

# LARGE DEVIATIONS FOR SYSTEMS WITH NON-UNIFORM STRUCTURE

VAUGHN CLIMENHAGA, DANIEL J. THOMPSON,  
AND KENICHIRO YAMAMOTO

ABSTRACT. We use a weak Gibbs property and a weak form of specification to derive level-2 large deviations principles for symbolic systems equipped with a large class of reference measures. This has applications to a broad class of symbolic systems, including  $\beta$ -shifts,  $S$ -gap shifts, and their factors. A crucial step in our approach is to prove a ‘horseshoe theorem’ for these systems.

## 1. INTRODUCTION

We introduce criteria for a symbolic system to satisfy the large deviations principle. These criteria are motivated by the ‘non-uniform’ structure of our main examples –  $\beta$ -shifts,  $S$ -gap shifts, and their factors – but apply more generally. We prove the following main result. (See §2 for precise definitions.)

**Theorem A.** *Let  $(X, \sigma)$  be a shift on a finite alphabet,  $m$  a Borel probability measure on  $X$ , and  $\varphi: X \rightarrow \mathbb{R}$  a continuous function. Let  $\mathcal{L}$  be the language of  $X$ . Suppose there exists a set  $\mathcal{G} \subset \mathcal{L}$  such that*

**[A.1]** *there exists  $\tau > 0$  such that for every  $v, w \in \mathcal{G}$  there exists  $u \in \mathcal{L}$  with  $|u| \leq \tau$  such that  $vuw \in \mathcal{G}$ ;*

**[A.2]**  *$\mathcal{L}$  is edit approachable by  $\mathcal{G}$ ;*

**[A.3]**  *$m$  is Gibbs for  $\varphi$  with respect to the collection  $\mathcal{G}$ .*

*Then  $(X, \sigma)$  satisfies a level-2 large deviations principle with reference measure  $m$  and rate function  $q^\varphi: \mathcal{M}(X) \rightarrow [-\infty, 0]$  given by*

$$(1.1) \quad q^\varphi(\mu) = \begin{cases} h(\mu) + \int \varphi d\mu - P(\varphi) & \mu \in \mathcal{M}_\sigma(X), \\ -\infty & \text{otherwise.} \end{cases}$$

Condition **[A.1]** is a form of the specification property for  $\mathcal{G}$ . Condition **[A.2]** means that any word  $w \in \mathcal{L}$  can be transformed into a

---

*Date:* July 19, 2015.

*2000 Mathematics Subject Classification.* 37A50, 60F10, 37D35, 37D25, 37B10.

V.C. was partially supported by NSF grant DMS-1362838. D.T. was partially supported by NSF grants DMS-1101576 and DMS-1259311.

word in  $\mathcal{G}$  without making too many edits. The condition **[A.3]** means that  $m$  satisfies an upper Gibbs bound on all cylinders, and a lower Gibbs bound on cylinders corresponding to words in  $\mathcal{G}$ . Level-2 large deviations gives an exponential decay rate for the measure of the set of points whose empirical averages are experiencing a ‘large deviation’ from their expected value (see §2.1).

The criteria we introduce can be verified for many shift spaces including  $S$ -gap shifts,  $\beta$ -shifts, and their factors, using a large class of equilibrium states as reference measures. For  $\beta$ -shifts, large deviations was previously known only in the case that the reference measure is the measure of maximal entropy (MME) [29]. For  $S$ -gap shifts, and subshift factors of  $\beta$ -shifts and  $S$ -gap shifts, these are the first results on large deviations.

The recent work of [8, 9] has provided the necessary groundwork to make our criteria verifiable for a large class of symbolic systems, by giving conditions for a potential  $\varphi$  to have a unique equilibrium state with the weak Gibbs property **[A.3]** – we state these in Theorem 3.1. In particular, for  $\beta$ -shifts, we proved in [9] that we have a unique equilibrium state with the weak Gibbs property for every Hölder continuous potential. In this paper, we show the same result for  $S$ -gap shifts.

Our approach belongs to the ‘orbit-gluing approach’ to large deviations, which relies on direct constructions based on the specification property (or one of its variants). Classic and recent references include [18, 15, 39, 29, 38]. The basic strategy is

- to obtain a (weak) Gibbs bound for the reference measure using constructive techniques;
- to establish the entropy density of ergodic measures.

Our approach is related to powerful general techniques of Pfister and Sullivan [29], who introduced two hypotheses from which large deviations follow: the approximate product property for the system, and the existence of (upper and lower) weak energy functions for the reference measure. Their results show that a uniform upper Gibbs property yields the upper large deviations bound, and we apply that result in this paper (see §4.1). However, their hypotheses for the lower large deviations bound are harder to verify; Pfister and Sullivan used an ad hoc argument in the case of the MME of a  $\beta$ -transformation, but this approach has so far not been extended to other settings. Our approach to the lower large deviations bound is similar in philosophy to Pfister and Sullivan but has some important novelties:

- (1) An axiomatic approach which explicitly shows how to derive large deviations results from a weak Gibbs property which is known to hold for a large class of equilibrium states;
- (2) Hypotheses which in many examples are more convenient to verify than those of [29];
- (3) A crucial step in our proof is to establish entropy density of the ergodic measures *supported on horseshoes* (see Theorem B). This is a point of independent interest in our approach, which is analogous to the Katok horseshoe theorem in smooth dynamics.

We discuss the relationship between our hypotheses and those of Pfister and Sullivan in §6. We expect that our approach can be adapted to non-symbolic topological dynamical systems, where analogues of some of the ideas in this paper have been introduced in [10, 11].

**A ‘horseshoe’ theorem.** The following result is a key step in our proof of Theorem A, and clarifies item (3) from the previous section. A more precise version of this statement is proved as Proposition 3.6.

**Theorem B.** *Let  $X$  be a shift space and suppose that  $\mathcal{G} \subset \mathcal{L}(X)$  satisfies [A.1] and [A.2]. Then there exists a family  $\{X_n\}_{n \in \mathbb{N}}$  of transitive sofic subshifts of  $X$  such that every invariant measure on  $X$  is entropy approachable by ergodic measures on  $X_n$ ; that is, for any  $\eta > 0$ , any  $\mu \in \mathcal{M}_\sigma(X)$ , and any neighborhood  $U$  of  $\mu$  in  $\mathcal{M}_\sigma(X)$ , there exist  $n \geq 1$  and  $\mu' \in \mathcal{M}_\sigma^e(X_n) \cap U$  such that  $h(\mu') > h(\mu) - \eta$ .*

**Comparison with other approaches.** Large deviations results in dynamics have received a great deal of attention since their introduction in the 1980’s [27, 28, 34, 35, 14, 21, 39, 1, 16, 17]. We recommend the introductions of [36, 26, 33, 12, 6] for extensive references and discussion. In particular, the ‘functional approach’ and the ‘tower approach’ are powerful alternatives to the ‘orbit-gluing’ approach developed here.

The ‘functional approach’ as described in [12] relies on differentiability of a certain functional on the space of observables and this approach has been used successfully in e.g. [21, 34, 35, 24, 12] for examples including rational maps of the Riemann sphere. The key requirements are a weak version of the Gibbs property, such as [12, (1.5)], and the existence of a dense subspace  $W \subset C(X)$  such that every  $\psi \in W$  has a unique equilibrium state. For  $\beta$ -shifts and  $S$ -gap shifts, the dense subspace condition is satisfied, but we do not see any obvious way to verify [12, (1.5)] for these examples.

The ‘tower approach’ relies on relating the original system to a countable state Markov shift via a tower construction, and yields results on both exponential and sub-exponential rates of decay. These results

typically use an SRB reference measure. The case of more general reference measures is largely unexplored using this approach. Classic and recent references include [33, 20, 32, 31, 26, 25, 5, 6].

**Layout of the paper.** In §2, we establish our definitions. In §3, we give various consequences of Theorem A using the thermodynamic results developed in [8, 9], including applications to  $\beta$ -shifts,  $S$ -gap shifts, and their factors. In §4, we prove Theorem A. In §5, we give proofs that the examples ( $\beta$ -shifts and  $S$ -gap shifts) satisfy the conditions of Theorem A. In §6, we make the connection between our hypotheses and the weak lower energy functions of Pfister and Sullivan. The proofs of lemmas which are not proved in the body of the text appear in §7.

**Acknowledgments.** We are grateful to Zheng Yin for pointing out an error in an early version of this paper. We would also like to thank Henri Comman for useful discussions, and the anonymous referee for helpful suggestions.

## 2. DEFINITIONS

Let  $A$  be a finite set and  $A^{\mathbb{N}}$  (resp.  $A^{\mathbb{Z}}$ ) be the set of all one-sided (resp. two-sided) infinite sequences on the alphabet  $A$ , with the standard metric  $d(x, y) = 2^{-t(x, y)}$ , where  $t(x, y) = \min\{|k| \mid x_k \neq y_k\}$ . The shift map on  $A^{\mathbb{N}}$  is  $\sigma: x_1x_2\cdots \mapsto x_2x_3\cdots$ , and the shift map on  $A^{\mathbb{Z}}$  is defined analogously. A subshift is a closed  $\sigma$ -invariant set  $X \subset A^{\mathbb{N}}$  or  $X \subset A^{\mathbb{Z}}$ . All of the results and proofs in this paper apply equally to one-sided and two-sided shifts, so we treat both cases simultaneously.

**2.1. Large deviations principles.** The large deviations principle referred to in Theorem A describes the rate of convergence of empirical averages relative to a fixed reference measure  $m$ . More precisely, one writes  $\delta_y$  for the Dirac measure concentrated at  $y$ , and  $\mathcal{E}_n(x) := \frac{1}{n} \sum_{j=0}^{n-1} \delta_{\sigma^j x}$  for the *empirical measure* associated to the orbit segment  $x, \sigma(x), \dots, \sigma^{n-1}(x)$ . If  $m$  is  $\sigma$ -invariant and ergodic, then  $\mathcal{E}_n(x) \rightarrow m$  for  $m$ -a.e.  $x$ , and one can quantify this convergence by studying the rate of decay of  $m\{x \mid \mathcal{E}_n(x) \in U\}$ , where  $U$  is a suitable subset of the space of all probability measures on  $X$ . For the systems studied here, this quantity decays exponentially in  $n$  whenever  $m \notin U$ ; thus the goal is to describe *rate functions*  $\underline{r}(U)$  and  $\bar{r}(U)$  that bound the lower and upper limits of  $\frac{1}{n} \log m\{x \mid \mathcal{E}_n(x) \in U\}$ . This type of result is called level-2 large deviations. One may also consider level-1 large deviations and study  $m\{x \mid \frac{1}{n} S_n \varphi(x) \in V\}$  for some fixed observable  $\varphi$  and some  $V \subset \mathbb{R}$ . For continuous  $\varphi$ , level-2 results imply level-1 results via the contraction principle (see [12]).

We now state the level-2 large deviations principle precisely. Denote by  $\mathcal{M}(X)$  the set of all Borel probability measures on  $X$  with the weak\* topology. This topology is induced by the metric

$$D(\mu, \nu) := \sum_{n=1}^{\infty} \frac{|\int \varphi_n d\mu - \int \varphi_n d\nu|}{2^{n+1} \|\varphi_n\|_{\infty}},$$

where  $\{\varphi_n\} \subset C(X)$  is a countable dense subset. Let  $\mathcal{M}_{\sigma}(X) \subset \mathcal{M}(X)$  be the set of  $\sigma$ -invariant measures, and let  $\mathcal{M}_{\sigma}^e(X) \subset \mathcal{M}_{\sigma}(X)$  be the set of ergodic measures.

**Definition 2.1.** We say that the system  $(X, \sigma)$  satisfies a *level-2 large deviations principle* with a reference measure  $m \in \mathcal{M}(X)$  and a rate function  $q: \mathcal{M}(X) \rightarrow [-\infty, 0]$  if  $q$  is upper semicontinuous,

$$\varliminf_{n \rightarrow \infty} \frac{1}{n} \log m(\{x \in X \mid \mathcal{E}_n(x) \in U\}) \geq \sup_{\mu \in U} q(\mu)$$

holds for any open set  $U \subset \mathcal{M}(X)$ , and

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log m(\{x \in X \mid \mathcal{E}_n(x) \in F\}) \leq \sup_{\mu \in F} q(\mu)$$

holds for any closed set  $F \subset \mathcal{M}(X)$ .

**2.2. Languages and decompositions.** The *language of  $X$* , denoted by  $\mathcal{L} = \mathcal{L}(X)$ , is the set of finite words that appear in some  $x \in X$  – that is,

$$\mathcal{L}(X) = \{w \in A^* \mid [w] \neq \emptyset\},$$

where  $A^* = \bigcup_{n \geq 0} A^n$  and  $[w]$  is the central cylinder for  $w$ , which in the one-sided case is the set of sequences  $x \in X$  that begin with the word  $w$ . Given  $w \in \mathcal{L}$ , let  $|w|$  denote the length of  $w$ . For any collection  $\mathcal{D} \subset \mathcal{L}$ , let  $\mathcal{D}_n$  denote  $\{w \in \mathcal{D} \mid |w| = n\}$ . Thus,  $\mathcal{L}_n$  is the set of all words of length  $n$  that appear in sequences belonging to  $X$ . Given words  $u, v$ , we use juxtaposition  $uv$  to denote the word obtained by concatenation.

