  $x$  is chosen according to an invariant probability measure. In the context of our paper, one can start with an ambient shift space  $X$ . Forbidding a word  $w$  from  $X$  is similar to creating a hole for the cylinder set corresponding to  $w$ . In this case,

the rate of “escape” into the hole is given exactly by the difference in entropy between the ambient shift and the subshift with the hole removed. One can also consider a system with multiple holes.

This line of research traces back to [PY79] for expansive maps. Subsequently this has been studied extensively by several authors in a large body of work: [BY11] (for systems like rotations, expanding maps on the interval and shift spaces), [KL09] (for expanding maps and any absolutely continuous invariant measure), [FP12] (for Gibbs measures), [CA23] (for one dimensional shifts of finite type) and [ACT24] (for Markov chains). One key difference between these works and ours is the following: In a lot of these papers, they study the ‘hitting time/escape rate’ of the hole given an ambient system. In our case, we are more interested in studying the dynamics on the space where all orbits touching the hole  $H$  are removed, that is, instead of studying the dynamical system  $(X, T)$  with an emphasis on the visits to the hole  $H$  we look at the system  $(X \setminus (\cup_{i \in \mathbb{Z}} T^i(H)), T)$ . For more about these directions of research we refer the reader to the given citations and the references within.

### ACKNOWLEDGEMENTS

The first author would like to thank Aditya Thorat for many discussions and brainstorming ideas around the problems. He would like to thank Sharvari N. Tikekar and Haritha Cheriyaath for first introducing these problems to him. He was supported by INSA Associate fellowship and the SERB Startup research grant. The third author was partially supported by ERC Synergy under Grant No. 810115 - DYNASNET and ERC Consolidator Grant 772466 “NOISE”.

We also thank Sophie MacDonald for helpful discussions and acknowledge support from NSERC.

### 2. EARLIER RESULTS IN ONE DIMENSION

Some of the notations and definitions used in this section, such as full shift, shift of finite type (SFT), the SFT  $X_{\{w\}}$ , and entropy  $h(X_{\{w\}})$  were given in the introduction.

A finite directed graph  $G$  is called *irreducible* if for every ordered pair of vertices  $I, J$  of  $G$  there is a directed path from  $I$  to  $J$  (and otherwise called *reducible*). An SFT  $X$  is called *irreducible* if for any two allowed words  $u, v$ , there is an allowed word  $w$  such that  $uwv$ , the concatenation of these three words, is also allowed in  $X$ . Otherwise,  $X$  is a *reducible* SFT.

A *presentation* of  $X_{\{w\}}$  is a finite directed graph  $G$ , together with a labeling of its edges, such that  $X_{\{w\}}$  is the set of all bi-infinite label sequences of bi-infinite walks in  $G$ . It is well-known that  $X_{\{w\}}$  is irreducible if and only if it has an irreducible presentation (see, for example [LM21, Proposition 3.3.11]). A presentation is called a *conjugacy presentation* if the mapping from bi-infinite edge sequences to bi-infinite label sequences in  $X_{\{w\}}$  is one-to-one (and hence a bijection). This mapping is a sliding block code with  $m = a = 0$  and thus is a conjugacy between  $X_{\{w\}}$  and the edge shift  $X_G$ , hence the name conjugacy presentation. An example of a conjugacy presentation for any  $X_{\{w\}}$  is the follower set graph described at the beginning of Section 3.

For an irreducible graph  $G$ , the adjacency matrix  $A_G$  has a unique maximum (in modulus) eigenvalue  $\lambda_G$  such that  $\lambda_G > 0$ , called the *Perron eigenvalue*. It has a unique right (Perron) eigenvector and a unique left (Perron) eigenvector, up to scale, and they are both strictly positive vectors. Let  $\chi_G(t) = t^{-d}p_G(t)$  where  $p_G(t)$  is the characteristic polynomial of  $A_G$  and  $d$  is the unique integer such that  $\chi_G(t)$  has a nonzero constant term. One can show that

all graphs  $G$  which have a labeling that is a conjugacy presentation of  $X_{\{w\}}$  have the same  $\lambda_G$  and the same  $\chi_G(t)$  which we shall respectively denote by  $\lambda_w$  and  $\chi_w(t)$ , and call the Perron eigenvalue and characteristic polynomial of  $X_{\{w\}}$ . ( $\lambda_w$  is the largest root of  $\chi_w(t)$ .) For more information on Perron-Frobenius theory, see [LM21, Chapter 4].

The *entropy* of a shift space  $X$  is defined as

$$h(X) = \lim_{n \rightarrow \infty} (1/n) \log |\mathcal{B}_n(X)|$$

where  $\mathcal{B}_n(X)$  is the set of words of length  $n$  that occur in elements of  $X$ . In the case of a single forbidden word  $w$ ,  $\mathcal{B}_n(X)$  was denoted by  $\mathcal{B}_n(w)$  in the introduction, and

$$h(X_{\{w\}}) = \lim_{n \rightarrow \infty} (1/n) \log |\mathcal{B}_n(w)|.$$

It is well known that  $h(X_{\{w\}}) = \log \lambda_w$ .

We now define the correlation polynomial, which we mentioned in the introduction. For any positive integer  $q$  and any word  $w = w_1 \cdots w_k$  of length  $k$  over the alphabet  $[q]$ , define the *self-overlap set* of  $w$  by

$$\text{overlap}(w, w) := \{i \in \{1, 2, \dots, k\} : w_1 w_2 \cdots w_i = w_{k-i+1} w_{k-i+2} \cdots w_k\}$$

and the *correlation polynomial* of  $w$  by

$$\phi_w(t) = \sum_{i \in \text{overlap}(w, w)} t^{i-1}.$$

For instance  $\phi_{100010}(t) = t^5 + t$ .

It can be derived from [Lin89, Theorem 1] that

$$(2) \quad \chi_w(t) = (t - q)\phi_w(t) + 1.$$

Let  $p_n(X_{\{w\}})$  be the number of periodic points of  $X_{\{w\}}$  with period  $n$ . The *zeta function* of  $X_{\{w\}}$  is defined as  $\zeta_w(t) = \exp\left(\sum_{n=1}^{\infty} \frac{p_n(X_{\{w\}})}{n} t^n\right)$ . It is known that for any conjugacy presentation of  $X_{\{w\}}$ ,  $\zeta_w(t) = \frac{1}{\det(Id - tA_G)}$  where  $Id$  is the identity matrix having the same size as  $A_G$  (see, for example, [LM21, Theorem 6.4.6]). Thus, it can be readily seen that  $\chi_w(t)$  determines  $\zeta_w(t)$ . It follows that the information contained in  $\zeta_w(t)$  is exactly the set of non-zero eigenvalues of  $A_G$ .

We use  $\tau_w$  to denote the *hitting time* random variable of  $w$  by an i.i.d. sequence of  $\text{Uniform}([q])$  random variables  $(Y_i)_{i \in \mathbb{N}}$ , i.e.,

$$\tau_w = \min\{t \in \mathbb{N} : Y_{t-k+1} Y_{t-k+2} \cdots Y_t = w\}.$$

Earlier results on the relations between quantities introduced above are summarized in the following two propositions. We made an effort to present these results in an organized way, and to give straightforward proofs of the various implications.

**Proposition 2.1.** *Let  $u, v$  be two strings of length  $k$  over the  $q$ -ary alphabet ( $q \geq 2$ ). Then the following are equivalent:*

- (1)  $\text{overlap}(u, u) = \text{overlap}(v, v)$ ;
- (2)  $\phi_u(t) = \phi_v(t)$  for all  $t$ ;
- (3)  $\phi_u(q) = \phi_v(q)$ ;
- (4)  $E\tau_u = E\tau_v$ ;
- (5)  $\zeta_u(t) = \zeta_v(t)$  for all  $t$ ;
- (6)  $|\mathcal{B}_n(u)| = |\mathcal{B}_n(v)|$  for all  $n$ ;

- (7)  $|\mathcal{B}_n(u)| = |\mathcal{B}_n(v)|$  for all sufficiently large  $n$ ;  
 (8)  $h(X_{\{u\}}) = h(X_{\{v\}})$ .

**Proof:** (1)  $\iff$  (2)  $\iff$  (3): (1)  $\Rightarrow$  (2)  $\Rightarrow$  (3) is obvious. To see (3)  $\Rightarrow$  (1), note that for any  $w$ , the coefficients of  $\phi_w(t)$ , when viewed as a string, is the  $q$ -ary expansion of  $\phi_w(q)$ . Thus,  $\phi_u(q) = \phi_v(q)$  implies  $\phi_u(t)$  and  $\phi_v(t)$  are the same polynomial, which in turn gives (1) by the definition of the correlation polynomial.

(3)  $\iff$  (4): This follows from the fact that  $E\tau_w = q\phi_w(q)$  for any string  $w$  (See [GO81a, pg. 189], [Sol66], [Li80]).

(1)  $\Rightarrow$  (6): This follows from [GO81a, pg. 185], where they showed that for any word  $w$ , the generating function of  $\mathcal{B}_n(w)$  is determined by  $\phi_w(t)$ . More precisely,

$$G_w(t) = \frac{t\phi_w(t)}{1 - (t - q)\phi_w(q)},$$

where  $G_w(t)$  is the generating function defined at the beginning of the introduction.

(6)  $\Rightarrow$  (7)  $\Rightarrow$  (8): Clear from the definition of entropy.

(8)  $\Rightarrow$  (1): The ideas in this proof are largely in Eriksson [Eri97]. It suffices to show (8)  $\Rightarrow$  (3). We prove the contrapositive. Assume WLOG that  $\phi_u(q) > \phi_v(q)$ . We now show  $\lambda_u > \lambda_v$ . To this end, we first write

$$\phi_u(t) = \sum_{i=0}^{k-1} a_i t^i, \quad \phi_v(t) = \sum_{i=0}^{k-1} b_i t^i$$

where  $a_i, b_i \in \{0, 1\}$ . Define  $r := \max\{i : a_i = 1, b_i = 0\}$ . Note that  $r \leq k - 2$  and  $\phi_u(t) - \phi_v(t) > t^r - \sum_{i=0}^{r-1} t^i$  for all  $t > 0$ . Let  $\lambda^*$  be the largest positive root of  $t^r - \sum_{i=0}^{r-1} t^i = 0$ .

We claim that  $\lambda_v > \lambda^*$ . To see this, let  $w$  be a word of length  $k$  with trivial self-overlap, i.e.,  $\phi_w(t) = t^{k-1}$ . Since  $\phi_v(t) \geq \phi_w(t)$  for all  $t > 0$ , we first deduce from equation (2) that  $\chi_v(\lambda_w) = \chi_v(\lambda_w) - \chi_w(\lambda_w) \leq 0$  and therefore  $\lambda_v \geq \lambda_w$ . It remains to show  $\lambda_w > \lambda^*$ . To this end, first observe that when  $q = 2$ , the characteristic polynomial of  $X_{\{w\}}$  is  $\chi_w(t) = (t - 2)t^{k-1} + 1 = (t - 1)(t^{k-1} - \sum_{i=0}^{k-2} t^i)$ . Thus, when  $q = 2$ ,  $\lambda_w$  is the largest root of  $t^{k-1} - \sum_{i=0}^{k-2} t^i = 0$  and therefore  $\lambda_w > \lambda^*$  (because the largest root of  $t^s - \sum_{i=0}^{s-1} t^i = 0$  is an increasing function of  $s$ ). Finally, noting from equation (2) that the value of  $\lambda_w$  increases as  $q$  increases, we have  $\lambda_w > \lambda^*$  for all  $q \geq 2$  and the claim is proven.

Now, since  $\lambda_v < q$ , we infer from equation (2) that

$$\chi_u(\lambda_v) - \chi_v(\lambda_v) = (\lambda_v - q)[\phi_u(\lambda_v) - \phi_v(\lambda_v)] < (\lambda_v - q)[(\lambda_v)^r - \sum_{i=0}^{r-1} (\lambda_v)^i].$$

Since  $q > \lambda_v > \lambda^*$ , we have  $\chi_u(\lambda_v) = \chi_u(\lambda_v) - \chi_v(\lambda_v) < 0$  (recalling that  $\chi_v(\lambda_v) = 0$ ) and therefore  $\lambda_u > \lambda_v$ .

(5)  $\iff$  (8): (5)  $\Rightarrow$  (8); this follows from the fact that the entropy is determined by the largest positive eigenvalue of  $A_G$  above and the zeta function is determined by all of the non-zero eigenvalues of  $A_G$ . To show (8)  $\Rightarrow$  (5), it suffices to show (2)  $\Rightarrow$  (5). But this follows from equation (2) and the fact that  $\chi_w(t)$  determines  $\zeta_w(t)$ .  $\square$

Some of the equivalent equalities in Proposition 2.1 can be strengthened to inequalities:

**Proposition 2.2.** *Let  $u, v, q$  be defined as in Proposition 2.1. Then, the following are equivalent:*



- (A)  $\phi_u(q) > \phi_v(q)$ ;
- (B)  $E\tau_u > E\tau_v$ ;
- (C)  $|\mathcal{B}_n(u)| \geq |\mathcal{B}_n(v)|$  for all  $n$  and  $|\mathcal{B}_n(u)| > |\mathcal{B}_n(v)|$  for all sufficiently large  $n$ ;
- (D)  $h(X_{\{u\}}) > h(X_{\{v\}})$ .

**Proof:** (A)  $\iff$  (B): This follows again from the fact that  $E\tau_w = q\phi_w(q)$  for any string  $w$  [GO81a, pg. 189].

(A)  $\Rightarrow$  (D): By the result in Section 7 of [GO81a], (A) implies  $|\mathcal{B}_n(u)| \geq |\mathcal{B}_n(v)|$  for all  $n$ . Thus,  $h(X_{\{u\}}) \geq h(X_{\{v\}})$ . By Proposition 2.1 (the equivalence between item (3) and item (8)), we must have  $h(X_{\{u\}}) > h(X_{\{v\}})$ .

(D)  $\Rightarrow$  (A): Suppose to the contrary that  $\phi_u(q) \leq \phi_v(q)$ , i.e., either  $\phi_u(q) < \phi_v(q)$  or  $\phi_u(q) = \phi_v(q)$ . In the former case, we have  $h(X_{\{u\}}) < h(X_{\{v\}})$  (by (A)  $\Rightarrow$  (D)), a contradiction; in the latter case, we have  $h(X_{\{u\}}) = h(X_{\{v\}})$  (by the equivalence between item (3) and item (8) in Proposition 2.1), which is again a contradiction.

(A)  $\Rightarrow$  (C): The first part is the result in Section 7 of [GO81a]. To see the second part, note that (A) implies (D), which immediately gives  $|\mathcal{B}_n(u)| > |\mathcal{B}_n(v)|$  for all sufficiently large  $n$ .

(C)  $\Rightarrow$  (A): Suppose to the contrary that  $\phi_u(q) \leq \phi_v(q)$ , i.e., either  $\phi_u(q) < \phi_v(q)$  or  $\phi_u(q) = \phi_v(q)$ . In the former case, we have  $|\mathcal{B}_n(u)| < |\mathcal{B}_n(v)|$  for all sufficiently large  $n$  (by (A)  $\Rightarrow$  (C)), a contradiction; in the latter case, we have  $|\mathcal{B}_n(u)| = |\mathcal{B}_n(v)|$  (by the equivalence between item (3) and item (6) in Proposition 2.1), which is again a contradiction.  $\square$

### 3. GRAPHICAL REPRESENTATION FOR 1D SFTs WITH ONE FORBIDDEN WORD

In this section we study a specific graph presentation for one dimensional SFTs  $X_{\{w\}}$ . The construction allows direct comparison of different such SFTs, using ideas from both probability, namely a coupling argument in Theorem 3.12, and from linear algebra, namely the Perron Frobenius theorem in Theorem 3.15. Together, these two theorems enrich the equivalence (B)  $\iff$  (D) from Proposition 2.2 under the natural assumption 3.11.

**3.1. Ambient shift is the full shift.** Fix  $q \in \{2, 3, \dots\}$ ,  $k \geq 2$ , and a word  $w = w_1w_2 \cdots w_k \in [q]^k$ . First we introduce a bit of notation:

**Definition 3.1.** For any  $a \in [q]$  and  $1 \leq i \leq k-1$ , define

$$(3) \quad d_i(a) := \max\{j : w_1w_2 \cdots w_j = w_{i-j+2}w_{i-j+3} \cdots w_ia\},$$

with  $d_i(a) = *a$  if the set of such  $j$  is empty.

The  $d_i$  can be viewed as a kind of self-overlap data: namely,  $d_i(a)$  is the length of the longest prefix of  $w$  that is a suffix of the word  $w_1w_2 \cdots w_ia$ .

**Definition 3.2.** Let  $L_{w;q}$  be the following directed, edge-labeled graph on the vertex set  $\{*a : a \in [q] \setminus \{w_1\}\} \cup \{1, 2, \dots, k-1\}$ . The edge set consists of:

- For each  $a, b \in [q] \setminus \{w_1\}$ , an edge  $*a \rightarrow *b$  labeled  $b$ ;
- For each  $a \in [q] \setminus \{w_1\}$ , an edge  $*a \rightarrow 1$  labeled  $w_1$ ;
- For each  $a \in [q]$  and  $1 \leq i \leq k-1$ , except for  $a = w_k$  and  $i = k-1$ , an edge  $i \rightarrow d_i(a)$  labeled  $a$ .

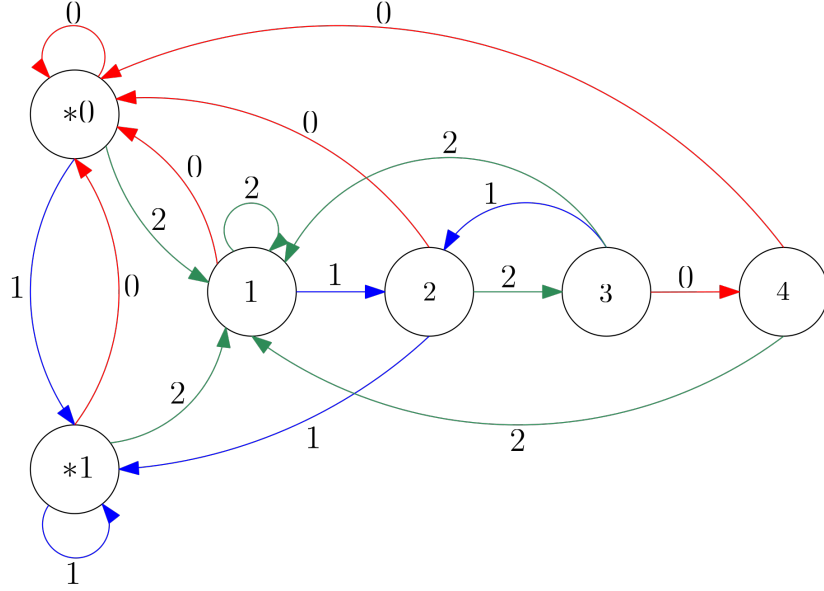


FIGURE 1. The graph  $L_{21201;3}$  for the word  $w = 21201$  over the alphabet  $[3] = \{0, 1, 2\}$ . Edges labeled 0, 1, 2 are colored in red, blue, and green, respectively.

For brevity we often suppress the dependence on  $q$  and simply write  $L_w$ . An example with  $q = 3$  is given in Figure 1.

The idea behind this construction is the following: imagine a one-sided infinite sequence  $x = (\dots, x_{-2}, x_{-1}, x_0) \in [q]^\mathbb{N}$  that we modify one step at a time by adding a new letter on the right. The vertices of  $L_w$  represent the length of the maximal suffix of  $x$  that is a prefix of  $w$ , and the edges of  $L_w$  represent appending that edge's label, say  $a$ , to  $x$  to obtain  $x' = (\dots, x_{-1}, x_0, a)$ .

The graph  $L_w$  is, up to a small modification (merging the states  $*a$  together), the *follower set graph*  $G_w$  of the shift  $X_{\{w\}}$ , a well-known object in symbolic dynamics [LM21, Section 3.2]. The follower set graph, described in the proof of Fact 3.3 below, is a standard way to present any SFT (or even any sofic shift, a generalization of SFTs). We stick with the graph  $L_w$ , rather than  $G_w$ , because a key lemma, Fact 3.7, does not hold for the latter.

