ISRAEL JOURNAL OF MATHEMATICS **232** (2019), 899–920
DOI: 10.1007/s11856-019-1891-5

POSITIVE ENTROPY EQUILIBRIUM STATES

BY

Vaughn Climenhaga*

Department of Mathematics, University of Houston
3551 Cullen Blvd., Houston, TX 77204-3008, USA
e-mail: climenha@math.uh.edu

AND

Van Cyr

Department of Mathematics, Bucknell University
1 Dent Dr., Lewisburg, PA 17837-2005, USA
e-mail: van.cyr@bucknell.edu

ABSTRACT

For transitive shifts of finite type, and more generally for shifts with specification, it is well-known that every equilibrium state for a Hölder continuous potential has positive entropy as long as the shift has positive topological entropy. We give a non-uniform specification condition under which this property continues to hold, and demonstrate that it does not necessarily hold for other non-uniform versions of specification that have been introduced elsewhere.

* The first author is partially supported by NSF grants DMS-1362838 and DMS-1554794.

Received September 8, 2017 and in revised form October 8, 2018

1. Introduction

Given a compact metric space X, a continuous map $f\colon X \to X$, and a continuous potential function $\varphi\colon X \to \mathbb{R}$, an equilibrium state for (X, f, φ) is an f-invariant measure realising the supremum in the variational principle

$$P(\varphi) = \sup_\mu \left(h_\mu(f) + \int \varphi \, d\mu \right).$$

It is often important to know under what conditions an equilibrium state is forced to have positive entropy, or equivalently, for which potentials we have

$$P(\varphi) > \sup_\mu \int \varphi \, d\mu. \tag{1.1}$$

Following [IRRL12], a potential satisfying (1.1) will be called hyperbolic.

If (X, σ) is a transitive subshift of finite type (SFT) with positive topological entropy, then every Hölder potential is hyperbolic. This also holds for all systems with the specification property [CFT19, Theorem 6.1].

The importance of (1.1) is discussed in [Buz04]; see [DKU90, Buz01] for its consequences regarding uniqueness of equilibrium states, and [Ryc83, Kel84, BK90] for its connection to quasi-compactness of the transfer operator, which has implications for the statistical properties of the system.

In [Buz04], Buzzi considers continuous piecewise monotonic interval maps f and shows that if f is topologically transitive and φ is Hölder continuous in the natural coding via the branch partition, then (1.1) holds. Buzzi conjectured that the result remains true without the assumption that the map f is continuous, but so far this question remains open.

We offer partial progress towards this conjecture by giving a general condition under which every Hölder potential satisfies (1.1). Our condition is formulated in terms of the symbolic representation of f, and can be thought of as a stronger version of the **almost specification** property [PS07, Tho12].

Given a shift space X, we write $\mathcal{L}$ for the **language** of X (the set of all finite words appearing in some element of X). A **prefix** of a word $w \in \mathcal{L}$ is any word of the form $w_1 \cdots w_k$ for some $k \leq |w|$; similarly, a **suffix** of w is any word of the form $w_k \cdots w_{|w|}$. We say that a subset $\mathcal{G} \subset \mathcal{L}$ has specification if there is $\tau \in \mathbb{N}$ such that for every $v, w \in \mathcal{G}$ there is $u \in \mathcal{L}$ with $|u| \leq \tau$ such that $v'uw' \in \mathcal{G}$ whenever v' is a suffix of v and w' is a prefix of w with $v', w' \in \mathcal{G}$. (This includes the case when $v' = v$ and $w' = w$.)

Given a non-decreasing function $g\colon \mathbb{N} \to \mathbb{N}$, the language $\mathcal{L}$ is said to be g-Hamming approachable by $\mathcal{G}$ if every sufficiently long $w \in \mathcal{L}$ can be transformed into a word in $\mathcal{G}$ by changing no more than $g(|w|)$ symbols.

THEOREM 1.1: *Let X be a shift space on a finite alphabet with positive topological entropy, and $\mathcal{L}$ its language. If there is a function $g\colon \mathbb{N} \to \mathbb{N}$ with $\lim_{n\to\infty} g(n)/\log(n) = 0$ and a set $\mathcal{G} \subset \mathcal{L}$ with specification such that $\mathcal{L}$ is g-Hamming approachable by $\mathcal{G}$, then every Hölder continuous potential on X is hyperbolic.*

An important class of shifts satisfying the conditions of the theorem is given by the β-shifts, which code the transformations $x \mapsto \beta x \pmod 1$ for $\beta > 1$. In this case $g(n) = 1$ for every n, and it was already shown in [CT13, Proposition 3.1] that every Hölder potential is hyperbolic. The proof there relied strongly on the lexicographic structure of the β-shifts; in particular it does not apply to their factors. Our approach here does pass to factors.

PROPOSITION 1.2: *Let X be a shift space satisfying the hypotheses of Theorem 1.1. Then every subshift factor of X satisfies them as well.*

Proof. By the proof of [CTY17, Lemma 2.12], if $g(n)$ works for X, and $\tilde{X}$ is a subshift factor obtained via an r-block code, then $\tilde{g}(n) = (4r+3)g(n+2r)+4r$ works for $\tilde{X}$. ∎

COROLLARY 1.3: *Let X be any subshift factor of a β-shift. Then every Hölder potential on X satisfies* (1.1)*, and has a unique equilibrium state, which has exponential decay of correlations and the central limit theorem for Hölder observables.*

Proof. Theorem 1.1 and Proposition 1.2 give (1.1); for the rest, see [Cli18, Theorem 1.4, Example 1.5, and §§1.7–1.8]. ∎

Remark 1.4: Another class of shift spaces studied in [CT12, CTY17] are the S-gap shifts, for which there is no function g as in Theorem 1.1; the best that can be done in general is $g(n) \approx \sqrt{n}$, see [CTY17, §5.1.2]. On the other hand, it was shown in [CTY17, (5.1)] that every Hölder potential for these shifts is hyperbolic. The corresponding question for their subshift factors remains open.

Remark 1.5: Another condition that appears in the literature to guarantee hyperbolicity of Hölder potentials is the 'local specification' condition of Hofbauer and Keller [HK82, Theorem 3], which can be stated as follows. Given $k \in \mathbb{N}$, let $\mathcal{F}_k$ be the set of $w \in \mathcal{L}$ such that for every $v \in \mathcal{L}$, there is $u \in \mathcal{L}$ with $|u| \leq k$ such that $wuv \in \mathcal{L}$. (Then $\mathcal{L}$ has specification iff there is k such that $\mathcal{F}_k = \mathcal{L}$.) The 'local specification' property from [HK82, Theorem 3] is equivalent to: for every $x \in X$ and every infinite $J \subset \mathbb{N}$, there is $k \in \mathbb{N}$ and an infinite $J' \subset J$ such that $x_1 \cdots x_j \in \mathcal{F}_k$ for every $j \in J'$.

Another result for interval maps was given in [LRL14], which showed that for a class of smooth interval maps with critical points and some non-uniformly expanding properties, (1.1) holds for every Hölder continuous potential (not just those that are Hölder in the natural coding).

Beyond β-transformations, it is natural to study the class of interval maps given by $x \mapsto \alpha + \beta x$ for $\alpha \in (0,1)$, $\beta > 1$. The coding spaces for these maps can be represented in terms of a countable graph using the general theory of Hofbauer [Hof79], but it is not clear what mistake function g these shifts admit, and so Buzzi's conjecture remains open for this class.

In light of Remark 1.4 above on S-gap shifts, and other results from [CTY17] in which $g(n)/n \to 0$ seems to be the relevant condition, it is natural to ask how sharp the sublogarithmic condition on g is. In fact, one cannot do much better, as the following family of examples shows.