A *decomposition* for  $\mathcal{L}$  is a choice of three sets of words  $\mathcal{C}^p, \mathcal{G}, \mathcal{C}^s \subset \mathcal{L}$ , together with a map from  $\mathcal{L}$  to  $\mathcal{C}^p \times \mathcal{G} \times \mathcal{C}^s$  which assigns to a word  $w \in \mathcal{L}$  a triple of words  $u^p \in \mathcal{C}^p, v \in \mathcal{G}, u^s \in \mathcal{C}^s$  such that  $w = u^p v u^s$ . We write  $\mathcal{L} = \mathcal{C}^p \mathcal{G} \mathcal{C}^s$  when the language can be decomposed in this way.<sup>1</sup> We make a standing assumption that  $\emptyset \in \mathcal{C}^p, \mathcal{G}, \mathcal{C}^s$  to allow for

<sup>1</sup>In [8], we defined a decomposition slightly differently – there, we assumed there exists (possibly multiple) ways to decompose each word. Our definition there did not carry the information on *which* decomposition to use for a given word like we do here. See Kwietniak, Oprocha and Rams [22] for a clarification of this issue.

words in  $\mathcal{L}$  that belong purely to one of the three collections (this is also implicit in [8, 9]).

Once a decomposition  $\mathcal{L} = \mathcal{C}^p \mathcal{G} \mathcal{C}^s$  has been fixed, we consider for each  $M \in \mathbb{N}$  the set

$$(2.1) \quad \mathcal{G}^M := \{w \mid \text{the decomposition } w = u^p v u^s \text{ has } |u^p|, |u^s| \leq M\}.$$

Note that  $\mathcal{L} = \bigcup_{M \in \mathbb{N}} \mathcal{G}^M$ , so this defines a filtration of the language.

**2.3. Entropy and pressure for shift spaces.** Given a collection  $\mathcal{D} \subset \mathcal{L}$ , the *entropy* of  $\mathcal{D}$  is

$$h(\mathcal{D}) := \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \#\mathcal{D}_n,$$

where  $\mathcal{D}_n = \{w \in \mathcal{D} \mid |w| = n\}$ . The entropy of an invariant measure  $\mu \in \mathcal{M}_\sigma(X)$  is  $h(\mu) := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{w \in \mathcal{L}_n} -\mu[w] \log \mu[w]$ . For a fixed potential function  $\varphi \in C(X)$ , the *pressure* of  $\mathcal{D} \subset \mathcal{L}$  is

$$P(\mathcal{D}, \varphi) := \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \Lambda_n(\mathcal{D}, \varphi),$$

where

$$\Lambda_n(\mathcal{D}, \varphi) = \sum_{w \in \mathcal{D}_n} e^{\sup_{x \in [w]} S_n \varphi(x)}$$

and  $S_n \varphi(x) = \sum_{k=0}^{n-1} \varphi(\sigma^k x)$ . We write  $P(\varphi) = P(\mathcal{L}, \varphi)$ .

We will be primarily concerned with potentials having some extra regularity: we say that  $\varphi$  has the *Bowen property on  $\mathcal{D}$*  if there is  $V \in \mathbb{R}$  such that for every  $n \in \mathbb{N}$ , every  $w \in \mathcal{D}_n$ , and every  $x, y \in [w]$ , we have  $|S_n \varphi(x) - S_n \varphi(y)| \leq V$ . In particular, if  $\varphi$  is Hölder continuous then it has the Bowen property on every  $\mathcal{D} \subset \mathcal{L}$ .

**2.4. Specification.** We define the specification properties that appear in this paper, and the relationships between them.

**Definition 2.2.** Given a shift space  $X$  and its language  $\mathcal{L}$ , consider a subset  $\mathcal{G} \subset \mathcal{L}$ . Given  $\tau \in \mathbb{N}$ , we say that  $\mathcal{G}$  has *(W)-specification with gap length  $\tau$*  if for every  $v, w \in \mathcal{G}$  there is  $u \in \mathcal{L}$  such that  $vwu \in \mathcal{G}$  and  $|u| \leq \tau$ .

In the case  $\mathcal{G} = \mathcal{L}$ , this is equivalent to the well-known weak specification property for the shift. If the gluing word  $u$  can always be taken to have length exactly  $\tau$ , we say that  $\mathcal{G}$  has (S)-specification. An important special case, which corresponds to specification with gap length 0, is the following.

**Definition 2.3.** We say that  $\mathcal{G} \subset \mathcal{L}$  has the *free concatenation* property if for all  $u, w \in \mathcal{G}$ , we have  $uw \in \mathcal{G}$ .

**Remark 2.4.** The definition of specification on  $\mathcal{G}$ , rather than on all of  $\mathcal{L}$ , was introduced by the first two authors in [8]. The definition above has an important difference from the (W)-specification property in [8]. There, we made the weaker requirement that a finite collection of words  $w^1, \dots, w^n \in \mathcal{G}$  can be glued to form a word in  $\mathcal{L}$ ; that is, there exist words  $u^1, \dots, u^{m-1}$  with length at most  $\tau$  such that  $w^1 u^1 w^2 u^2 \dots u^{m-1} w^m \in \mathcal{L}$ . We require the stronger property of Definition 2.2 for the arguments in §4.2, which allow us to replace  $\mathcal{G}$  with a collection  $\mathcal{F}$  having the free concatenation property; this is necessary for our construction of horseshoes. The stronger property is natural in the symbolic setting, and is satisfied for all our motivating examples.

**2.5. Edit Approachability.** First we introduce the edit metric (sometimes known as the Damerau–Levenshtein metric) on  $\mathcal{L}$ .

**Definition 2.5.** Define an *edit* of a word  $w = w_1 \dots w_n \in \mathcal{L}$  to be a transformation of  $w$  by one of the following actions, where  $u^j \in \mathcal{L}$  are arbitrary words and  $a, a' \in A$  are arbitrary symbols.

- (1) *Substitution:*  $w = u^1 a u^2 \mapsto w' = u^1 a' u^2$ .
- (2) *Insertion:*  $w = u^1 u^2 \mapsto w' = u^1 a' u^2$ .
- (3) *Deletion:*  $w = u^1 a u^2 \mapsto w' = u^1 u^2$ .

Given  $v, w \in \mathcal{L}$ , define the *edit distance* between  $v$  and  $w$  to be the minimum number of edits required to transform the word  $v$  into the word  $w$ : we will denote this by  $\hat{d}(v, w)$ .

The following lemma about the size of balls in the edit metric will be crucial for our entropy estimates.

**Lemma 2.6.** *There is  $C > 0$  such that given  $n \in \mathbb{N}$ ,  $w \in \mathcal{L}_n$ , and  $\delta > 0$ , we have*

$$(2.2) \quad \#\{v \in \mathcal{L} \mid \hat{d}(v, w) \leq \delta n\} \leq C n^C (e^{C\delta} e^{-\delta \log \delta})^n.$$

Now we can introduce our key definition, which requires that any word in  $\mathcal{L}$  can be transformed into a word in  $\mathcal{G}$  with a relatively small number of edits.

**Definition 2.7.** Say that a non-decreasing function  $g: \mathbb{N} \rightarrow \mathbb{N}$  is a *mistake function* if  $\frac{g(n)}{n}$  converges to 0. We say that  $\mathcal{L}$  is *edit approachable* by  $\mathcal{G}$ , where  $\mathcal{G} \subset \mathcal{L}$ , if there is a mistake function  $g$  such that for every  $w \in \mathcal{L}$ , there exists  $v \in \mathcal{G}$  with  $\hat{d}(v, w) \leq g(|w|)$ .

If  $\mathcal{L}$  is edit approachable by  $\mathcal{G}$ , and  $\mathcal{G}$  satisfies the specification property, then it is easy to see that the symbolic space satisfies the following global specification property, which we could call the *almost specification after edits* property: there exists a mistake function  $g$  so that

for any words  $u, v \in \mathcal{L}$ , there are words  $u', v' \in \mathcal{L}$  so that  $u'v' \in \mathcal{L}$ ,  $\hat{d}(u, u') \leq g(|u|)$  and  $\hat{d}(v, v') \leq g(|v|)$ .

Edit approachability allows us to replace sufficiently long words in  $\mathcal{L}$  with words in  $\mathcal{G}$  in such a way that estimates on Birkhoff averages, and thus estimates on empirical measures, can be well controlled, while at the same time, (2.2) guarantees that not much entropy is lost this way.

Control on the Birkhoff averages is given by the following lemma.

**Lemma 2.8.** *For any continuous function  $\varphi: X \rightarrow \mathbb{R}$  and any mistake function  $g(n)$ , there is a sequence of positive numbers  $\delta_n \rightarrow 0$  such that if  $x, y \in X$  and  $m, n \in \mathbb{N}$  are such that  $\hat{d}(x_1 \cdots x_n, y_1 \cdots y_m) \leq g(n)$ , then  $|\frac{1}{n}S_n\varphi(x) - \frac{1}{m}S_m\varphi(y)| \leq \delta_n$ .*

Although we do not use it in this paper, we prove the following consequence of Lemma 2.6 and 2.8 in §7.

**Proposition 2.9.** *If  $\mathcal{L}$  is edit approachable by  $\mathcal{G}$ , then  $P(\mathcal{G}, \varphi) = P(\varphi)$  for every  $\varphi \in C(X)$ .*

**2.6. Hamming Approachability.** For  $v, w \in \mathcal{L}_n$ , let  $d_{\text{Ham}}$  denote the Hamming distance between  $v$  and  $w$ . This is the number of substitutions it takes to transform  $v$  into  $w$ ; insertions and deletions are not allowed, and in particular  $v, w$  must have the same length.

**Definition 2.10.** We say that  $\mathcal{L}$  is *Hamming approachable* by  $\mathcal{G}$ , where  $\mathcal{G} \subset \mathcal{L}$ , if there is a mistake function  $g$  such that for every  $w \in \mathcal{L}$ , there exists  $v \in \mathcal{G}$  with  $d_{\text{Ham}}(v, w) \leq g(|w|)$ .

Clearly, if  $\mathcal{L}$  is Hamming approachable by  $\mathcal{G}$ , then  $\mathcal{L}$  is edit approachable by  $\mathcal{G}$ . If  $\mathcal{L}$  is Hamming approachable by  $\mathcal{G}$ , and  $\mathcal{G}$  satisfies (S)-specification, then it is easy to see that the symbolic space satisfies the *almost specification* property of [8, §3.3]: there exists a mistake function  $g$  so that for any words  $u, v \in \mathcal{L}$ , there are words  $u', v' \in \mathcal{L}$  so that  $u'v' \in \mathcal{L}$ ,  $d_{\text{Ham}}(u, u') \leq g(|u|)$  and  $d_{\text{Ham}}(v, v') \leq g(|v|)$ .

**2.7. Gibbs properties.** The standard Gibbs property for a shift space says that a measure  $m \in \mathcal{M}(X)$  is *Gibbs* if there are constants  $K, K' > 0$  such that

$$(2.3) \quad K \leq \frac{m[x_1 \cdots x_n]}{e^{-nP(\varphi) + S_n\varphi(x)}} \leq K'$$

for all  $x \in X$  and  $n \in \mathbb{N}$ . We will require that the upper bound hold uniformly, while the lower bound will only be required to hold when  $x_1 \cdots x_n \in \mathcal{G}$ . More precisely, we make the following definition for a collection  $\mathcal{G} \subset \mathcal{L}$ .



**Definition 2.11.** A measure  $m \in \mathcal{M}(X)$  is *Gibbs for  $\varphi$  with respect to  $\mathcal{G}$*  if there are constants  $K, K' > 0$  such that

$$(2.4) \quad m[x_1 \cdots x_n] \leq K' e^{-nP(\varphi) + S_n \varphi(x)}$$

for every  $x \in X$  and  $n \in \mathbb{N}$ , and

$$(2.5) \quad m[x_1 \cdots x_n] \geq K e^{-nP(\varphi) + S_n \varphi(x)}$$

whenever  $x \in X$  and  $n \in \mathbb{N}$  are such that  $x_1 \cdots x_n \in \mathcal{G}$ .

Definition 2.11 is property **[A.3]** of Theorem A. Theorem 3.1 provides examples of measures satisfying this definition.

**2.8. Properties under factors.** One advantage of our techniques is that they behave well under factors. We let  $\Sigma$  be a shift space, and we let  $\mathcal{G} \subset \mathcal{L}(\Sigma)$ . Suppose that  $X$  is a subshift factor of  $\Sigma$ , that is, there exists a continuous surjective map  $\pi: \Sigma \rightarrow X$  such that  $\sigma \circ \pi = \pi \circ \sigma$ . By [23, Theorem 6.29],  $\pi$  is a block code: there exist  $r \in \mathbb{N}$  and  $\psi: \mathcal{L}_{2r+1} \rightarrow \mathcal{A}$ , where  $\mathcal{A}$  is the alphabet of  $X$ , such that

$$(2.6) \quad (\pi x)_n = \psi(x_{n-r} x_{n-r+1} \cdots x_{n+r-1} x_{n+r}).$$

This induces a surjective map  $\Psi: \mathcal{L}(\Sigma)_{n+2r} \rightarrow \mathcal{L}(X)_n$  by

$$\Psi(w_1 \cdots w_{n+2r}) = \psi(w_1 \cdots w_{2r+1}) \psi(w_{2r+2} \cdots w_{4r+2}) \cdots \psi(w_n \cdots w_{n+2r}).$$

We set  $\tilde{\mathcal{G}} = \Psi(\mathcal{G})$ . The key to our study of  $X$  is that  $\tilde{\mathcal{G}}$  inherits a number of good properties of  $\mathcal{G}$ , including in particular **[A.1]**, **[A.2]**, and the condition **(I)** that appears in Theorem 3.1 below.