**Fact 3.3.**  $L_w$  is a conjugacy presentation of  $X_{\{w\}}$ . That is, the mapping from bi-infinite edge sequences to bi-infinite label sequences is a bijection onto  $X_{\{w\}}$ .

*Proof.* Say that a word  $u$  is *allowed* if it does not contain  $w$ . For an allowed word  $u$ , define the *follower set*

$$(4) \quad F(u) = \{\text{words } v : uv \text{ is allowed}\}.$$

We claim that there are only finitely many follower sets, in particular that  $F(u) = F(u^*)$  where  $u^*$  is the longest suffix of  $u$  that is a proper prefix of  $w$ . Clearly  $F(u) \subseteq F(u^*)$ . For the reverse inclusion, let  $v \in F(u^*)$ . If  $v \notin F(u)$ , then  $w$  is contained in  $uv$  but not in  $u^*v$ . This means that  $u$  has a longer (than  $u^*$ ) suffix that is a proper prefix of  $w$ , a contradiction.

The *follower set graph*  $G_w$  is a labeled directed graph whose vertices are  $\{F(u)\}$  over all allowed  $u$ , and for each  $a \in [q]$  s.t.  $ua$  is allowed, there is an edge, labeled  $a$ , from  $F(u)$  to  $F(ua)$ . Note that by the above,  $G_w$  is a finite labeled directed graph.



We claim that  $G_w$  is a conjugacy presentation of  $X_{\{w\}}$ . It is clear that the set of label sequences of  $G_w$  is contained in  $X_{\{w\}}$ .

Observe that any edge sequence in  $G_w$  with label sequence  $x = \dots, a_n, a_{n+1}, \dots$  is of the form  $\dots F(u^n)F(u^{n+1})\dots$  where each  $F(u^{n+1}) = F(u^n a_n)$ . But then for fixed  $n$ ,  $F(u^{n+k}) = F(u^n a_n \dots a_{n+k}) = F(u^*)$  where  $u^*$  is the smallest prefix of  $w$  that is a suffix of  $u^n a_n \dots a_{n+k}$ ; but clearly  $u^*$  is contained in  $a_n \dots a_{n+k}$  and thus  $F(u^{n+k}) = F(u^*)$  can be uniquely recovered from  $x$ . Thus, every element of  $X_{\{w\}}$  is the label sequence of a unique edge sequence in  $G_w$ , and so  $G_w$  is indeed a conjugacy presentation of  $X_{\{w\}}$ .

Finally, we see that  $L_w$  is an *in-splitting* of  $G_w$  (for the definition, see [LM21, Definition 2.4.7]): identify each vertex  $F(w_1 \dots w_i)$  in  $G_w$  with the vertex  $i$  in  $L_w$ ; in  $G_w$ , the vertex  $F(\emptyset)$  has  $q-1$  self-loops, each labeled by a unique element of  $[q] \setminus \{w_1\}$ , and, to form  $L_w$  from  $G_w$ , that vertex is split into  $q-1$  vertices, denoted  $*a$ ,  $a \in [q] \setminus \{w_1\}$ . Since in-splittings define conjugacies ([LM21, Theorem 2.4.10]),  $L_w$  is also a conjugacy presentation of  $X_{\{w\}}$ .  $\square$

We now note some nice properties of  $L_w$  which will be useful throughout this section. By convention, we write  $*a < i$  for any  $a \in [q]$  and  $1 \leq i \leq k-1$ .

**Proposition 3.4.** *Fix  $w \in [q]^k$  for  $k \geq 2$  and let  $L_w$  and  $d_i$  be as in Definitions 3.1 and 3.2.*

- i. *All vertices  $i \leq k-2$  and  $*a$  for  $a \in [q] \setminus \{w_1\}$  have out-degree  $q$ , and state  $k-1$  has out-degree  $q-1$ .*
- ii. *Incoming edges to a vertex  $i \geq 1$  have label  $w_i$ , and incoming edges to vertex  $*a$  have label  $a$ .*
- iii. *For all  $i \in \{1, \dots, k-2\}$  and  $a \neq w_{i+1}$ ,  $d_i(a) \leq i$ ; and  $d_i(w_{i+1}) = i+1$ .*
- iv. *If  $d_i(a), d_{i'}(a') \geq 1$ , then  $i - d_i(a) = i' - d_{i'}(a')$  only if  $i = i'$ ,  $a = w_{i+1}$  and  $a' = w_{i'+1}$ .*
- v. *Any prefix  $u = w_1 w_2 \dots w_r$  of  $w$  with  $r \geq 2$  has graph  $L_u$  identical to the induced subgraph of  $L_w$  on the vertices*

$$(5) \quad \{ *a : a \in [q] \setminus \{w_1\} \} \cup \{1, 2, \dots, r-1\}$$

*Proof.* Parts i, ii, and v are clear from the definition. For part iii, it follows from the definition of  $d_i(a)$  that  $d_i(a) \leq i+1$  for all  $a$ , and from the edge-labeling that  $d_i(a) \neq d_i(a')$  if  $a \neq a'$ . For part iv, write  $d_i = d_i(a)$  and  $d_{i'} = d_{i'}(a')$  for short, and suppose for the sake of contradiction that  $i - d_i = i' - d_{i'}$  for some  $i \neq i'$  with  $i, i' \geq 1$ . Then also  $d_i \neq d_{i'}$  by re-arranging, so assume WLOG  $d_i > d_{i'}$ . If  $a = w_{i+1}$ , then by part iii,  $a' = w_{i'+1}$ , and similarly with the roles of  $a$  and  $a'$  reversed, so assume  $a \neq w_{i+1}$  and  $a' \neq w_{i'+1}$ . Unraveling the definition of  $d_i$  gives

$$(6) \quad w_{i-d_i+j+1} = w_j \text{ for } j = 1, 2, \dots, d_i-1, \text{ and } w_{i+1} \neq w_{d_i}.$$

Since  $d_i > d_{i'} \geq 1$  by assumption, setting  $j = d_{i'}$  gives

$$(7) \quad w_{d_{i'}} = w_{i-d_i+d_{i'}+1} = w_{i'-d_{i'}+d_{i'}+1} = w_{i'+1},$$

contradicting the definition of  $d_{i'}$ .  $\square$

We remark that the properties in Proposition 3.4 are not sufficient to characterize the set of graphs that arise as  $L_w$  for some word  $w$  and integer alphabet size  $q \geq 2$ , and we are not aware of such a set of sufficient conditions. We now give a few applications of Proposition 3.4, starting with a characterization of which of the  $L_w$  graphs are irreducible. Recall that a graph is *irreducible* if for any vertices  $x, y$ , there is a (directed) path from  $x$  to  $y$  (and otherwise it is *reducible*).

**Proposition 3.5.** *The graph  $L_w$  is reducible if and only if  $q = 2$  and  $w$  is one of the words  $10^{k-1}, 1^{k-1}0, 01^{k-1}, 0^{k-1}1$  for some  $k \geq 2$ .*

*Proof.* Since the graph  $L_w$  always has the paths  $*a \rightarrow 1 \rightarrow \dots \rightarrow k-1$  for  $a \in [q] \setminus \{w_1\}$  and contains an induced complete graph on the vertices labeled  $*a$ , it suffices to show that for each  $i \geq 1$ ,  $d_i(a) < i$  for some letter  $a \in [q]$ . We start with the easier case  $q \geq 3$ . Proposition 3.4 parts *i* and *iii* together imply that there is exactly one edge  $i \rightarrow i+1$  and at most one edge  $i \rightarrow i$ , and thus at least one edge  $i \rightarrow d_i(a) < i$ .

So assume  $q = 2$ . Write  $\bar{a} = 1 - a$  and  $d_i = d_i(\overline{w_{i+1}})$  for the binary alphabet. Suppose that  $L_w$  is reducible, i.e. that there is no path  $k-1 \rightarrow *a$  for any  $a$ . Then there exists  $1 \leq i \leq k-1$  such that  $d_j \geq i$  for all  $j = i, i+1, \dots, k-1$ : let  $I$  be the largest such  $i$ . If  $I = k-1$ , then  $d_{k-1} = k-1$ , and unraveling the definition of  $d_{k-1}$  gives  $w_1 = w_2 = \dots = w_{k-1} = \overline{w_k}$ , i.e.  $w = 1^{k-1}0$  or  $0^{k-1}1$ .

So suppose  $I < k-1$ . By part *iv* of Proposition 3.4, since  $I \geq 1$ , the values  $j - d_j$  must all be distinct for  $j \in \{I, I+1, \dots, k-1\}$ . In particular, by part *iii* of Proposition 3.4,  $d_I = I$ , and by induction over  $j$ ,  $d_j = I$  for all  $j \in \{I, I+1, \dots, k-1\}$ . Writing out the definition of  $d_j$ , we obtain that for  $j \in \{I, I+1, \dots, k-1\}$ ,

$$(8) \quad w_m = w_{j+1-I+m} \text{ for } m \in \{1, \dots, I-1\}, \text{ and also } w_I = \overline{w_{I+1}}.$$

If  $I \geq 2$ , then taking  $m = I-1$  with  $j = I$  and separately with  $I+1$  in Equation 8 immediately gives the contradiction  $w_{I-1} = w_I = w_{I+1}$  and  $w_I = \overline{w_{I+1}}$ . Thus  $I = 1$  and  $w_1 = \overline{w_2} = \overline{w_3} = \dots = \overline{w_k}$ , i.e.  $w = 10^{k-1}$  or  $01^{k-1}$ , as desired.  $\square$

**Remark 3.6.** Using Proposition 3.5, one can easily show that the corresponding shift space  $X_{\{w\}}$  is reducible in the same cases, and irreducible otherwise.

Next we show that the graph  $L_w$  is essentially enough information to determine the word  $w$ . In what follows, we say a map  $\varphi$  between graphs  $L_w, L_{w'}$  is a *graph isomorphism* if it is an isomorphism of the underlying unlabeled graphs, regardless of how  $\varphi$  transforms the edge/vertex labels.

**Fact 3.7.** *Let  $q \geq 2$ . If  $w, w' \in [q]^k$  are two words such that  $L_w, L_{w'}$  are isomorphic by an isomorphism  $\varphi$  that preserves the vertex names, then  $w = w'$ . ( $\varphi$  ‘preserving the vertex names’ means: viewed as a map between the vertex sets,  $\varphi(i) = i$  for all  $i$  and  $\varphi(*a) = *a$  for all  $*a$  vertices.)*

*Proof.* It suffices to show that given the graph  $L_w$  with all edge labels erased, one can recover the word  $w$ . We give a simple recursive algorithm to recover  $w$  along with the edge labels of  $L_w$ . For each vertex  $*a$  that appears in  $L_w$ , assign all the incoming edges to it the label  $a$ . By the construction of  $L_w$ ,  $w_1$  must be the unique digit in  $[q]$  not appearing among the  $*a$  vertices. Assume by induction we have determined the values  $w_1, w_2, \dots, w_j$  for some  $1 \leq j \leq k-1$  and the labels of all incoming edges to vertices  $1, 2, \dots, j$ . Then by the construction of  $L_w$ ,  $w_{j+1}$  is the unique digit in  $[q]$  not appearing in the set of labels of the outgoing edges from vertex  $j$ , which by Proposition 3.4 parts *ii* and *iii* and induction are among the known values.  $\square$

One can easily strengthen Fact 3.7 to the following:

**Theorem 3.8.** *Let  $q \geq 2$ . If  $w, w' \in [q]^k$  are such that  $L_w$  and  $L_{w'}$  are isomorphic, then there is a permutation  $\pi$  of the alphabet  $[q]$  such that  $\pi(w_i) = w'_i$  for  $i = 1, 2, \dots, k$ .*

Theorem 3.8 implies:

**Corollary 3.9.** *Let  $q \geq 2$ . For any words  $w, w' \in [q]^k$ ,  $L_w$  and  $L_{w'}$  are isomorphic if and only if there is a permutation  $\pi$  of the alphabet  $[q]$  such that  $\pi(w_i) = w'_i$  for  $i = 1, 2, \dots, k$ .*

For example, in the binary alphabet case  $q = 2$ , the word  $w$  can be recovered from the vertex labels on  $L_w$  from the following simple recursive formula. In this case, we can write  $*a$  for the unique  $*$  vertex in  $L_w$ , so that  $w_1 = \bar{a}$  (recalling the notation from the proof of Proposition 3.5). Then for each  $i \in \{1, \dots, k\}$ , one easily finds

$$w_i = \begin{cases} w_1, & \text{if } d_{i-1} = *a \\ \overline{w_{d_{i-1}}}, & \text{otherwise} \end{cases}$$

*Proof of Theorem 3.8.* We will first show that the isomorphism of the graphs must preserve certain labels. Indeed if  $\varphi$  is the isomorphism from  $L_w$  to  $L_{w'}$  then  $\varphi(k-1) = k-1$ . This is true since  $k-1$  is the only vertex with degree  $q-1$  in both the graphs. We will proceed by induction to show that  $\varphi(t) = t$  for all  $1 \leq t \leq k-1$ . Suppose we know this for  $t \geq j$  for some  $j$ . Now the only incoming edge to the vertex labeled  $j$  which is not amongst  $\{j, j+1, j+2, \dots, k-1\}$  must originate at  $j-1$ . Thus  $\varphi(j-1) = j-1$ . This completes the induction step and implies that  $\varphi(t) = t$  for all  $1 \leq t \leq k-1$ . Without loss of generality assume that  $w_1 = w'_1 = 0$ . Then  $\varphi$  is a bijection from vertices labeled  $\{*s : 1 \leq s \leq q-1\}$  in  $L_w$  to the same set of vertices in  $L_{w'}$ . This induces a bijection on  $[q]$  which by Fact 3.7 takes  $w$  to  $w'$ .  $\square$

By Fact 3.3, an isomorphism of the graphs  $L_w$  and  $L_{w'}$  implies conjugacy of the corresponding shift spaces. As we will learn in Theorem 4.3 of Section 4, many pairs of words  $w, w'$ , are not equivalent under a permutation of  $[q]$  and yet give rise to conjugate shift spaces  $X_{\{w\}}, X_{\{w'\}}$ .

Also, we note that Corollary 3.9 implies that the correlation polynomial  $\phi_w$  is uniquely determined by the graph  $L_w$ . Despite this, we do not know an explicit map between the family of possible correlation polynomials and the family of possible graphs  $L_w$ . See section 6 for further discussion.

We now turn to the main results of this section, which obtain the same conclusions as in [GO81a], but under a different assumption that is adapted to Perron Frobenius theory. We first make a short excursion into discrete probability. First recall the stochastic dominance order for random variables:

**Definition 3.10.** *For any random variables  $A, B$ , possibly defined on different probability spaces, write  $A \prec_{st} B$  if  $\mathbb{P}(A \geq t) \leq \mathbb{P}(B \geq t)$  for all  $t \in \mathbb{R}$ . Equivalently,  $A \prec_{st} B$  if and only if there is a probability space  $(\Omega, \mathbb{P})$  and a pair of random variables  $(X, Y)$  on  $(\Omega, \mathbb{P})$  such that  $X$  and  $Y$  have the distributions of  $A$  and  $B$  respectively, and  $\mathbb{P}(X \leq Y) = 1$ .*

Recall the random variable  $\tau_w$ , the hitting time of word  $w$  in an iid  $\text{Uniform}([q])$  sequence. It follows easily from the results of [GO81a, pg. 204] that  $h(X_{\{w\}}) \geq h(X_{\{w'\}})$  if and only if  $\tau_{w'} \prec_{st} \tau_w$ . In Theorems 3.12 and 3.15 we prove the same result under the following natural assumption on the edges  $d_i$ . Fix words  $w, w' \in [q]^k$  and let  $L = L_w, L' = L_{w'}$  and  $d_i, d'_i$  as in Definition 3.1.

**Assumption 3.11.** For words  $w, w' \in [q]^k$ , we say condition  $D(w, w')$  holds if for each  $i \in \{1, 2, \dots, k-1\}$ , there exists a permutation  $\pi_i : [q] \rightarrow [q]$  such that  $d_i(a) \leq d'_i(\pi_i(a))$  for all  $a \in [q]$ .

In the binary alphabet case  $q = 2$ , the assumption  $D(w, w')$  is equivalent to  $d_i \leq d'_i$  for all  $i$  (recalling the notation from the proof of Proposition 3.5). For example,  $D(w, w')$  holds for the pair  $w = 01010, w' = 01000$ ; also,  $D(w, 0^k)$  always holds for  $w \in [q]^k$ . Put simply,  $D(w, w')$  says that the edges of  $L$  go further back towards the  $*$  vertices than the edges of  $L'$ .

While this assumption is natural in the context of the graphs  $L_w$ , we do not know any explicit formulation of Assumption 3.11 in terms of autocorrelation polynomials. See Section 6 for further discussion.

We note that this type of ordering has already been considered in more detail in [DSS16], under the name ‘skip-free’ Markov chains, which in particular contains a more general version of Theorem 3.12. We include a simplified proof here for the sake of completeness.

**Theorem 3.12.** Let  $w, w' \in [q]^k$  for  $k \geq 3$  be any words with graphs  $L = L_w, L' = L_{w'}$ . If  $D(w, w')$  holds, then  $\tau_{w'} \prec_{st} \tau_w$ . Additionally, if for some  $1 \leq i \leq k-1$  and  $a \in [q]$  we have a strict inequality  $d_i(a) < d'_i(\pi_i(a))$ , then  $\tau_w$  and  $\tau_{w'}$  are not equal in distribution.

*Proof.* Consider the simple random walk on the graph  $G_w$  obtained from  $L_w$  by identifying all vertices  $\{*a : a \in [q] \setminus \{w_1\}\}$  to a single vertex labeled  $*$ , and adding an extra vertex  $k$  is added at the end of  $L_w$ , i.e. there is an edge  $k-1 \rightarrow k$  corresponding to hitting the word  $w$ . Let  $X, X'$  denote simple random walk on the graphs  $G_w, G_{w'}$  respectively, stopped at the first time  $\tau$  (resp.  $\tau'$ ) that  $X$  (resp.  $X'$ ) hits state  $k$ . That is,  $X$  is a random sequence of vertices  $X = (* = X_0, X_1, X_2, \dots, X_{\tau_k} = k)$  such that: given  $X_i, X_{i+1}$  a uniformly random neighbor of  $X_i$  for each  $1 \leq i < \tau_k$ . (For completeness one can add outgoing edges from vertex  $k$  corresponding to appending any digit to  $w$ , so  $G_w$  is a conjugacy representation for the full shift, and simple random walk on  $G_w$  is the corresponding maximum entropy Markov chain.) Note that  $\tau, \tau'$  are distributed as  $\tau_w$  and  $\tau_{w'}$ , respectively.

We couple the pairs  $(X, \tau)$  and  $(X', \tau')$  as follows. Generate  $X$  and  $X'$  along with a random function  $\sigma$  recursively, initialized at  $X(0) = X'(0) = \sigma(0) = *$ . For  $s \geq 1$ , given  $X(s), X'(s)$  and  $\sigma(s)$  for  $s \leq s$ , let  $X(s+1)$  have the distribution of an independent simple random walk step from  $X(s)$  (in the graph  $G$ ), and define  $(X'(s), \sigma(s))$  by

$$(X'(s+1), \sigma(s+1)) = \begin{cases} (X'(s) + 1, \sigma(s) + 1) & \text{if } X'(s) = X(s) \text{ and } X(s+1) = X(s) + 1 \\ (d'_{X'(s)}(\pi_{X'(s)}(a)), \sigma(s) + 1) & \text{if } X'(s) = X(s) \text{ and } X(s+1) = d_{X(s)}(a) \\ (X'(s), \sigma(s)) & \text{if } X'(s) > X(s) \end{cases}$$

(Note that we interpret  $d_i(a) = *$  if  $d_i(a) = *b$  for some  $b \in [q]$ .) In words, when  $X$  and  $X'$  are at the same level, they move together (after translating by the permutation  $\pi_X$ ), and when they are at different levels,  $X$  moves independently while  $X'$  remains frozen. The (random) counter  $\sigma(s)$  is the number of steps taken by  $X'$  when  $X$  has taken  $s$  steps. The process ends at time  $\tau = \min\{s : X(s) = k\} = \min\{s : X'(s) = k\}$ . By induction, using the assumption  $D(w, w')$ , and because the only simple path in both graphs from  $i \rightarrow j$  with  $i < j$  is the path  $i \rightarrow i+1 \rightarrow \dots \rightarrow j-1 \rightarrow j$ , for all  $s \geq 0$  we have  $X(s) \leq X'(\sigma(s))$  and  $\sigma(s) \leq s$ .