THEOREM 1.6: *Let $f\colon \mathbb{N} \to \mathbb{N}$ be non-decreasing, and suppose that there is $n_1 \in \mathbb{N}$ such that $1 \leq f(n) \leq n/2$ for all $n \geq n_1$. Let*

$$G = \{0^a 1^b \mid a, b \geq f(a+b)\},$$

and let X be the coded shift generated by G. Then for $\varphi = -\mathbf{1}_{[1]}$, the potentials $t\varphi$ have $P(t\varphi) \geq 0$ for all $t \in \mathbb{R}$, and $t \mapsto P(t\varphi)$ is non-increasing. Writing

$$t_0 = \inf\{t \mid P(t\varphi) = 0\} = \sup\{t \mid P(t\varphi) > 0\} \tag{1.2}$$

for the first root of Bowen's equation (possibly $+\infty$), the following are true.

(i) *For the function $g(n) = 2n_1 + 2\max(f(n), n_1)$, $\mathcal{L} = \mathcal{L}(X)$ is g-Hamming approachable by $\mathcal{G} = G^*$.*
(ii) *Given $t \geq 0$, the potential $t\varphi$ is hyperbolic if and only if $t < t_0$.*
(iii) *If $0 \leq t < t_0$, then there is a unique equilibrium state for $t\varphi$, and it has positive entropy.*
(iv) *If $t > t_0$, then δ_0 is the unique equilibrium state for $t\varphi$.*
(v) *$t_0 < \infty$ if and only if there exists $\gamma > 0$ such that $\sum_{n\in\mathbb{N}} \gamma^{f(n)} < \infty$.*

Remark 1.7: The examples in Theorem 1.6 are modifications of the coded shift generated by $G = \{0^n 1^n : n \in \mathbb{N}\}$, which was studied by Conrad [Con], who showed that for sufficiently large values of t, the potential $t\varphi$ has the delta measure δ_0 as its unique equilibrium state, and in particular is not hyperbolic.

The last statement in Theorem 1.6 allows us to give a class of shifts for which there is a Hölder potential that is not hyperbolic.

Corollary 1.8: *If* $\liminf g(n)/\log(n) > 0$*, then the conclusion of Theorem 1.1 fails in the following sense: there is a shift* X *with language* $\mathcal{L}$ *and a collection* $\mathcal{G} \subset \mathcal{L}$ *such that* $\mathcal{G}^* \subset \mathcal{G}$ *and* $\mathcal{L}$ *is* g*-Hamming approachable by* $\mathcal{G}$*, but there is a locally constant potential function with a delta measure as its unique equilibrium state.*

Remark 1.9: In fact, Theorem 1.6 shows that hyperbolicity can fail for some error functions g with $\liminf g(n)/\log(n) = 0$ and $\limsup g(n)/\log(n) > 0$, as long as there is $\gamma > 0$ such that $\sum_n \gamma^{g(n)} < \infty$. This does not cover all functions g with $\liminf = 0$ and $\limsup > 0$; it would be interesting to know if Theorem 1.1 can be extended to include functions g where $\limsup > 0$ but $\sum_n \gamma^{g(n)} = \infty$ for all $\gamma > 0$.

Acknowledgments. We are grateful to the anonymous referee for pointing out an error in the original version of §4.1.

2. Background definitions

2.1. Shift spaces. Given a finite set A, let $\sigma\colon A^{\mathbb{N}} \to A^{\mathbb{N}}$ denote the left shift map.[1] Equip $A^{\mathbb{N}}$ with the product topology; equivalently, define a metric on A by $d(x, y) = 2^{-\min\{n\in\mathbb{N} | x_n \neq y_n\}}$. A **shift space over the alphabet** A is a closed σ-invariant subset $X \subset A^{\mathbb{N}}$.

Write $A^* = \bigcup_{n=0}^{\infty} A^n$ for the collection of all finite words over A. Given a shift space X, the **language** of X is

$$\mathcal{L} = \mathcal{L}(X) = \{w \in A^* \mid x_1 \cdots x_n = w \text{ for some } x \in X \text{ and } n \in \mathbb{N}\}.$$

Given $\mathcal{D} \subset \mathcal{L}$, write $\mathcal{D}_n = \mathcal{D} \cap A^n$ for the set of all words of length n in $\mathcal{D}$. In particular, $\mathcal{L}_n$ denotes the set of all words of length n in the language of X.

[1] Our results all remain true for two-sided shifts ($\sigma\colon A^{\mathbb{Z}} \to A^{\mathbb{Z}}$).

Given $w \in \mathcal{L}_n$, let

$$[w] = \{x \in X \mid x_1 \cdots x_n = w\}$$

be the corresponding **cylinder** in X.

2.2. Thermodynamic formalism and equilibrium states. Let X be a shift space and $\mathcal{L}$ its language. Given a continuous function $\varphi\colon X \to \mathbb{R}$, which we call a **potential**, consider for each $w \in \mathcal{L}_n$ the quantity

$$\Phi(w) := \sup_{x\in[w]} S_n\varphi(x),$$

where

$$S_n\varphi(x) = \sum_{k=0}^{n-1} \varphi(\sigma^k x).$$

Given $\mathcal{D} \subset \mathcal{L}$, the nth **partition sum** associated to $\mathcal{D}$ and φ is

$$\Lambda_n(\mathcal{D},\varphi) := \sum_{w\in\mathcal{D}_n} e^{\Phi(w)}.$$

The **pressure** of $\mathcal{D}$ with respect to φ is

$$P(\mathcal{D},\varphi) := \overline{\lim_{n\to\infty}} \frac{1}{n} \log \Lambda_n(\mathcal{D},\varphi).$$

In the specific case $\varphi = 0$, this reduces to the **entropy** of $\mathcal{D}$:

$$h(\mathcal{D}) := \overline{\lim_{n\to\infty}} \frac{1}{n} \log \#\mathcal{D}_n.$$

When $\mathcal{D} = \mathcal{L}(X)$, we write $P(X,\varphi) = P(\mathcal{L}(X),\varphi)$. Let $\mathcal{M}_\sigma(X)$ denote the set of σ-invariant Borel probability measures on X. The **variational principle** [Wal82, Theorem 9.10] says that

$$P(X,\varphi) = \sup\left\{h_\mu(\sigma) + \int \varphi\,d\mu \mid \mu \in \mathcal{M}_\sigma(X)\right\}.$$

A measure achieving this supremum is called an **equilibrium state**.

Write

$$I(\varphi) = \left\{\int \varphi\,d\mu : \mu \in \mathcal{M}_\sigma(X)\right\}.$$

Following [IRRL12], we call a potential function **hyperbolic** if it satisfies (1.1); that is, if $P(X,\varphi) > \sup I$. Given $\varepsilon > 0$, there is $n \in \mathbb{N}$ such that

$$\frac{1}{n} S_n\varphi(x) < \sup I + \varepsilon$$

for all $x \in X$; consequently, φ is hyperbolic if and only if there is $n \in \mathbb{N}$ such that

$$P(X, \varphi) > \sup_{x \in X} \frac{1}{n} S_n \varphi(x). \tag{2.1}$$

Equivalently, one may observe that φ and $\frac{1}{n} S_n \varphi(x)$ are cohomologous,[2] and so φ is hyperbolic if and only if there is a potential ψ cohomologous to φ such that

$$P(X, \varphi) = P(X, \psi) > \sup_{x \in X} \psi(x). \tag{2.2}$$

2.3. Specification, decompositions, and uniqueness. Following the definition in [CTY17, Cli18], say that $\mathcal{G} \subset \mathcal{L}$ has **specification** if there is $\tau > 0$ such that for every $v, w \in \mathcal{G}$ there exists $u \in \mathcal{L}$ with length $|u| \leq \tau$ such that $v'uw' \in \mathcal{G}$ whenever v' is a suffix of v and w' is a prefix of w with $v', w' \in \mathcal{G}$. This is a version of a condition that appeared in [CT12, CT13] and generalises the classical specification property of Bowen [Bow75], which corresponds roughly to this definition with $\mathcal{G} = \mathcal{L}$.