**Lemma 2.12.** *Let  $\mathcal{G} \subset \mathcal{L}(\Sigma)$  and  $\tilde{\mathcal{G}} \subset \mathcal{L}(X)$  be as above.*

- (1) *If  $\mathcal{G}$  satisfies **[A.1]**, then  $\tilde{\mathcal{G}}$  satisfies **[A.1]**.*
- (2) *If  $\mathcal{G}$  satisfies **[A.2]**, then  $\tilde{\mathcal{G}}$  satisfies **[A.2]**.*
- (3) *If  $\mathcal{G}$  satisfies **(I)**, then  $\tilde{\mathcal{G}}$  satisfies **(I)**.*

Furthermore, if  $\mathcal{C}^p \mathcal{G} \mathcal{C}^s$  is a decomposition for  $\mathcal{L}(\Sigma)$ , there is a natural decomposition for  $\mathcal{L}(X)$ . We define  $\tilde{\mathcal{C}}^p$  by taking  $\Psi(\mathcal{C}^p \mathcal{L}_{2k}(\Sigma))$ , and  $\tilde{\mathcal{C}}^s$  by taking  $\Psi(\mathcal{L}_{2k}(\Sigma) \mathcal{C}^s)$ . It is easy to check the following lemma.

**Lemma 2.13.** *If  $\mathcal{C}^p \mathcal{G} \mathcal{C}^s$  is a decomposition for  $\mathcal{L}(\Sigma)$ , then  $\tilde{\mathcal{C}}^p \tilde{\mathcal{G}} \tilde{\mathcal{C}}^s$  is a decomposition for  $\mathcal{L}(X)$ . If  $h(\mathcal{C}^p \cup \mathcal{C}^s) = 0$ , then  $h(\tilde{\mathcal{C}}^p \cup \tilde{\mathcal{C}}^s) = 0$ .*

### 3. CONSEQUENCES OF THEOREM A

**3.1. Unique equilibrium states.** The following result from [8, 9] provides unique equilibrium states which satisfy the weak Gibbs property **[A.3]**, and is our primary tool for finding reference measures to which Theorem A applies. Roughly speaking, Conditions **(I)**–**(II)** state

that  $\mathcal{C}^p$  and  $\mathcal{C}^s$  contain all obstructions to specification (for the system) and regularity (for the potential), while Condition **(III)** states that these obstructions carry smaller pressure than the whole system.

**Theorem 3.1** ([9], Theorem C and Remark 2.2). *Let  $(X, \sigma)$  be a subshift on a finite alphabet and  $\varphi \in C(X)$  a potential. Suppose there exist collections of words  $\mathcal{C}^p, \mathcal{G}, \mathcal{C}^s \subset \mathcal{L}$  such that  $\mathcal{C}^p \mathcal{G} \mathcal{C}^s = \mathcal{L}$  and the following conditions hold:*

- (I)  $\mathcal{G}^M$  has  $(W)$ -specification for every  $M \in \mathbb{N}$ ;
- (II)  $\varphi$  has the Bowen property on  $\mathcal{G}$ ;
- (III)  $P(\mathcal{C}^p \cup \mathcal{C}^s, \varphi) < P(\varphi)$ .

*Then  $\varphi$  has a unique equilibrium state  $m_\varphi$ , and  $m_\varphi$  is Gibbs for  $\varphi$  with respect to  $\mathcal{G}$ . In particular,  $m_\varphi$  satisfies [A.3].*

We also note that Theorem B of [11] is a non-symbolic version of this result whose hypotheses are weaker than those of Theorem 3.1.

**Theorem 3.2.** *Let  $X$  be a subshift on a finite alphabet and  $\varphi \in C(X)$  a potential. Suppose  $\mathcal{L}$  has a decomposition  $\mathcal{L} = \mathcal{C}^p \mathcal{G} \mathcal{C}^s$  satisfying [A.2] and **(I)**–**(III)**. Then writing  $m_\varphi$  for the unique equilibrium state of  $\varphi$ , the system  $(X, \sigma)$  satisfies a level-2 large deviation principle with reference measure  $m_\varphi$  and rate function  $q^\varphi$  given by (1.1).*

*Proof.* Condition **(I)** implies [A.1]. By Theorem 3.1 there is a unique equilibrium state  $m_\varphi$ , and moreover  $m_\varphi$  satisfies [A.3]. Thus, Theorem A gives the result.  $\square$

For a shift space  $X$  and a collection of words  $\mathcal{C} \subset \mathcal{L}$ , it is typically much easier to verify  $h(\mathcal{C}) < h(X)$  than  $P(\mathcal{C}, \varphi) < P(X, \varphi)$ . For  $\beta$ -shifts, it was shown in [9, Proposition 3.1] that **(III)** holds for every Bowen potential  $\varphi$ , and we show in §5 that this is also true for  $S$ -gap shifts. However, for other shift spaces where no analogous argument is available yet, the following lemma is a convenient way to ensure that **(III)** holds for a large class of functions.

**Lemma 3.3.** *Suppose  $X$  is a shift space and  $\mathcal{C} \subset \mathcal{L}(X)$  is a collection of words such that  $h(\mathcal{C}) < h(X)$ , and let  $\varphi: X \rightarrow \mathbb{R}$ . If  $\varphi$  satisfies the bounded range condition*

$$(BR) \quad \sup \varphi - \inf \varphi < h(X) - h(\mathcal{C}),$$

*then  $P(\mathcal{C}, \varphi) < P(X, \varphi)$ .*

Often we can take  $h(\mathcal{C}) = 0$ , in which case the condition (BR) on  $\varphi$  reduces to the condition

$$(BR_0) \quad \sup \varphi - \inf \varphi < h(X).$$

**3.2. Factors.** Our results are well behaved under the operation of passing to a subshift factor.

**Theorem 3.4.** *Let  $\Sigma$  be a subshift on a finite alphabet, and suppose that  $\mathcal{L} = \mathcal{L}(\Sigma)$  has a decomposition  $\mathcal{L} = \mathcal{C}^p \mathcal{G} \mathcal{C}^s$  satisfying [A.2] and (I). Assume further that  $h(\mathcal{C}) = 0$ , where  $\mathcal{C} = \mathcal{C}^p \cup \mathcal{C}^s$ . Let  $X$  be a subshift factor of  $\Sigma$  and  $\varphi: X \rightarrow \mathbb{R}$  be a continuous function satisfying  $(BR_0)$  and the Bowen property. Then  $\varphi$  has a unique equilibrium state  $m_\varphi$ , and the system  $(X, \sigma)$  satisfies a level-2 large deviation principle with reference measure  $m_\varphi$  and rate function  $q^\varphi$  given by (1.1).*

*Proof.* By Lemmas 2.12 and 2.13, the language  $\tilde{\mathcal{L}}$  of  $X$  has a decomposition  $\tilde{\mathcal{C}}^p \tilde{\mathcal{G}} \tilde{\mathcal{C}}^s$  such that  $\tilde{\mathcal{G}}$  satisfies (I) and [A.2], and  $h(\tilde{\mathcal{C}}^p \cup \tilde{\mathcal{C}}^s) = 0$ . The Bowen property for  $\varphi$  implies (II), and Lemma 3.3 gives (III). Thus Theorem 3.2 gives the result.  $\square$

**3.3.  $\beta$ -shifts and  $S$ -gap shifts.** Our main examples are the  $\beta$ -shifts, the  $S$ -gap shifts, and their subshift factors. For all of these examples we can take  $h(\mathcal{C}) = 0$ , and for the  $\beta$ -shifts (in [9, Proposition 3.1]) and  $S$ -gap shifts (in §5.1), we can show that every Bowen potential satisfies

$$(3.1) \quad P(\varphi) > \sup_{\mu \in \mathcal{M}_\sigma(X)} \int \varphi d\mu,$$

which in turn implies  $P(\mathcal{C}, \varphi) < P(\varphi)$  and removes the need for the bounded range condition. For factors of  $\beta$ -shifts and  $S$ -gap shifts, we do require the additional assumption  $(BR_0)$  on the potential  $\varphi$  at present.

**3.3.1.  $S$ -gap shifts.** An  $S$ -gap shift  $\Sigma_S$  is a subshift of  $\{0, 1\}^{\mathbb{Z}}$  defined by the rule that for a fixed  $S \subset \{0, 1, 2, \dots\}$ , the number of 0's between consecutive 1's is an integer in  $S$ . That is, the language of  $\Sigma_S$  is

$$\{0^{n_1} 1 0^{n_2} 1 \dots 1 0^{n_k} 10^m \mid n_i \in S \text{ for all } 1 \leq i \leq k \text{ and } n, m \in \mathbb{N}\},$$

together with  $\{0^n \mid n \in \mathbb{N}\}$ , where we assume that  $S$  is infinite (when  $S$  is finite,  $\Sigma_S$  is sofic and can be analysed without the techniques of this paper). The language for  $\Sigma_S$  admits the following decomposition:

$$\begin{aligned} \mathcal{G} &= \{0^{n_1} 1 0^{n_2} 1 0^{n_3} 1 \dots 1 0^{n_k} 1 \mid n_i \in S \text{ for all } 1 \leq i \leq k\}, \\ \mathcal{C}^p &= \{0^n 1 \mid n \notin S\}, \\ \mathcal{C}^s &= \{0^n \mid n \in \mathbb{N}\}, \end{aligned}$$

which was first studied in [8]. We verify in §5.1 that this decomposition satisfies Conditions [A.2] and (I)–(III) for every Bowen potential  $\varphi$ .

3.3.2.  *$\beta$ -shifts.* Fix  $\beta > 1$ , write  $b = \lceil \beta \rceil$ , and let  $\omega^\beta \in \{0, 1, \dots, b-1\}^{\mathbb{N}}$  be the greedy  $\beta$ -expansion of 1. Then  $\omega^\beta$  satisfies  $\sum_{j=1}^{\infty} \omega_j^\beta \beta^{-j} = 1$ , and has the property that  $\sigma^j(\omega^\beta) \preceq \omega^\beta$  for all  $j \geq 1$ , where  $\preceq$  denotes the lexicographic ordering. The  $\beta$ -shift is defined by

$$\Sigma_\beta = \{x \in \{0, 1, \dots, b-1\}^{\mathbb{N}} \mid \sigma^j(x) \preceq \omega^\beta \text{ for all } j \geq 1\}.$$

The first and second author showed in [8, 9] that the language for  $\Sigma_\beta$  admits a decomposition  $\mathcal{L}(\Sigma_\beta) = \mathcal{GC}^s$  that satisfies **(I)**–**(III)** for every Bowen potential  $\varphi$ . In §5.2 we briefly review the construction and show that condition **[A.2]** is also satisfied.

3.3.3. *Results for examples.* We collect our results as applied to these examples in the following theorem. We say that a subshift is non-trivial if it is not a single periodic orbit. We proved in [8, Proposition 2.4] that a non-trivial subshift factor of a  $\beta$ -shift or  $S$ -gap shift has positive entropy.

**Theorem 3.5.** *Let  $X$  and  $\varphi$  be one of the following:*

- (1)  *$X$  is a  $\beta$ -shift or an  $S$ -gap shift, and  $\varphi$  has the Bowen property;*
- (2)  *$X$  is a non-trivial subshift factor of a  $\beta$ -shift or an  $S$ -gap shift, and  $\varphi$  satisfies  $(BR_0)$  and the Bowen property.*

*Then  $\varphi$  has a unique equilibrium state  $m_\varphi$ , and  $(X, \sigma)$  satisfies a level 2 large deviation principle with reference measure  $m_\varphi$  and rate function  $q^\varphi: \mathcal{M}(X) \rightarrow [-\infty, 0]$  given by (1.1).*

*Proof.* The case when  $X$  is an  $S$ -gap shift is proved in §5.1. The case when  $X$  is a  $\beta$ -shift is proved in §5.2. The result for factors of  $\beta$ -shifts and  $S$ -gap shifts follows from Theorem 3.4.  $\square$

To the best of our knowledge, the above statement was only previously known in the case when  $X$  is a  $\beta$ -shift and  $m_0$  is the measure of maximal entropy [29] (apart from the exceptional set of special cases above where  $X$  has specification, see [4, 13]).

**3.4. Horseshoe theorem: precise statement.** We now state a more precise version of our ‘horseshoe result’, which is a key step in the proof of Theorem A, and may be of independent interest.

**Proposition 3.6.** *Let  $X$  be a shift space and suppose that  $\mathcal{G} \subset \mathcal{L}$  satisfies **[A.1]** and **[A.2]**. Then there exists an increasing sequence  $\{X_n\}$  of compact  $\sigma$ -invariant subsets of  $X$  with the following properties.*

- (1) *Each  $X_n$  is a topologically transitive sofic shift.*
- (2) *There is  $T \in \mathbb{N}$  such that for every  $n$  and every  $w \in \mathcal{L}(X_n)$ , there are  $u, v \in \mathcal{L}$  with  $|u|, |v| \leq n + T$  such that  $uwv \in \mathcal{G}$ .*

- (3) *Every invariant measure on  $X$  is entropy approachable by ergodic measures on  $X_n$ : for any  $\eta > 0$ , any  $\mu \in \mathcal{M}_\sigma(X)$ , and any neighborhood  $U$  of  $\mu$  in  $\mathcal{M}_\sigma(X)$ , there exist  $n \geq 1$  and  $\mu' \in \mathcal{M}_\sigma^e(X_n) \cap U$  such that  $h(\mu') > h(\mu) - \eta$  holds.*

By the variational principle and the entropy approachability in Proposition 3.6, we have the further result that  $\lim_{n \rightarrow \infty} h(X_n) = h(X)$ , and more generally

$$P(X, \varphi) = \lim_{n \rightarrow \infty} P(X_n, \varphi) = \sup_{n \in \mathbb{N}} P(X_n, \varphi)$$

for every  $\varphi \in C(X)$ . Thus, we can interpret the sets  $X_n$  as well behaved ‘horseshoes’ which can be used to approximate the original space  $X$ , revealing a structure reminiscent of Katok horseshoes [19].