It follows immediately from the construction that  $X(\cdot)$  has the distribution of simple random walk on  $G_w$  stopped on hitting vertex  $k$ , and  $\tau$  has the distribution of  $\tau_w$ . Similarly,  $X'(\gamma(\cdot))$  has the distribution of simple random walk on  $G_{w'}$  stopped on hitting vertex  $k$ , where  $\gamma(t) = \min \sigma^{-1}(t)$  are the jump times for  $X'$ , and thus  $\sigma(\tau)$  has the distribution of  $\tau_{w'}$ . Combining this with the previous paragraph,  $\tau \geq \tau'$  under this coupling. Additionally,  $\tau > \sigma(\tau)$  if  $\sigma(s) < s$  for some  $s < \tau$ , and the latter occurs with positive probability if  $d_i(a) < d'_i(\pi_i(a))$  for some  $a$  and  $i$ .  $\square$

Theorem 3.12 says that under condition  $D(w, w')$ , the stochastic ordering of the hitting times  $\tau_w$  can be read off from the graphs  $L_w$ . It turns out that entropies can also be compared easily when condition  $D(w, w')$  holds. Recall from the Perron-Frobenius Theorem that the adjacency matrix of an irreducible graph has a unique maximum (in modulus) real eigenvalue  $\lambda > 0$  (the Perron eigenvalue), and the corresponding left and right eigenvectors  $\ell$  and  $r$  are both positive. Thus:

**Lemma 3.13.** *For any irreducible graph  $L_w$ , the corresponding left and right eigenvectors  $\ell, r$  are strictly positive. For the reducible cases appearing in Lemma 3.5 one can check by direct computation (omitted) that the Perron eigenvalue and corresponding left and right eigenvectors are strictly positive, with the exception that  $\ell_{*a}$  and  $r_{k-1}$  may be 0.*

We will use the following property of the right Perron eigenvector of  $L_w$ :

**Proposition 3.14.** *Fix  $w \in [q]^k$ . Let  $A = A_w$  denote the adjacency matrix of  $L_w$ , and write  $r = r_w$  for the right eigenvector of  $A$  corresponding to the Perron-Frobenius eigenvalue  $\lambda = \lambda_w$  of  $A$ . Then  $0 < \lambda - q + 1 < 1$  and  $r$  has strictly decreasing entries. In particular, the entries of  $r$  decrease at least exponentially: for  $i = 1, \dots, k-2$ ,*

$$(9) \quad r_{i+1} \leq (\lambda - q + 1)r_i,$$

and for all  $a \in [q] \setminus \{w_1\}$ ,

$$(10) \quad r_1 = (\lambda - q + 1)r_{*a}$$

*Proof.* First, for any  $a \in [q] \setminus \{w_1\}$ ,

$$(11) \quad \lambda r_{*a} = r_1 + \sum_{b \in [q] \setminus \{w_1\}} r_{*b}.$$

Since the right hand side does not depend on  $a$ , we have  $r_{*a} = r_{*a'}$  for any  $a, a' \in [q] \setminus \{w_1\}$ . Thus

$$(12) \quad \lambda r_{*a} = (q-1)r_{*a} + r_1,$$

which gives Equation 10. Note that by positivity of  $r$ ,  $\lambda - q + 1 > 0$ . Combining this with  $\lambda < q$  (since the shift space  $X_{\{w\}}$  is a proper subshift of the full  $q$ -shift, which has entropy  $\log q$ ), we get  $0 < \lambda - q + 1 < 1$ . We proceed by induction: assume that Equation (9) holds for  $i = 1, \dots, j-1$ . Then

$$(13) \quad \lambda r_j = \sum_{a \in [q]} r_{d_j(a)} \geq r_{j+1} + (q-1)r_j.$$

Here the equality uses that  $r$  is a right eigenvector, and the inequality follows from Proposition 3.4 part *iii*, the induction hypothesis, and the fact that  $\lambda < q$ . Re-arranging gives Equation 9. The same calculation, along with equation 10, proves the base case  $i = 1$ .  $\square$

The proof actually shows something a bit stronger: namely that for  $i = 1, \dots, k - 2$ ,

$$(14) \quad \frac{r_{i+1}}{r_i} \leq \lambda - \sum_{a \in [q] \setminus \{w_{i+1}\}} (\lambda - q + 1)^{-i+d_i(a)}.$$

We can now describe the ordering of the  $h(X_{\{w\}})$  under the same assumption as in Theorem 3.12.

**Theorem 3.15.** *Fix  $w, w' \in [q]^k$  for  $k \geq 2$  and let  $r$  be as in Proposition 3.14. If  $D(w, w')$  holds, then  $h(X_{\{w\}}) \geq h(X_{\{w'\}})$ . If  $d_i(a) < d'_i(\pi_i(a))$  for some  $1 \leq i \leq k - 2$  and  $a \in [q]$ , or if  $L_w, L_{w'}$  are both irreducible and  $d_{k-1}(a) < d_{k-1}(\pi_{k-1}(a))$  for some  $a \in [q]$ , then the entropy inequality is strict.*

*Proof.* Let  $A, A'$  denote the adjacency matrices of  $L, L'$ , and write  $r, r'$  for the right eigenvectors of  $A, A'$  corresponding to the top eigenvalues  $\lambda, \lambda'$  (where all matrices and vectors are indexed by the vertex labels of the corresponding graphs). Recall that  $h(X_{\{w\}}) = \log \lambda_w$  and  $h(X_{\{w'\}}) = \log \lambda_{w'}$  [LM21, Theorem 4.3.3]. Combining Assumption  $D(w, w')$  and Proposition 3.14, we obtain that for all  $i \in \{1, \dots, k - 2\}$

$$(15) \quad \sum_{a \in [q]} r_{d'_i(a)} \leq \sum_{a \in [q]} r_{d_i(a)},$$

and that

$$(16) \quad \sum_{a \in [q] \setminus \{w_k\}} r_{d'_{k-1}(a)} \leq \sum_{a \in [q] \setminus \{w'_k\}} r_{d_{k-1}(a)}.$$

Thus for  $i = 1, \dots, k - 2$ ,

$$(17) \quad (A'r)_i = \sum_{a \in [q]} r_{d'_i(a)} \leq \sum_{a \in [q]} r_{d_i(a)} = (Ar)_i = \lambda r_i,$$

and similarly,  $(A'r)_{k-1} \leq \lambda r_{k-1}$ . Also, we may assume without loss of generality (by applying a permutation of the alphabet) that  $w_1 = w'_1$ . Note that in this case the outgoing edges of the vertices  $\{*a : a \in [q] \setminus \{w_1\}\}$  in  $L$  and  $L'$  have a bijective correspondence which preserves labels, so for all  $a \in [q] \setminus \{w_1\}$  we have by an easy computation that  $(A'r)_{*a} = \lambda r_{*a}$ .

Combining these observations gives  $A'r \leq \lambda r$ . Take a left eigenvector  $\ell'$  for the Perron-Frobenius eigenvalue  $\lambda'$  in  $A'$  and calculate:

$$(18) \quad \lambda' \ell' \cdot r = \ell' A'r \leq \ell' (\lambda r) = \lambda \ell' \cdot r.$$

(Here  $\cdot$  is the dot product.) The entries  $\ell'_i$  and  $r_i$  are strictly positive for  $1 \leq i \leq k - 2$ , and by Lemma 3.13, the first and last entries are non-negative. It follows that  $\lambda' \leq \lambda$ . By Lemma 3.13, if some strict inequality  $d_i(a) < d'_i(\pi_i(a))$  holds for  $1 \leq i \leq k - 2$ , then  $(A'r)_i < \lambda r_i$ , the same argument yields  $\lambda < \lambda'$ ; and the same conclusion follows if  $L_w, L_{w'}$  are irreducible and  $d_{k-1}(a) < d'_{k-1}(\pi_{k-1}(a))$  for some  $a$ .  $\square$



**3.2. The ambient shift is the golden mean shift.** In this section we generalize the  $L_w$  graph construction and Theorem 3.8 to the case where the underlying shift space is not the full  $q$ -shift, but the golden mean shift. Let us set things up first. Let  $q \geq 2$  and  $\mathcal{F} \subset [q]^2$ . If  $w \in \mathcal{B}_n(X_{\mathcal{F}})$ , we are broadly interested in learning the effect of forbidding a further word  $w$ . We will denote by  $(X_{\mathcal{F}})_{\{w\}}$  the SFT with the forbidden set  $\mathcal{F} \cup \{w\}$ .

**Definition 3.16.** For  $w \in \mathcal{B}_n(\mathcal{F})$  of length  $k$ , start with the graph  $L_w$  exactly as in Definition 3.1. Define a new graph  $L_w^{\mathcal{F}}$  by deleting edges as follows: for any path of length 2 consisting of edges  $e, f$  labeled  $a_e, a_f$  with  $a_e a_f \in \mathcal{F}$ , delete edge  $f$ .

**Fact 3.17.** *The resulting graph  $L_w^{\mathcal{F}}$  is a conjugacy presentation of the shift space  $(X_{\mathcal{F}})_{\{w\}}$ .*

This follows from Fact 3.3.

Note that in the binary case  $q = 2$ , except for trivial shift spaces, we can take  $\mathcal{F} = \{11\}$  without loss of generality. The graph  $L_w^{\mathcal{F}}$  is identical to  $L_w$  except that each vertex  $i$  with  $w_i = 1$  (and the vertex  $*1$  if it is present in  $L$ ) only has one outgoing edge  $i \rightarrow i+1$ . As a result, the word  $w$  is determined in a simple way from the labeled graph  $L_w^{\{11\}}$  (analogous to Fact 3.7):

**Fact 3.18.** *Given the graph  $L = L_w^{\{11\}}$  for an allowable word  $w \in X_{\{11\}}$  of length  $k$ , for  $i \in \{1, \dots, k-2\}$ ,*

$$(19) \quad w_i = \begin{cases} 1, & \text{outdegree}_L(i) = 1 \\ 0, & \text{outdegree}_L(i) = 2, \end{cases}$$

and

$$(20) \quad w_{k-1}w_k = \begin{cases} 10, & \text{outdegree}_L(k-1) = 0 \\ 0\overline{w_{d_{k-1}}}, & \text{outdegree}_L(k-1) = 1 \end{cases}$$

With a bit more work, we also get the same conclusion as in Theorem 3.8 for the golden mean shift.

**Theorem 3.19.** *For any  $k \geq 3$  and  $w, w' \in \mathcal{B}_k(\{11\})$ ,  $w = w'$  if and only if  $L_w^{\{11\}}$  and  $L_{w'}^{\{11\}}$  are isomorphic as unlabeled graphs.*

In order to prove this theorem, we need the following facts.

**Fact 3.20.** *There is at most one 1-cycle and at most one 2-cycle in  $L_w^{\{11\}}$ .*

*Proof.* This follows immediately from the fact that  $(X_{\{11\}})_{\{w\}}$  contains only one fixed point (i.e.,  $0^\infty$ ) and only one orbit of length 2 (i.e.,  $\{(01)^\infty, (10)^\infty\}$ ), and the fact that  $L_w^{\{11\}}$  is a conjugacy presentation of  $(X_{\{11\}})_{\{w\}}$ .  $\square$

**Fact 3.21.** *If  $L_w^{\{11\}}$  has a self-loop at some vertex  $j$  where  $j \geq 1$ , then we must have  $w_1w_2 \cdots w_jw_{j+1} = 0^j1$ , i.e., the  $(j+1)$ -prefix of  $w$  must be  $0^j1$ .*

*Proof.* First note that the self-loop must be labeled 0 (which means  $\overline{w_{j+1}} = 0$ ). Now, having a self-loop at vertex  $j$  exactly means that the longest possible overlap between  $w_1w_2w_3 \cdots w_j\overline{w_{j+1}}$  and  $w_1w_2 \cdots w_k$  is of length  $j$ , which in particular implies  $w_2w_3 \cdots w_j\overline{w_{j+1}} = w_1w_2 \cdots w_j$ . Thus, one immediately obtains  $w_1 = w_2 = w_3 = \cdots = w_j = \overline{w_{j+1}} = 0$ .  $\square$

**Fact 3.22.** *If  $L_w^{\{11\}}$  has a 2-cycle at the two vertices  $j, j+1$  where  $j \geq 2$ , then we must have*

$$w_1 w_2 \cdots w_j w_{j+1} w_{j+2} = \begin{cases} 10101 \cdots 0100 & \text{if } j \text{ is odd} \\ 010101 \cdots 0100 & \text{if } j \text{ is even.} \end{cases}$$

*Proof.* The edge from  $j$  to  $j+1$  must be labeled 0. Thus, the 2-cycle must be labeled 00 or 01. Since  $L_w^{\{11\}}$  is a conjugacy presentation it follows that the 2 cycle must be labeled 01 (i.e.,  $w_{j+1} \overline{w_{j+2}} = 01$ ).

Now, on the other hand, the existence of the 2-cycle at  $j, j+1$  implies that the longest possible overlap between  $w_1 w_2 \cdots w_j w_{j+1} \overline{w_{j+2}}$  and  $w_1 w_2 \cdots w_k$  is of length  $j$ , which in particular gives

$$w_3 w_4 \cdots w_j w_{j+1} \overline{w_{j+2}} = w_1 \cdots w_j.$$

The result then follows by recalling  $w_{j+1} \overline{w_{j+2}} = 01$ .  $\square$

Now we can prove Theorem 3.19.

*Proof of Theorem 3.19:* In this proof we will call the vertex labeled  $*$  in  $L_w^{\{11\}}$  by 0 for convenience. When  $k = 3$ , there are only five possible choices of  $w$ : 000, 001, 010, 100 and 101. One can easily check that the corresponding unlabeled versions of  $L_w^{\{11\}}$  for these five words are never isomorphic.

So we assume  $k \geq 4$  in the following. We will prove the following equivalent statement: given any (directed) graph  $G = (V, E)$  that arises as an unlabeled version of  $L_w^{\{11\}}$  for some  $w \in \mathcal{B}_k(\{11\})$ , we can uniquely determine the word  $w$ , or equivalently, the corresponding labeling of the graph  $G$ .

Let  $G = (V, E)$  be such a graph. Suppose we identified the vertex named  $k-1$ . Then we can use the induction argument in the beginning of the proof of Theorem 3.8 to identify the vertices names  $j$  for  $1 \leq j \leq k-1$ . Finally by Fact 3.18, we would have uniquely determined the word  $w$ . Thus all we have to do is to identify the vertex  $v^*$  with name  $k-1$ . vertex  $v^* \in V$  which must have label  $k-1$ .

For  $i = 0, 1$ , define

$$(21) \quad V_i = \{v \in V : \text{outdegree}_G(v) = 2 - i\}.$$

If there exists a vertex  $v \notin V_0 \cup V_1$ , i.e. if there is a vertex with out-degree zero, which must be the only such vertex by the construction of  $L_w^{\{11\}}$ , then by Fact 3.18  $v$  must have label  $k-1$ , as desired. So assume that  $V_0, V_1$  are a partition of  $V$ . By Fact 3.18, each vertex  $x \in V_0$  must have vertex label  $i(x)$  with  $w_{i(x)} = 0$ , and each vertex  $y \in V_1$  – except possibly the vertex with label  $k-1$  – must have vertex label  $i(y)$  with  $w_{i(y)} = 1$ . Thus, since  $w$  is an allowed word in  $X^{\{11\}}$ , each vertex  $y \in V_1$  with  $w_{i(y)} = 1$  must have  $w_{i(y)+1} = 0$ , so the outgoing edge at  $y$  has endpoint  $x \in V_0$ . If there exists  $v \in V_1$  such that the outgoing edge  $v \rightarrow z$  has  $z \notin V_0$ , then  $v$  must be the unique vertex with label  $k-1$ , and again we are done.

So assume that all outgoing edges from  $V_1$  have endpoints in  $V_0$  (and  $V_0, V_1$  form a partition of  $V$ ). Consider the following six cases:

- (Case 1):  $w_1 w_2 w_3 = 001$ ; (Case 2):  $w_1 w_2 w_3 = 100$ ; (Case 3):  $w_1 w_2 w_3 = 101$ ;  
 (Case 4):  $w_1 w_2 w_3 = 000$ ; (Case 5):  $w_1 w_2 w_3 w_4 = 0100$ ; (Case 6):  $w_1 w_2 w_3 w_4 = 0101$ .

One easily checks that, because  $w \in \mathcal{B}_k(\{11\})$ , these are the only possible length-3 prefixes of  $w$  (note that Case 5 and 6 combined corresponds to  $w_1 w_2 w_3 = 010$ ).

Now, the subgraph (of  $L_w^{\{11\}}$ ) we see at vertices 0, 1, 2 (and vertices 0, 1, 2, 3 for cases (5.1) and (5.2)) are all given in Figure 2.

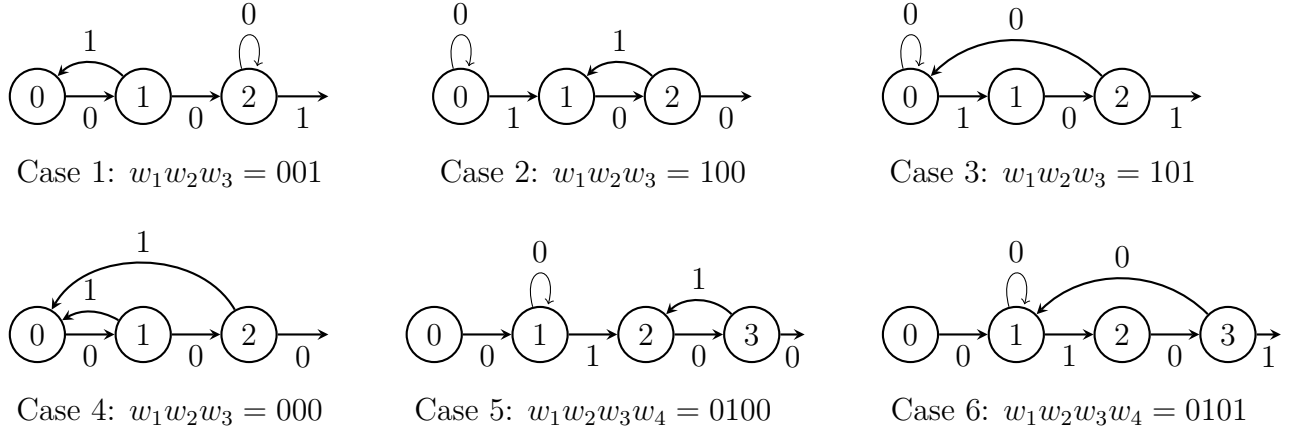


FIGURE 2. All six cases

Let  $S_1, S_2, S_3, S_4$  be the unlabeled subgraphs shown in Figure 3, which will be called the “base graphs”. Then, one can see from Figure 2 that for any  $w$ , at least one of the base graphs will appear in  $L_w^{\{11\}}$ .