If $\mathcal{G}$ has specification with $\tau = 0$, then we have $vw \in \mathcal{G}$ whenever $v, w \in \mathcal{G}$, and in this case we say that $\mathcal{G}$ has the **free concatenation property**.

When $\mathcal{L}$ has specification, it was proved by Bertrand [Ber88] that $\mathcal{L}$ contains a **sychronising word**; that is, a word $s \in \mathcal{L}$ with the property that if $vs \in \mathcal{L}$ and $sw \in \mathcal{L}$, then $vsw \in \mathcal{L}$. In this case the collection $\{sw : sws \in \mathcal{L}\}$ has the free concatenation property. The following generalisation of this fact was proved in [Cli18, Proposition 7.3 and §7.1.2].

Proposition 2.1: *If $\mathcal{G} \subset \mathcal{L}$ has specification, then there is a collection $\mathcal{F} \subset \mathcal{L}$ and a number $N \in \mathbb{N}$ such that*

1. *$\mathcal{F}$ has the free concatenation property, and*
2. *given any $w \in \mathcal{G}$, there are $u, v \in \mathcal{L}$ with $|u|, |v| \leq N$ and $uwv \in \mathcal{F}$.*

See [Cli18] for a more explicit description of the collection $\mathcal{F}$; all we will need are the properties listed above. Writing

$$d = \gcd\{|v| : v \in \mathcal{F}\},$$

[2] Put $\xi(x) = \frac{1}{n} \sum_{k=0}^{n-1} (n-k) \varphi(\sigma_k x)$, then $\xi(x) - \xi(\sigma x) = \frac{1}{n} S_n \varphi(x) - \varphi(x)$.

it follows from the free concatenation property that we can choose $N \in \mathbb{N}$ large enough that $\mathcal{F}_n \neq \emptyset$ whenever $n \geq N$ is a multiple of d. Thus Proposition 2.1 has the following consequence.

Corollary 2.2: *Given $\mathcal{G}, \mathcal{F}$ as in Proposition 2.1 and d as in the previous paragraph, there is $N \in \mathbb{N}$ such that given any $w \in \mathcal{G}$ and any $n \geq |w| + 2N$ that is a multiple of d, there are $u, v \in \mathcal{L}$ with $|u| < N$, $uwv \in \mathcal{F}$, and $|uwv| = n$.*

Proof. Let N_0 be given by Proposition 2.1, and N_1 by the previous paragraph; then choose N large enough that $N > N_0$ and $N - 2N_0 \geq N_1$. Given any $w \in \mathcal{G}$, Proposition 2.1 gives $u, v' \in \mathcal{L}$ with $|u|, |v'| \leq N_0 < N$ such that $uwv' \in \mathcal{F}$. Let $n \geq |w| + 2N$ be a multiple of d. By definition, $|uwv'|$ is a multiple of d, and thus $n - |uwv'|$ is also a multiple of d. Moreover,

$$n - |uwv'| \geq (|w| + N) - |w| - 2N_0 \geq N_1,$$

so there is $v'' \in \mathcal{F}$ with $|v''| = n - |uwv'|$, hence $uwv'v'' \in \mathcal{F}$ and $|uwv'v''| = n$. ∎

If $\mathcal{G}$ is 'large enough', then specification for $\mathcal{G}$ can be used to deduce uniqueness of the equilibrium state. More precisely, a **decomposition** of $\mathcal{L}$ is a choice of $\mathcal{C}^p, \mathcal{G}, \mathcal{C}^s \subset \mathcal{L}$ such that for every $w \in \mathcal{L}$ there are $u^p \in \mathcal{C}^p$, $v \in \mathcal{G}$, and $u^s \in \mathcal{C}^s$ with $w = u^p v u^s$.

Theorem 2.3 ([Cli18], Theorem 1.1): *Suppose that $\mathcal{G}$ has specification and is closed under intersections and unions in the following sense: if $u, v, w \in \mathcal{L}$ are such that $uvw \in \mathcal{L}$, $uv \in \mathcal{G}$, and $vw \in \mathcal{G}$, then we have $v, uvw \in \mathcal{G}$. Let φ be a Hölder potential and $\mathcal{C}^p\mathcal{G}\mathcal{C}^s$ a decomposition of $\mathcal{L}$ with $P(\mathcal{C}^p \cup \mathcal{C}^s, \varphi) < P(\varphi)$. Then φ has a unique equilibrium state μ, and μ has exponential decay of correlations (up to a finite period) and satisfies the central limit theorem for Hölder observables.*

One can also use the results of [CT13] to deduce uniqueness (but not the statistical properties) under extremely similar hypotheses.

Remark 2.4: For β-shifts and their factors, one can find a decomposition with $h(\mathcal{C}^p \cup \mathcal{C}^s) = 0$, and then the pressure gap condition in Theorem 2.3 can be verified by proving hyperbolicity of the potential function, since an easy argument shows that $P(\mathcal{D}, \varphi) \leq h(\mathcal{D}) + \sup_\mu \int \varphi \, d\mu$ for every $\mathcal{D} \subset \mathcal{L}$.

2.4. Hamming approachability and asymptotic estimates. Given a function $g\colon \mathbb{N} \to \mathbb{N}$, we say that $\mathcal{L}$ is **g-Hamming approachable** by $\mathcal{G} \subset \mathcal{L}$ if there is $n_0 \in \mathbb{N}$ such that for every $n \geq n_0$ and $w \in \mathcal{L}_n$, there is $v \in \mathcal{G}_n$ with

$$d_{\mathrm{Ham}}(v,w) := \#\{1 \leq i \leq |w| : v_i \neq w_i\} \leq g(|w|). \tag{2.3}$$

This follows [CTY17, Definition 2.10], with the difference that we include the function g in the notation, and will ultimately require that g be sublogarithmic, not just sublinear. We assume without loss of generality that g is non-decreasing.

We will also need to use the fact that for any $k \leq m \in \mathbb{N}$ and any $w \in \mathcal{L}_m$, we have

$$\#\{v \in \mathcal{L}_m : d_{\mathrm{Ham}}(v,w) \leq k\} \leq \binom{m}{k}(\#A)^k. \tag{2.4}$$

This becomes more useful with an estimate for $\binom{m}{k}$. Recall from Stirling's formula that $\log(n!) = n\log n - n + O(\log n)$, and thus

$$\begin{aligned}\log\binom{m}{k} =&(m\log m - m) - (k\log k - k)\\ &- ((m-k)\log(m-k) - (m-k)) + O(\log m)\\ =&k\log\frac{m}{k} + (m-k)\log\frac{m}{m-k} + O(\log m).\end{aligned}$$

Writing

$$h(t) = -t\log t - (1-t)\log(1-t)$$

for the bipartite entropy function, this gives

$$\log\binom{m}{k} = h\Big(\frac{k}{m}\Big)m + O(\log m), \tag{2.5}$$

and so there is a constant Q such that (2.4) gives

$$\#\{v \in \mathcal{L}_m : d_{\mathrm{Ham}}(v,w) \leq k\} \leq e^{mh(k/m)}m^Q(\#A)^k. \tag{2.6}$$

Lemma 2.5: *Suppose $\mathcal{D} \subset \mathcal{L}$ has $h(\mathcal{D}) > 0$, and let $\beta > 0$ be small enough that $h(\beta) + \beta\log(\#A) < h(\mathcal{D})$. Then for every $N \in \mathbb{N}$ there are arbitrarily large $m \in \mathbb{N}$ with the following property: given any $w_1, \dots, w_N \in \mathcal{D}_m$, there is $v \in \mathcal{D}_m$ with $d_{\mathrm{Ham}}(v, w_i) > \beta m$ for all $1 \leq i \leq N$.*