In the proof of the main results, we will use the following consequence of the second property in Proposition 3.6: If a measure  $m \in \mathcal{M}(X)$  is Gibbs with respect to  $\mathcal{G}$ , then  $m$  has the following Gibbs property on the family of subshifts  $\{X_n\}$ : there exist constants  $K_n, K' > 0$  such that for every  $x \in X_n$  and  $k \in \mathbb{N}$ , we have

$$(3.2) \quad K_n \leq \frac{m[x_1 \cdots x_k]}{e^{-kP(\varphi) + S_k \varphi(x)}} \leq K'.$$

This follows from the fact that  $x_1 \cdots x_k$  can be extended to a word in  $\mathcal{G}$  by adding a word to each end whose length is bounded by a constant depending only on  $n$ .

#### 4. PROOF OF THEOREMS A AND B

The large deviations property in Definition 2.1 comprises an upper bound and a lower bound. We establish the upper bound first, by verifying criteria given by Pfister and Sullivan in [29].

**4.1. Upper bound.** Given  $\mu \in \mathcal{M}_\sigma(X)$ , let  $q^\varphi(\mu) = h(\mu) + \int \varphi d\mu - P(\varphi)$ , as in (1.1). We show that for any closed set  $F \subset \mathcal{M}(X)$ , we have

$$(4.1) \quad \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log m(\mathcal{E}_n^{-1}(F)) \leq \sup_{\mu \in F \cap \mathcal{M}_\sigma(X)} q^\varphi(\mu).$$

Our key tool is the following result of Pfister and Sullivan.

**Theorem 4.1.** [29, Theorem 3.2 and Proposition 4.2] *Let  $(X, \sigma)$  be a subshift,  $m \in \mathcal{M}(X)$ ,  $\psi \in C(X)$ , and assume that the equation*

$$(4.2) \quad \overline{\lim}_{n \rightarrow \infty} \sup_{w \in \mathcal{L}_n} \left( \frac{1}{n} \log m([w]) + \frac{1}{n} \sup_{x \in [w]} S_n \psi(x) \right) \leq 0$$

holds. Then

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log m(\mathcal{E}_n^{-1}(F)) \leq \sup_{\mu \in F \cap \mathcal{M}_\sigma(X)} \left( h(\mu) - \int \psi d\mu \right).$$

We will apply Theorem 4.1 with  $\psi = P(\varphi) - \varphi$ . The upper Gibbs bound in [A.3] (see (2.4)) yields a constant  $K'$  such that

$$m([w]) \leq K' e^{-nP(\varphi) + S_n \varphi(x)}$$

for every  $x \in [w]$ . Thus

$$\frac{1}{n} \log m([w]) + \frac{1}{n} \sup_{x \in [w]} S_n(P(\varphi) - \varphi)(x) \leq \frac{1}{n} \log(K') \rightarrow 0,$$

for every  $x \in [w]$ . This establishes (4.2), so Theorem 4.1 provides the desired upper bound.

**4.2. Free concatenation and the gluing map.** The following result is [7, Proposition 3.7].

**Proposition 4.2.** *If  $\mathcal{G}$  has  $(W)$ -specification (Definition 2.2), then there are  $r, s \in \mathcal{G}$  and  $c \in \mathcal{L}$  such that writing*

$$\mathcal{B} = \mathcal{L}r \cap s\mathcal{L} \cap \mathcal{G} = \{w \in \mathcal{G} \mid w_1 \cdots w_{|s|} = s \text{ and } w_{|w|-|r|+1} \cdots w_{|w|} = r\}$$

and  $\mathcal{F} := c\mathcal{B} = \{cw \mid w \in \mathcal{B}\}$ , the collection  $\mathcal{F} \subset \mathcal{L}$  has the free concatenation property. Moreover, a measure  $m$  has the Gibbs property for  $\varphi$  w.r.t.  $\mathcal{G}$  if and only if it has the Gibbs property for  $\varphi$  w.r.t.  $\mathcal{F}$ .

*Sketch of proof.* We outline the main ideas; details are in [7, §6.2]. The basic argument is inspired by Bertrand's proof that shifts with specification have a synchronizing word [3]. Let  $\tau$  be the gap length for the specification property for  $\mathcal{G}$ . Given  $r, s \in \mathcal{G}$ , let  $C(r, s) = \{c \in \mathcal{L} \mid |c| \leq \tau \text{ and } rcs \in \mathcal{G}\}$  be the set of "short words which connect  $r$  to  $s$ "; this set is finite and non-empty. If  $u, v \in \mathcal{L}$  are such that  $r' = ur \in \mathcal{G}$  and  $s' = sv \in \mathcal{G}$ , then  $C(r', s') \subset C(r, s)$ . If this inclusion is strict for some choice of  $u, v$  then replace  $r, s$  with  $r', s'$ . Iterate this process; since each  $C(r, s)$  is finite and non-empty it must terminate, and we obtain  $r, s \in \mathcal{G}$  such that  $C(r', s') = C(r, s)$  for every  $r' = ur \in \mathcal{G}$  and  $s' = sv \in \mathcal{G}$ . Pick any  $c \in C(r, s)$ , then it is not hard to see from the characterisation of  $r, s$  that  $\mathcal{F}$  has the free concatenation property. Equivalence of the Gibbs properties follows since words in  $\mathcal{F}$  can be extended to  $\mathcal{G}$  with a bounded number of symbols, and vice versa.  $\square$

**Lemma 4.3.**  $\mathcal{L}$  is edit approachable by  $\mathcal{F}$ .

*Proof.* Given  $w \in \mathcal{G}$ , by (W)-specification there are  $u, v \in \mathcal{L}$  with  $|u|, |v| \leq \tau$  such that  $suwvr \in \mathcal{G}$ , hence  $csuwvr \in \mathcal{F}$ . Thus every word in  $\mathcal{G}$  can be turned into a word in  $\mathcal{F}$  with at most  $|c| + |s| + |r| + 2\tau$  edits. Since  $\mathcal{L}$  is edit approachable by  $\mathcal{G}$ , this suffices.  $\square$

Write  $\mathcal{F}^*$  for the set of all finite sequences  $(w^1, \dots, w^m)$  where each  $w^i \in \mathcal{F}$ . Let  $\Phi: \mathcal{F}^* \rightarrow \mathcal{F}$  be the concatenation map  $(w^1, \dots, w^m) \mapsto w^1 \cdots w^m$ . This extends to a map  $\mathcal{F}^{\mathbb{N}} \rightarrow \mathcal{F}$  in the natural way. We note that for each  $n_1, \dots, n_k \in \mathbb{N}$ , the restriction of  $\Phi$  to  $\prod_{i=1}^k \mathcal{F}_{n_i}$  is clearly injective.

**4.3. Proof of Proposition 3.6.** We prove Proposition 3.6, and thus Theorem B. This is crucial for our large deviation lower bounds.

**Step 0: Definition and basic properties of  $X_n$ .** First, we define the sequence of shift spaces  $X_n$  which will meet our requirements. Let  $\mathcal{F}_{\leq n} = \bigcup_{i=0}^n \mathcal{F}_i$ , and consider the set of words

$$(4.3) \quad \Phi(\mathcal{F}_{\leq n}^*) := \bigcup_{m=1}^{\infty} \{w^1 \cdots w^m \mid w^i \in \mathcal{F}_{\leq n} \text{ for all } 1 \leq i \leq m\}.$$

We can turn this set into the language of a shift space by including all subwords to obtain

$$(4.4) \quad \mathcal{L}(X_n) := \{\text{all subwords of elements of } \Phi(\mathcal{F}_{\leq n}^*)\}.$$

Then,  $X_n$  is defined as the shift whose language is  $\mathcal{L}(X_n)$ . That  $X_n$  is well defined is verified trivially using [23, Proposition 1.3.4].

**Lemma 4.4.** *The shift space  $X_n$  has the following properties.*

- (1) *For every  $w \in \mathcal{L}(X_n)$ , there are  $u, v \in \mathcal{L}_{\leq n}$  such that  $uwv \in \mathcal{F}$ .*
- (2)  *$X_n$  is a sofic shift and has (W)-specification with gap size  $2n$ .*

*Proof.* To check the first property claimed for  $X_n$ , we observe that if  $w \in \mathcal{L}(X_n)$  is a subword of  $w^1 \cdots w^m$ , then by appending at most  $n$  symbols to either end of  $w$ , we can obtain a word of the form  $w^i \cdots w^j \in \mathcal{F}$ . The (W)-specification property for  $X_n$  follows immediately since words in  $\mathcal{F}$  can be freely concatenated. To see that  $X_n$  is sofic, we note that it can be presented by a loop graph, with each (of the finitely many) loops corresponding to a word in  $\mathcal{F}_{\leq n}^*$ .  $\square$

To get the property of  $X_n$  claimed in Proposition 3.6(2), we observe that fixing  $u \in \mathcal{G}$  and putting  $T = |u|$ , any word in  $\mathcal{F}$  can be extended to a word in  $\mathcal{G}$  by adding  $u$  to its beginning.

**Remark 4.5.** If every word in  $\mathcal{L}$  can be extended to a word in  $\mathcal{G}$ , then it is easy to show that  $X = \overline{\bigcup X_n}$ .

The rest of the proof of Proposition 3.6 is an extension of the approach used by Pfister and Sullivan in [30]:

- (1) Construct a subshift  $Y \subset X_n$  for some  $n \geq 1$  such that every  $\nu \in \mathcal{M}_\sigma(Y)$  is weak\*-close to  $\mu$ .
- (2) Use edit approachability of  $\mathcal{L}$  by  $\mathcal{F}$  to explicitly build a subshift  $H \subset Y$  with a rich structure.
- (3) Show that  $H$  (and hence  $Y$ ) has entropy close to  $h(\mu)$  by using this structure.
- (4) Obtain the measures  $\mu'$  as maximal entropy measures for  $Y$ .

In preparation for the above steps, fix  $\eta > 0$  and use the ergodic decomposition of  $\mu$  together with affinity of the entropy map to find  $\lambda = \sum_{i=1}^p a_i \mu_i$  such that

- the  $\mu_i$  are ergodic;
- the  $a_i$  are rational numbers in  $[0, 1]$  such that  $\sum_{i=1}^p a_i = 1$ ;
- $D(\mu, \lambda) \leq \eta$ ;
- $h(\lambda) > h(\mu) - \eta$ .

Let  $h_i = 0$  when  $h(\mu_i) = 0$ , and  $\max(0, h(\mu_i) - \eta) < h_i < h(\mu_i)$  otherwise.

**Definition 4.6.** Given  $\nu \in \mathcal{M}(X)$  and  $\zeta > 0$ , let

$$\mathcal{L}^{\nu, \zeta} := \{w \in \mathcal{L} \mid D(\mathcal{E}_{|w|}(x), \nu) < \zeta \text{ for all } x \in [w]\}.$$

Combining [29, Propositions 2.1 and 4.1], we have the following.

**Lemma 4.7.** [29, Propositions 2.1 and 4.1] *There exists  $N \in \mathbb{N}$  such that for  $n \geq N$  and  $1 \leq i \leq p$ , we have  $\#\mathcal{L}_n^{\mu_i, \eta} \geq e^{nh_i}$ .*

Because  $\mathcal{L}$  is edit approachable by  $\mathcal{F}$ , there is a mistake function  $g$  such that every  $w \in \mathcal{L}$  has  $v \in \mathcal{F}$  with  $\hat{d}(v, w) \leq g(|w|)$ . By Lemma 2.8, we can choose  $N$  large enough so that, in addition to the cardinality estimates in Lemma 4.7, we have the following property.

- If  $n \geq N$  and  $x, y \in X$  are such that  $\hat{d}(x_1 \cdots x_n, y_1 \cdots y_m) \leq g(n)$ , then  $D(\mathcal{E}_n(x), \mathcal{E}_m(y)) \leq \eta$ .

Without loss of generality, assume that  $0 < a_i < 1$  for each  $i$ . Choose  $n$  such that we have  $n_i := a_i n \in \mathbb{N}$ ,  $n_i + g(n_i) \leq n$ , and  $n_i \geq N$  for every  $i$ , and moreover

$$(4.5) \quad \frac{n}{n + \sum_{i=1}^p g(n_i)} (h(\lambda) - \eta) \geq h(\lambda) - 2\eta.$$

To prove the proposition, we will follow the steps listed above to show that there exists  $\mu' \in \mathcal{M}_\sigma^e(X_n)$  such that  $D(\mu, \mu') \leq 6\eta$  and  $h(\mu') > h(\mu) - 4\eta$ .



**Step 1: Definition of  $Y \subset X_n$ .** Fix  $K \in \mathbb{N}$  such that  $4/K \leq \eta$ . Now let

$$(4.6) \quad Y := \{x \in X_n \mid x_t x_{t+1} \cdots x_{t+Kn-1} \in \mathcal{L}_{Kn}^{\mu, 5\eta} \text{ for all } t \geq 0\}.$$

Then  $Y \subset X_n$  is compact and  $\sigma$ -invariant. Moreover, the following holds.