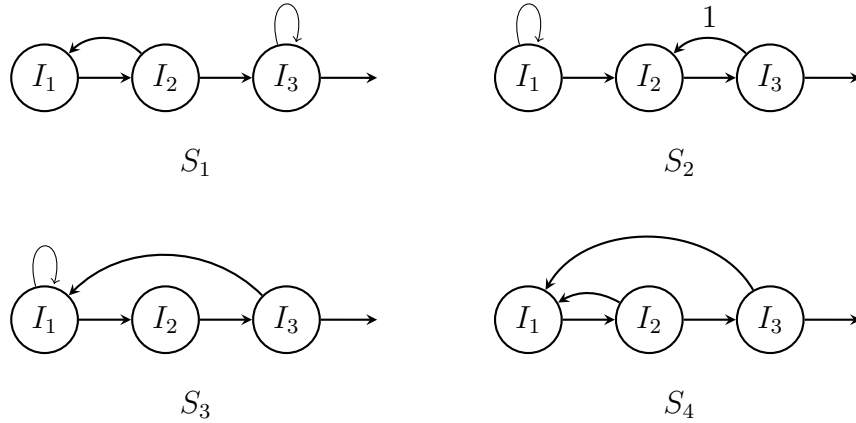


FIGURE 3. “Base graphs”  $S_1, S_2, S_3, S_4$

Now, given  $G$ , we build the following algorithm to identify the vertex 0, which in turn helps us identify the vertex  $k-1$  (because  $k-1$  is the terminal vertex of the only Hamiltonian path of  $L_w^{\{11\}}$  starting at the vertex 0). The proof of the validity of the algorithm, which is little tedious, can be found in the Appendix.

The algorithm is:

**Step 1:** First, we search for  $S_1$  in  $G$ . If we see  $S_1$  as a subgraph of  $G$  then the vertex  $I_1$  (in  $S_1$ ) must be labeled 0 (and then we are done).

**Step 2:** If we can not find  $S_1$  in  $G$ , then we search for  $S_2$ . If we see  $S_2$  somewhere, we consider two subcases:

- if there is a vertex (call this vertex  $T$ ) going into  $I_1$  (of  $S_2$ ) and  $T$  has in-degree 0, then we are in Case 5, and the vertex  $T$  must be labeled 0 and we are done;
- otherwise, if all vertices going into  $I_1$  have in-degree at least 1, then we must be in Case 2, where the vertex with a self-loop must be labeled 0.

**Step 3:** If neither  $S_1$  nor  $S_2$  is found in  $G$ , then we search for  $S_3$ . If we see  $S_3$  somewhere, then again we consider two subcases:

- if there is a vertex (call this vertex  $T$ ) going into  $I_1$  and  $T$  has in-degree 0, then we are in Case 6 and the vertex  $T$  must be labeled 0;
- otherwise, if all vertices going into  $I_1$  have in-degree at least 1, then we must be in Case 3, and the vertex with a self-loop must be labeled 0.

**Step 4:** If we fail to find  $S_1, S_2$  and also  $S_3$  in  $G$ , then we must see  $S_4$ . In this case,  $I_1$  (in  $S_4$ ) must be labeled 0 and again we are done.  $\square$

#### 4. CONJUGACY OF ONE-DIMENSIONAL SFTs OBTAINED BY FORBIDDING ONE WORD

In this section, we consider the relation between  $X_{\{u\}}$  and  $X_{\{v\}}$  from a symbolic dynamics point of view. In particular, we seek conditions that imply that  $X_{\{u\}}$  and  $X_{\{v\}}$  are conjugate (conjugacy was defined in Section 1). Note that if there is a permutation on the alphabet  $[q]$  which takes  $u$  to  $v$ , then  $X_{\{u\}}$  is trivially conjugate to  $X_{\{v\}}$ . In Proposition 4.1 we identify a large class of non-trivial examples.

One simple way to get a conjugacy between  $X_{\{u\}}$  and  $X_{\{v\}}$  is to find a 1-block conjugacy with 1-block inverse from  $X_{L_{\{u\}}}$  to  $X_{L_{\{v\}}}$ . Unfortunately, Corollary 3.9 tells us that conjugacies of this type do not exist if  $u$  and  $v$  are essentially different. However, a different type of conjugacy between  $X_{\{u\}}$  and  $X_{\{v\}}$  may exist. Indeed, our main result in this section shows that, if  $u$  and  $v$  both have only trivial self-overlap, then, other than the four exceptional cases of Proposition 3.5,  $X_{\{u\}}$  is always conjugate to  $X_{\{v\}}$ . Moreover, it turns out that a version of this result also holds when the ambient shift space is the golden mean shift.

We need a definition of overlap between two different words. For two words  $u, v \in [q]^k$ , the *overlap set of  $u$  and  $v$*  is defined by

$$\text{overlap}(u, v) = \{i \in [1, k] : v_{[1, i]} = u_{[k-i+1, k]}\}.$$

Note that when  $u = v$ , this becomes the self-overlap of  $u$  defined in Section 2.

**4.1. The ambient shift is the full shift.** We start with the following proposition which says that if  $u$  and  $v$  both have trivial self-overlap and they do not overlap with each other, then  $X_{\{u\}}$  and  $X_{\{v\}}$  are conjugate to each other.

**Proposition 4.1.** *Let  $u, v \in [q]^k$ . If*

$$\text{overlap}(u, u) = \text{overlap}(v, v) = \{k\} \quad \text{and} \quad \text{overlap}(u, v) = \text{overlap}(v, u) = \emptyset,$$

*then  $X_{\{u\}}$  is conjugate to  $X_{\{v\}}$ .*

*Proof.* Denote  $u = u_1 u_2 \cdots u_k$  and  $v = v_1 v_2 \cdots v_k$ . Define a  $(2k-1)$ -block code  $\Phi : [q]^{2k-1} \rightarrow [q]$  with memory  $k-1$  and anticipation  $k-1$  by

$$(22) \quad \Phi(x_{[-k+1, k-1]}) = \begin{cases} u_{1-i} & \text{if } x_{[i, i+k-1]} = v \text{ for some } i \in [-k+1, 0] \\ v_{1-i} & \text{if } x_{[i, i+k-1]} = u \text{ for some } i \in [-k+1, 0] \\ x_0 & \text{otherwise.} \end{cases}$$

Note that  $\Phi$  is well-defined because  $u, v$  both have trivial self-overlap and no cross-correlation. Let  $\phi$  be the sliding block code induced by  $\Phi$ . Thus, we deduce from (22) that for all  $i$ ,

$$(23) \quad \phi(x)_{[i, i+k-1]} = \begin{cases} u & \text{if } x_{[i, i+k-1]} = v \\ v & \text{if } x_{[i, i+k-1]} = u, \end{cases}$$

i.e.,  $\phi$  replaces appearances of  $u$  in  $x$  with  $v$  and vice versa.

We first claim that

$$(24) \quad \phi(x)_{[i, i+k-1]} \notin \{u, v\} \text{ if } x_{[i, i+k-1]} \notin \{u, v\} \quad \text{for all } i.$$

To see this, for any fixed  $i$ , consider two cases: if there exists  $j \in [i-k+1, i+k-1]$  such that  $x_{[j, j+k-1]} \in \{u, v\}$ , then by (23),  $\phi(x)_{[j, j+k-1]} \in \{u, v\}$  and therefore  $\phi(x)_{[i, i+k-1]} \notin \{u, v\}$  since  $\text{overlap}(u, u) = \text{overlap}(v, v) = \{k\}$  and  $\text{overlap}(u, v) = \text{overlap}(v, u) = \emptyset$ ; on the other hand, if for all  $j \in [i-k+1, i+k-1]$ ,  $x_{[j, j+k-1]} \notin \{u, v\}$ , then we directly infer from (22) that  $\phi(x)_{[i, i+k-1]} = x_{[i, i+k-1]} \notin \{u, v\}$ .

Now, by (23) and (24),  $\phi$  is an involution and therefore a conjugacy on  $[q]^{\mathbb{Z}}$ . It remains to show  $\phi(X_{\{u\}}) = X_{\{v\}}$ . We first show  $\phi(X_{\{u\}}) \subset X_{\{v\}}$ . To see this, let  $x \in X_{\{u\}}$ . We claim that  $\phi(x) \in X_{\{v\}}$ . If not, there exists  $i$  such that  $\phi(x)_{[i, i+k-1]} = v$ . Then, according to (23) and (24), we must have  $x_{[i, i+k-1]} = u$ , which contradicts  $x \in X_{\{u\}}$ . A similar argument gives  $\phi(X_{\{v\}}) \subset X_{\{u\}}$ , which is equivalent to  $X_{\{v\}} \subset \phi(X_{\{u\}})$  since  $\phi$  is an involution. Thus,  $\phi(X_{\{u\}}) = X_{\{v\}}$  and the proof is complete.  $\square$

**Remark 4.2.** The conjugacy we constructed in the proof is given by replacing any appearance of  $v$  by  $u$  and vice versa. It is not hard to check that this type of conjugacy between  $X_{\{u\}}$  and  $X_{\{v\}}$  extends to a conjugacy on the full  $q$ -shift. We also note that if  $u$  can be obtained from  $v$  by permuting the alphabet, then there is another type of conjugacy between  $X_{\{u\}}$  and  $X_{\{v\}}$ . We call these two types of conjugacies together *swap conjugacies*.

To present our main results, we start with the binary alphabet case. Suppose  $q = 2$ . Let  $k$  be a positive integer and define

$$\mathcal{C}(k) := \{w \in \{0, 1\}^k \setminus \{10^{k-1}, 01^{k-1}, 1^{k-1}0, 0^{k-1}1\} : \text{overlap}(w, w) = \{k\}\}.$$

Note that here we have to exclude the four words  $10^{k-1}, 01^{k-1}, 1^{k-1}0, 0^{k-1}1$  since according to Proposition 3.5, each of these words gives an reducible SFT which can not be conjugate to any irreducible SFT.

Our first main result in this section is

**Theorem 4.3.** *Let  $q = 2$ . Given any two distinct words  $u, v \in \mathcal{C}(k)$ , there exist an integer  $N \leq 6$  and a sequence of words  $u = u^{(1)}, u^{(2)}, \dots, u^{(N-1)}, u^{(N)} = v$  in  $\mathcal{C}(k)$  such that for any  $1 \leq i < N$ , either  $u^{(i+1)}$  is obtained from  $u^{(i)}$  by flipping all bits or*

$$\text{overlap}(u^{(i)}, u^{(i+1)}) = \text{overlap}(u^{(i+1)}, u^{(i)}) = \emptyset.$$

*Consequently, there is a chain of swap conjugacies from  $X_{\{u\}}$  to  $X_{\{v\}}$ , and the length of the chain is upper bounded by 6.*

We need a series of lemmas to prove Theorem 4.3. To introduce these lemmas, we first define

$$\mathcal{C}_1(k) = \{w \in \mathcal{C}(k) : w_1 = 1, w_k = 0\}.$$

For each  $2 \leq j \leq k-2$  and  $2 \leq l \leq k-2$ , we also define sets

$$D(j) := \{w \in \mathcal{C}_1(k) : w_{[1, \lceil k/2 \rceil + j]} = 1^j 0^{\lceil k/2 \rceil}\},$$

$$E(l) := \{w \in \mathcal{C}_1(k) : w_{[\lceil k/2 \rceil - l + 1, k]} = 1^{\lceil k/2 \rceil} 0^l\}.$$

Let  $z := 1010^{k-3}$  and for each  $2 \leq j \leq k-2$ , let  $p(j) := 1^j 0^{k-j}$  be the “test words”. Note that  $z \in \mathcal{C}_1(k)$  if  $k > 4$  and  $p(j) \in D(j) \subset \mathcal{C}_1(k)$  if  $k \geq 4$  and  $2 \leq j \leq \lfloor k/2 \rfloor$ . The following fact can be verified immediately.

**Fact 4.4.** *Let  $k > 4$ . Then, for any  $2 \leq j \leq k-2$ ,*

$$\begin{aligned} \text{overlap}(z, z) &= \text{overlap}(p(j), p(j)) = \{k\} \\ \text{and} \quad \text{overlap}(p(j), z) &= \text{overlap}(z, p(j)) = \emptyset. \end{aligned}$$

We then have the following lemma.

**Lemma 4.5.** *Let  $k > 4$ . Then, for any  $w \in \mathcal{C}_1(k)$ , we have*

- (1) *If  $\text{overlap}(p(\lfloor k/2 \rfloor), w) \neq \emptyset$ , then  $w \in D(j)$  for some  $2 \leq j \leq \lfloor k/2 \rfloor$ ;*
- (2) *If  $\text{overlap}(w, p(\lfloor k/2 \rfloor)) \neq \emptyset$ , then  $w \in E(l)$  for some  $2 \leq l \leq \lceil k/2 \rceil$ .*

*Proof.* We prove item 1 here and the proof of item 2 is similar.

Suppose  $\text{overlap}(p(\lfloor k/2 \rfloor), w) \neq \emptyset$ . Since  $w$  starts with 1 and ends with 0,  $p(\lfloor k/2 \rfloor)$  must overlap with  $w$  in at least  $\lceil k/2 \rceil + 1$  many bits. On the one hand, this overlapping substring is a suffix of  $p(\lfloor k/2 \rfloor)$  and therefore it is of the form  $1^j 0^{\lceil k/2 \rceil}$  for some  $1 \leq j \leq \lfloor k/2 \rfloor$ ; on the other hand, this overlapping substring is also a prefix of  $w$ , indicating that  $w$  must start with  $1^j 0^{\lceil k/2 \rceil}$  for some  $1 \leq j \leq \lfloor k/2 \rfloor$ . Now to complete the proof, it remains to show  $j \neq 1$ . To see this, suppose to the contrary that  $j = 1$ . Then  $w$  starts with  $10^{\lceil k/2 \rceil}$ . Noting that  $w$  has trivial self-overlap, we must have  $w = 10^{k-1}$ , which is impossible.  $\square$

An immediate consequence of Lemma 4.5 is the following.

**Corollary 4.6.** *Let  $k > 4$  and  $w \in \mathcal{C}_1(k)$ . If  $w \notin D(j)$  for any  $2 \leq j \leq \lfloor k/2 \rfloor$  and  $w \notin E(l)$  for any  $2 \leq l \leq \lceil k/2 \rceil$ , then*

$$\text{overlap}(p(\lfloor k/2 \rfloor), w) = \text{overlap}(w, p(\lfloor k/2 \rfloor)) = \emptyset.$$

Lemma 4.7 and Fact 4.8 below deal with strings in  $D(j)$  and  $E(l)$ , respectively.

**Lemma 4.7.** *Let  $k > 4$ . Then, for any  $2 \leq j \leq \lfloor k/2 \rfloor$  and any  $w \in D(j)$ , either  $w = p(j)$  or*

$$\text{overlap}(w, p(j)) = \text{overlap}(p(j), w) = \emptyset.$$

*Proof.* Suppose  $w \neq p(j)$ . We will show  $\text{overlap}(w, p(j)) = \text{overlap}(p(j), w) = \emptyset$ .

Now,  $\text{overlap}(p(j), w) = \emptyset$  is straightforward. To see  $\text{overlap}(w, p(j)) = \emptyset$ , suppose to the contrary that  $\text{overlap}(w, p(j)) \neq \emptyset$ . Recalling that  $w_k = 0$ , we know  $w$  must have a suffix  $1^j 0^l$  for some  $l \geq 1$ ; on the other hand,  $w \in D(j)$  implies that  $1^j 0^{\lceil k/2 \rceil}$  is a prefix of  $w$ . Since  $w \neq p(j)$ , we must have  $l < \lceil k/2 \rceil$ . But then  $1^j 0^l$  is both a prefix and a suffix of  $w$ , contradicting the fact that  $w$  has only trivial self-overlap.  $\square$

The following fact is also easy to verify.



**Fact 4.8.** *Let  $k > 4$ . Then, for any  $2 \leq l \leq \lfloor k/2 \rfloor$  and any  $w \in E(l)$ ,*  

$$\text{overlap}(w, z) = \text{overlap}(z, w) = \emptyset.$$

We are now ready to prove Theorem 4.3.

**Proof of Theorem 4.3:** We only prove the first claim of Theorem 4.3, since the claim on the existence of the swap-conjugacy chain then follows from Proposition 4.1.

We first note that  $\mathcal{C}(k) = \emptyset$  if  $1 \leq k \leq 3$  and  $\mathcal{C}(k) = \{0011, 1100\}$  if  $k = 4$ . Therefore, Theorem 4.3 trivially holds when  $k \leq 4$ .

Now consider  $k > 4$ . Let  $p(j), z, \mathcal{C}_1(k), D(j), E(l)$  be define as before, where  $2 \leq j, l \leq k-2$ . Let

$$\mathcal{C}_2(k) = \{w \in \mathcal{C}(k) : w_1 = 0, w_k = 1\}.$$

Note that  $\mathcal{C}(k) = \mathcal{C}_1(k) \cup \mathcal{C}_2(k)$  and there is a one-to-one correspondence between words in  $\mathcal{C}_1(k)$  and words in  $\mathcal{C}_2(k)$  given by flipping all bits. Thus, it suffices to prove Theorem 4.3 for  $u, v \in \mathcal{C}_1(k)$  and for  $N \leq 5$ .

To this end, let  $u \in \mathcal{C}_1(k)$  be arbitrary. We first claim that there exist  $N \leq 3$  and a sequence of words  $u = u^{(1)}, u^{(2)}, \dots, u^{(N)} = z$  such that for any  $1 \leq i < N$ ,  $u^{(i)} \in \mathcal{C}_1(k)$  and  $\text{overlap}(u^{(i)}, u^{(i+1)}) = \text{overlap}(u^{(i+1)}, u^{(i)}) = \emptyset$ . Indeed, if  $u = p(j)$  for some  $2 \leq j \leq k-2$ , then, according to Fact 4.4, the claim holds with  $u^{(1)} = u, u^{(2)} = z$ ; if  $u \notin D(j)$  for any  $2 \leq j \leq \lfloor k/2 \rfloor$  and  $u \notin E(l)$  for any  $2 \leq l \leq \lfloor k/2 \rfloor$ , then, by Corollary 4.6 and Fact 4.4, the claim holds with  $u^{(1)} = u, u^{(2)} = p(\lfloor k/2 \rfloor), u^{(3)} = z$ ; if  $u \in D(j)$  for some  $2 \leq j \leq \lfloor k/2 \rfloor$  with  $u \neq p(j)$ , then, by Lemma 4.7 and Fact 4.4, the claim holds with  $u^{(1)} = u, u^{(2)} = p(j), u^{(3)} = z$ ; if  $u \in E(l)$  for some  $2 \leq l \leq \lfloor k/2 \rfloor$ , then by Fact 4.8, the claim again holds with  $u^{(1)} = u, u^{(2)} = z$ . Thus, the claim is proved.

Now, by applying the argument above to any  $v \in \mathcal{C}_1(k)$ , we deduce that there exist  $M \leq 3$  and a sequence of words  $z = v^{(1)}, v^{(2)}, \dots, v^{(M-1)}, v^{(M)} = v$  such that for any  $1 \leq i < M$ ,  $v^{(i)} \in \mathcal{C}_1(k)$  and  $\text{overlap}(v^{(i)}, v^{(i+1)}) = \text{overlap}(v^{(i+1)}, v^{(i)}) = \emptyset$ . The proof is then completed by combining the two sequences  $u^{(1)}, \dots, u^{(N)}$  and  $v^{(1)}, \dots, v^{(M)}$ .  $\square$

We now generalize the conjugacy result in Theorem 4.3 to larger alphabets.

**Theorem 4.9.** *Suppose  $q > 2$ . Let  $u, v \in [q]^k$  such that  $\text{overlap}(u, u) = \text{overlap}(v, v) = \{k\}$ . Then, there exist an integer  $N \leq 11$  and a sequence of words  $u = u^{(1)}, u^{(2)}, \dots, u^{(N-1)}, u^{(N)} = v$  with trivial self-overlap such that for any  $1 \leq i < N$ , either  $u^{(i+1)}$  is obtained from  $u^{(i)}$  by permuting the alphabet or*

$$\text{overlap}(u^{(i)}, u^{(i+1)}) = \text{overlap}(u^{(i+1)}, u^{(i)}) = \emptyset.$$

*Consequently, there is a chain of swap conjugacies from  $X_{\{u\}}$  to  $X_{\{v\}}$ , and the length of the chain is upper bounded by 11.*

**Remark 4.10.** Compared to Theorem 4.3, words of the form  $a^{k-1}b$  and  $ab^{k-1}$  are not excluded in Theorem 4.9 because by Remark 3.6 the SFTs given by them are irreducible (rather than reducible as in the  $q = 2$  case) when  $q > 2$ .