Proof. Choose $\eta, \xi > 0$ such that

$$h(\beta) + \beta\log(\#A) + \xi < \eta < h(\mathcal{D}).$$

Given $m \in \mathbb{N}$ and $w_1, \dots, w_N \in \mathcal{D}_m$, (2.6) gives

$$\# \bigcup_{i=1}^{N} \{v \in \mathcal{L}_m : d_{\mathrm{Ham}}(v, w_i) \le \beta m\} \le N e^{mh(\beta)} m^Q (\#A)^{\beta m} < N m^Q e^{(\eta - \xi)m}.$$

The right-hand side is $< \#\mathcal{D}_m$ whenever $N m^Q < e^{m\xi}$ and $\#\mathcal{D}_m \ge e^{m\eta}$; this happens infinitely often. ∎

2.5. Coded systems. Given a finite alphabet A and a collection of words $G \subset A^*$, write G^* for the set of all finite concatenations of words in G. The **coded shift** generated by G is the subshift X over the alphabet A whose language consists of all subwords of elements of G^*. We refer to G as a **generating set** for X. The generating set is said to be **uniquely decipherable** if whenever $u^1 u^2 \cdots u^m = v^1 v^2 \cdots v^n$ with $u^i, v^j \in G$, we have $m = n$ and $u^j = v^j$ for all j [LM95, Definition 8.1.21].

Theorem 2.6: [Cli18, Theorem 1.8] *Let X be a coded shift on a finite alphabet and φ a Hölder potential on X. If X has a uniquely decipherable generating set G such that*

$$\mathcal{D} = \mathcal{D}(G) := \{w \in \mathcal{L} : w \text{ is a subword of some } g \in G\}$$

satisfies $P(\mathcal{D}, \varphi) < P(\varphi)$, then φ has a unique equlibrium state μ, and μ has exponential decay of correlations (up to a finite period) and satisfies the central limit theorem for Hölder observables.

3. Proof of Theorem 1.1

In §3.1 we establish some preliminary results that are needed in order to describe precisely (in §3.2) the mechanism by which we generate entropy.

3.1. Preliminaries for the proof. We start with the following consequence of Corollary 2.2.

Lemma 3.1: *Under the hypotheses of Theorem 1.1, there are $N \in \mathbb{N}$ and $\mathcal{F} \subset \mathcal{L}$ with the free concatenation property such that writing $d = \gcd\{|v| : v \in \mathcal{F}\}$, the following is true: for every $w \in \mathcal{L}$ such that $|w| \ge 2N$ and $|w|$ is a multiple of d, there is some $w' \in \mathcal{F}$ such that $|w| = |w'|$ and*

$$d_{\mathrm{Ham}}(w_{[1,|w|-i]}, w'_{(i,|w'|]}) \le g(|w|) + 2N \quad \text{for some } 0 \le i \le N-1. \tag{3.1}$$

Proof. Let $\mathcal{F}$ be as in Proposition 2.1 and N as in Corollary 2.2. Then $x = w_{[1,|w|-2N]}$ has $y \in \mathcal{G}_{|w|-2N}$ such that $d_{\mathrm{Ham}}(x,y) \le g(|w|-2N) \le g(|w|)$, where we use the fact that g is non-decreasing. Corollary 2.2 gives $u, v \in \mathcal{L}$ such that $|u| < N$, $uyv \in \mathcal{F}$ and $|uyv| = |w|$. Let $w' = uyv$ and $i = |u|$; then writing $w = xzz'$ where $|z'| = i$, we have

$$\begin{aligned} d_{\mathrm{Ham}}(w_{[1,|w|-i]}, w'_{(i,|w'|]}) &= d_{\mathrm{Ham}}(xz, yv) = d_{\mathrm{Ham}}(x,y) + d_{\mathrm{Ham}}(z,v) \\ &\le g(|w|) + |z| \le g(|w|) + 2N. \qquad \blacksquare \end{aligned}$$

Consider the map $\mathcal{L}_n \to \mathcal{F}_n$ given by $w \mapsto w'$ as in Lemma 3.1. By (2.6), the multiplicity of this map is at most $N e^{nh(\frac{g(n)+2N}{n-N})} n^Q (\#A)^{g(n)+2N}$. Writing c_n for this quantity we observe that $\#\mathcal{F}_n \ge (\#\mathcal{L}_n)/c_n$ whenever n is a multiple of d, and that $\lim_{n\to\infty} \frac{1}{n} \log c_n = 0$, so $h(\mathcal{F}) = h(\mathcal{L}) = h_{\mathrm{top}}(X) > 0$. Thus we can take $\beta > 0$ small enough that $h(\beta) + \beta \log(\#A) < h(\mathcal{F})$, and fix some $m \ge \max(3N, n_0)$ such that the conclusion of Lemma 2.5 holds, where n_0 is as in the paragraph preceding (2.3). Note that m must be a multiple of $d = \gcd\{|v| : v \in \mathcal{F}\}$.

Now we fix several more parameters that will be used in the proof. First we will find $V > 0$ that controls $|\Phi(v) - \Phi(w)|$ in terms of $d_{\mathrm{Ham}}(v,w)$; then we will choose $\gamma > 0$ small relative to m, V; then we choose a large $L > 0$ that helps us control $\sum_i g(n_i)$; and finally we will choose $\delta > 0$ small enough that a certain entropy estimate later on is positive.

Let $\alpha > 0$ be the Hölder exponent of φ, and write

$$|\varphi|_\alpha = \sup_{x \ne y} \frac{|\varphi(x) - \varphi(y)|}{d(x,y)^\alpha}.$$

Then for every $n \in \mathbb{N}$, $w \in \mathcal{L}_n$, and $x, y \in [w]$, we have

$$|S_n\varphi(x) - S_n\varphi(y)| \le \sum_{k=0}^{n-1} |\varphi(\sigma^k x) - \varphi(\sigma^k y)| \le \sum_{k=0}^{n-1} |\varphi|_\alpha 2^{-(n-k)\alpha} < \frac{|\varphi|_\alpha}{1 - 2^{-\alpha}}.$$

In particular, writing $V := |\varphi|_\alpha (1 - 2^{-\alpha})^{-1}$, we have

$$|S_n\varphi(x) - \Phi(w)| \le V \quad \text{for all } n \in \mathbb{N}, w \in \mathcal{L}_n, \text{ and } x \in [w]. \tag{3.2}$$

This has the corollary that for every $v, w \in \mathcal{L}$ with $|v| = |w|$, we have

$$|\Phi(v) - \Phi(w)| \le V d_{\mathrm{Ham}}(v,w). \tag{3.3}$$

Lemma 3.2: *For every $\gamma > 0$ there is $L > 0$ such that for every $n_1, \dots, n_\ell \in \mathbb{N}$ we have*

$$\sum_{i=1}^{\ell} g(n_i) \le \ell\Big(L + \gamma \log \frac{\sum n_i}{\ell}\Big). \tag{3.4}$$

Proof. Since $g(n)/\log n \to 0$, there exists $K \in \mathbb{N}$ such that

$$g(n) < \gamma \log(n) \quad \text{for all } n > K. \tag{3.5}$$

Let $L := \max\{g(n) : 1 \le n \le K\}$. Then we have the following estimate: given any $n > K$, $\ell \in \mathbb{N}$, and $n_1, \dots, n_\ell \in \mathbb{N}$ such that $\sum_{i=1}^{\ell} n_i = n$, we have

$$\begin{aligned} \sum_{i=1}^{\ell} g(n_i) &\le \sum_{\{i:n_i \le K\}} g(n_i) + \sum_{\{i:n_i > K\}} g(n_i) \\ &\le L\#\{i : n_i \le K\} + \sum_{\{i:n_i>K\}} \gamma \log n_i \\ &\le L\ell + \gamma\ell \log(n/\ell) = \ell(L + \gamma \log(n/\ell)). \end{aligned} \tag{3.6}$$

The last inequality uses convexity; the function $(x_1, \dots, x_\ell) \mapsto \sum_i \log x_i$ is maximized (subject to the constraint $\sum x_i = n$) when $x_1 = \cdots = x_\ell = n/\ell$, for which values we have $\sum_i \log x_i = \ell \log(n/\ell)$. ∎

For the duration of the proof, we fix $0 < \gamma < (16m^2V)^{-1}$, and let L be given by Lemma 3.2. Without loss of generality, we assume that $L \ge 2m$. Finally, with V, β, m, γ, L fixed, we choose $\delta > 0$ small enough that

$$\frac{|\log \delta|}{8m^2} > 2 \log\Big(\frac{2L + \gamma|\log \delta|}{\beta m}\Big) + 4VL. \tag{3.7}$$

3.2. Construction of nearby words. To prove hyperbolicity of φ it suffices to show that for every $x \in X$, we have

$$P(\varphi) > \varlimsup_{n\to\infty} \frac{1}{n} S_n\varphi(x).$$

To this end, we take $w \in \mathcal{L}$ to be a (sufficiently long) word, and estimate $\Lambda_{|w|}(\mathcal{L}, \varphi)$ in terms of $e^{\Phi(w)}$.

