CRITICAL HAUSDORFF MEASURE FOR HIGHER-DIMENSIONAL BEDFORD-MCMULLEN SUBSHIFTS

HUA QIU AND QI WANG

ABSTRACT. Let $\Lambda = \mathrm{diag}(m_1, \dots, m_d)$ be a diagonal expanding integer matrix with $m_1 \geq \cdots \geq m_d \geq 2$ and at least two distinct diagonal entries, and let $X \subset \mathcal{A}^{\mathbb{N}}$ be a subshift with weak specification over $\mathcal{A} = \prod_{r=1}^{d} \{0, 1, \dots, m_r - 1\}$. Let $K = R(X)$ be the corresponding self-affine set. Writing $\gamma = \dim_{\mathrm{H}} K$, we prove a sharp critical-measure dichotomy:

$$\dim_{\mathrm{H}} K = \dim_{\mathrm{B}} K \iff 0 < \mathcal{H}^{\gamma}(K) < \infty,$$
$$\dim_{\mathrm{H}} K < \dim_{\mathrm{B}} K \iff \mathcal{H}^{\gamma}(K) = \infty.$$

In the second case, $\mathcal{H}^{\gamma}$ is not σ-finite on K. This resolves Peres' 1994 conjecture for planar systems and extends to higher dimensions. For sofic subshifts, we prove the existence of the box dimension and show that dimension equality implies that the critical Hausdorff measure is positive and σ-finite, although it may be infinite. We also construct two examples with $\dim_{\mathrm{H}} K < \dim_{\mathrm{B}} K$: in one example the critical Hausdorff measure is positive and finite, whereas in the other it is infinite but σ-finite.

CONTENTS

1. Introduction . . . 2
2. Preliminaries and reduction of the main theorem . . . 4
2.1. Factor maps and weighted pressure . . . 4
2.2. A reduction of Theorem 1.1 . . . 7
3. Approximate cubes and the density estimates . . . 8
4. Partition of admissible words . . . 9
5. Proof of Proposition 2.2 . . . 14
5.1. Measure construction . . . 15
5.2. The mass for approximate cubes . . . 19
5.3. Completion of the proof . . . 24
6. Proof of Proposition 2.3 . . . 24
6.1. Common equilibrium states . . . 24
6.2. Recursive-function estimate . . . 26
6.3. Fibre estimates . . . 30
7. Proof of Theorem 1.2 . . . 31
7.1. Box dimension . . . 32
7.2. The case $\dim_{\mathrm{H}} K = \dim_{\mathrm{B}} K$ . . . 36
7.3. The case $\dim_{\mathrm{H}} K < \dim_{\mathrm{B}} K$ . . . 37
References . . . 39

2020 *Mathematics Subject Classification.* Primary 28A80; Secondary 37B10, 37A35, 28A78.

Key words and phrases. Bedford–McMullen subshift, weak specification, critical Hausdorff measure, factor map.

The research of Qiu was supported by the National Natural Science Foundation of China, grants 12471087 and 12531004. The research of Wang was supported by the Nanjing University PhD Student Zhujian Program, grant ZJJH2026B08.

1. Introduction

Fix a diagonal expanding integer matrix and its digit alphabet

$$\Lambda = \operatorname{diag}(m_1, \dots, m_d), \qquad m_1 \geq \cdots \geq m_d \geq 2, \qquad \mathcal{A} = \prod_{r=1}^{d} \{0, 1, \dots, m_r - 1\}.$$

Assume that $m_1, \dots, m_d$ take at least two distinct values. Let $n_1 > \cdots > n_s$ be these distinct values, so that $s \geq 2$. Choose $0 = d_0 < d_1 < \cdots < d_s = d$ so that