*Proof.* Again, we only prove the first part of the theorem, since the existence of the swap-conjugacy chain then follows from Proposition 4.1.

For any  $a, b \in [q]$  with  $a \neq b$ , define

$$\mathcal{C}_{ab}(k) := \{w \in [q]^k : w_1 = a, w_k = b, \text{overlap}(w, w) = \{k\}\}.$$

Observe that for any such pair  $a$  and  $b$ , there is a one-to-one correspondence between words in  $\mathcal{C}_{10}(k)$  and  $\mathcal{C}_{ab}(k)$  by permuting the alphabet. Thus, it suffices to prove Theorem 4.9 for

$u, v \in \mathcal{C}_{10}(k)$  and for  $N \leq 9$ . We remind the reader that words in  $\mathcal{C}_{10}(k)$  are not necessarily binary words.

We first consider the case  $k \leq 4$ . Suppose  $k = 3$ . To prove Theorem 4.9 holds for any  $u, v \in \mathcal{C}_{10}(k)$ , it suffices to show that for any word  $w \in \mathcal{C}_{10}(k)$ , there is a sequence of words  $w = w^{(1)}, w^{(2)}, \dots, w^{(N)} = 1^{k-1}0$  with trivial self-overlap such that for any  $1 \leq i < N$ , either  $w^{(i+1)}$  is obtained from  $w^{(i)}$  by permuting the alphabet or  $\text{overlap}(w^{(i)}, w^{(i+1)}) = \text{overlap}(w^{(i+1)}, w^{(i)}) = \emptyset$ . Indeed, if  $w$  uses all but the letter  $c \in [q]$ , then, by choosing  $w^{(1)} = w, w^{(2)} = c^{k-1}0, u^{(3)} = 1^{k-1}0$ , we have  $\text{overlap}(w^{(1)}, w^{(2)}) = \text{overlap}(w^{(2)}, w^{(1)}) = \emptyset$  and  $w^{(3)}$  is obtained from  $w^{(2)}$  by permuting the alphabet; if  $w \in \mathcal{C}_{10}(k)$  uses all the letters from  $[q]$ , then  $q = 3$  and we must have  $w = 120$ . Thus, choosing  $w^{(1)} = w = 120, w^{(2)} = 110$ , we have  $\text{overlap}(w^{(1)}, w^{(2)}) = \text{overlap}(w^{(2)}, w^{(1)}) = \emptyset$ . This proves the theorem for  $k = 3$ . The case  $k = 4$  can be handled similarly.

Now consider  $k > 4$ . Let  $u \in \mathcal{C}_{10}(k)$  be arbitrary. Let  $z := 1010^{k-3}$  which in particular is in  $\mathcal{C}_{10}(k)$ . We claim that, there exist  $N \leq 5$  and a sequence of words  $u = u^{(1)}, u^{(2)}, \dots, u^{(N-1)}, u^{(N)} = z$  with trivial self-overlap such that for any  $1 \leq i < N$ , either  $u^{(i+1)}$  is obtained from  $u^{(i)}$  by permuting the alphabet or  $\text{overlap}(u^{(i)}, u^{(i+1)}) = \text{overlap}(u^{(i+1)}, u^{(i)}) = \emptyset$ .

To prove this claim, we consider three cases:

- If  $u$  is binary and  $u \notin \{10^{k-1}, 01^{k-1}, 1^{k-1}0, 0^{k-1}1\}$ , then by the proof of Theorem 4.3, the claim holds with  $N = 3$ .
- If  $u \in \{10^{k-1}, 01^{k-1}, 1^{k-1}0, 0^{k-1}1\}$ , we assume without loss of generality that  $u = 10^{k-1}$ . Let  $a \in [q] \setminus \{0, 1\}$  and choose  $u^{(1)} = u = 10^{k-1}, u^{(2)} = aa0^{k-2}, u^{(3)} = 110^{k-2}, u^{(4)} = z$  (note that these words all have trivial self-overlap). Then, one can check that  $\text{overlap}(u^{(1)}, u^{(2)}) = \text{overlap}(u^{(2)}, u^{(1)}) = \emptyset$ ,  $u^{(3)}$  is obtained from  $u^{(2)}$  by permuting the alphabet, and  $\text{overlap}(u^{(3)}, u^{(4)}) = \text{overlap}(u^{(4)}, u^{(3)}) = \emptyset$ . Thus, the claim holds with  $N = 4$ .
- if  $u$  contains a symbol in  $[q] \setminus \{0, 1\}$ , let  $x(u) = 1^{l+1}0^{k-l-1}$  where  $l$  is the longest all-1 substring appearing in  $u$ . Then,  $x(u) \in \mathcal{C}_{10}(k)$ ,  $x(u)$  is binary and one check that  $\text{overlap}(x(u), u) = \text{overlap}(u, x(u)) = \emptyset$ . Moreover, according to the two cases we just proved above, there exist a sequence of words  $x(u) = u^{(1)}, u^{(2)}, \dots, u^{(K)} = z$  with trivial self-overlap such that  $K \leq 4$  and for any  $1 \leq i < K$ , either  $u^{(i+1)}$  is obtained from  $u^{(i)}$  by permuting the alphabet or  $\text{overlap}(u^{(i)}, u^{(i+1)}) = \text{overlap}(u^{(i+1)}, u^{(i)}) = \emptyset$ . Therefore, by adding  $u$  to this sequence of words, we conclude that the same thing holds for the new sequence  $u, u^{(1)}, \dots, u^{(K)}$ , i.e., the claim holds with  $N = 5$ .

Now that the claim is proved, we can apply it to any  $v \in \mathcal{C}_{10}(k)$  to deduce that there exist  $M \leq 5$  and a sequence of words  $z = v^{(1)}, v^{(2)}, \dots, v^{(M-1)}, v^{(M)} = v$  with trivial self-overlap such that for any  $1 \leq i < M$ , either  $v^{(i+1)}$  is obtained from  $v^{(i)}$  by permuting the alphabet or  $\text{overlap}(v^{(i)}, v^{(i+1)}) = \text{overlap}(v^{(i+1)}, v^{(i)}) = \emptyset$ . The proof is then completed by combining the two sequences  $u^{(1)}, \dots, u^{(N)}$  and  $v^{(1)}, \dots, v^{(M)}$ .  $\square$

Finally, it is natural to ask whether we can extend Proposition 4.1 and Theorem 4.3 to the case when  $u$  and  $v$  have nontrivial self-overlap. Indeed, the extension of the former is possible while the extension of the latter is not.

**Proposition 4.11.** (An extension of Proposition 4.1) Let  $u, v \in [q]^k$  such that

$$\text{overlap}(u, u) = \text{overlap}(v, v).$$

If

$$\text{overlap}(u, v) = \text{overlap}(v, u) = \text{overlap}(u, u) \setminus \{k\},$$

then there is a swap conjugacy between  $X_{\{u\}}$  and  $X_{\{v\}}$ .

*Proof.* The proof is similar to that of Proposition 4.1. Indeed, one can check from the proof there that  $\text{overlap}(u, v) = \text{overlap}(v, u) = \text{overlap}(u, u) \setminus \{k\}$  is sufficient to define a swap conjugacy which replaces any appearance of  $u$  by  $v$  and vice versa.  $\square$

The extension of Theorem 4.3 to words with non-trivial self-overlap is however not true, as can be seen from the following example.

**Example 4.12.** Suppose  $q = 2$ . Let  $k = 6$  and  $\text{overlap}(u, u) = \text{overlap}(v, v) = \{3, 6\}$ . Observe that 110110, 011011, 100100, 001001 are the only four binary strings of length 6 having self-overlap set  $\{3, 6\}$ . If we take, for example,  $u = 110110$  and  $v = 011011$ , then one can check that no matter how you choose the intermediate words in between  $u$  and  $v$  (so that they form a chain of words from  $u$  to  $v$ ), there always exist two adjacent words in this chain such that Proposition 4.11 fails. Therefore there does not exist a chain of swap conjugacies between  $X_{\{u\}}$  and  $X_{\{v\}}$ .

**4.2. The ambient shift is the golden mean shift.** Results in the previous subsection can be extended to the case when the ambient shift is the golden mean shift by a similar argument. We present them in this subsection.

Let  $X_{\{11\}}$  denote the golden mean shift, i.e., the set of binary bi-infinite sequences that do not contain 11. Let  $\mathcal{B}_k(\{11\})$  be the set of allowed words of length  $k$  in  $X_{\{11\}}$ . For any  $w \in \mathcal{B}_k(\{11\})$  let  $(X_{\{11\}})_{\{w\}}$  be the SFT obtained by forbidding  $w$  from the golden mean shift.

The following proposition is an analog of Proposition 4.1 for the golden mean ambient shift case.

**Proposition 4.13.** *Let  $u, v \in \mathcal{B}_k(\{11\})$  be such that*

- (1) *They have the same initial symbol and terminal symbol, i.e.,  $u_1 = v_1$  and  $u_k = v_k$ ;*
- (2) *They both have trivial self-overlap and they do not overlap with each other, that is,*

$$\text{overlap}(u, u) = \text{overlap}(v, v) = \{k\} \quad \text{and} \quad \text{overlap}(u, v) = \text{overlap}(v, u) = \emptyset.$$

*Then there is a swap conjugacy between  $(X_{\{11\}})_{\{u\}}$  and  $(X_{\{11\}})_{\{v\}}$ .*

We omit the proof since it is similar to the proof of Proposition 4.1. Item 1 of Proposition 4.13 exactly means  $u$  and  $v$  have the same “extender set” (see, for example, Section 2 of [GRP19] for the definition of extender sets). This condition is introduced here to ensure the point obtained from the replacement operation is in the golden mean shift as long as the original point is. We also point out that though the extender set condition is crucial for our techniques, it does not seem to be necessary for the conjugacy between SFTs obtained by forbidding one word (from a general ambient SFT).

We introduce some notations before generalizing Theorem 4.3. Let

$$\mathcal{R} := \{w \in \mathcal{B}_k(\{11\}) : w \in \{10^{k-1}, 0^{k-1}1, (10)^{\frac{k-1}{2}}0, 0(01)^{\frac{k-1}{2}}\}\}$$

where we regard  $(10)^{\frac{k-1}{2}}0$  and  $0(01)^{\frac{k-1}{2}}$  as empty words when  $k$  is even. One directly checks from  $L_w^{\{11\}}$  that  $(X_{\{11\}})_{\{w\}}$  is reducible when  $w \in \mathcal{R}$  and  $k \geq 3$ , and the SFTs  $\{(X_{\{11\}})_{\{w\}} :$

$w \in \mathcal{R}\}$  are never conjugate to each other. Define

$$\mathcal{G}(k) := \{w \in \mathcal{B}_k(\{11\}) : \text{overlap}(w, w) = \{k\} \text{ and } w \notin \mathcal{R}\};$$

$$\mathcal{G}_1(k) := \{w \in \mathcal{G}(k) : w_1 = 1, w_k = 0\}, \quad \mathcal{G}_2(k) := \{w \in \mathcal{G}(k) : w_1 = 0, w_k = 1\}.$$

Our main theorem in this subsection is the following, which is an analog of Theorem 4.3.

**Theorem 4.14.** *For any  $u, v \in \mathcal{G}_1(k)$  (resp.  $u, v \in \mathcal{G}_2(k)$ ), there exist an integer  $N \leq 5$  and a sequence of words  $u = u^{(1)}, u^{(2)}, \dots, u^{(N-1)}, u^{(N)} = v$  in  $\mathcal{G}_1(k)$  (resp.  $\mathcal{G}_2(k)$ ) such that for any  $1 \leq i < N$ ,*

$$\text{overlap}(u^{(i)}, u^{(i+1)}) = \text{overlap}(u^{(i+1)}, u^{(i)}) = \emptyset.$$

*Consequently, there is a chain of swap conjugacies between  $(X_{\{11\}})_{\{u\}}$  and  $(X_{\{11\}})_{\{v\}}$ , and the length of the chain is upper bounded by 5.*

We give a sketch of the proof of Theorem 4.14 since it is similar to that of Theorem 4.3. The main difference between these two proofs is the choice of “test words”.

*Proof.* We only prove the case when  $u, v \in \mathcal{G}_1(k)$  since there is a one-to-one correspondence between elements in  $\mathcal{G}_1(k)$  and elements in  $\mathcal{G}_2(k)$  given by reversing the words, and for two words  $u$  and  $v$ , condition 2 in Proposition 4.13 is preserved after reversing them.

We assume  $k \geq 7$  in the following since one can readily check that  $\mathcal{G}_1(k)$  contains at most one element when  $k < 7$ . We also assume  $k$  is even to avoid notational cumbersomeness.<sup>1</sup>

Let  $z := 10010^{k-4}$  and define the “test words”  $p(j) := (10)^j 0^{k-2j}$  for all  $2 \leq j \leq \frac{k-2}{2}$ . Note that these words are all in  $\mathcal{G}_1(k)$ . Since there are at least three consecutive 0’s at the end of each  $p(j)$ , we have for any  $2 \leq j \leq \frac{k-2}{2}$ ,

$$\text{overlap}(z, p(j)) = \text{overlap}(p(j), z) = \emptyset.$$

Now let  $m := \lceil \frac{k}{4} \rceil$  and define sets

$$D(j) := \{w \in \mathcal{G}_1(k) : w_{[1, k+2j-2m]} = (10)^j 0^{k-2m}\} \quad \text{for all } 2 \leq j \leq m$$

$$E(l) := \{w \in \mathcal{G}_1(k) : w_{[k-2m-l+1, k]} = (10)^m 0^l\} \quad \text{for all } 2 \leq l \leq k-2m.$$

Then, by a similar argument as in the proof of Lemma 4.5, we have for any  $w \in \mathcal{G}_1(k)$ , the following hold:

- (1) If  $\text{overlap}(p(m), w) \neq \emptyset$ , then  $w \in D(j)$  for some  $2 \leq j \leq m$ ;
- (2) If  $\text{overlap}(w, p(m)) \neq \emptyset$ , then  $w \in E(l)$  for some  $2 \leq l \leq k-2m$ .

Therefore, for any  $w \in \mathcal{G}_1(k)$  such that  $w \notin D(j)$  for any  $2 \leq j \leq m$  and  $w \notin E(l)$  for any  $2 \leq l \leq k-2m$ , we must have

$$\text{overlap}(w, p(m)) = \text{overlap}(p(m), w) = \emptyset.$$

We then deal with words in  $D(j)$  and  $E(l)$ . Indeed, these two cases can be handled similarly as in Lemma 4.7 and Fact 4.8: if  $w \in D(j)$  for some  $2 \leq j \leq m$ , then a similar proof to Lemma 4.7 gives that either  $w = p(j)$  or  $\text{overlap}(w, p(j)) = \text{overlap}(p(j), w) = \emptyset$ ; if  $w \in E(l)$  for some  $2 \leq l \leq k-2m$ , then, similar to Fact 4.8, we have  $\text{overlap}(w, z) = \text{overlap}(z, w) = \emptyset$ .

Combining all the cases above, we conclude that for any  $w \in \mathcal{G}_1(k)$ , there is a sequence of words  $w = w^{(1)}, \dots, w^{(N)} = z$  in  $\mathcal{G}_1(k)$  such that  $N \leq 3$  and for any  $1 \leq i < N$ ,  $\text{overlap}(w^{(i)}, w^{(i+1)}) = \text{overlap}(w^{(i+1)}, w^{(i)}) = \emptyset$ , from which the theorem follows.  $\square$

<sup>1</sup>The case  $k$  is odd, just as in the proof of Theorem 4.3, can be handled by a similarly argument.

**Remark 4.15.** It is worth noting that in Theorem 4.14 we need to consider the two disjoint sets  $\mathcal{G}_1(k)$  and  $\mathcal{G}_2(k)$  separately. This is because when the ambient shift is the golden mean shift, words in  $\mathcal{G}_1(k)$  can not be obtained from words in  $\mathcal{G}_2(k)$  by simply flipping the bits (while this is always possible when the ambient is the full 2-shift). As a simple example, let  $k = 5$ . Then  $u = 10100$  is in  $\mathcal{G}_1(k)$  but the word  $v = 01011$ , obtained by flipping all the bits of  $u$ , is not in  $\mathcal{G}_2(k)$  because it is not allowed in the golden mean shift.

For  $k \geq 7$ , choose  $w = 10010^{k-4} \in \mathcal{G}_1(k)$ . One can check from the graph  $L_w^{\{11\}}$  that  $(X_{\{11\}})_{\{w\}}$  is an irreducible SFT. Then, according to Theorem 4.14,  $(X_{\{11\}})_{\{u\}}$  is irreducible for all  $u \in \mathcal{G}_1(k)$ . A similar argument also applies to  $\mathcal{G}_2(k)$ . Thus, we have the following corollary.

**Corollary 4.16.** *Let  $w \in \mathcal{B}_k(\{11\})$  be such that  $\text{overlap}(w, w) = \{k\}$ . Then,  $(X_{\{11\}})_{\{w\}}$  is reducible if and only if  $w \in \mathcal{R}$ .*

## 5. HIGHER DIMENSIONAL SHIFT SPACES

In this section, we will have a look at  $\mathbb{Z}^d$  for  $d \geq 2$ . For patterns on a higher dimensional lattice it is already a question what the set of overlaps mean and what kind of overlap sets can arise. Further there is no nice analogue of Perron-Frobenius theory which will work in our context. Consequently everything becomes significantly more complicated and it is very rare to have general results for higher dimensional actions. This makes the results of this section, though special, significant.

Given two sets  $S, T \subset \mathbb{Z}^d$  and patterns  $u \in [q]^S, u' \in [q]^T$ , we will say that  $u$  and  $u'$  agree if for all  $i \in S \cap T$ , we have that  $u_i = u'_i$ . Observe that patterns with disjoint support agree because they don't overlap anywhere.

Given a pattern  $u$  on a set  $S \subset \mathbb{Z}^d$ , its translate the pattern  $\sigma^{-i}(u)$  has a chance to non-trivially agree with  $u$  only if their supports, the sets  $S + i$  and  $S$  intersect. If  $j$  is a point in the intersection then  $j = k + i$  for some  $k \in S$ . Rewriting this we find that  $i = j - k \in S - S$ . Thus trivially we have that the translate of a shape  $S$  by a vector  $i$  intersects itself if and only if  $i \in S - S$ . In addition if the pattern and its translate by  $i$  agree then we must have  $u_j = u_k = u_{j-i} = (\sigma^{-i}(u))_j$ . With this in mind we have the following definition of self-overlap set in higher dimension. To distinguish from the  $d = 1$  case we will call them agreement/self-agreement sets thereafter.

**Definition 5.1.** *Given a finite set  $S \subset \mathbb{Z}^d$  and a pattern  $u \in [q]^S$ , we define*

$$\text{agree}(u) = \{i \in S - S : \text{ for all } j \in S \cap (S + i) \text{ for which } u_j = u_{j-i}\}.$$

*In other words,  $\text{agree}(u)$  is the set of sites  $i \in S - S$  for which  $u$  and its translate  $\sigma^{-i}(u)$  agree on  $S \cap (S + i)$ .*

**Remark 5.2.** (1) Recall the definition of the overlap sets introduced in Section 2 for words  $w \in A^{\{1,2,\dots,k\}}$ . It is easy to see that

$$\text{agree}(w) = \{0\} \cup (k - \text{overlap}(w, w)) \cup (\text{overlap}(w, w) - k).$$

In  $d = 1$  it was better to consider overlap sets because it is more intuitive to think of the matching suffixes and prefixes and because of its relationship with the correlation polynomial. In higher dimensions, we think about agreement sets instead because there is no natural way of talking about prefixes and suffixes.

- (2) There is some redundancy in the definition: Indeed  $i \in \text{agree}(u)$  if and only if  $-i \in \text{agree}(u)$ . But beyond this, we do not have a very deep understanding of what  $\text{agree}(u)$  could look like for a given set  $S$  (see Section 6.4). When  $S$  is an interval in  $\mathbb{Z}$  then they have been characterised by Guibas and Oldyko [GO81b].