Let $m \in \mathbb{N}$ be as above. Given $n \gg m$ with $(2m)|n$, fix $k_n \in [\delta n, 2\delta n] \cap \mathbb{N}$, and let

$$\mathcal{J}_n = \Big\{\mathbf{n} = (n_1, \dots, n_{k_n}) : \sum n_i = n \text{ and } (2m)|n_i \text{ for all } i\Big\}.$$

Given $\mathbf{n} \in \mathcal{J}_n$, let $N_j = n_1 + n_2 + \cdots + n_{j-1}$ be the partial sums. For a fixed $w \in \mathcal{L}_n$, we will associate to each $\mathbf{n} \in \mathcal{J}_n$ a word $\psi(\mathbf{n}) \in \mathcal{L}_n$ such that

(1) $\psi(\mathbf{n})$ is Hamming-close to w on the intervals $(N_i, N_{i+1} - m]$;
(2) $\psi(\mathbf{n})$ is Hamming-far from w on the intervals $(N_i - m, N_i]$.

This will allow us to decipher $\mathbf{n}$ from $\psi(\mathbf{n})$ up to some (controllable) error; that is, we will be able to control the multiplicity of the map $\psi\colon \mathcal{J}_n \to \mathcal{L}_n$. Moreover, each $\psi(\mathbf{n})$ will have ergodic sum $\Phi(\psi(\mathbf{n}))$ that is close to $\Phi(w)$. These two facts, together with an estimate on $\#\mathcal{J}_n$, will give us the desired lower bound on $\Lambda_n(\mathcal{L}, \varphi)$.

Let us make this more precise. Given $\mathbf{n}$, we have $n_i \geq 2m \geq m + 2N$ for all i, and so applying Lemma 3.1 to $w_{(N_i, N_{i+1}-m]} \in \mathcal{L}_{n_i - m}$ gives $v^i \in \mathcal{F}_{n_i - m}$ such that

$$d_{\mathrm{Ham}}(w_{(N_i, N_{i+1}-m-a_i]}, v^i_{(a_i, n_i - m]}) \leq g(n_i) + 2N \quad \text{for some } 0 \leq a_i < N. \tag{3.8}$$

Consequently, we have

$$d_{\mathrm{Ham}}(v^i, w_{(N_i - a_i, N_{i+1} - m - a_i])}) \leq g(n_i) + 3N \leq g(n_i) + m. \tag{3.9}$$

Moreover, by Lemma 2.5 there are words $s^i \in \mathcal{F}_m$ such that

$$d_{\mathrm{Ham}}(s^i, w_{(N_i - m - a, N_i - a]}) \geq \beta m \quad \text{for all } 1 \leq a \leq N. \tag{3.10}$$

Now we can define the map $\psi = \psi_w \colon \mathcal{J}_n \to \mathcal{L}_n$ by

$$\psi(\mathbf{n}) = v^1 s^1 v^2 s^2 \cdots v^{k_n} s^{k_n}. \tag{3.11}$$

Summing over all $\mathbf{n} \in \mathcal{J}_n$ gives

$$\log \Lambda_n(\mathcal{L}, \varphi) \geq \Phi(w) + \log \#\mathcal{J}_n - \max_{\mathbf{n} \in \mathcal{J}_n} |\Phi(\psi(\mathbf{n})) - \Phi(w)| - \max_{u \in \mathcal{L}_n} \#\psi^{-1}(u).$$

If we divide both sides by n, send $n \to \infty$, and write

$$\begin{aligned} h_{\mathcal{J}} &:= \varliminf_{n\to\infty} \frac{1}{n} \log \#\mathcal{J}_n, \\ \Delta_\Phi &:= \varlimsup_{n\to\infty} \frac{1}{n} \max_{w \in \mathcal{L}_n} \max_{\mathbf{n} \in \mathcal{J}_n} |\Phi(\psi_w(\mathbf{n})) - \Phi(w)|, \\ h_\psi &:= \varlimsup_{n\to\infty} \frac{1}{n} \max_{w \in \mathcal{L}_n} \max_{u \in \mathcal{L}_n} \#\psi_w^{-1}(u), \end{aligned}$$

we get

$$P(\varphi) \geq \sup I + h_{\mathcal{J}} - \Delta_\Phi - h_\psi, \tag{3.12}$$

where we recall that

$$I = \left\{ \int \varphi \, d\mu : \mu \in \mathcal{M}_\sigma(X) \right\} = \Big[\inf_{x \in X} \varliminf_{n\to\infty} \frac{1}{n} S_n\varphi(x), \sup_{x\in X} \varlimsup_{n\to\infty} \frac{1}{n} S_n\varphi(x) \Big].$$

To complete the proof of Theorem 1.1, it suffices to show that $h_{\mathcal{J}} > \Delta_\Phi + h_\psi$, which we do in the next section.

3.3. Estimates on errors and entropy.

3.3.1. *Entropy gained from $\mathcal{J}$.* Using (2.5) and the definition of $\mathcal{J}_n$, we have

$$\log \#\mathcal{J}_n = \log \binom{\frac{n}{2m}}{k_n} \geq h\Big(\frac{\delta}{2m}\Big)\frac{n}{2m} + O(\log n),$$

and thus

$$h_{\mathcal{J}} \geq \frac{\delta}{4m^2} \Big| \log \frac{\delta}{2m} \Big| \geq \frac{\delta}{4m^2} |\log \delta|. \tag{3.13}$$

3.3.2. *Errors in ergodic sums.* Given any $w \in \mathcal{L}_n$ and $\mathbf{n} \in \mathcal{J}_n$, with v^i as in the definition of ψ we see from (3.3) and (3.8) that

$$|\Phi(w_{(N_i, N_{i+1}-m]}) - \Phi(v^i)| \leq (g(n_i) + 3N)V \leq (g(n_i)+m)V,$$

and hence $|\Phi(w_{(N_i,N_{i+1}]} - \Phi(v^i s^i)| \leq (g(n_i)+2m)V$. Summing over all i and using Lemma 3.2 gives

$$|\Phi(\psi(\mathbf{n})) - \Phi(w)| \leq \sum_{i=1}^{k_n} (g(n_i)+2m)V \leq k_n(L + 2m + \gamma \log(n/k_n))V,$$

and since $L \geq 2m$ we get

$$\max_{w\in\mathcal{L}_n} \max_{\mathbf{n}\in\mathcal{J}_n} |\Phi(\psi(\mathbf{n})) - \Phi(w)| \leq k_n(2L + \gamma\log(n/k_n))V. \tag{3.14}$$