**Lemma 4.8.** *We have  $D(\mu, \nu) \leq 6\eta$  for any  $\nu \in \mathcal{M}_\sigma^e(Y)$ .*

*Proof.* Since  $\nu$  is ergodic, there exists a generic point  $x \in Y$ , that is,  $\mathcal{E}_m(x)$  converges to  $\nu$ . We choose  $L$  so that  $nK/L \leq \eta$  holds, take an arbitrary integer  $m \geq L$  and choose integers  $s$  and  $0 \leq q < Kn$  so that  $m = sKn + q$  holds. Then, using (4.6) and the inequalities  $\frac{q}{m} \leq \frac{Kn}{L} \leq \eta$ , we have

$$\begin{aligned} D(\mathcal{E}_m(x), \mu) &\leq \sum_{i=0}^{s-1} \frac{Kn}{m} D(\mathcal{E}_{Kn}(\sigma^{iKn}x), \mu) + \frac{q}{m} D(\mathcal{E}_q(\sigma^{sKn}x), \mu) \\ &\leq 5\eta + \eta = 6\eta. \end{aligned}$$

Thus taking  $m \rightarrow \infty$ , we have the lemma.  $\square$

**Step 2: Construction of  $H$ .** For brevity of notation we write  $\mathcal{D}^i = \mathcal{L}_{n_i}^{\mu_i, \eta}$ . Extend the definitions of  $n_i, \mathcal{D}^i, \mu_i, a_i$  to indices  $i > p$  by repeating periodically: that is, if  $i = pq + r$ ,  $1 \leq r \leq p$ , then  $n_i = n_r$ ,  $\mathcal{D}^i = \mathcal{D}^r$ ,  $\mu_i = \mu_r$  and  $a_i = a_r$ .

By the assumption that  $\mathcal{L}$  is edit approachable by  $\mathcal{F}$ , we can define a map  $\phi_{\mathcal{F}}: \mathcal{L} \rightarrow \mathcal{F}$  such that  $\hat{d}(w, \phi_{\mathcal{F}}(w)) \leq g(|w|)$ . We extend the map  $\Phi: \mathcal{F}^* \rightarrow \mathcal{F}$  to a map  $\Phi: \mathcal{L}^* \rightarrow \mathcal{F}$  by ‘editing then gluing’. That is, given  $(w^1, \dots, w^n) \in \mathcal{L}^*$ , we put  $\Phi(w^1, \dots, w^n) = \phi_{\mathcal{F}}(w^1) \cdots \phi_{\mathcal{F}}(w^n)$ . The map  $\Phi$  extends to subsets of  $\mathcal{L}^{\mathbb{N}}$  in the natural way, and we consider it here with the following domain:

$$(4.7) \quad \Phi: \prod_{j=1}^{\infty} \mathcal{D}^j \rightarrow X.$$

In other words, given  $\mathbf{w} = \{w^j\} \in \prod_{j=1}^{\infty} \mathcal{D}^j$ , let  $v^j = \phi_{\mathcal{F}}(w^j) \in \mathcal{F}$  and  $\Phi(\mathbf{w}) = v^1 v^2 \cdots$ . Let  $H = \Phi(\prod_{j=1}^{\infty} \mathcal{D}^j)$ . Then we have  $H \subset X_n$  since  $n_i + g(n_i) \leq n$ .

A sort of periodicity is built into the definition of the sequences  $\Phi(\mathbf{w})$ : the word  $v^i$  is an approximation of a suitable generic point for the measure  $\mu_i$ , and the measures  $\mu_i$  repeat periodically ( $\mu_{i+p} = \mu_i$ ). The following lemma states that following  $\Phi(\mathbf{w})$  for a single ‘cycle’ of this periodic behaviour gives a good approximation to  $\mu$ . We write

$\ell_j = \ell_j(\mathbf{w}) = |v^j|$  for the length of the words associated to the index  $j$ , and observe that  $|\ell_j - n_j| \leq g(n_j) \leq g(n)$ .

**Lemma 4.9.** *Fix  $\mathbf{w} \in \prod_{j=1}^{\infty} \mathcal{D}^j$ . For  $q \geq 0$ , let  $c_q = c_q(\mathbf{w}) = \sum_{r=1}^p \ell_{qp+r}$  be the length of the  $q$ th “cycle” in  $\Phi(\mathbf{w})$  and let  $b_m = b_m(\mathbf{w}) = \sum_{q=0}^{m-1} c_q$ . Then we have  $D(\mathcal{E}_{c_m}(\sigma^{b_m}\Phi(\mathbf{w})), \mu) \leq 3\eta$ .*

*Proof.* Choose  $x^j \in [w^j]$  for each  $j \in \mathbb{N}$ , so that by the definition of  $\mathcal{D}^j$ , we have  $D(\mathcal{E}_{n_j}(x^j), \mu_j) \leq \eta$ . Let  $y = \sigma^{b_m}\Phi(\mathbf{w})$  and let  $d_j = \sum_{i=0}^{j-1} \ell_{mp+i}$  for  $1 \leq j \leq p$ . By the definition of  $\Phi$  and the property following Lemma 4.7, we have

$$D(\mathcal{E}_{\ell_j}(\sigma^{d_j}y), \mu_j) \leq D(\mathcal{E}_{\ell_j}(\sigma^{d_j}y), \mathcal{E}_{n_j}(x^{mp+j})) + D(\mathcal{E}_{n_j}(x^{mp+j}), \mu_j) \leq 2\eta.$$

Observe that  $c_q \approx n$ : more precisely, we have

$$(4.8) \quad |c_q - n| \leq \sum_{r=1}^p |\ell_{qp+r} - n_r| \leq pg(n).$$

Taking convex combinations gives

$$\begin{aligned} D(\mu, \mathcal{E}_{c_m}(y)) &\leq D\left(\sum_{j=1}^p a_j \mu_j, \sum_{j=1}^p \frac{\ell_j}{c_m} \mathcal{E}_{\ell_j}(\sigma^{d_j}y)\right) + \eta \\ &\leq \left(\sum_{j=1}^p \left|a_j - \frac{\ell_j}{c_m}\right|\right) + 2\eta \leq 3\eta, \end{aligned}$$

provided  $N$  is chosen large enough such that  $n \geq n_j \geq N$ , and such that (4.8) guarantees we have  $\sum_{j=1}^p |a_j - \frac{\ell_j}{c_m}| \leq \eta$ .  $\square$

We are now in a position to show that  $H \subset Y$ . Given  $y = \Phi(\mathbf{w}) \in H$  and  $t \in \mathbb{N}$ , we can choose  $m_1, m_2$  such that

$$b_{m_1-1} \leq t < b_{m_1} < b_{m_2} \leq t + Kn < b_{m_2+1},$$

and so

$$\mathcal{E}_{Kn}(\sigma^t y) = \left(\sum_{q=m_1}^{m_2-1} \frac{c_q}{Kn} \mathcal{E}_{c_q}(\sigma^{b_q} y)\right) + \xi_1 \mathcal{E}_{b_{m_1}-t}(\sigma^t y) + \xi_2 \mathcal{E}_{t+Kn-b_{m_2}}(\sigma^{b_{m_2}} y),$$

where  $0 \leq \xi_1, \xi_2 \leq \frac{n+pg(n)}{Kn} \leq \eta$ . Each of the empirical measures in the large sum is within  $3\eta$  of  $\mu$ , by Lemma 4.9, and thus we have

$$D(\mathcal{E}_{Kn}(\sigma^t y), \mu) \leq 5\eta.$$

In particular, this shows that  $y \in Y$ .

**Step 3. Estimation of entropy of  $H$ .** Now we use the definition of  $H$  to estimate its topological entropy. Our key tool will be the estimate obtained in Lemmas 2.6.

**Lemma 4.10.** *The topological entropy of  $H$  is at least  $h(\mu) - 4\eta$ .*

*Proof.* Fix  $m \in \mathbb{N}$  and set  $b' = n + \sum_{j=1}^p g(n_j)$ . Note that

$$(4.9) \quad mb' \geq \sup_{\mathbf{w} \in \prod_{j=1}^{\infty} \mathcal{D}^j} b_m(\mathbf{w}) \text{ and } \frac{n}{b'}(h(\lambda) - \eta) \geq h(\lambda) - 2\eta$$

holds, where  $b_m(\mathbf{w})$  is as in Lemma 4.9. Moreover, since  $n = \sum_{j=1}^p n_j$  and each  $n_j \geq N$ , we have  $b' \geq pN$ .

Let  $\mathbf{z} \in \prod_{j=1}^{\infty} \mathcal{D}^j$  be arbitrary, and given  $\mathbf{w} \in \prod_{j=1}^{mp} \mathcal{D}^j$ , let  $\mathbf{wz}$  denote the concatenation of  $\mathbf{w}$  and  $\mathbf{z}$ , so that  $(\mathbf{wz})^j = w^j$  if  $1 \leq j \leq mp$  and  $z^{j-mp}$  otherwise.

Let  $\phi_m: \prod_{j=1}^{mp} \mathcal{D}^j \rightarrow \mathcal{L}_{mb'}$  be the map that takes  $w^1, \dots, w^{mp}$  to the first  $mb'$  symbols of  $\Phi(\mathbf{wz})$ , where  $\Phi$  is the ‘edit and glue’ map from Step 2. Note that  $\phi_m(\prod_{j=1}^{mp} \mathcal{D}^j) \subset \mathcal{L}_{mb'}(H)$ .

Now in order to estimate the entropy of  $H$ , we will use our estimates on the cardinality of  $\mathcal{D}^j$  together with a bound on  $\#\phi_m^{-1}(v)$  for  $v \in \mathcal{L}_{mb'}$ . Recall that  $\phi_{\mathcal{F}}: \mathcal{L} \rightarrow \mathcal{F}$  is a map which satisfies  $\hat{d}(w, \phi_{\mathcal{F}}(w)) \leq g(|w|)$ . First we use Lemma 2.6, recalling that  $\frac{g(n)}{n} \rightarrow 0$ , to fix  $N_0$  sufficiently large so that

$$(4.10) \quad \#\{w \in \mathcal{L} \mid \phi_{\mathcal{F}}(w) = v\} \leq e^{\eta|v|/2}$$

for every  $v \in \mathcal{F}$  with  $|v| \geq N_0$ .

Now as  $w$  ranges over  $\mathcal{D}^j$ , the word  $\phi_{\mathcal{F}}(w)$  may vary in length; however, since its  $\hat{d}$ -distance from  $w$  is at most  $g(|w|)$ , the number of different lengths it can take is at most  $2g(|w|) + 1$ . As above, given  $\mathbf{w} \in \prod_{j=1}^{mp} \mathcal{D}^j$  we write  $\ell_j = \ell_j(\mathbf{w}) = |\phi_{\mathcal{F}}(w^j)|$ , so  $\ell_j \in [n_j - g(n_j), n_j + g(n_j)]$ .

We see that as  $\mathbf{w}$  ranges over  $\prod_{j=1}^{mp} \mathcal{D}^j$ , the number of different values taken by  $(\ell_1, \dots, \ell_{mp})$  is bounded above by

$$(4.11) \quad (2g(n) + 1)^{mp} = e^{mp \log(2g(n)+1)} \leq e^{mb' \frac{\log(2g(n)+1)}{\min_j n_j}} \leq e^{\eta mb'/2},$$

where the last inequality follows from observing that  $n_j \approx a_j n$  and choosing  $N$  sufficiently large (since each  $n_j \geq N$ ).

Given  $u \in \mathcal{L}_{mb'}(H)$  and a fixed choice of  $(\ell_1, \dots, \ell_{mp})$ , it follows from (4.10) that the number of  $\mathbf{w} \in \prod_{j=1}^{mp} \mathcal{D}^j$  with  $\phi_m(\mathbf{w}) = v$  and  $\ell_j(\mathbf{w}) = \ell_j$  for each  $1 \leq j \leq mp$  is at most

$$\prod_{j=1}^{mp} e^{\eta \ell_j/2} = e^{\frac{\eta}{2} \sum_{j=1}^{mp} \ell_j} \leq e^{\frac{\eta}{2} mb'}.$$

Combining this with (4.11), we see that  $\#\phi_m^{-1}(v) \leq e^{\eta mb'}$ , and thus we obtain the estimate

$$\#\mathcal{L}_{mb'}(H) \geq e^{-\eta mb'} \prod_{j=1}^{mp} (\#\mathcal{D}^j).$$

Using Lemma 4.7, it follows that

$$\begin{aligned} h(H) &\geq \left( \lim_{m \rightarrow \infty} \frac{1}{mb'} \sum_{j=1}^{mp} \log \#\mathcal{D}^j \right) - \eta \\ &\geq \left( \lim_{m \rightarrow \infty} \frac{1}{mb'} \sum_{j=1}^{mp} n_j h_j \right) - \eta \geq h(\lambda) - 3\eta \geq h(\mu) - 4\eta. \quad \square \end{aligned}$$

**Step 4: End of the proof of Proposition 3.6.** Let  $\mu'$  be an ergodic measure of maximal entropy for  $Y$ . Lemma 4.8 shows that  $D(\mu', \mu) \leq 6\eta$ , and Lemma 4.10 shows that  $h(\mu') \geq h(\mu) - 4\eta$ . Since  $Y \subset X_n$  by definition, this completes the proof of Proposition 3.6.

**4.4. Lower bounds.** Now we complete the proof of Theorem A by showing that the lower bound

$$(4.12) \quad \varliminf_{n \rightarrow \infty} \frac{1}{n} \log m(\{x \in X : \mathcal{E}_n(x) \in U\}) \geq \sup_{\mu \in U} q^\varphi(\mu)$$

holds for any open set  $U \subset \mathcal{M}(X)$ , where  $q^\varphi(\mu)$  is as in (1.1).

To show (4.12), it is sufficient to show that for any  $\mu \in \mathcal{M}(X)$  and any open neighborhood  $U \subset \mathcal{M}(X)$  of  $\mu$ ,

$$(4.13) \quad \varliminf_{n \rightarrow \infty} \frac{1}{n} \log m(\{x \in X : \mathcal{E}_n(x) \in U\}) \geq q^\varphi(\mu).$$

If  $\mu$  is not  $\sigma$ -invariant, then  $q^\varphi(\mu) = -\infty$  and so the equation (4.13) is trivial. Thus, we will prove the equation (4.13) for  $\mu \in \mathcal{M}_\sigma(X)$ .