Suppose that we are given two patterns  $u^{(0)}$  and  $u^{(1)}$  where we want to replace all the appearances of  $u^{(0)}$  by  $u^{(1)}$ . This does not necessarily make sense. For instance suppose  $u^{(0)} = 101$  and  $u^{(1)} = 000$  and that we want to replace the appearances of  $u^{(1)}$  by  $u^{(0)}$  in the word  $w = 1000010101$ . Now  $u^{(1)}$  appears starting from the second and third letter but we cannot concurrently place  $u^{(0)}$  starting at the second and third position since the third letter can't concurrently be both 0 and 1. On the other hand, we can concurrently replace appearances of  $u^{(0)}$  by  $u^{(1)}$ . This motivates the following proposition.

**Proposition 5.3.** *Let  $S, T \subset \mathbb{Z}^d$  and  $u^{(0)}, u^{(1)} \in [q]^S$  be such that  $\text{agree}(u^{(0)}) \subset \text{agree}(u^{(1)})$ . Suppose we are given a pattern  $w \in [q]^T$  and  $I \subset \mathbb{Z}^d$  be such that  $w|_{i+S} = \sigma^{-i}(u^{(0)})$  for all  $i \in I$ . Then there exists  $w' \in [q]^T$  such that  $w'|_{i+S} = \sigma^{-i}(u^{(1)})$  for all  $i \in S$  and  $w'|_{T \setminus \cup_{i \in I}(i+S)} = w|_{T \setminus \cup_{i \in I}(i+S)}$ .*

*Proof.* For this proposition we need to check that the definition of  $w'$  prescribed in the proposition makes sense. In particular for all  $i_1, i_2 \in I$  we need to check that the patterns  $\sigma^{-i_1}(u^{(1)})$  and  $\sigma^{-i_2}(u^{(1)})$  agree on  $(i_1+S) \cap (i_2+S)$ . This is trivially true if  $(i_1+S) \cap (i_2+S) = \emptyset$ . Suppose instead that  $j \in (i_1+S) \cap (i_2+S)$ . We need to verify that  $(u^{(1)})_{j-i_1} = (u^{(1)})_{j-i_2}$ . Now since  $w|_{i_1+S} = \sigma^{-i_1}(u^{(0)})$  and  $\sigma^{-i_2}(u^{(0)}) = w|_{i_2+S}$  and  $(i_1+S) \cap (i_2+S) \neq \emptyset$  it follows that  $i_2 - i_1 \in \text{agree}(u^{(0)})$ . By our hypothesis,  $i_2 - i_1 \in \text{agree}(u^{(1)})$  as well. Thus  $(u^{(1)})_{j-i_1} = (u^{(1)})_{j-i_1-i_2+i_1} = (u^{(1)})_{j-i_2}$ . This completes the proof.  $\square$

Here on we will not be very careful about the support of the patterns. In particular when we say that  $u \in A^S$  and  $w|_{i+S} = u$ , we just mean that  $w|_{i+S} = \sigma^{-i}(u)$ . This is just to simplify the notation.

As before we will be interested in the properties of the shift space

$$X_{\{u\}} := \{x \in [q]^{\mathbb{Z}^d} : u \text{ does not appear in } x\}.$$

Let us strengthen this assumption of containment a bit further so that it may apply to a variety of contexts where the ambient shift space is not the full shift. Let  $X_{\mathcal{F}} \subset [q]^{\mathbb{Z}^d}$  be a shift space with a forbidden list  $\mathcal{F}$ . Given a finite set  $S \subset \mathbb{Z}^d$  consider the set of *globally allowed patterns* in  $X_{\mathcal{F}}$  given by

$$\mathcal{B}_S(X_{\mathcal{F}}) = \{x|_S : x \in X_{\mathcal{F}}\}$$

Let  $(X_{\mathcal{F}})_{\{u\}}$  be obtained by forbidding a pattern  $u \in \mathcal{B}_S(X_{\mathcal{F}})$ :

$$(X_{\mathcal{F}})_{\{u\}} := \{x \in X_{\mathcal{F}} : u \text{ does not appear in } x\}.$$

Let  $u^{(0)}, u^{(1)} \in \mathcal{B}_S(X_{\mathcal{F}})$ . We say that  $u^{(0)}$  can be *replaced* by  $u^{(1)}$  in  $X_{\mathcal{F}}$  if for all  $x \in X_{\mathcal{F}}$  and a set of sites  $I \subset \mathbb{Z}^d$  for which  $x|_{i+S} = u^{(0)}$  for all  $i \in I$ , there is a configuration  $y \in X_{\mathcal{F}}$  such that

$$y|_{i+S} = u^{(1)} \text{ for all } i \in I \text{ and } y|_{\mathbb{Z}^d \setminus \cup_{i \in I}(i+S)} = x|_{\mathbb{Z}^d \setminus \cup_{i \in I}(i+S)}.$$

Let us make some elementary observations.



- (1) We claim that a pattern  $u^{(0)}$  can be replaced by a pattern  $u^{(1)}$  in a full shift if and only if  $\text{agree}(u^{(0)}) \subset \text{agree}(u^{(1)})$ . To see this, first observe that if  $\text{agree}(u^{(0)}) \subset \text{agree}(u^{(1)})$  then  $u^{(0)}$  can be replaced by  $u^{(1)}$  in the full shift by Proposition 5.3. For the other direction suppose that  $i \in \text{agree}(u^{(0)}) \setminus \text{agree}(u^{(1)})$ . Take any configuration  $x \in [q]^{\mathbb{Z}^d}$  such that  $x|_S = x|_{i+S} = u^{(0)}$ . There exists no  $y \in [q]^{\mathbb{Z}^d}$  such that  $y|_S = y|_{i+S} = u^{(1)}$  since  $i \notin \text{agree}(u^{(1)})$ . Thus it follows that  $u^{(0)}$  cannot be replaced by  $u^{(1)}$  in the full shift.
- (2) The interpretation of this replacement condition already becomes subtle if  $X_{\mathcal{F}}$  is a non-trivial shift of finite type. Let us take a simple example. Suppose the alphabet is  $[3]$  and consider  $\mathcal{F} = \{0101, 1010\}$ ,  $u^{(1)} = 101$  and  $u^{(2)} = 122$ . Though  $\text{agree}(u^{(1)}) \not\subset \text{agree}(u^{(2)})$  (in fact  $\text{agree}(u^{(2)}) \subset \text{agree}(u^{(1)})$ ) we have that  $u^{(1)}$  is replaceable by  $u^{(2)}$ . Let us see why. Let  $x \in X_{\mathcal{F}}$  and  $I \subset \mathbb{Z}$  such that  $x|_{i+S} = u^{(1)}$  for all  $i \in I$ . First notice that if  $i_1, i_2 \in I$  then  $(i_1 + S) \cap (i_2 + S) = \emptyset$ . This is because any non-trivial agreement forms the pattern 10101 which is forbidden in the shift space. This implies that replacing  $u^{(1)}$  on the set  $I$  by  $u^{(2)}$  forms a valid point  $y \in [3]^{\mathbb{Z}}$ . Further since it only replaces some 0's and 1's by 2's it cannot possibly form a forbidden word which wasn't already there (which are composed of 0's and 1's). Thus  $y \in X_{\mathcal{F}}$ .

How and when can patterns  $u^{(0)}$  and  $u^{(1)}$  be replaced to give a bijection is an interesting question. A little harder is to understand, when can we ensure that these replacements lead to conjugacy? This is related not only to self-agreements of the patterns but also agreements of the pattern  $u^{(0)}$ ,  $u^{(1)}$  and the ambient shift. We have discussed some of these issues in Section 4 (look at Propositions 4.1 and 4.13).

We will be interested in patterns  $u^{(0)}$  and  $u^{(1)}$  which are *mutually replaceable* in  $X$ , meaning that  $u^{(0)}$  can be replaced by  $u^{(1)}$  and  $u^{(1)}$  can be replaced by  $u^{(0)}$  in  $X$ .

Many similar properties can be found in the literature. For instance Meyerovitch had a notion of exchangeability in [Mey13] which allows for the replacement of a pattern  $u$  by a pattern  $v$  and vice versa one at a time rather than simultaneously and used it to prove that for all measures of maximal entropy  $\mu$ ,  $\mu([u]) = \mu([v])$ . Meyerovitch's condition is weaker and thus his conclusions follows for us as well. However our results do not hold under his weaker assumptions. Indeed, according to his definition, (say) the patterns 11 and 10 would be exchangeable in the full shift but  $h(X_{\{11\}}) > h(X_{\{10\}})$  by Proposition 2.2. A similar property was also defined by Garcia-Ramos and Pavlov which allowed for replacements of a pattern  $u$  by a pattern  $v$  (and not vice versa) one at a time [GRP19] in points  $x \in X$ . However in our case, their agreement sets are different and that is why they are not mutually replaceable and by Proposition 2.2 we have that  $|\mathcal{B}_{[1,n]}(X_{\{11\}})| > |\mathcal{B}_{[1,n]}(X_{\{10\}})|$  for large enough  $n$  (in fact all  $n \geq 3$ ).

We will compare the size of the languages of various shift spaces in this section. Given a finite set  $S \subset \mathbb{Z}^d$  and a shift space  $X_{\mathcal{F}}$  and a pattern  $u \in \mathcal{B}_U(X_{\mathcal{F}})$  for some  $U$ , we write

$$\mathcal{B}_S(X_{\mathcal{F}}, u) = \{x|_S : x \in X_{\mathcal{F}} \text{ and } u \text{ does not appear in } x|_S\}.$$

Note that if  $\mathcal{F}$  is the empty set,  $S$  is the interval  $[1, n]$  and  $u$  is a word, then  $\mathcal{B}_S(X_{\mathcal{F}}, u)$  is just  $\mathcal{B}_n(u)$  defined for the 1-dimensional case.

Now we are prepared to state the main results of this section.

**Theorem 5.4.** *Let  $X_{\mathcal{F}}$  be a shift space and  $u^{(0)}, u^{(1)}$  be mutually replaceable in  $X_{\mathcal{F}}$ . Then for all finite sets  $S \subset \mathbb{Z}^d$  there is a bijection between  $\mathcal{B}_S(X_{\mathcal{F}}, u^{(0)})$  and  $\mathcal{B}_S(X_{\mathcal{F}}, u^{(1)})$ .*

**Remark 5.5.** (1) This generalises the implication  $(1 \implies 6)$  in Proposition 2.1.  
 (2) We believe that the same argument works for countable groups as well however we preferred to keep the simpler context which already illustrates the key issues.

*Proof of Theorem 5.4.* The proof is a simple application of the inclusion-exclusion principle. Assume that  $u^{(0)}, u^{(1)} \in [q]^U$  and that  $0 \in U$ . Given a subset  $T \subset S$  such that  $T + U \subset S$  and  $k = 0, 1$  consider

$$\mathcal{B}_S^{T,k}(X_{\mathcal{F}}) = \{w \in \mathcal{B}_S(X_{\mathcal{F}}) : w|_{i+U} = u^{(k)} \text{ for } i \in T\}.$$

Now consider the map  $\phi_{T,k} : \mathcal{B}_S^{T,k}(X_{\mathcal{F}}) \rightarrow \mathcal{B}_S^{T,1-k}(X_{\mathcal{F}})$  obtained by replacing appearances of  $u^{(k)}$  by  $u^{(1-k)}$ , namely,

$$(\phi_{T,k}(w))|_{i+U} = u^{(1-k)} \text{ for } i \in T \text{ and } (\phi_{T,k}(w))|_{S \setminus \cup_{i \in T}(i+U)} = w|_{S \setminus \cup_{i \in T}(i+U)}.$$

This gives an injective map from  $\mathcal{B}_S^{T,k}(X_{\mathcal{F}})$  to  $\mathcal{B}_S^{T,1-k}(X_{\mathcal{F}})$  for  $k = 0, 1$  proving that the cardinalities of the two sets are the same.

Now notice that the set  $\mathcal{B}_S(X_{\mathcal{F}}, u^{(k)})$  is the set of patterns from  $X_{\mathcal{F}}$  on  $S$  in which  $u^{(k)}$  does not appear even once. By inclusion-exclusion we have that

$$|\mathcal{B}_S(X_{\mathcal{F}}, u^{(k)})| = \sum_{T+U \subset S} (-1)^{|T|} |\mathcal{B}_S^{T,k}(X_{\mathcal{F}})|.$$

We remark that the summation on the right hand side also includes  $T = \emptyset$ , in which case,  $\mathcal{B}_S^{T,k}(X_{\mathcal{F}}) = \mathcal{B}_S(X_{\mathcal{F}})$ . Since  $|\mathcal{B}_S^{T,0}(X_{\mathcal{F}})| = |\mathcal{B}_S^{T,1}(X_{\mathcal{F}})|$  for all  $T \subset S$ , this completes the proof.  $\square$

**Remark 5.6.** One may be tempted to conjecture here that if  $u^{(0)}$  is replaceable by  $u^{(1)}$  in  $X_{\mathcal{F}}$  then there is an injection from  $\mathcal{B}_S(X_{\mathcal{F}}, u^{(0)})$  and  $\mathcal{B}_S(X_{\mathcal{F}}, u^{(1)})$ . This is not clear! Indeed we only get an injection from  $\mathcal{B}_S^{T,0}(X_{\mathcal{F}})$  to  $\mathcal{B}_S^{T,1}(X_{\mathcal{F}})$  and we cannot seem to use inclusion-exclusion directly.

**Remark 5.7.** Note that the map does not give a direct bijection between  $\mathcal{B}_S(X_{\mathcal{F}}, u^{(0)})$  and  $\mathcal{B}_S(X_{\mathcal{F}}, u^{(1)})$  when  $u^{(0)}$  is replaceable by  $u^{(1)}$ . We have observed a similar phenomenon in Section 4. More directly one can observe this when one tries to replace appearance of  $u^{(0)} = 1100$  by  $u^{(1)} = 1000$  in the full shift. Though both of them have the same (trivial) agreement set, notice that if one is to carry out the replacement in 11100 then one obtains 11000 which still has the word  $u^{(0)}$ . One can obviously apply the same map again but it is not clear if it would lead to a bijection. The reason is that the two patterns  $u^{(0)}$  and  $u^{(1)}$  have non-trivial agreement with each other (see Proposition 4.1).

The following corollary will follow easily from some fundamental properties of entropy.

**Corollary 5.8.** *Let  $X_{\mathcal{F}}$  be a shift space and  $u^{(0)}, u^{(1)}$  be mutually replaceable in  $X_{\mathcal{F}}$ . Then*

$$h((X_{\mathcal{F}})_{\{u^{(0)}\}}) = h((X_{\mathcal{F}})_{\{u^{(1)}\}}).$$

The proof will make use of the following concept. Given a forbidden list of patterns  $\mathcal{G}$ , the set of *locally allowed* patterns in  $X_{\mathcal{G}}$  on the set  $S$  is given by

$$\mathcal{B}_S^{loc}(\mathcal{G}) := \{a \in A^S : \text{patterns from } \mathcal{G} \text{ do not appear in } a\}.$$

While this list of patterns is dependent on the specific choice of forbidden patterns  $\mathcal{G}$ , it follows that unlike the elements of globally allowed patterns, that is,  $\mathcal{B}(X_{\mathcal{G}})$ , the set of locally allowed patterns is algorithmically computable if  $\mathcal{G}$  is finite.

*Proof.* We have that

$$\mathcal{B}_S(X_{\mathcal{F}, u^{(k)}}) \subset \mathcal{B}_S(X_{\mathcal{F}}, u^{(k)}) \subset \mathcal{B}_S^{loc}(\mathcal{F} \cup \{u^{(k)}\}).$$

It follows from [Rue04, Equation 3.9], [Fri97] that

$$\begin{aligned} h((X_{\mathcal{F}})_{\{u^{(k)}\}}) &= \lim_{n \rightarrow \infty} \frac{1}{|[1, n]^d|} \log (|\mathcal{B}_{[1, n]^d}(X_{\mathcal{F}, u^{(k)}})|) \\ &= \lim_{n \rightarrow \infty} \frac{1}{|[1, n]^d|} \log (|\mathcal{B}_{[1, n]^d}^{loc}(\mathcal{F} \cup \{u^{(k)}\})|) \end{aligned}$$

Thus it follows that

$$h((X_{\mathcal{F}})_{\{u^{(k)}\}}) = \lim_{n \rightarrow \infty} \frac{1}{|[1, n]^d|} \log (|\mathcal{B}_{[1, n]^d}(X_{\mathcal{F}}, u^{(k)})|).$$

The corollary now follows from Theorem 5.4.  $\square$

A similar statement also follows for periodic points. Given a subgroup  $\Lambda \subset \mathbb{Z}^d$ , we denote the set of periodic points with period  $\Lambda$  in a shift space  $X_{\mathcal{F}}$  by

$$\text{per}^{\Lambda}(X_{\mathcal{F}}) = \{x \in X_{\mathcal{F}} : \sigma^i(x) = x \text{ for all } i \in \Lambda\}$$

and for all  $S \subset \mathbb{Z}^d$  its language

$$\text{per}_S^{\Lambda}(X_{\mathcal{F}}) = \{x|_S : x \in \text{per}^{\Lambda}(X_{\mathcal{F}}), \sigma^i(x) = x \text{ for all } i \in \Lambda\}.$$

For instance if a sequence  $x \in A^{\mathbb{Z}}$  has period 10, the corresponding subgroup is  $\Lambda = 10\mathbb{Z}$ . Let  $D \subset \mathbb{Z}^d$  be a fundamental domain for  $\Lambda$ , that is, for all  $i \in \mathbb{Z}^d$  there exists a unique  $\lambda \in \Lambda$  and  $d \in D$  such that  $d + \lambda = i$ . For instance in the case  $\Lambda = 2\mathbb{Z} \times \{0\}$ , a fundamental domain corresponding to it is  $D = \{0, 1\} \times \mathbb{Z}$ . Let  $X_{\mathcal{F}}$  be a shift space and  $u \in \mathcal{B}_U(X_{\mathcal{F}})$  be a pattern for some finite  $U \subset \mathbb{Z}^d$ .

Define for all  $S \subset D$ , the restricted language of periodic points in which  $u$  does not appear on  $S$ . More symbolically, we denote it by

$$\text{per}_S^{\Lambda}(X_{\mathcal{F}}, u) = \{x|_S : x \in \text{per}^{\Lambda}(X_{\mathcal{F}}) \text{ and } x_{i+U} \neq u \text{ for all } i \in S\}.$$

Note the subtle difference between definition of  $\text{per}_S^{\Lambda}(X_{\mathcal{F}}, u)$  and  $\mathcal{B}_S(X_{\mathcal{F}}, u)$ . In the former we look at patterns on  $S + U$  while in the latter we look at patterns on  $S$ . This is simply because in the former we are restricting to  $S \subset D$  and even if  $u$  appears in a periodic point  $x \in \text{per}^{\Lambda}(X_{\mathcal{F}})$  it may not appear in  $x|_D$ . The following proposition generalises Theorem 5.4.

**Proposition 5.9.** *Let  $\Lambda \subset \mathbb{Z}^d$  be a subgroup and  $D \subset \mathbb{Z}^d$  be a fundamental domain for  $\Lambda$ . Let  $u^{(0)}$  and  $u^{(1)}$  be patterns in a shift space  $X_{\mathcal{F}}$  which are mutually replaceable. Then for all finite sets  $S \subset D$  there is a bijection between  $\text{per}_S^{\Lambda}(X_{\mathcal{F}}, u^{(0)})$  and  $\text{per}_S^{\Lambda}(X_{\mathcal{F}}, u^{(1)})$ .*

The proof is very similar to the proof of Theorem 5.4 and thus we only sketch the idea. For the set  $T \subset S$  consider the set

$$\text{per}_S^{\Lambda, T, k}(X_{\mathcal{F}}) = \{x|_S : x \in \text{per}^{\Lambda}(X_{\mathcal{F}}) \text{ and } x_{i+U} = u^{(k)} \text{ for } i \in T\}.$$

By using a map which replaces  $u^{(0)}$  by  $u^{(1)}$  on  $i + \lambda + U$  for all  $i \in T, \lambda \in \Lambda$  and similarly another one which replaces  $u^{(1)}$  by  $u^{(0)}$  we have that there is a bijection between the sets  $\text{per}_S^{\Lambda, T, 0}(X_{\mathcal{F}})$  and  $\text{per}_S^{\Lambda, T, 1}(X_{\mathcal{F}})$ . Finally by inclusion-exclusion we have a bijection between the sets  $\text{per}_S^{\Lambda}(X_{\mathcal{F}}, u^{(0)})$  and  $\text{per}_S^{\Lambda}(X_{\mathcal{F}}, u^{(1)})$ .