Dividing by n and using $k_n \in [\delta n, 2\delta n]$ gives

$$\Delta_\Phi \leq 2\delta V(2L + \gamma|\log\delta|). \tag{3.15}$$

3.3.3. *Multiplicity of ψ.* Given $u \in \mathcal{L}_n$, let

$$R_u = \{j \in [1,n] : m|j \text{ and } d_{\mathrm{Ham}}(u_{[j,j+m)}, w_{[j-a,j+m-a)}) \geq \beta m \text{ for all } 0 \leq a < N\}.$$

It follows from (3.10) that $\{N_i\}_{i=1}^{k_n} \subset R_{\psi(\mathbf{n})}$ for all $\mathbf{n} \in \mathcal{J}_n$. Moreover, given $\mathbf{n} \in \mathcal{J}_n$ we see from (3.9) that $u = \psi(\mathbf{n})$ has

$$\sum_{j=N_i/m}^{(N_{i+1}/m)-1} d_{\mathrm{Ham}}(u_{(jm,(j+1)m]}, w_{(jm-a_i,(j+1)m-a_i]}) \leq g(n_i) + 2m \tag{3.16}$$

for every $1 \leq i \leq n_k$, and summing over i gives

$$\begin{aligned} \beta m \cdot \#R_u &\leq \sum_{j=1}^{n/m} \min_{0 \leq a < N} d_{\mathrm{Ham}}(u_{[jm,jm+m)}, w_{[jm-a,jm+m-a)}) \\ &\leq \sum_{i=1}^{k_n} (g(n_i) + 2m) \leq k_n(2L + \gamma|\log\delta|), \end{aligned} \tag{3.17}$$

where the last inequality again uses Lemma 3.2 and the inequalities $L \geq 2m$, $k_n \geq \delta n$. Thus we have

$$\#R_u \leq k_n \cdot \frac{2L + \gamma|\log\delta|}{\beta m},$$

and since $\mathbf{n} \in \mathcal{J}_n$ is determined by a choice of k_n elements from R_u, we conclude from (2.5) that

$$\log \#\psi^{-1}(u) \leq h\Big(\frac{\beta m}{2L + \gamma|\log\delta|}\Big)\frac{2\delta n}{\beta m}(2L + \gamma|\log\delta|) + O(\log n),$$

and so

$$\begin{aligned} h_\psi &\leq \frac{\beta m}{2L + \gamma|\log\delta|} \log\Big(\frac{2L + \gamma|\log\delta|}{\beta m}\Big)\frac{2\delta}{\beta m}(2L + \gamma|\log\delta|) \\ &= 2\delta \log\Big(\frac{2L + \gamma|\log\delta|}{\beta m}\Big). \end{aligned} \tag{3.18}$$

3.3.4. *Completion of the proof.* Combining (3.13), (3.15), and (3.18), we get

$$\frac{h_{\mathcal{J}} - \Delta_\Phi - h_\psi}{\delta} \geq \frac{|\log\delta|}{4m^2} - 4VL - 2V\gamma|\log\delta| - 2\log\Big(\frac{2L + \gamma|\log\delta|}{\beta m}\Big).$$

Since we chose γ to be smaller than $(16m^2V)^{-1}$, we have

$$\frac{|\log\delta|}{8m^2} - 2V\gamma|\log\delta| > 0,$$

and thus

$$\frac{h_{\mathcal{J}} - \Delta_\Phi - h_\psi}{\delta} > \frac{|\log \delta|}{8m^2} - 4VL - 2\log\Big(\frac{2L + \gamma|\log\delta|}{\beta m}\Big).$$

The right-hand side is positive by our choice of δ in (3.7), and we conclude that $h_{\mathcal{J}} > \Delta_\Phi + h_\psi$. By (3.12), this gives $P(\varphi) > \sup I$, which completes the proof of Theorem 1.1.

4. Proof of Theorem 1.6

Now we consider the shift space X described in Theorem 1.6. Write $\mathcal{L}$ for the language of X and $f\colon \mathbb{N} \to \mathbb{N}$ for the function used to define

$$G = \{0^a 1^b : a, b \geq f(a+b)\}.$$

Recall that $\varphi = -\mathbf{1}_{[1]}$. Before we prove the five statements listed in the theorem, we demonstrate that $P(t\varphi)$ is non-negative and non-increasing. Let δ_0 be the δ-measure on the fixed point $0 \in X$. Then for every $t \in \mathbb{R}$ we have

$$P(t\varphi) \geq h_{\delta_0}(\sigma) + t\int \varphi\, d\delta_0 = t\varphi(0) = 0.$$

Since $\varphi \leq 0$ it follows from basic properties of pressure that whenever $s < t$, we have

$$P(t\varphi) = P(s\varphi + (t-s)\varphi) \leq P(s\varphi + (t-s)0) = P(s\varphi),$$

so the pressure function is non-increasing.

4.1. Hamming approachability. Let n_1 be such that $f(n) \leq n/2$ for all $n \geq n_1$; in particular, for all $n \geq n_1$ there are $a, b \geq f(n)$ such that $a + b = n$, and thus $0^a 1^b \in G$. We need the following lemma.

Lemma 4.1: *Given $n \geq n_1$ and $w \in \mathcal{L}_n$, suppose that w can be written as $w = u0^a1^bv$ for some $u, v \in \mathcal{L}$ with $|u|, |v| \leq n_1$ and $a, b \geq 0$. (Note that u, v are allowed to be empty.) Then there is $\tilde{w} \in G$ such that*

$$d_{\mathrm{Ham}}(w, \tilde{w}) \leq n_1 + \max(f(n), n_1).$$

Proof. If $n_1 + a < f(n)$, then $\tilde{w} = 0^{f(n)}1^{n-f(n)} \in G$ satisfies

$$\begin{aligned} d_{\mathrm{Ham}}(w, \tilde{w}) \leq & d_{\mathrm{Ham}}(w_{[1,f(n)]}, 0^{f(n)}) + d_{\mathrm{Ham}}(w_{(f(n),n]}, 1^{n-f(n)}) \\ = & d_{\mathrm{Ham}}(u0^a1^{f(n)-a-|u|}, 0^{f(n)}) + d_{\mathrm{Ham}}(1^{n-f(n)-|v|}v, 1^{n-f(n)}) \\ \leq & f(n) + n_1. \end{aligned}$$

Similarly, if $n_1 + a > n - f(n)$, then $\tilde{w} = 0^{n-f(n)}1^{f(n)} \in G$ satisfies

$$d_{\mathrm{Ham}}(w, \tilde{w}) \leq n_1 + f(n).$$

Finally, if $f(n) \leq n_1 + a \leq n - f(n)$, then $\tilde{w} = 0^{n_1+a}1^{n-n_1-a}$ satisfies

$$d_{\mathrm{Ham}}(w, \tilde{w}) \leq 2n_1. \qquad \blacksquare$$

Now given any $w \in \mathcal{L}$ with $|w| \geq 2n_1$, there are integers

$$0 = \ell_0 < \ell_1 < \cdots < \ell_m = n$$

such that

$$\begin{aligned} w_{(\ell_{i-1},\ell_i]} &= 0^{a_i}1^{b_i} \quad \text{for all } 1 \leq i \leq m, \\ a_i, b_i &\geq f(a_i + b_i) \quad \text{for all } 1 < i < m, \\ a_1, b_1, a_m, b_m &\geq 0. \end{aligned}$$

Choose $0 \leq j \leq k \leq m$ such that

$$n_1 \in (\ell_{j-1}, \ell_j] \quad \text{and} \quad n - n_1 \in (\ell_{k-1}, \ell_k].$$

If $j = k$ then w has the form required for Lemma 4.1, and thus there is $\tilde{w} \in G$ such that $d_{\mathrm{Ham}}(w, \tilde{w}) \leq n_1 + \max(f(n), n_1)$. If $j < k$, then we can write $w = w^p w^c w^s$, where

$$w^p := w_{(0,\ell_{j+1}]}, \quad w^c := w_{(\ell_{j+1},\ell_k]}, \quad w^s := w_{(\ell_k,n]}.$$

Note that $w^c \in \mathcal{F}$, and w^p, w^s both have the form required for Lemma 4.1, so taking $\tilde{w}^p$ and $\tilde{w}^s$ as given by that lemma, we have $\tilde{w}^p w^c \tilde{w}^s \in G^*$ and

$$d_{\mathrm{Ham}}(w, \tilde{w}^p w^c \tilde{w}^s) \leq d_{\mathrm{Ham}}(w^p, \tilde{w}^p) + d_{\mathrm{Ham}}(w^s, \tilde{w}^s) \leq 2n_1 + 2\max(f(n), n_1).$$

This proves the first item in Theorem 1.6.