Let  $\mu \in \mathcal{M}_\sigma(X)$  and  $\eta > 0$ . Then by Proposition 3.6, there exists an ergodic measure  $\nu \in U \cap \mathcal{M}_\sigma^e(X_k)$  for some  $k$  such that  $h(\nu) > h(\mu) - \eta$  and  $\int \varphi d\nu > \int \varphi d\mu - \eta$ . We use  $\nu$  to build a subset of  $\mathcal{E}_n^{-1}(U)$ , as follows.

Take  $\zeta > 0$  so small that  $\mathbb{B}(\nu, 2\zeta) \subset U$  and every measure  $\nu'$  in this neighbourhood has  $|\int \varphi d\nu' - \int \varphi d\nu| \leq \eta$ . In particular, for every  $w \in \mathcal{L}^{\nu, \zeta}$ , we have  $[w] \subset \mathcal{E}_n^{-1}(U)$ . Then, again by [29, Propositions 2.1 and 4.1], for all sufficiently large  $n$  we have

$$(4.14) \quad \#(\mathcal{L}_n^{\nu, \zeta} \cap \mathcal{L}(X_k)) \geq e^{n(h(\mu) - \eta)}.$$

We note that by the Gibbs property (3.2), we have

$$m[w] \geq K_k e^{-nP(\varphi) + S_n \varphi(x)}$$

for all  $w \in \mathcal{L}(X_k)_n$  and  $x \in [w]$ . In particular, when  $w \in \mathcal{L}^{\nu, \zeta}$  this yields

$$m[w] \geq K_k e^{-nP(\varphi) + n \int \varphi d\nu - n\eta} \geq K_k e^{n(-P(\varphi) + \int \varphi d\mu - 2\eta)}.$$

Using the estimate (4.14) and the fact that  $[w] \subset \mathcal{E}_n^{-1}(U)$  for every  $w \in \mathcal{L}^{\nu, \zeta}$ , we obtain

$$m(\mathcal{E}_n^{-1}(U)) \geq K_k e^{n(h(\mu) - P(\varphi) + \int \varphi d\mu - 3\eta)}.$$

Since  $\eta > 0$  was arbitrary, this establishes the lower bound (4.13).

## 5. APPLICATIONS

**5.1.  $S$ -gap shifts.** The family of  $S$ -gap shifts were introduced in [23], and have received a recent increase in attention [13, 8, 2]. To check the conditions of Theorem 3.2 for a Bowen potential  $\varphi$ , we verify the specification properties **[A.1]** and **(I)** on  $\mathcal{G}$  and  $\mathcal{G}^M$ , the edit approachability property **[A.2]**, and the estimate **(III)** on  $P(\mathcal{C}^p \cup \mathcal{C}^s, \varphi)$ .

**5.1.1. Specification properties.** It is immediate that  $\mathcal{G}$  has the free concatenation property, and thus Condition **[A.1]** is satisfied. Condition **(I)** holds because a word in  $\mathcal{G}^M$  has the form  $0^n 10^{n_1} 10^{n_2} 1 \dots 10^{n_k} 10^m$ , where  $n_i \in S$  for all  $1 \leq i \leq k$ , and  $n, m \leq M$ . Thus, any word in  $\mathcal{G}^M$  can be extended to a word in  $\mathcal{G}$  by adding a uniformly bounded number of symbols at each end (the number of symbols to be added depends on  $M$ , but not on the length of the word), and this implies that  $\mathcal{G}^M$  has the (W)-specification property.

**5.1.2. Edit approachability.** Because  $S$  is infinite, we can choose for every  $n \in \mathbb{N}$  some  $s_n \in S$  such that  $\frac{s_n}{n} \rightarrow 0$  and  $s_n \rightarrow \infty$ . (Note that the same element of  $S$  may appear as  $s_n$  for multiple values of  $n$ .) Now define  $g: \mathbb{N} \rightarrow \mathbb{N}$  by  $g(n) := 2(\lceil n/s_n \rceil + s_n)$ , and observe that  $g$  is a mistake function.

Let  $z \in \mathcal{L}(X)_n$  and write  $z = 0^k 10^{n_1} 10^{n_2} \dots 0^{n_i} 10^\ell$ .

$$z = 0^k 10^{n_1} 10^{n_2} \dots 0^{n_i} 10^\ell.$$

We now change at most  $k/s$  of the symbols  $0^k$  to form the word  $z^p := 0^i 10^s 10^s \dots 0^s 10^s$  ( $0 \leq i \leq s$ ). We also change at most  $\ell/s$  of the symbols  $0^\ell$  to form the word  $z^s := 0^s 10^s 10^s \dots 0^s 10^j$  ( $0 \leq j \leq s$ ). We set  $w := z^p 10^{n_1} 10^{n_2} \dots 0^{n_i} 1z^s$ ,  $u := 0^{s-i}$  and  $v := 0^{s-j} 1$ . Then we have  $\hat{d}(z, uvw) \leq 2(\lceil n/s_n \rceil + s_n) = g(n)$ , and  $uvw \in \mathcal{G}$  by the definition of  $\mathcal{G}$ . This shows that  $\mathcal{G}$  satisfies **[A.2]**.

**Remark 5.1.** When  $\Sigma_S$  is mixing – that is, when  $\gcd(S+1) = 1$  – it is possible to show that  $\mathcal{L}$  is Hamming approachable by  $\mathcal{G}$ . The idea is to combine the argument for edit approachability with some additional combinatorial estimates. It immediately follows that mixing  $S$ -gap shifts have the almost specification property.<sup>2</sup> It is also not hard to show the following spectral decomposition result: if  $d = \gcd(S+1) > 1$ , then  $(\Sigma_S, \sigma^d)$  is topologically conjugate to a union of  $d$  disjoint mixing  $S$ -gap shifts.

5.1.3. *Estimating  $P(\mathcal{C}^p \cup \mathcal{C}^s, \varphi)$ .* Now we show that if  $\varphi$  is any potential with the Bowen property on an  $S$ -gap shift, then  $P(\mathcal{C}^p \cup \mathcal{C}^s, \varphi) < P(\varphi)$ , verifying Condition **(III)**. It is easy to see that  $h(\mathcal{C}^p \cup \mathcal{C}^s) = 0$ , so it suffices to show that

$$(5.1) \quad P(\varphi) > \overline{\lim}_{n \rightarrow \infty} \sup_{x \in X} \frac{1}{n} S_n \varphi(x),$$

which is equivalent to every equilibrium state for  $\varphi$  having positive entropy. Our strategy is to produce a large number of admissible words that are close (in the edit metric) to a given word, so that no single word can carry full pressure. This strategy was also used to establish (5.1) for  $\beta$ -shifts in [9, Proposition 3.1]. For  $S$ -gap shifts, we must deal with a difficulty which does not occur for  $\beta$ -shifts: if  $x \in \Sigma_S$  is such that positions  $i$  and  $j$  both admit edits yielding new words  $x', x'' \in \Sigma_S$ , it may not be possible to make both edits simultaneously. This lack of independence between the possible edits means that it is more difficult to produce nearby words than in the case of  $\beta$ -shifts. Here, we state a sequence of lemmas which prove (5.1), whose proofs are given in §7.

**Lemma 5.2.** *We have  $P(\varphi) > \varphi(0)$ .*

In the following lemma, we use Lemma 5.2 to control words which have a small frequency of occurrence of the symbol 1.

**Lemma 5.3.** *There exists  $\epsilon > 0$  and a constant  $L = L(\epsilon)$  so that if  $x_1 \cdots x_n$  contains fewer than  $\epsilon n$  occurrences of the symbol 1, then  $\frac{1}{n} S_n \varphi(x) \leq \varphi(0) + L < P(\varphi) - L$ .*

We now control words which do not have a small frequency of occurrence of the symbol 1. This is where we use our strategy of creating a large number of new words by making edits. We need the following estimate, which is a consequence of Stirling’s formula.

<sup>2</sup>In [8], the first two authors claimed to give an example of an  $S$ -gap shift without the almost specification property. That example was in error – there was an elementary mistake in the computation, and the argument here does in fact yield the almost specification property for that example.

**Lemma 5.4.** *If  $\delta n \leq k \leq \frac{n}{2}$ , then  $\log \binom{n}{k} \geq -n\delta \log \delta - 2 \log n$ .*

This estimate can be used to give a lower bound on the cardinality of a set of words where we can control the Birkhoff averages of  $\varphi$ , and we can use this to estimate the pressure from below.

**Lemma 5.5.** *Given  $\epsilon$  as in Lemma 5.3, there exists  $L' > 0$  such that whenever  $n$  is sufficiently large and  $x_1 \cdots x_n \in \mathcal{L}$  contains  $m \geq \epsilon n$  occurrences of the symbol 1, we have  $\frac{1}{n} S_n \varphi(x) < P(\varphi) - L'$ .*

We conclude from Lemma 5.3 and Lemma 5.5 that

$$\overline{\lim}_{n \rightarrow \infty} \sup_{x \in X} \frac{1}{n} S_n \varphi(x) \leq \max\{P(\varphi) - L, P(\varphi) - L'\} < P(\varphi),$$

and it is easy to verify **(III)** from this together with  $h(\mathcal{C}^p \cup \mathcal{C}^s) = 0$ .

**5.2.  $\beta$ -shifts.** Every  $\beta$ -shift can be presented by a countable state directed labelled graph with vertices  $v_1, v_2, \dots$ . For every  $i \geq 1$ , we draw an edge from  $v_i$  to  $v_{i+1}$ , and label it with the value  $\omega_i^\beta$ . Next, whenever  $\omega_i^\beta > 0$ , for each integer from 0 to  $\omega_i^\beta - 1$ , we draw an edge from  $v_i$  to  $v_1$  labelled by that value.

The  $\beta$ -shift can be characterised as the set of sequences given by the labels of infinite paths through the directed graph which start at  $v_1$ . For our set  $\mathcal{G}$ , we take the collection of words labelling a path that begins and ends at the vertex  $v_1$ . Thus,  $\mathcal{G}$  automatically satisfies the free concatenation property, and in particular, **[A.1]** holds.

Let  $\varphi: \Sigma^\beta \rightarrow \mathbb{R}$  be a continuous function satisfying the Bowen property. It is shown in [9, §3.1] that conditions **(I)**–**(III)** in Theorem 3.1 hold, so it only remains to check condition **[A.2]**.

We now show that  $\mathcal{L}$  is Hamming approachable (and thus edit approachable) by  $\mathcal{G}$  with mistake function  $g \equiv 1$ . Let  $z \in \mathcal{L}_n$ . We set  $j := \max\{1 \leq i \leq n : z_i \neq 0\}$  and define a new word  $w \in \mathcal{L}_n$  by

$$w_i = \begin{cases} z_i & (1 \leq i \leq n, i \neq j); \\ z_j - 1 & (i = j). \end{cases}$$

It is easy to see that  $d_{\text{Ham}}(z, w) = \hat{d}(z, w) = 1 = g(n)$  and  $w \in \mathcal{G}$ , which implies **[A.2]**. It follows that  $(\Sigma^\beta, \sigma)$  satisfies the level-2 large deviations principle with reference measure  $m_\varphi$ , and rate function  $q^\varphi$  given by (1.1).

## 6. HAMMING APPROACHABILITY AND WEAK LOWER ENERGY FUNCTIONALS

For completeness, we describe the connection between our hypotheses and the hypotheses of Pfister and Sullivan [29] for the large deviations lower bound. As remarked in §2.6, if  $\mathcal{L}$  is Hamming approachable by  $\mathcal{G}$ , and  $\mathcal{G}$  satisfies (S)-specification, then the symbolic space satisfies the almost specification property, and thus the approximate product property of [29]. So, in this setting, lower large deviations follow from [29] by finding a lower weak energy function.

**Proposition 6.1.** *Let  $(X, \sigma)$  be a shift on a finite alphabet,  $m$  a Borel probability measure on  $X$ , and  $\varphi: X \rightarrow \mathbb{R}$  a continuous function. Let  $\mathcal{L}$  be the language of  $X$ . Suppose that  $\mathcal{G} \subset \mathcal{L}$  is such that*

- ( $\diamond 1$ )  *$m$  has the lower Gibbs property (2.5) for  $\varphi$  with respect to  $\mathcal{G}$ ;*
- ( $\diamond 2$ )  *$\mathcal{L}$  is Hamming approachable by  $\mathcal{G}$ ;*
- ( $\diamond 3$ ) *the function  $\psi := P(\varphi) - \varphi$  is non-negative.*

*Then  $\psi$  is a weak lower energy function in the sense of Pfister and Sullivan [29, Definition 3.3].*

*Proof.* By [29, Proposition 4.3], it is sufficient to show that for any  $\delta > 0$ , there exists  $N$  so that  $n \geq N$  implies that for each  $v \in \mathcal{L}_n$  there exists  $w \in \mathcal{L}_n$  satisfying  $d_{\text{Ham}}(v, w) \leq \delta n$  and

$$(6.1) \quad \frac{1}{n} \log m([w]) + \inf_{x \in [w]} S_n \psi(x) \geq -\delta.$$

Fix any  $\delta > 0$  and let  $K$  be the constant from the lower Gibbs property (2.5) given by ( $\diamond 1$ ). Then there is  $N_1$  so that  $n \geq N_1$  implies  $\frac{1}{n} \log K \geq -\delta$ . By ( $\diamond 2$ ), there exists  $N_\delta \geq N_1$  so that  $n \geq N_\delta$  implies that for each  $v \in \mathcal{L}_n$ , there exists  $w \in \mathcal{G}_n$  such that  $d_H(v, w) \leq \delta n$ .