Since  $\mathbb{Z}^d$  acts on  $\text{per}^\Lambda(X_{\mathcal{F}})$  with stabiliser  $\Lambda$  we have a natural action of the group  $\mathbb{Z}^d/\Lambda$  on  $\text{per}^\Lambda(X_{\mathcal{F}})$ . It follows that the topological entropy of the action of  $\mathbb{Z}^d/\Lambda$  is simply equal to

$$h_{\mathbb{Z}^d/\Lambda}(X_{\mathcal{F}}) = \lim_{n \rightarrow \infty} \frac{\log(|\text{per}_{[1,n]^d}^\Lambda(X_{\mathcal{F}})|)}{\text{number of cosets of } \Lambda \text{ intersecting } [1,n]^d}.$$

The following is an easy and direct corollary.

**Corollary 5.10.** *Let  $\Lambda \subset \mathbb{Z}^d$  be a subgroup. Let  $u^{(0)}$  and  $u^{(1)}$  be patterns in a shift space  $X_{\mathcal{F}}$  which are mutually replaceable. Then the topological entropy of  $\text{per}^\Lambda((X_{\mathcal{F}})_{\{u^{(0)}\}})$  equals the topological entropy of  $\text{per}^\Lambda((X_{\mathcal{F}})_{\{u^{(1)}\}})$  for the action of the group  $\mathbb{Z}^d/\Lambda$ . In addition the cardinality of  $\text{per}^\Lambda((X_{\mathcal{F}})_{\{u^{(0)}\}})$  equals the cardinality of  $\text{per}^\Lambda((X_{\mathcal{F}})_{\{u^{(1)}\}})$ .*

**Remark 5.11.** *This generalises the implication  $1 \implies 5$  from Proposition 2.1 and generalises a consequence of a theorem of Lind [Lin89, Theorem 1].*

*Proof of Corollary 5.10.* The proof of the entropy statement is analogous to the proof of Corollary 5.8. By Proposition 5.9 for all  $S \subset D$  there is a bijection between  $\text{per}_S^\Lambda(X_{\mathcal{F}}, u^{(0)})$  and  $\text{per}_S^\Lambda(X_{\mathcal{F}}, u^{(1)})$ . Now as in Corollary 5.8 we can use ideas from [Rue04, Equation 3.9], [Fri97] to prove that the topological entropies must be equal. Finally note that as  $S \uparrow D$ ,

$$|\text{per}_S^\Lambda(X_{\mathcal{F}}, u^{(k)})| \uparrow |\text{per}^\Lambda((X_{\mathcal{F}})_{\{u^{(k)}\}})|$$

for both  $k = 1, 2$ . Given the bijection in Proposition 5.9, we see that the cardinalities must be equal. This completes the proof.  $\square$

## 6. DISCUSSION

**6.1. The conjugacy problem for shift spaces with one forbidden word.** One of the major question which we weren't able to address in this paper is that of conjugacy.

**Question 6.1.** *Given two words  $w, w' \in A^k$ , is there an algorithm to determine whether the shift spaces  $X_{\{w\}}$  and  $X_{\{w'\}}$  are conjugate?*

It follows from Proposition 2.1 that a necessary condition is that the correlation polynomials for  $w$  and  $w'$  are identical. In the case that  $w$  and  $w'$  have only trivial self-overlaps, Theorem 4.9 provides a conjugacy. Beyond that, we don't know much. For instance, though we know from Example 4.12 that there does not exist a chain of swap conjugacies between the shift spaces  $X_{\{110110\}}$  and  $X_{\{011011\}}$ , we don't know whether they are conjugate. In this case, the correlation polynomials are identical and one can check that the shift spaces agree on the main known conjugacy invariants (see [LM21, Section 12.3, Exercise 12.3.8]).

**6.2. From follower set graph to correlation polynomial.** One significant question left open by our results is the implicit connection between the correlation polynomial  $\phi_w$  and the graphs  $L_w$ , or more concretely, the  $d_i(a)$  edge configurations. On one hand, the correlation polynomial is a natural object in terms of both hitting times and entropies. Namely, the explicit formula  $\mathbb{E}\tau_w = q\phi_w(q)$  shows that each self-overlap of  $w$  contributes to the expected hitting time of  $w$  in an explicit way; and the connection between  $\phi_w$  and the entropy  $h(X_{\{w\}})$  is also concrete, if slightly less explicit, the entropy being the largest real root less than  $q$  of the polynomial  $1 + (t - q)\phi_w(t)$ . On the other hand, the condition  $D(w, w')$  is also natural when comparing hitting times and entropies of two candidate shifts  $X_{\{w\}}$  and  $X_{\{w'\}}$ , as Propositions

3.12 and 3.15 show. These theorems are also more intuitive than the corresponding facts for the correlation polynomial:  $D(w, w')$  guarantees that ‘more edges go further back in the chain  $L$ ,’ which heuristically suggests the Markov chain should take longer to hit the terminal state, and also that the graph should have a larger space of allowable paths (and thus higher entropy).

As a result, one expects to be able to convert between these two viewpoints. Indeed, by Theorem 3.8, each graph  $L_w$  corresponds to a unique word  $w$ , and thus a unique correlation polynomial  $\phi_w$ . However, the proof of Theorem 3.8 constructs this correspondence implicitly. Let  $\mathcal{P}$  denote the family of polynomials that occur as correlation polynomials for some alphabet  $[q]$  and word  $w$  over  $[q]$ , and let  $\mathcal{L}$  denote the set of graphs arising as follower set graphs of some such  $w$ . The family  $\mathcal{P}$  can be described by explicit (but complex) ‘forward/backward propagation’ rules, or alternatively by a recursive algorithm [GO81b]. Is it possible to give a similar characterization of the family  $\mathcal{L}$ ? One attempt is to observe that the family  $\mathcal{L}$  is closed under taking prefixes, i.e. if  $L \in \mathcal{L}$ , then the subgraph of  $L$  obtained by deleting the vertex with label  $k - 1$ , if  $L = L_w$  for  $w \in [q]^k$  is also in  $\mathcal{L}$ . Thus, one could hope that there exists an algorithm to decide which extensions of a given  $L \in \mathcal{L}$  are also in  $\mathcal{L}$ . Realizing this sketch would require finding additional properties of  $\mathcal{L}$  beyond those in Proposition 3.4.

Another issue with condition  $D(w, w')$  is that it induces an incomplete partial ordering on the set of words  $w \in [q]^k$ , in the sense that there are pairs  $w, w'$  such that neither  $D(w, w')$  nor  $D(w', w)$  holds, but  $h(X_{\{w\}}) \neq h(X_{\{w'\}})$  ( $w = 1001$  and  $w' = 1100$  is an example). Additionally, the resulting poset is not ‘closed from below:’ there exist words  $w \in [q]^k$  whose entropy is strictly larger than the minimum possible entropy among words of length  $k$ , but there is no word  $w'$  such that  $D(w, w')$  holds. (A minimal example for the binary alphabet is  $w = 1011$ .) The poset is ‘closed from above,’ in that  $D(1^k, w)$  holds for all words  $w \in [q]^k$ . Is it possible to weaken the condition  $D(w, w')$  to something which is compatible with follower set graphs, and which induces the same ordering on words  $w \in [q]^k$  as ordering by entropy of the shifts  $X_{\{w\}}$ , or equivalently, the ordering of the hitting times  $\tau_w$  by stochastic domination? Perhaps the alternative assumptions appearing in the main theorems of [DSS16] are a good starting point.

**6.3. When the ambient shift space is an SFT with memory one.** When the ambient shift space is the full  $q$ -shift, given words  $u$  and  $v$  of the same length,  $h(X_{\{u\}})$  and  $h(X_{\{v\}})$  can be easily compared using the correlation polynomial:  $h(X_{\{u\}}) > h(X_{\{v\}})$  iff  $\phi_u(q) > \phi_v(q)$ ; see Proposition 2.2. It is natural to ask if this result can be extended to some cases where the ambient shift space  $X$  is not the full shift, by replacing  $q$  with  $\lambda$  where  $h(X) = \log(\lambda)$ . Numerical evidence suggests that at least a weak inequality of this type may be true (under suitable conditions). We show that this is actually true in some very special cases below (Proposition 6.2).

Let  $X_{\mathcal{F}}$  be an irreducible SFT with memory one over  $[q]$ . Let  $T$  be the associated matrix of  $X_{\mathcal{F}}$ , i.e., it is a  $q$ -by- $q$  matrix and  $T_{ij} = 1$  if and only if  $ij$  is allowed in  $X_{\mathcal{F}}$ . Let  $\chi_T(t)$  denote the characteristic polynomial of  $T$  and  $\lambda_T$  be the largest root of  $\chi_T(t) = 0$ . For an allowed word  $w$  in  $X_T$  of length  $k$ , let  $\phi_w(t)$  denote the correlation polynomial of the 2-block representation of  $w$  (the 2-block representation of  $w = w_1w_2 \cdots w_k$  is  $w^{[2]} = (w_1w_2)(w_2w_3) \cdots (w_{k-1}w_k)$ , which is a  $(k-1)$ -block over the alphabet  $[q] \times [q]$ ). The *extender set* of  $w$  is the set of words that can be combined with  $w$  to form a new allowed word in  $X_T$ .

Let  $(X_{\mathcal{F}})_{\{w\}}$  be the SFT obtained from  $X_T$  by forbidding  $w$ . Just as in the case of the full shift, there is an analogous conjugacy presentation of  $(X_{\mathcal{F}})_{\{w\}}$ , meaning a labeled

graph s.t. the set of all bi-infinite label sequences is  $(X_{\mathcal{F}})_{\{w\}}$  and the map from bi-infinite edge sequences to bi-infinite label sequences is one-to-one. Moreover, there is an analogous characteristic polynomial  $\chi_w(t)$ , with nonzero constant term, whose largest root  $\lambda_w$  satisfies  $h(X_{\mathcal{F}})_{\{w\}} = \log \lambda_w$  (for any conjugacy presentation).

Letting  $i, j$  denote the initial, terminal symbols of  $w$ , one easily derives from [Lin89, Theorem 1] that

$$(25) \quad \chi_w(t) = \chi_T(t)\phi_w(t) + \text{cof}_{ij}(tI - T)$$

where  $\text{cof}_{ij}(t - IT)$  equals  $(-1)^{i+j}$  times the determinant of the matrix  $t - IT$  with its  $i$ th row and  $j$ th column deleted, and

**Proposition 6.2.** *Let  $X_{\mathcal{F}}$  be an ambient irreducible SFT with memory one, and  $u, v$  be allowed words in  $X_{\mathcal{F}}$  with length  $k$  and the same extender set. Let  $\eta_T$  be the second largest real root of  $\chi_T(t) = 0$ . Then any of the following implies  $\phi_u(\lambda_T) > \phi_v(\lambda_T) \iff h((X_{\mathcal{F}})_{\{u\}}) > h((X_{\mathcal{F}})_{\{v\}})$ :*

- (a)  $\lambda_T \geq 2$ ,  $\eta_T \geq 2$  and  $\text{cof}_{ij}(\eta_T I - T) < 0$ ;
- (b)  $\lambda_T > 2$ ,  $\eta_T \leq 2$  and  $\text{cof}_{ij}(2I - T) < 0$ ; (or  $\lambda_T > 2$ ,  $\eta_T < 2$  and  $\text{cof}_{ij}(2I - T) \leq 0$ )
- (c)  $\lambda_T > 2$ ,  $\eta_T < 2$ ,  $\text{cof}_{ij}(2I - T) > 0$  and  $k \geq \log \frac{-\text{cof}_{ij}(2I - T)}{\chi_T(2)} + 1$ .
- (d)  $\lambda_T \geq 2$  and  $k$  is large enough.

*Proof.* It suffices to show  $\phi_u(\lambda_T) > \phi_v(\lambda_T) \iff \lambda_u > \lambda_v$ . Note that we have  $\lambda_T \geq 2$  in all the four cases. We first write

$$\phi_u(t) = \sum_{i=0}^{k-1} a_i t^i, \quad \phi_v(t) = \sum_{i=0}^{k-1} b_i t^i$$

where  $a_i, b_i \in \{0, 1\}$ . Define  $r := \max\{i : a_i = 1, b_i = 0\}$ . This  $r$  is well-defined since  $\lambda_T > 2$ . Now  $\phi_u(t) - \phi_v(t) > t^r - \sum_{i=0}^{r-1} t^i$ . Let  $\lambda^*$  be the largest root of  $t^r - \sum_{i=0}^{r-1} t^i = 0$ .

( $\Rightarrow$ ) : To prove this direction, we first claim it suffices to show  $\lambda_v > \max\{\eta, \lambda^*\}$ . For if  $\lambda_v > \max\{\eta_T, \lambda_r\}$ , then  $\chi_T(\lambda_v) < 0$  and  $\phi_u(\lambda_v) - \phi_v(\lambda_v) > (\lambda_v)^r - \sum_{i=0}^{r-1} (\lambda_v)^i > 0$ . Thus, by (25),

$$\chi_u(\lambda_v) = \chi_u(\lambda_v) - \chi_v(\lambda_v) = \chi_T(\lambda_v)(\phi_u(\lambda_v) - \phi_v(\lambda_v)) < 0$$

and therefore  $\lambda_u > \lambda_v$ .

Now, it is not hard to show that all the four condition implies  $\lambda_v > \{\eta_T, \lambda_r\}$ :

- Case (a): Recall from (25) that  $\chi_v(\eta_T) = \chi_T(\eta_T)\phi_v(\eta_T) + \text{cof}_{ij}(\eta_T I - T)$ . Now the conditions in (a) imply  $\chi_v(\eta_T) < 0$ , which gives  $\lambda_v > \eta_T$ . Observing that  $\lambda^* < 2$ , we indeed have  $\lambda_v > \max\{\eta_T, \lambda^*\}$ .
- Case (b): By a similar argument, it is not hard to derive  $\chi_v(2) < 0$ . So  $\lambda_v > 2$  and therefore  $\lambda_v > \max\{\eta_T, \lambda^*\}$ .
- Case (c): Note that  $\chi_T(2) < 0$  in this case. Since  $\phi_v(2) \geq 2^{k-1}$ , we derive from (25) that  $\chi_v(2) < \chi_T(2)2^{k-1} + \text{cof}_{ij}(2I - T) \leq 0$  because  $k \geq \log \frac{-\text{cof}_{ij}(2I - T)}{\chi_T(2)} + 1$  by assumption. Therefore,  $\lambda_v > 2$  which immediately gives  $\lambda_v > \max\{\eta_T, \lambda^*\}$ .
- Case (d): Irreducibility of  $X$  implies  $\eta_T < \lambda_T$ . Also  $\lambda^* < 2 \leq \lambda_T$ . When  $k$  is large enough, the entropy of  $(X_{\mathcal{F}})_{\{v\}}$  is close to the entropy of  $X_{\mathcal{F}}$ , i.e.,  $\lambda_v$  is close to  $\lambda_T$ , which implies  $\lambda_v > \max\{\eta_T, \lambda^*\}$ .

( $\Leftarrow$ ): Let  $\lambda_u > \lambda_v$ . We assume to the contrary that  $\phi_u(\lambda_T) \leq \phi_v(\lambda_T)$ . If  $\phi_u(\lambda_T) < \phi_v(\lambda_T)$ , then the proof of the other direction implies  $\lambda_u < \lambda_v$ , a contradiction; if  $\phi_u(\lambda_T) = \phi_v(\lambda_T)$ ,



then  $\phi_u(t) = \phi_v(t)$  for all  $t$  and therefore (25) tells us  $\chi_u(t) = \chi_v(t)$  for all  $t$  (where we use the fact that the term  $\text{cof}_{ij}(tI - T)$  is the same for  $u$  and  $v$  due to the same extender set assumption). Thus,  $\lambda_u = \lambda_v$ , a contradiction.  $\square$

**6.4. Agreement sets in higher dimensions.** We would like to discuss the question: Is there a “nice” characterisation of agreement sets for patterns in  $\mathbb{Z}^d$ ? To illustrate what we mean by “nice”, in the case when  $d = 1$  and the underlying shape is an interval, patterns are just words. The agreement sets of words are well-understood [GO81b, Theorem 5.1]. This theorem gives us two facts.

- (1) A nice algorithmic way of generating agreement sets.
- (2) Two conditions: Forward and backward propagation which characterise what subsets of  $\mathbb{Z}$  are agreement sets.

As far as our study goes such a characterisation seems to be missing when  $d \geq 2$ . For us, this question is critical because it is the first step to understand how the agreement set of a pattern influences the drop in entropy when it is forbidden. It also seems to be very important in other contexts, for instance, in pattern matching [ABK<sup>+</sup>22]. The second motivation lead to extensive enquiry on agreement sets starting from the papers: [AB92, RR93]. Nevertheless a characterisation of agreement sets seems to be open. A nice survey on this and related problems can be found in [Smi17].

We will end with a couple of remarks:

- (1) In  $d = 1$ , [GO81b] showed that the collection of agreement sets of words in the binary alphabet equals the collection for arbitrarily sized alphabet. It is not enough to look at the binary alphabet to characterise agreement sets in higher dimensions. Indeed a quick enumeration will show that there exists no  $w \in \{0, 1\}^{[1,2]^2}$  such that  $\text{agree}(w) = \{(0, 0)\}$  but clearly the agreement set of the pattern  $\frac{1}{3} \frac{2}{4} \in \{1, 2, 3, 4\}^{[1,2]^2}$  has only  $\{(0, 0)\}$  as its self-agreement.
- (2) If the shape  $S \subset \mathbb{Z}^d$  is not convex then it may also happen for  $w \in A^S$  that  $u \in \text{agree}(w)$  but  $2u \notin \text{agree}(w)$ . One can already see this in  $d = 1$  when the set  $S$  is not an interval but here is a connected example in 2 dimensions. Consider the pattern

$$w = \begin{array}{cc} 1 & 1 \\ 1 & 0 \end{array}$$

It is clear  $(0, 1) \in \text{agree}(w)$  but  $(0, 2) \notin \text{agree}(w)$ . This example shows that forward propagation (see [GO81b] for the definition) is not true for the agreement sets for general shapes in  $\mathbb{Z}^2$ .

**6.5. Higher dimensional SFTs.** In Section 5 we discussed how the equality of the agreement sets implies equality of entropy. However a much more interesting and fundamental question is to understand how the agreement sets can cause a drop in entropy.

For instance if we are given  $w, w' \in A^{[-n,n]^d}$ , is there an easy way to decide whether or not  $h(X_{\{w\}}) > h(X_{\{w'\}})$ ? Towards this direction here is a conjecture that we were unable to prove.

**Conjecture 6.3.** *Let  $w, w' \in A^{[-n,n]^d}$  be such that  $\text{agree}(w) \subsetneq \text{agree}(w')$ . Then  $h(X_{\{w\}}) < h(X_{\{w'\}})$ .*

The conjecture even if proved brings us only a tiny bit closer to understanding how the agreement sets influence entropy. The first step of this program is to first understand what agreement sets look like; this has already been mentioned before in Section 6.4. A baby

version of the more general question (which does not satisfy the hypothesis of the conjecture), is to determine which of the following has larger entropy  $X_{\{q^{(1)}\}}$  or  $X_{\{q^{(2)}\}}$  where

$$q^{(1)} = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix} \text{ and } q^{(2)} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Intuitively, we have that  $q^{(1)}$  has ‘smaller’ agreements than  $q^{(2)}$  thus as in  $d = 1$  we conjecture that  $h(X_{\{q^{(1)}\}}) > h(X_{\{q^{(2)}\}})$ . The more general version of such a conjecture will require a deeper understanding of agreement sets.