4.2. Hyperbolicity when $P(t\varphi) > 0$. Let $I_t = \{\int t\varphi \, d\mu : \mu \in \mathcal{M}_\sigma(X)\}$. The second statement in Theorem 1.6 is equivalent to the claim that when $t \geq 0$, we have $P(t\varphi) > \sup I_t$ if and only if $t < t_0$, where t_0 is the first root of Bowen's equation (1.2). Since $t \mapsto P(t\varphi)$ is non-increasing, we see that $t < t_0$ if and only if $P(t\varphi) > 0$. On the other hand, since $\int \varphi \, \delta_1 = -1 \leq \varphi \leq 0 = \int \varphi \, \delta_0$, we have $I_t = [-t, 0]$ for all $t \geq 0$, and so $\sup I_t = 0$, which proves the desired equivalence.

4.3. Unique equilibrium state when $t < t_0$. To deduce uniqueness of the equilibrium state for $t\varphi$ when $0 \le t < t_0$, we apply Theorem 2.6. (Positive entropy of the equilibrium state will then follow since $t\varphi$ is hyperbolic.) The shift X is coded with generating set $G = \{0^a 1^b : a, b \ge f(a+b)\}$. This is uniquely decipherable because if $w = u^1 u^2 \cdots u^m$ with $u^i \in G$, then we can recover u^1 from w as the longest initial segment of the form $0^a 1^b$ with $a, b \ge 1$, then u^2 from the remainder of w by the same procedure, and so on. Moreover, the set

$$\mathcal{D} = \mathcal{D}(G) := \{w \in \mathcal{L} : w \text{ is a subword of some } g \in G\}$$

is easily seen to satisfy $\mathcal{D} \subset \{0^a 1^b : a, b \ge 0\}$, and hence $\#\mathcal{D}_n \le n+1$, so $h(\mathcal{D}) = 0$. We conclude that

$$P(\mathcal{D}, t\varphi) \le h(\mathcal{D}) + \sup I_t = \sup I_t \quad \text{for all } t,$$

and since we showed that $t\varphi$ is hyperbolic whenever $0 \le t < t_0$, we conclude that $P(\mathcal{D}, t\varphi) < P(t\varphi)$ for this range of t, and so we can apply Theorem 2.6.

4.4. Only the delta measure past t_0. Since $t \mapsto P(t\varphi)$ is non-increasing and nonnegative, we have $P(t\varphi) = P(t_0\varphi) = 0$ for all $t \ge t_0$. Thus δ_0 is an equilibrium state for all $t \ge t_0$. When $t > t_0$, we observe that every other $\mu \in \mathcal{M}_\sigma(X)$ has $\mu[1] > 0$ and hence $\int \varphi \, d\mu < 0$, so

$$\begin{aligned} h_\mu(\sigma) + \int t\varphi \, d\mu &= h_\mu(\sigma) + \int t_0 \varphi \, d\mu + \int (t - t_0)\varphi \, d\mu \\ &\le P(t_0\varphi) + (t - t_0) \int \varphi \, d\mu < 0, \end{aligned}$$

which shows that δ_0 is the unique equilibrium state on this range of t.

4.5. Bowen's equation has a root if and only if $\sum \gamma^{f(n)} < \infty$. For the final statement in Theorem 1.6, we fix $t > 0$ and study the power series

$$F(x) := \sum_{n=1}^{\infty} \Lambda_n(G, t\varphi) x^n \quad \text{and} \quad H(x) := 1 + \sum_{n=1}^{\infty} \Lambda_n(G^*, t\varphi) x^n.$$

Proposition 4.2: *For the shift space in Theorem 1.6 and $t > 0$, the following are equivalent:*

(a) $P(t\varphi) = 0$.
(b) *The power series $H(x)$ converges for every $0 \le x < 1$.*
(c) *The power series $F(x)$ converges for every $0 \le x < 1$, with $F(x) < 1$.*
(d) *The power series $F(x)$ converges for $x = 1$, with $F(1) \le 1$.*

Proof. (a)⇔(b). Consider the power series $A(x) = \sum_{n=0}^{\infty} \Lambda_n(X, t\varphi)x^n$ (here $\Lambda_0(X, t\varphi) = 1$). Since $\lim \sqrt[n]{\Lambda_n(X, t\varphi)} = e^{P(t\varphi)}$, the root test tells us that the radius of convergence of $A(x)$ is $e^{-P(t\varphi)} \le 1$. In particular, $P(t\varphi) = 0$ if and only if $A(x)$ converges for every $0 \le x < 1$, so to prove the first equivalence it suffices to show that the power series $A(x)$ and $H(x)$ converge for the same values of $x \in [0, 1)$. To this end, consider the sets of words

$$\mathcal{P} = \{0^a 1^b : a < f(a+b)\} \quad \text{and} \quad \mathcal{S} = \{0^a 1^b : b < f(a+b)\}.$$

Every $w \in \mathcal{L}$ admits a unique decomposition as $w = u^p v u^s$ for some $u^p \in \mathcal{P}$, $v \in G^*$, and $u^s \in \mathcal{S}$, and since $\Phi(u^p v u^s) = \Phi(u^p) + \Phi(v) + \Phi(u^s)$, we have

$$\sum_{n=0}^{N} \Lambda_n(X, t\varphi)x^n = \sum_{\substack{a,b,c\ge 0\\ a+b+c\le N}} \Lambda_a(\mathcal{P}, t\varphi)x^a \Lambda_b(G^*, t\varphi)x^b \Lambda_c(\mathcal{S}, t\varphi)x^c. \tag{4.1}$$

Consider the power series associated to $\mathcal{P}$ and $\mathcal{S}$:

$$C^{\mathcal{P}}(x) := 1 + \sum_{n=1}^{\infty} \Lambda_n(\mathcal{P}, \varphi)x^n \quad \text{and} \quad C^{\mathcal{S}}(x) := 1 + \sum_{n=1}^{\infty} \Lambda_n(\mathcal{S}, \varphi)x^n.$$

Write $H_N, A_N, C_N^{\mathcal{P}}, C_N^{\mathcal{S}}$ for the partial sums (over $n \le N$) of the respective power series; then (4.1) gives

$$C_N^{\mathcal{P}}(x)H_N(x)C_N^{\mathcal{S}}(x) \le A_{3N}(x) \le C_{3N}^{\mathcal{P}}(x)H_{3N}(x)C_{3N}^{\mathcal{S}}(x). \tag{4.2}$$

We claim that $C^{\mathcal{P}}(x)$ and $C^{\mathcal{S}}(x)$ both converge for all $0 \le x < 1$. For $C^{\mathcal{S}}(x)$ we have

$$C^{\mathcal{S}}(x) = 1 + \sum_{n=1}^{\infty} \Big(\sum_{k=0}^{f(n)-1} e^{-tk} \Big) x^n = 1 + \sum_{n=1}^{\infty} \Big(\frac{1 - e^{-tf(n)}}{1 - e^{-t}} \Big) x^n,$$

which has radius of convergence $x = 1$ since the coefficients lie in the interval $(0, 1]$. Similarly for $C^{\mathcal{P}}(x)$, we have

$$C^{\mathcal{P}}(x) = 1 + \sum_{n=1}^{\infty} \Big(\sum_{k=0}^{f(n)-1} e^{-t(n-k)} \Big) x^n = 1 + \sum_{n=1}^{\infty} \Big(\frac{e^{-t(n-f(n))} - e^{-tn}}{e^t - 1} \Big) x^n,$$

and since $1 \le f(n) \le n/2$ for all sufficiently large n, the coefficients converge to 0 and the radius of convergence of $C^{\mathcal{P}}(x)$ is greater than or equal to $x = 1$. Thus $C^{\mathcal{P}}(x)$ and $C^{\mathcal{S}}(x)$ both converge for all $0 \le x < 1$, and it follows from (4.2) that for every such x, $H(x)$ converges if and only if $A(x)$ converges. This proves the equivalence of (a) and (b).