Fix any  $n \geq N_\delta$ ,  $v \in \mathcal{L}_n$ . By condition ( $\diamond 2$ ), there exists  $w \in \mathcal{G}_n$  such that  $d_n^H(v, w) \leq \delta n$ . Let  $x \in [w]$ . Then it follows from ( $\diamond 1$ ) that

$$\begin{aligned} \frac{1}{n} \log m([w]) + \frac{1}{n} S_n \psi(x) &= \frac{1}{n} \log m([w]) + P(\varphi) - \frac{1}{n} S_n \varphi(x) \\ &\geq \frac{1}{n} \log K \geq -\delta, \end{aligned}$$

which implies equation (6.1). □

**Remark 6.2.** Given Theorem 3.1, [29, Propositions 3.1 and 3.2], and §5, Proposition 6.1 gives another approach to the lower large deviations bound for  $\beta$ -shifts and  $S$ -gap shifts. In particular, Proposition 6.1 illuminates the mechanism that is implicitly used in the work of Pfister and Sullivan to obtain large deviations for the measure of maximal entropy for the  $\beta$ -transformation - the proof of existence of the weak lower energy function is presented in an ad hoc way in their study.



## 7. PROOFS OF LEMMAS

*Proof of Lemma 2.6.* We obtain an upper bound on the number of words that can be obtained by making at most  $m$  edits to  $w$  as follows. We introduce an additional symbol  $e$  (for ‘edit’), and construct a new word  $w'$  of length  $n+m$  which contains exactly  $m$  of the symbols  $e$ , and so that  $w_1 = w'_1$ . Note that  $\binom{n+m}{n}$  is an upper bound on the number of such words  $w'$ . Now obtain a new word  $v \in \mathcal{L}$  from  $w'$  by performing exactly one of the following actions at each symbol  $e$ , and then deleting the  $e$ .

- (1) Change the symbol immediately before  $e$  to a different symbol.
- (2) Insert a symbol immediately before  $e$ .
- (3) Delete the symbol immediately before  $e$ .
- (4) Leave the symbol immediately before  $e$  unchanged.

Note that every word  $v$  which satisfies  $\hat{d}(v, w) \leq m$  can be produced by this procedure. At each symbol  $e$ , there are a total of  $2\#A + 2$  possible actions, so we see that

$$\#\{v \mid \hat{d}(v, w)\} \leq (2\#A + 2)^m \binom{n+m}{n}.$$

From Stirling’s formula there is a constant  $C'$  such that

$$|\log n! - (n \log n - n)| \leq C' \log n$$

for every  $n \in \mathbb{N}$ , and so when  $m \leq \delta n$  we have

$$\begin{aligned} \log \binom{n+m}{n} &= \log(n+m)! - \log n! - \log m! \\ &\leq ((n+m) \log(n+m) - n \log n - m \log m) \\ &\quad + C'(\log(m+n) + \log m + \log n) \\ &= \left( n \log \frac{n+m}{n} + m \log \frac{n+m}{m} \right) + 3C' \log(m+n) \\ &\leq n(\log(1+\delta) + \delta \log(1+\delta^{-1})) + 3C' \log((1+\delta)n) \\ &= n((1+\delta) \log(1+\delta) - \delta \log \delta) + 3C' \log((1+\delta)n). \end{aligned}$$

Using the inequalities  $1 + \delta \leq 2$  and  $\log(1 + \delta) \leq \delta$ , we see that the left-hand side of (2.2) admits the bound

$$\begin{aligned} \#\{v \in \mathcal{L} \mid \hat{d}(v, w) \leq \delta n\} &\leq (2\#A + 2)^{\delta n} e^{n((1+\delta) \log(1+\delta) - \delta \log \delta)} e^{3C' \log((1+\delta)n)} \\ &\leq (2\#A + 2)^{\delta n} e^{2n\delta} e^{n(-\delta \log \delta)} (1 + \delta)^{3C'} n^{3C'}, \end{aligned}$$

which completes the proof.  $\square$

*Proof of Lemma 2.8.* Let  $\hat{g}(n) = g(n) + 1$ , so that  $\hat{g}$  is also a mistake function. Take  $x, y$  and  $m, n$  as in the hypothesis of the lemma, and let  $k = \hat{d}(x_1 \cdots x_n, y_1 \cdots y_m) \leq g(n)$ .

Following the set-up of the proof of the previous lemma, we obtain a new word  $w'$  by inserting the symbol  $e$  into  $k$  positions of  $x_1 \cdots x_n$  to mark where an insertion, deletion or substitution will take place to obtain  $y_1 \cdots y_m$ . We write  $w = w^1 w^2 \cdots w^{k+1}$  so that the last symbol of each  $w^i$  with  $1 \leq i \leq k$  is  $e$  (note that  $w^{k+1}$  may be the empty word). Let  $w_r^i$  be the word obtained by omitting the last two symbols from  $w^i$ , and form the word  $w_r = w_r^1 w_r^2 \cdots w_r^{k+1}$  (where  $r$  stands for ‘reduced’, and if  $|w^i| \leq 2$ , then  $w_r^i$  is the empty word). For  $n \geq 0$ , let

$$V(n) = \sup\{|S_{m'}\varphi(x) - S_m\varphi(y)| \mid x_1 \cdots x_n = y_1 \cdots y_n \\ \text{and } m, m' \in \{n, n+1, n+2\}\}.$$

Note that continuity of  $\varphi$  implies that  $\frac{1}{n}V(n) \rightarrow 0$ . In particular, for  $z \geq 1$ , we may write  $\epsilon(z) = \sup_{m \geq z} \frac{1}{m}V(m)$  and obtain  $\epsilon(z) \rightarrow 0$ . We will use this fact for “long” words, while for “short” words we will use the bound  $V(n) \leq 2(n+2)\|\varphi\| \leq 4(n+1)\|\varphi\|$ .

Both  $x_1 \cdots x_n$  and  $y_1 \cdots y_m$  can be obtained from  $w_r$  by inserting at most two symbols at the end of each  $w_r^i$ , and so  $|S_n\varphi(x) - S_m\varphi(y)| \leq \sum_{j=1}^{k+1} V(n_j)$ , where  $n_j = |w_r^j|$ . To bound this sum, we let  $C_n = \sqrt{\frac{n}{\hat{g}(n)}}$  and break the sum into two parts, corresponding to  $n_j < C_n$  and  $n_j \geq C_n$ . We have

$$(7.1) \quad |S_n\varphi(x) - S_m\varphi(y)| \leq \sum_{n_j < C_n} V(n_j) + \sum_{n_j \geq C_n} V(n_j) \\ \leq \sum_{n_j < C_n} 4(n_j + 1)\|\varphi\| + \sum_{n_j \geq C_n} n_j\epsilon(C_n) \leq 4C_n\|\varphi\|\hat{g}(n) + n\epsilon(C_n),$$

where the last inequality uses the fact that there are  $k+1 \leq \hat{g}(n)$  values of  $j$  in total, and that  $\sum n_j \leq n$ .

Now we can estimate the difference in Lemma 2.8 as

$$\left| \frac{1}{n}S_n\varphi(x) - \frac{1}{m}S_m\varphi(y) \right| \leq \frac{1}{n}|S_n\varphi(x) - S_m\varphi(y)| + \left| \frac{1}{n} - \frac{1}{m} \right| |S_m\varphi(y)| \\ \leq 4\|\varphi\|C_n \frac{\hat{g}(n)}{n} + \epsilon(C_n) + \frac{|m-n|}{n} \frac{1}{m} |S_m\varphi(y)| \\ \leq 4\|\varphi\| \sqrt{\frac{\hat{g}(n)}{n}} + \epsilon(C_n) + \frac{\hat{g}(n)}{n} \|\varphi\|.$$

Because  $\hat{g}$  is a mistake function, the first and third terms go to 0 as  $n \rightarrow \infty$ , while  $C_n \rightarrow \infty$  and so the second term goes to 0 as well. This completes the proof of Lemma 2.8.  $\square$

*Proof of Proposition 2.9.* Clearly  $P(\mathcal{G}, \varphi) \leq P(\varphi)$ , so it suffices to prove the other inequality. We compare  $\Lambda_n(\mathcal{L}, \varphi)$  and  $\Lambda_n(\mathcal{G}, \varphi)$  using Lemmas 2.6 and 2.8. By edit approachability, for each  $w \in \mathcal{L}_n$  there exists  $v = v(w) \in \mathcal{G}$  such that  $\hat{d}(v, w) \leq g(|w|)$ . Lemma 2.6 tells us that given  $v \in \mathcal{G}$ , the number of words  $w \in \mathcal{L}_n$  for which  $v = v(w)$  is at most

$$Cn^C (e^{C\delta} e^{-\delta \log \delta})^n,$$

where  $\delta = g(n)/n$ . In particular, for all sufficiently large  $n$  this expression is bounded above by  $e^{\delta'_n n}$ , where  $\delta'_n \rightarrow 0$ .

It follows from Lemma 2.8 that there is  $\delta_n \rightarrow 0$  such that for every  $v, w$  as above and any  $x \in [v]$ ,  $y \in [w]$ , we have

$$|S_n \varphi(x) - S_{|w|} \varphi(y)| \leq n\delta_n.$$

Together the above estimates imply that

$$\Lambda_n(\mathcal{L}, \varphi) \leq \sum_{m=n-g(n)}^{n+g(n)} \sum_{w \in \mathcal{G}_m} e^{\delta'_n n} e^{n\delta_n + \sup_{y \in [w]} S_m \varphi(y)},$$

and so in particular there is  $m \in [n - g(n), n + g(n)]$  such that

$$\Lambda_m(\mathcal{G}, \varphi) \geq \frac{1}{2g(n)} e^{-(\delta'_n + \delta_n)n} \Lambda_n(\mathcal{L}, \varphi).$$

Since  $g(n)$  is sublinear and  $\delta_n, \delta'_n \rightarrow 0$ , this implies the result.  $\square$

*Proof of Lemma 2.12.* Items 1) and 3) can be obtained by making minor modifications to the proof of Proposition 2.2 in [8, §6.2], so we omit these arguments and prove only item 2).

Let  $\mathcal{G} \subset \mathcal{L}(\Sigma)$  and  $\tilde{\mathcal{G}} \subset \mathcal{L}(X)$  be as in Lemma 2.12 and assume that  $\mathcal{G}$  satisfies [A.2]. Let  $g: \mathbb{N} \rightarrow \mathbb{N}$  be a mistake function as in [A.2] for  $\mathcal{G}$ . Then we define a mistake function  $\tilde{g}: \mathbb{N} \rightarrow \mathbb{N}$  by  $\tilde{g}(n) = (4r + 3)g(n + 2r) + 4r$ . Take a  $\tilde{z} \in \mathcal{L}(X)_n$ . Since  $\Psi$  is surjective, there exists  $z \in \mathcal{L}(\Sigma)_{n+2r}$  so that  $\Psi(z) = \tilde{z}$ . Since  $\mathcal{G}$  satisfies [A.2], we can find  $w \in \mathcal{G}$  so that  $\hat{d}(z, w) \leq g(n + 2r)$  holds, where we recall that  $r$  is the length of the block code. We set  $\tilde{w} = \Psi(w)$ .

Because  $\hat{d}(z, w) \leq g(n + 2r)$ , there exist an integer  $K \geq n - ((2r + 1)g(n + 2r) + 2r)$  and two increasing sequences  $m_1 < \dots < m_K$ ,  $n_1 < \dots < n_K$  so that

$$z_{m_i-r} \dots z_{m_i+r} = w_{n_i-r} \dots w_{n_i+r}$$

for each  $1 \leq i \leq K$ . Because  $\Psi$  is a block code with length  $r$ , we have  $\tilde{z}_{m_i} = \tilde{w}_{n_i}$  for  $1 \leq i \leq K$ . This implies that

$$\hat{d}(\tilde{z}, \tilde{w}) \leq (n - K) + (|w| - K) \leq 2(n - K) + ||w| - n| \leq \tilde{g}(n).$$

Thus  $\mathcal{G}$  satisfies **[A.2]**.  $\square$

*Proof of Lemma 3.3.* We have

$$\Lambda_n(\mathcal{C}, \varphi) = \sum_{w \in \mathcal{C}_n} e^{\sup_{x \in [w]} S_n \varphi(x)} \leq e^{n(\sup \varphi)} \Lambda_n(\mathcal{C}, 0),$$

and so  $P(\mathcal{C}, \varphi) \leq h(\mathcal{C}) + \sup \varphi$ . By the variational principle and the assumption (BR), we have

$$P(X, \varphi) \geq h(X) + \inf \varphi > h(\mathcal{C}) + \sup \varphi \geq P(\mathcal{C}, \varphi),$$

which proves the lemma.  $\square$

*Proof of Lemma 5.2.* Let  $V$  be such that  $|S_n \varphi(x) - S_n \varphi(y)| \leq V$  whenever  $x_1 \cdots x_n = y_1 \cdots y_n$ , and in particular  $S_n \varphi(x) \geq n\varphi(0) - V'$  for every  $x \in [0^{n-1}1]$ , where  $V' = V + \varphi(0) - (\inf \varphi)$ .