$$m_r = n_i \qquad d_{i-1} < r \leq d_i.$$

Thus the i-th group consists of coordinates $d_{i-1} + 1, \dots, d_i$, and the first i groups together consist of coordinates $1, \dots, d_i$. A subshift $X \subset \mathcal{A}^{\mathbb{N}}$ is a nonempty closed set satisfying $\sigma(X) \subseteq X$, where $\sigma(x_1 x_2 x_3 \cdots) = x_2 x_3 \cdots$ is the left shift. A *finite word* of length N over $\mathcal{A}$ is a finite sequence $I = i_1 \cdots i_N \in \mathcal{A}^N$, and its *length* is denoted by $|I| = N$. The empty word has length 0. If $I = i_1 \cdots i_N$ and $J = j_1 \cdots j_M$, their *concatenation* is

$$IJ = i_1 \cdots i_N j_1 \cdots j_M, \qquad |IJ| = N + M.$$

Write $\mathcal{L}(X)$ for the set of finite words that occur in sequences of X, together with the empty word. The elements of $\mathcal{L}(X)$ are called *admissible words*. If $\mathcal{S} \subset \mathcal{A}$ is nonempty and $X = \mathcal{S}^{\mathbb{N}}$, then X is the *full shift* on the digit set $\mathcal{S}$. The subshift X has *weak specification* [Fen11] if there is an integer $p \geq 0$ such that, for every $I, J \in \mathcal{L}(X)$, there exists an admissible word W with $|W| \leq p$ such that $IWJ \in \mathcal{L}(X)$. Thus any two admissible words can be joined, although the joining word may depend on the pair. For $x = (x_k)_{k \geq 1} \in X$, each $x_k \in \mathcal{A} \subset \mathbb{Z}^d$ is a d-dimensional digit vector. The associated self-affine set is

$$K = R(X), \qquad R(x) = \sum_{k=1}^{\infty} \Lambda^{-k} x_k. \tag{1.1}$$

We use $\mathcal{H}^t$, $\dim_{\mathrm{H}}$, $\underline{\dim}_{\mathrm{B}}$, and $\overline{\dim}_{\mathrm{B}}$ for Hausdorff measure, Hausdorff dimension, lower box dimension, and upper box dimension, respectively; see [Fal14]. For a bounded set E, when its lower and upper box dimensions agree, their common value is denoted by $\dim_{\mathrm{B}} E$. For a nonempty set E, we call $\mathcal{H}^{\dim_{\mathrm{H}} E}(E)$ its *critical Hausdorff measure*.

The Hausdorff dimension gives the threshold at which Hausdorff measure changes from infinity to zero, but it does not determine the measure at the threshold itself. More precisely, if $\gamma = \dim_{\mathrm{H}} E$, then $\mathcal{H}^t(E) = \infty$ for $t < \gamma$ and $\mathcal{H}^t(E) = 0$ for $t > \gamma$, whereas $\mathcal{H}^\gamma(E)$ may be zero, positive and finite, or infinite. For a self-similar set satisfying the open set condition, Hutchinson's theorem gives $0 < \mathcal{H}^\gamma(E) < \infty$ [Hut81]. The situation is more delicate for the diagonal self-affine sets considered here: symbolic cylinders have different side lengths in different coordinate directions, so their distribution among the coordinate fibres contains geometric information that is not recorded by the Hausdorff dimension alone. Thus the critical Hausdorff measure is a genuinely finer quantity than the dimension.

The Bedford–McMullen family is the basic non-conformal setting in which this distinction can be seen explicitly. When $d = 2$ and $X = \mathcal{S}^{\mathbb{N}}$ is a full shift, the set K is a *Bedford–McMullen carpet*. In 1984, Bedford and McMullen independently computed its Hausdorff and box dimensions [Bed84, McM84]. If all nonempty fibres of the projection of $\mathcal{S}$ onto the second coordinate have the same cardinality, then $\dim_{\mathrm{H}} K = \dim_{\mathrm{B}} K$ and the critical Hausdorff measure is positive and finite. McMullen asked about the critical measure when these dimensions are different. In 1994, Peres answered this question by proving that, in that

case, the critical Hausdorff measure is infinite and not σ-finite [Per94]. Thus, for a planar Bedford-McMullen full shift and $\gamma = \dim_{\mathrm{H}} K$,

$$\dim_{\mathrm{H}} K = \dim_{\mathrm{B}} K \iff 0 < \mathcal{H}^{\gamma}(K) < \infty,$$
$$\dim_{\mathrm{H}} K < \dim_{\mathrm{B}} K \iff \mathcal{H}^{\gamma}(K) = \infty.$$