(b)⇔(c). Since X is uniquely decipherable we have

$$\Lambda_n(G^*, t\varphi) = \sum_{j=1}^{n} \sum_{n_1+\cdots+n_j=n} \prod_{i=1}^{j} \Lambda_{n_i}(G, t\varphi).$$

It follows that whenever $|F(x)| < 1$ we have

$$H(x) = 1 + \sum_{k=1}^{\infty} F(x)^k = \frac{1}{1 - F(x)} \tag{4.3}$$

and if $0 \le x < 1$ is such that $F(x) \ge 1$, then $H(x)$ does not converge.

(c)⇔(d). Suppose $F(1)$ converges. Then $F(x)$ converges for all $|x| < 1$ by standard facts on power series, and since all the coefficients are nonnegative (and not all of them vanish), the function F is strictly increasing on $[0, 1]$, so $0 \le F(x) < F(1)$ for all $x \in [0, 1)$, which proves (d)⇒(c).

Now we prove (c)⇒(d). Suppose that for all $0 \le x < 1$ we have $F(x) < 1$. Then the partial sums $F_N(x)$ also satisfy $F_N(x) < 1$ for all $x \in [0, 1)$ and $N \in \mathbb{N}$, since the coefficients are nonnegative. By continuity we get $F_N(1) \le 1$ for all $N \in \mathbb{N}$, and thus $F(1) \le 1$. ∎

By Proposition 4.2, in order to complete the proof of Theorem 1.6(v) it suffices to show that there is $t > 0$ with $F(1) \le 1$ if and only if there is $\gamma > 0$ such that $\sum_n \gamma^{f(n)} < \infty$. Observe that

$$\Lambda_n(G, t\varphi) = \sum_{k=f(n)}^{n-f(n)} e^{-tk} = \frac{e^{-t(f(n)-1)} - e^{-t(n-f(n))}}{e^t - 1} \tag{4.4}$$

whenever $f(n) \le n/2$, and $\Lambda_n(G, t\varphi) = 0$ otherwise. Since $f(n) \le n/2$ for all sufficiently large n, we have

$$\sum \frac{e^{-t(n-f(n))}}{e^t - 1} < \infty,$$

implying that $F(1) < \infty$ if and only if $\sum_{n=1}^{\infty} e^{-t(f(n)-1)}/(e^t - 1) < \infty$. In particular, if $F(1) \le 1$ then $\sum \gamma^{f(n)} < \infty$ for $\gamma = e^{-t}$.

For the converse direction, suppose that $\gamma > 0$ is such that $\sum \gamma^{f(n)} < \infty$. Then for all $t \ge -\log \gamma$, (4.4) gives

$$\sum_{n=1}^{\infty} \Lambda_n(G, t\varphi) \le \sum_{n=1}^{\infty} \frac{e^{-t(f(n)-1)}}{e^t - 1} \le \sum_{n=1}^{\infty} \frac{\gamma^{f(n)-1}}{e^t - 1} \le \frac{1}{\gamma(e^t - 1)} \sum_{n=1}^{\infty} \gamma^{f(n)}.$$

By taking t sufficiently large, the right-hand side can be made ≤ 1, so for this value of t we have $F(1) \leq 1$, which completes the proof of Theorem 1.6.

References

[Ber88] A. Bertrand, *Specification, synchronisation, average length*, in *Coding Theory and Applications (Cachan, 1986)*, Lecture Notes in Computer Science, Vol. 311, Springer, Berlin, 1988, pp. 86–95.

[BK90] V. Baladi and G. Keller, *Zeta functions and transfer operators for piecewise monotone transformations*, Communications in Mathematical Physics **127** (1990), 459–477.

[Bow75] R. Bowen, *Some systems with unique equilibrium states*, Mathematical Systems Theory **8** (1974/75), 193–202.

[Buz01] J. Buzzi, *Thermodynamical formalism for piecewise invertible maps: absolutely continuous invariant measures as equilibrium states*, in *Smooth Ergodic Theory and its Applications (Seattle, WA, 1999)*, Proceedings of Symposia in Pure Mathematics, Vol. 69, American Mathematical Society, Providence, RI, 2001, pp. 749–783.

[Buz04] J. Buzzi, *Entropy of equilibrium measures of continuous piecewise monotonic maps*, Stochastics and Dynamics **4** (2004), 84–94.

[CFT19] V. Climenhaga, T. Fisher and D. J. Thompson, *Equilibrium states for Mañé diffeomorphisms*, Ergodic Theory and Dynamical Systems, to appear, `https://doi.org/10.1017/etds.2017.125`.

[Cli18] V. Climenhaga, *Specification and towers in shift spaces*, Communications in Mathematical Physics **364** (2018), 441–504.

[Con] S. Conrad, *A coded shift with a Hölder potential that is not hyperbolic*, preprint.

[CT12] V. Climenhaga and D. J. Thompson, *Intrinsic ergodicity beyond specification: β-shifts, S-gap shifts, and their factors*, Israel Journal of Mathematics **192** (2012), 785–817.

[CT13] V. Climenhaga and D. J. Thompson, *Equilibrium states beyond specification and the Bowen property*, Journal of the London Mathematical Society **87** (2013), 401–427.

[CTY17] V. Climenhaga, D. J. Thompson and K. Yamamoto, *Large deviations for systems with non-uniform structure*, Transactions of the American Mathematical Society **369** (2017), 4167–4192.

[DKU90] M. Denker, G. Keller and M. Urbański, *On the uniqueness of equilibrium states for piecewise monotone mappings*, Studia Mathematica **97** (1990), 27–36.

[HK82] F. Hofbauer and G. Keller, *Equilibrium states for piecewise monotonic transformations*, Ergodic Theory and Dynamical Systems **2** (1982), 23–43.

[Hof79] F. Hofbauer, *On intrinsic ergodicity of piecewise monotonic transformations with positive entropy*, Israel Journal of Mathematics **34** (1979), 213–237.

[IRRL12] I. Inoquio-Renteria and J. Rivera-Letelier, *A characterization of hyperbolic potentials of rational maps*, Bulletin of the Brazilian Mathematical Society **43** (2012), 99–127.

[Kel84] G. Keller, *On the rate of convergence to equilibrium in one-dimensional systems*, Communications in Mathematical Physics **96** (1984), 181–193.

[LM95] D. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding*, Cambridge University Press, Cambridge, 1995.

[LRL14] H. Li and J. Rivera-Letelier, *Equilibrium states of weakly hyperbolic one-dimensional maps for Hölder potentials*, Communications in Mathematical Physics **328** (2014), 397–419.

[PS07] C.-E. Pfister and W. G. Sullivan, *On the topological entropy of saturated sets*, Ergodic Theory and Dynamical Systems **27** (2007), 929–956.

[Ryc83] M. Rychlik, *Bounded variation and invariant measures*, Studia Mathematica **76** (1983), 69–80.

[Tho12] D. J. Thompson, *Irregular sets, the β-transformation and the almost specification property*, Transactions of the American Mathematical Society **364** (2012), 5395–5414.

[Wal82] P. Walters, *An Introduction to Ergodic Theory*, Graduate Texts in Mathematics, Vol. 79, Springer, New York–Berlin, 1982.