Choose  $k$  large (just how large will be determined later) and let  $n_1, n_2, \dots, n_k \in S$  be distinct. Let  $\pi$  be any permutation of the integers  $\{1, \dots, k\}$ , and let  $w_\pi$  be the word  $0^{n_{\pi(1)}}10^{n_{\pi(2)}}1 \cdots 0^{n_{\pi(k)}}1$  of length  $N = \sum_{j=1}^k (n_j + 1)$ . The estimates in the previous paragraph give

$$S_N \varphi(y) \geq N\varphi(0) - kV'$$

for every  $y \in [w_\pi]$ . Now let  $\vec{\pi} = (\pi_1, \dots, \pi_m)$  be any sequence of  $m$  such permutations, and let  $v_{\vec{\pi}} = w_{\pi_1} \cdots w_{\pi_m}$ . Choosing any  $y_{\vec{\pi}} \in [v_{\vec{\pi}}]$ , we obtain the estimate

$$(7.2) \quad \Lambda_{mN}(\mathcal{L}, \varphi) \geq \sum_{\vec{\pi}} e^{S_{mN} \varphi(y_{\vec{\pi}})} \geq (k!)^m e^{mN\varphi(0) - mkV'}.$$

We have the general bound

$$(7.3) \quad \log(k!) = \sum_{j=1}^k \log j \geq \int_1^k \log t \, dt = k \log k - k - 1,$$

which yields

$$\log \Lambda_{mN}(\mathcal{L}, \varphi) \geq m(k \log k - k - 1) + mN\varphi(0) - mkV',$$

so that dividing by  $mN$  and sending  $m \rightarrow \infty$  we have

$$P(\varphi) \geq \varphi(0) + \frac{k}{N} \left( \log k - 1 - \frac{1}{k} - V' \right).$$

Taking  $k$  large gives the result.  $\square$

*Proof of Lemma 5.3.* By Lemma 5.2, there exists  $\epsilon > 0$  such that  $\varphi(0) + 2\epsilon V' < P(\varphi)$ , where  $V'$  is the constant from the proof of the previous lemma. Note that if  $x_1 \cdots x_n$  contains fewer than  $\epsilon n$  occurrences of the symbol 1, then  $S_n \varphi(x) \leq n\varphi(0) + \epsilon n V'$ , and in particular

$$(7.4) \quad \frac{1}{n} S_n \varphi(x) \leq \varphi(0) + \epsilon V' < P(\varphi) - \epsilon V'.$$

Setting  $L = \epsilon V'$  gives the result.  $\square$

*Proof of Lemma 5.4.* We use the upper bound

$$\begin{aligned} \log(k!) &= \sum_{j=1}^k \log j \leq \int_1^{k+1} \log t \, dt = (k+1) \log(k+1) - k \\ &= (k \log k - k) + k \log\left(1 + \frac{1}{k}\right) + \log(k+1) \\ &\leq (k \log k - k) + (1 + \log(k+1)), \end{aligned}$$

which together with (7.3) gives, for all large  $n$ ,

$$\begin{aligned} \log \binom{n}{k} &= \log(n!) - \log(k!) - \log((n-k)!) \\ &\geq (n \log n - n - 1) - (k \log k - k) - (1 + \log(k+1)) \\ &\quad - ((n-k) \log(n-k) - (n-k)) - (1 + \log(n-k+1)) \\ &\geq nh \left(\frac{k}{n}\right) - 2 \log n, \end{aligned}$$

where  $h(\delta) = -\delta \log \delta - (1-\delta) \log(1-\delta)$ .  $\square$

*Proof of Lemma 5.5.* Now assume that  $x_1 \cdots x_n$  contains  $m \geq \epsilon n$  occurrences of the symbol 1. By considering a smaller collection of indices where the entry is 1 if necessary, we may assume that  $m \leq 2\epsilon n$ .

Given  $\delta > 0$  small (just how small will be determined later), let  $\delta m < k < 2\delta m$ . Let  $R$  be the set of indices in which  $x_1 \cdots x_n$  has a nonzero symbol, and let  $\mathcal{Z}$  be the collection of subsets of  $R$  with exactly  $k$  elements.

We define a map  $\phi: \mathcal{Z} \rightarrow X$  as follows. Fix  $n_1 \neq n_2 \in S$ . Given  $Z \in \mathcal{Z}$ , at each index  $k \in Z$  insert the word  $0^{n_1}1$  into  $x$ , unless  $x_{k+1} \cdots x_{k+n_1+1} = 0^{n_1}1$ , in which case insert the word  $0^{n_2}1$ . This is allowed by the definition of the  $S$ -gap shift, and we note that  $\phi$  is 1-1.

Let  $\ell = \max\{n_1, n_2\} + 1$ , and observe that  $\phi(Z)$  is obtained from  $x$  by inserting at most  $k\ell$  symbols, so that if  $p$  is the size of the alphabet, then the map  $\Phi: \mathcal{Z} \rightarrow \mathcal{L}_n$  obtained by truncating  $\phi(Z)$  to the first  $n$  symbols has the property that  $\#\Phi^{-1}(w) \leq p^{k\ell}$  for each  $w \in \mathcal{L}_n$ .

We conclude that the map  $\Phi$  yields at least  $\binom{m}{k} p^{-k\ell}$  words  $w$  in  $\mathcal{L}_n$  with the property that

$$S_n \varphi(y) \geq S_n \varphi(x) - k\ell V' \geq S_n \varphi(x) - 4\epsilon \delta n V'$$

for every  $y \in [w]$ . In particular, together with Lemma 5.4 and the conditions on  $m$  and  $k$ , this gives the estimate

$$\begin{aligned} \log \Lambda_n(\mathcal{L}, \varphi) &\geq -m\delta \log \delta - 2 \log m - k\ell \log p + S_n \varphi(x) - 4\epsilon\delta n V' \\ &\geq (\epsilon n)(-\delta \log \delta) - 4 \log(\epsilon n) - 4\epsilon\delta \ell \log p + S_n \varphi(x) - 4\epsilon\delta n V'. \end{aligned}$$

Dividing by  $n$  gives

$$\frac{1}{n} \log \Lambda_n(\Lambda, \varphi) \geq \frac{1}{n} S_n \varphi(x) + \epsilon\delta(-\log \delta - 4\ell \log p - 4V') - 4 \frac{\log(\epsilon n)}{n},$$

which yields the desired result when  $\delta$  is chosen sufficiently small and  $n$  is chosen sufficiently large.  $\square$

## REFERENCES

- [1] V. Araújo and M.J. Pacifico. Large deviations for non-uniformly expanding maps. *J. Stat. Phys.* 125 (2006), no. 2, 415–457.
- [2] S. Baker and A. Ghenciu, Dynamical properties of  $S$ -gap shifts and other shift spaces, preprint (2014), arXiv:1411.7167
- [3] Anne Bertrand, *Specification, synchronisation, average length*, Coding theory and applications (Cachan, 1986), Lecture Notes in Comput. Sci., vol. 311, Springer, Berlin, 1988, pp. 86–95. MR 960710 (89i:94022)
- [4] Anne Bertrand-Mathis, *Développement en base  $\theta$* , Bull. Soc. Math. France, 114 (1986) 271–323.
- [5] Y. M. Chung. Large deviations on Markov towers. *Nonlinearity* 24 (2011), 1229–1252.
- [6] Y. M. Chung and H. Takahasi. Large deviation principle for Benedicks-Carleson quadratic maps: *Comm. Math. Phys.* 315 (2012) 803–826.
- [7] V. Climenhaga, Specification and towers in shift spaces, preprint (2015), arXiv:1502.00931v1.
- [8] V. Climenhaga and D. J. Thompson, Intrinsic ergodicity beyond specification:  $\beta$ -shifts,  $S$ -gap shifts, and their factors, *Israel J. of Math.* 192, no. 2, 785–817 (2012)
- [9] V. Climenhaga and D. J. Thompson, Equilibrium states beyond specification and the Bowen property, *J. London Math. Soc.* 87, no. 2 (2013) 401–427
- [10] V. Climenhaga and D. J. Thompson, Intrinsic ergodicity via obstruction entropies, *Ergod. Th. & Dynam. Sys.* 34, no. 6 (2014) 1816–1831.
- [11] V. Climenhaga and D. J. Thompson, Unique equilibrium states for flows and homeomorphisms with non-uniform structure, preprint (2015), arXiv:1505.03803.
- [12] H. Comman and J. Rivera-Letelier. Large deviation principles for non-uniformly hyperbolic rational maps. *Ergod. Th. & Dynam. Sys.* 31 (2011), 321–349.
- [13] D. Ahmadi Dastjerdi and S. Jangjoo, Dynamics and topology for  $S$ -gap shifts, *Topology and its Applications*, 159 (2012), 2654–2661.
- [14] M. Denker, Large deviation and the pressure function, *Transactions of the 11th Prague Conference on Information theory, Statistical Decision Functions,*

- Prague, 1990*, Academia Publ. House of the Czechoslovak Acad. of Science, 21–33 (1992).
- [15] A. Eizenberg, Y. Kifer and B. Weiss, Large deviations for  $\mathbb{Z}^d$ -actions, *Comm. Math. Phys.* 164, 433–454 (1994).
  - [16] R.S. Ellis, *Entropy, Large deviations and Statistical Mechanics*, Grundlehren der Mathematischen Wissenschaften 271, Springer-Verlag, New York, 1985.
  - [17] R.S. Ellis, The Theory of Large Deviations and Applications to Statistical Mechanics, in “Long-Range Interacting Systems: Les Houches 2008 Session XC, Oxford University Press (New York), 227–277 (2010).
  - [18] H. Föllmer and S. Orey. Large deviations for the empirical field of a Gibbs measure. *Ann. Probab.* 16 (1988), no. 3, 961–977.
  - [19] A. Katok. Lyapunov exponents, entropy and periodic orbits for diffeomorphisms. *Inst. Hautes tudes Sci. Publ. Math.* No. 51 (1980), 137–173.
  - [20] G. Keller and T. Nowicki. Spectral theory, zeta functions and the distribution of periodic points for Collet-Eckmann maps. *Comm. Math. Phys.* 149 (1992), no. 1, 31–69.
  - [21] Y. Kifer, Large deviations in dynamical systems and stochastic processes, *Trans. Amer. Math. Soc.* 321, 505–524 (1990).
  - [22] D. Kwietniak, P. Oprocha and M. Rams, On entropy of dynamical systems with almost specification, preprint (2014), arXiv:1411.1989
  - [23] D. Lind and B. Marcus, *An introduction to symbolic dynamics and coding*, Cambridge University Press, 1995.
  - [24] A.O. Lopes. Entropy and large deviation. *Nonlinearity* 3 (1990), no. 2, 527–546.
  - [25] I. Melbourne, Large and moderate deviations for slowly mixing dynamical systems, *Proc. Amer. Math. Soc.*, 37, no.5, (2009), 1735–1741.
  - [26] I. Melbourne and M. Nicol, Large deviations for nonuniformly hyperbolic systems, *Trans. Amer. Math. Soc.*, 360, no.12 (2008), 6661–6676.
  - [27] S. Orey and S. Pelikan, Large deviation principles for stationary processes, *Ann. Probab.* 16, no. 4, 1481–1495 (1988).
  - [28] S. Orey and S. Pelikan, Deviation of trajectory averages and the defect in Pesin’s formula for Anosov diffeomorphisms, *Trans. Amer. Math. Soc.* 315 (1989), 741–753.
  - [29] C.-E. Pfister and W.G. Sullivan, Large Deviations Estimates for Dynamical Systems without the Specification Property. Application to the  $\beta$ -shifts, *Nonlinearity* 18 (2005), 237–261.
  - [30] C.-E. Pfister, W.G. Sullivan, On the Topological Entropy of Saturated Sets, *Ergodic Theory Dynam. Systems* 27, 929–956 (2007).
  - [31] M. Pollicott and R. Sharp. Large deviations for intermittent maps. *Nonlinearity* 22 (2009), no. 9, 2079–2092.
  - [32] M. Pollicott, R. Sharp, and M. Yuri. Large deviations for maps with indifferent fixed points. *Nonlinearity* 11 (1998), no. 4, 1173–1184.
  - [33] L. Rey-Bellet and L.-S Young, Large deviations in non-uniformly hyperbolic dynamical systems, *Ergodic Theory Dynam. Systems*, 28 (2008), 578–612.
  - [34] Y. Takahashi, Entropy functional (free energy) for dynamical systems and their random perturbations. *Stochastic analysis (Katata/Kyoto, 1982)*, North-Holland Math. Library 32, North-Holland, Amsterdam, 437–467 (1984).

- [35] Y. Takahashi, Two aspects of large deviation theory for large time, in “Probabilistic methods in mathematical physics (Katata/Kyoto, 1985)”, Academic Press, Boston 363–384 (1987).
- [36] P. Varandas, Non-uniform specification and large deviations for weak Gibbs measures, *J. Statist. Phys.*, 146, 330–358 (2012).
- [37] P. Walters. Equilibrium states for  $\beta$ -transformations and related transformations. *Math. Z.* 159 (1978), no. 1, 65–88.
- [38] K. Yamamoto, On the weaker forms of the specification property and their applications, *Proc. Amer. Math. Soc.* **137**, no. 11, 3807–3814 (2009).
- [39] L.-S. Young, Some large deviation results for dynamical systems, *Trans. Amer. Math. Soc.* 318, no. 2, 525–543 (1990).

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF HOUSTON, HOUSTON, TEXAS  
77204, USA

*E-mail address:* `climenha@math.uh.edu`

DEPARTMENT OF MATHEMATICS, THE OHIO STATE UNIVERSITY, 100 MATH  
TOWER, 231 WEST 18TH AVENUE, COLUMBUS, OHIO 43210, USA

*E-mail address:* `thompson@math.osu.edu`

DEPARTMENT OF GENERAL EDUCATION, NAGAOKA UNIVERSITY OF TECH-  
NOLOGY, NIIGATA 940-2188, JAPAN

*E-mail address:* `k.yamamoto@vos.nagaokaut.ac.jp`