To show how much more complicated obtaining strict inequalities of entropy in higher dimensions tends to be, we will now give a proof sketch in a simple case which we know how to handle. In the following we assume that the reader is familiar with the basics of higher dimensional symbolic dynamics. Let  $p^{(1)}$  be an  $n \times n$  pattern with a single 1 and rest as 0’s while  $p^{(2)}$  is an  $n \times n$  pattern full of 0’s.

**Claim 6.4.**  $h(X_{\{p^{(1)}\}}) < h(X_{\{p^{(2)}\}})$ .

Let us prove the claim. For this we will use the *lexicographic order* on  $\mathbb{Z}^2$ :  $(i, j) > (k, l)$  if  $j > l$  or  $j = l$  and  $i > k$ . The reverse of this order is called the *reverse lexicographic order*.

We begin by constructing an injective map from  $\mathcal{B}_{[1,m]^2}(X_{\{p^{(1)}\}})$  to  $\mathcal{B}_{[1,m]^2}(X_{\{p^{(2)}\}})$ . Let  $p$  be a pattern in  $\mathcal{B}_{[1,m]^2}(X_{\{p^{(1)}\}})$ . In lexicographic order of appearance, replace the appearances of  $p^{(2)}$  by  $p^{(1)}$  one by one in  $p$ . Observe

- (1) The replacement cannot create another  $p^{(1)}$  except where the replacement takes place.
- (2) The replacement cannot create an appearance of  $p^{(2)}$ .

One can prove that this procedure is reversible because one can now replace appearances of  $p^{(1)}$  by  $p^{(2)}$  in the reverse lexicographic order to get back the pattern  $p$ . By continuing the replacement until all appearances of  $p^{(2)}$  have been replaced by  $p^{(1)}$  we get a pattern  $p' \in \mathcal{B}_{[1,m]^2}(X_{\{p^{(2)}\}})$ . Since the procedure is reversible, we have that the map  $p \rightarrow p'$  is injective and

$$|\mathcal{B}_{[1,m]^2}(X_{\{p^{(1)}\}})| \leq |\mathcal{B}_{[1,m]^2}(X_{\{p^{(2)}\}})|.$$

Now we will observe that certain patterns from  $\mathcal{B}_{[1,n+1]^2}(X_{\{p^{(2)}\}})$  cannot appear in such a  $p'$ . Start with the  $(n+1) \times (n+1)$  pattern full of 0’s. In the reverse lexicographic order, one by one, replace the appearances of  $p^{(2)}$  by  $p^{(1)}$  to get a pattern  $q$ . Observe that  $q \in \mathcal{B}(X_{\{p^{(2)}\}})$ . However  $q$  cannot appear in a pattern  $p'$  constructed as above. For the sake of contradiction suppose that there did appear such a  $p'$ . The proof splits into three cases:

- (1) *The 1 appearing in  $p^{(1)}$  appears at the upper right corner:* In this case  $q$  is the pattern with four appearances of 1 clumped in the upper right corner. To obtain  $p$  from  $p'$ , we can now replace the appearance of  $p^{(1)}$  by  $p^{(2)}$  in  $p'$  in the reverse lexicographic order. Only one of the appearances of 1 in  $q$  will be replaced by 0, and so the resulting pattern  $p$  will still have a  $p^{(1)}$ . Thus  $p \notin \mathcal{B}_{[1,m]^2}(X_{\{p^{(1)}\}})$  which leads to a contradiction.
- (2) *1 appears in top or right edge in  $p^{(1)}$ :* This is similar to the case above and here too we obtain  $p \notin \mathcal{B}_{[1,m]^2}(X_{\{p^{(1)}\}})$ .
- (3) *1 does not appear in top or right edge on  $p^{(1)}$ :* Suppose the 1 appears at the  $(i, j)$  position in  $p^{(1)}$  where  $i < n$  and  $j < n$ . In this case  $q$  is the pattern supported on the  $(n+1) \times (n+1)$  box (say  $B$ ) which is full of 0’s except at  $(i+1, j+1)$  where it

is 1. One can easily see that going from  $p'$  to  $p$ , the  $q$  is replaced by a pattern full of 0's on  $B$ . Now to reconstruct  $p'$  we would have to replace the appearance of  $p^{(2)}$  by  $p^{(1)}$  in the lexicographic order. But then it follows that there is  $(i', j') \in B$  which is lexicographically smaller than  $(i + 1, j + 1)$  such that  $p'_{(i', j')} = 1$ , a contradiction.

Thus we find that in fact

$$|\mathcal{B}_{[1, m]^2}(X_{\{p^{(1)}\}})| \leq |\mathcal{B}_{[1, m]^2}((X_{\{p^{(2)}\}})_{\{q\}})|.$$

It follows that  $h(X_{\{p^{(1)}\}}) \leq h((X_{\{p^{(2)}\}})_{\{q\}})$ . To prove the claim it is enough to prove that  $h((X_{\{p^{(2)}\}})_{\{q\}}) < h(X_{\{p^{(2)}\}})$ . This follows because 1 is a safe symbol<sup>2</sup> in  $X_{\{p^{(2)}\}}$  which implies that  $X_{\{p^{(2)}\}}$  is entropy minimal, that is, forbidding any pattern from  $X_{\{p^{(2)}\}}$  causes a loss in entropy [QS03]. This completes the sketch.

We see that the argument here is of a very different nature from those in  $d = 1$ . At the same time it is somewhat special for the patterns  $p^{(1)}$  and  $p^{(2)}$  which have been especially chosen to demonstrate the method. The only other result known to us in this context is by Pavlov [Pav11] who gives upper and lower bounds in the drop in entropy when a single pattern is forbidden, depending on the shape of the pattern, in the case of strongly irreducible subshifts.

## REFERENCES

- [AB92] Amihood Amir and Gary Benson. Two-dimensional periodicity and its applications. SODA '92, page 440–452, USA, 1992. Society for Industrial and Applied Mathematics.
- [ABK<sup>+</sup>22] Amihood Amir, Ayelet Butman, Eitan Kondratovsky, Avivit Levy, and Dina Sokol. Multidimensional period recovery. *Algorithmica*, 84(6):1490–1510, 2022.
- [ACT24] N. Agarwal, H. Cheriya, and S. N. Tikekar. On escape rate for subshift with Markov measure. *arXiv preprint arXiv:2401.05118*, 2024.
- [BY11] Leonid A. Bunimovich and Alex Yurchenko. Where to place a hole to achieve a maximal escape rate. *Israel J. Math.*, 182:229–252, 2011.
- [CA23] Haritha Cheriya and Nikita Agarwal. Subshifts of finite type with a hole. *J. Aust. Math. Soc.*, 115(1):73–98, 2023.
- [CHR24] J. Carrigan, I. Hollars, and E. Rowland. A natural bijection for contiguous pattern avoidance in words. *Discrete Mathematics*, 347(3):113793, 2024.
- [DSS16] Emilio De Santis and Fabio L. Spizzichino. Some sufficient conditions for stochastic comparisons between hitting times for skip-free markov chains. *Methodology And Computing In Applied Probability*, 18, 12 2016.
- [Eri97] Kimmo Eriksson. Autocorrelation and the enumeration of strings avoiding a fixed string. *Combin. Probab. Comput.*, 6(1):45–48, 1997.
- [FP12] Andrew Ferguson and Mark Pollicott. Escape rates for Gibbs measures. *Ergodic Theory Dynam. Systems*, 32(3):961–988, 2012.
- [Fri97] Shmuel Friedland. On the entropy of zd subshifts of finite type. *Linear algebra and its applications*, 252(1-3):199–220, 1997.
- [GO81a] L. J. Guibas and A. M. Odlyzko. String overlaps, pattern matching, and nontransitive games. *J. Combin. Theory Ser. A*, 30(2):183–208, 1981.
- [GO81b] Leo J Guibas and Andrew M Odlyzko. Periods in strings. *Journal of Combinatorial Theory, Series A*, 30(1):19–42, 1981.

---

<sup>2</sup>1 is a safe symbol is a  $\mathbb{Z}^d$  shift space  $X$  if for all  $x \in X$  and  $k \in \mathbb{Z}^d$  the configuration  $y$  given by

$$y_l = \begin{cases} x_l & \text{if } l \neq k \\ 1 & \text{if } l = k \end{cases}$$

is an element of  $X$

- [GRP19] Felipe García-Ramos and Ronnie Pavlov. Extender sets and measures of maximal entropy for subshifts. *Journal of the London Mathematical Society*, 100(3):1013–1033, 2019.
- [JS15] Philippe Jacquet and Wojciech Szpankowski. *Analytic pattern matching: from DNA to Twitter*. Cambridge University Press, 2015.
- [KL09] Gerhard Keller and Carlangelo Liverani. Rare events, escape rates and quasistationarity: some exact formulae. *J. Stat. Phys.*, 135(3):519–534, 2009.
- [Li80] Shuo-Yen Robert Li. A Martingale Approach to the Study of Occurrence of Sequence Patterns in Repeated Experiments. *The Annals of Probability*, 8(6):1171 – 1176, 1980.
- [Lin89] D. A. Lind. Perturbations of shifts of finite type. *SIAM J. Discrete Math.*, 2(3):350–365, 1989.
- [LM21] Douglas Lind and Brian Marcus. *An introduction to symbolic dynamics and coding*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, second edition, 2021.
- [Mey13] Tom Meyerovitch. Gibbs and equilibrium measures for some families of subshifts. *Ergodic Theory and Dynamical Systems*, 33(3):934–953, 2013.
- [Pav11] Ronnie Pavlov. Perturbations of multidimensional shifts of finite type. *Ergodic Theory and Dynamical Systems*, 31(2):483–526, 2011.
- [Pen69] Walter Penney. *Journal of Recreational Mathematics*, page 241, 1969.
- [PY79] Giulio Pianigiani and James A. Yorke. Expanding maps on sets which are almost invariant. Decay and chaos. *Trans. Amer. Math. Soc.*, 252:351–366, 1979.
- [QŞ03] Anthony Quas and Ayşe A Şahin. Entropy gaps and locally maximal entropy in subshifts. *Ergodic Theory and Dynamical Systems*, 23(4):1227–1245, 2003.
- [RR93] Mireille Régnier and Ladan Rostami. A unifying look at d-dimensional periodicities and space coverings. In *Annual Symposium on Combinatorial Pattern Matching*, pages 215–227. Springer, 1993.
- [Rue04] David Ruelle. *Thermodynamic formalism*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, second edition, 2004. The mathematical structures of equilibrium statistical mechanics.
- [Smi17] Taylor Smith. Properties of two-dimensional words. Master’s thesis, University of Waterloo, 2017.
- [Sol66] A. D. Solov’ev. A combinatorial identity and its application to the problem concerning the first occurrence of a rare event. *Theory of Probability & Its Applications*, 11(2):276–282, 1966.

## APPENDIX: VALIDITY OF THE ALGORITHM IN THE PROOF OF THEOREM 3.19

*Proof of Step 1:* Suppose  $S_1$  appears in  $G$ . We first show that the labels of  $I_1, I_2, I_3$  must be  $\mathcal{L}(I_1) = j, \mathcal{L}(I_2) = j + 1, \mathcal{L}(I_3) = j + 2$  for some  $j \geq 0$  (here, for any vertex  $J \in \mathcal{V}(G)$ ,  $\mathcal{L}(J)$  denotes the label of  $J$ ). To see this, first observe that the labels of  $I_1, I_2$  must be two adjacent integers since  $I_1, I_2$  are in the same 2-cycle. Now if  $\mathcal{L}(I_1) = j + 1, \mathcal{L}(I_2) = j$  for some  $j \geq 0$ , then the labeled version of  $S_1$  is given in Figure 4 (where we use the fact that outgoing edges at  $I_2$  have different labels). But (as you can see from the edge labeling in Figure 4) the forward edge from  $I_2$  to  $I_1$  is labeled 1, which, by Fact 3.18, means  $I_1$  should not have a “going back edge”, a contradiction. Thus, we must have  $\mathcal{L}(I_1) = j, \mathcal{L}(I_2) = j + 1$ . But then, the edge  $I_2 \rightarrow I_3$  must be a “forward edge” and therefore  $\mathcal{L}(I_3) = j + 2$ .

Now we claim that  $j = 0$ . If not, then there is a forward edge from  $j - 1$  to  $j$ . But then, Fact 3.21 implies this edge must be labeled 0, and Fact 3.22 implies this edge must be labeled 1, which is a contradiction.

Therefore, if we see  $S_1$ , then the vertex  $I_1$  in  $S_1$  (cf. Figure 3) must be labeled 0.

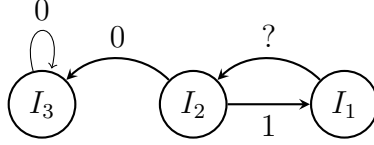


FIGURE 4. If  $\mathcal{L}(I_1) = j + 1, \mathcal{L}(I_2) = j, \mathcal{L}(I_3) = l$  for  $S_1$ , where  $l < j$ .

*Proof of Step 2:* If  $S_2$  appears in  $G$ :

Again, we first consider the label of the vertices  $I_1, I_2, I_3$  in  $S_2$ . The existence of the 2-cycle at  $I_2, I_3$  implies that the labels of  $I_2, I_3$  must be two adjacent integers. We claim that we must have  $\mathcal{L}(I_2) < \mathcal{L}(I_3)$ . If not, then the edge  $I_3 \rightarrow I_2$  must be a “forward edge”, which means  $\mathcal{L}(I_3) = \mathcal{L}(I_2) - 1$ . On the other hand, noting that the edge  $I_1 \rightarrow I_2$  must also be a “forward edge”, we have  $\mathcal{L}(I_1) = \mathcal{L}(I_2) - 1$ , which means  $I_1$  and  $I_3$  are the same vertex, a contradiction.

So assume  $\mathcal{L}(I_2) = j + 1, \mathcal{L}(I_3) = j + 2$  for some  $j$ . Since  $I_1$  has a self-loop, the edge from  $I_1$  to  $I_2$  must be a “forward edge”. Thus, we have  $\mathcal{L}(I_1) = j$ .

We then consider what value  $j$  can take. By Fact 3.21 and Fact 3.22, one have  $j \in \{0, 1\}$ . Indeed  $j = 0$  corresponds to Case 2 of Figure 2,  $j = 1$  corresponds to Case 5 of Figure 2, and the way to distinguish these two cases is to check the in-degree of vertices that can go to  $I_1$ : for case 2, any vertex that goes into the self-loop-vertex must have in-degree  $\geq 1$ ; for case 5, the vertex 0 goes into the self-loop-vertex and we claim that the vertex 0 must have in-degree 0. This is because if there is another vertex  $l$  such that  $d_l = 0$ , then  $w_l \overline{w_{l+1}} = 01$  so the “going back edge” at  $l$  will never go to 0 (it at least goes to the vertex 2, see Figure 5).

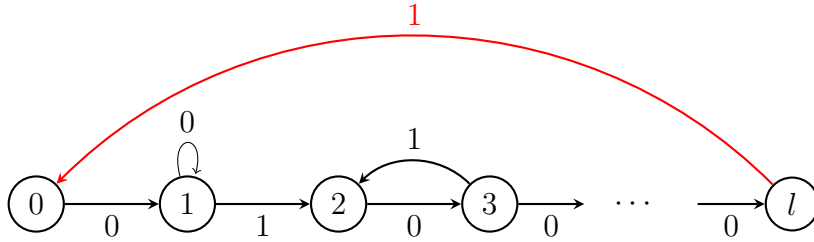


FIGURE 5. An illustration of why the in-degree of the vertex 0 is 0 for Case 5. Here, the red edge is invalid.

Having known how to distinguish Case 2 from Case 5, identifying the vertex labeled 0 is straight forward: when we are in Case 2, then the vertex with a self-loop will be labeled 0; if we are in Case 5, then there is a unique vertex with in-degree 0 and there is an edge from this vertex to  $I_1$ ; this vertex must be labeled 0.

*Proof of Step 3:* Suppose  $S_1$  and  $S_2$  do not appear in  $G$ , and  $S_3$  appears. We first claim that the vertices of  $S_3$  must be labeled  $\mathcal{L}(I_1) = j, \mathcal{L}(I_2) = j + 1, \mathcal{L}(I_3) = j + 2$  for some  $j \geq 0$ . To see this, noting that  $I_1$  has a self-loop, the edge  $I_1 \rightarrow I_2$  must be a forward edge, which implies  $\mathcal{L}(I_1) = j$  and  $\mathcal{L}(I_2) = j + 1$  for some  $j \geq 0$ . Now, if  $\mathcal{L}(I_3) \neq j + 2$ , then the edge  $I_2 \rightarrow I_3$  must be a “going back edge,” from which, we infer that  $\mathcal{L}(I_3) = j - 1$ . Now,

by Fact 3.21, the forward edge  $I_1 \rightarrow I_2$  must be labeled 1. But then, by Fact 3.18,  $I_2$  should not have a “going back edge,” a contradiction.

For the value of  $j$ , we first claim that  $j \leq 2$ . For if not, then the self-loop at the vertex  $j$  implies  $w_1 w_2 \cdots w_j = 0^j$  where  $j > 2$ . But then, one check that  $w_{j+1} w_{j+2} \overline{w_{j+3}} = 100$  and  $w_{j-2} w_{j-1} w_j = 000$ , which means the “going back edge”  $j + 2 \rightarrow j$  should not exist.

Thus, there are at most three possibilities:  $j = 0$ ,  $j = 1$  or  $j = 2$ . We claim that  $j = 2$  is impossible. Indeed, if  $j = 2$ , then  $w_1 w_2 w_3 w_4 w_5 = 00101$ , and the unlabeled version of the subgraph involving vertices 0, 1, 2 is exactly  $S_1$  (see Figure 6), which contradicts the assumption we made at the beginning of this case.

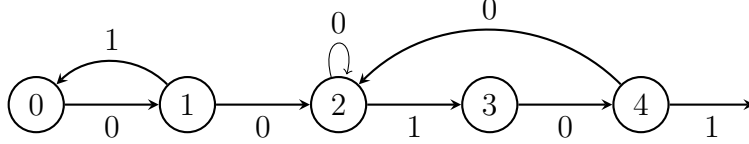


FIGURE 6. An illustration of why  $j \neq 2$  in the proof of Step 3.

Thus, we must have  $j \in \{0, 1\}$ . Indeed, one directly verifies that  $j = 0$  corresponds to Case 3 of Figure 2, and  $j = 1$  corresponds to Case 6 of Figure 2, and the way to distinguish these two cases is to check the in-degree of vertices that can go to the vertex  $I_2$ : for Case 3, any vertex that goes into the self-loop vertex must have in-degree  $\geq 1$ ; for Case 6, the vertex 0 goes into the self-loop vertex and, for a similar reason as in the proof of Step 2, the vertex 0 must have in-degree 0.

Having known how to distinguish Case 3 from Case 6, identifying the vertex labeled 0 is straight forward: if we are in Case 3, then the vertex with a self-loop will be labeled 0; if we are in Case 6, then there is a unique vertex with in-degree 0 and goes into  $I_1$ , this vertex must be labeled 0.

*Proof of Step 4:* If  $S_1, S_2, S_3$  do not appear in  $G$ , then  $S_4$  must appear. Now, by checking all the six cases in Figure 2, we conclude that we must be in Case 4. But in Case 4,  $S_4$  must appear at vertices 0, 1 and 2. Thus, the vertex  $I_1$  in  $S_4$  must be labeled 0.

TATA INSTITUTE OF FUNDAMENTAL RESEARCH - CENTRE FOR APPLICABLE MATHEMATICS, BENGALURU

Email address: nishant.chandgotia@gmail.com

THE UNIVERSITY OF BRITISH COLUMBIA, VANCOUVER

Email address: marcus@math.ubc.ca

THE ALFRÉD RÉNYI INSTITUTE OF MATHEMATICS, BUDAPEST

Email address: jfrichey001@gmail.com

THE UNIVERSITY OF BRITISH COLUMBIA, VANCOUVER

Email address: chengyuw@connect.hku.hk