At the end of the same paper, Peres considered sets coded by a primitive transition matrix and formulated the following conjecture. Let $\mathcal{S} \subset \mathcal{A}$ be a digit set and let $A = (A_{uv})_{u,v\in\mathcal{S}}$ be a primitive zero-one matrix; that is, A^q is a positive matrix for some integer $q \geq 1$. The corresponding subshift is

$$X_A = \{x = (x_k)_{k\geq 1} \in \mathcal{S}^{\mathbb{N}} : A_{x_k x_{k+1}} = 1 \text{ for every } k \geq 1\}, \qquad K_A = R(X_A).$$

The subshift X_A is called a *primitive subshift of finite type*.

Conjecture ([Per94]). *If $\gamma_A = \dim_{\mathrm{H}} K_A < \dim_{\mathrm{B}} K_A$, then $\mathcal{H}^{\gamma_A}(K_A) = \infty$.*

The motivation for this paper is to prove this two-dimensional conjecture and to extend it to higher-dimensional diagonal affine-invariant sets coded by subshifts satisfying weak specification. The main result is the following.

Theorem 1.1. *Let $X \subset \mathcal{A}^{\mathbb{N}}$ be a subshift satisfying weak specification and set $K = R(X)$. Let $\gamma = \dim_{\mathrm{H}} K$. Then*

$$\dim_{\mathrm{H}} K = \dim_{\mathrm{B}} K \iff 0 < \mathcal{H}^{\gamma}(K) < \infty, \tag{1.2}$$
$$\dim_{\mathrm{H}} K < \dim_{\mathrm{B}} K \iff \mathcal{H}^{\gamma}(K) = \infty. \tag{1.3}$$

In the second case, $\mathcal{H}^{\gamma}$ is not σ-finite on K.

Every primitive subshift of finite type satisfies weak specification, so Theorem 1.1 resolves Peres' conjecture. Under weak specification, Z. Feng proved the existence of the box dimension and that dimension equality implies positive and finite critical Hausdorff measure [Fen26]. Our contribution is the converse measure statement: a strict dimension gap implies that the critical Hausdorff measure is infinite and, in fact, not σ-finite. In particular, the critical Hausdorff measure can never vanish under weak specification.

The dimension theory surrounding this problem developed in several directions. In the plane, the constructions of Lalley–Gatzouras and Barański allow more general rectangular patterns and contraction ratios than the Bedford–McMullen construction [LG92, Bar07]. For higher-dimensional Bedford–McMullen sets, Kenyon and Peres established the Hausdorff-dimension variational principle and studied measures of full dimension [KP96a, KP96b]. In the present setting, approximate cubes encode different coordinate groups over different proportions of symbolic time, so the Hausdorff-dimension variational principle involves a weighted sum of factor entropies. D.-J. Feng's weighted thermodynamic formalism provides the corresponding equilibrium states and entropy estimates [Fen11].

Several finer aspects of diagonal self-affine geometry have also been studied, including dimension gaps for sponges, Assouad-type dimensions, L^q spectra, random sponges, and dimension formulas for diagonal self-affine sets and measures [DS17, FH17, Kol23, BB25, Fen25, Rap26]. The critical Hausdorff measure is a different refinement: it asks for the size of a set at its already known Hausdorff dimension. For Barański carpets, the authors obtained a dichotomy in terms of uniform fibre conditions [QW24]. Related recent work on Hausdorff measure and finer geometry for planar self-affine sets includes [Bar25, BKY26].

To understand how the conclusions in Theorem 1.1 can change when weak specification fails, we next consider sofic subshifts, which contain all subshifts of finite type. A sofic subshift

$X \subset \mathcal{A}^{\mathbb{N}}$ can be represented by a finite vertex-labelled directed graph and is therefore closely connected with graph-directed constructions; compare [MW88]. It is *irreducible* if, for every $I, J \in \mathcal{L}(X)$, there is a word $W \in \mathcal{L}(X)$ such that $IWJ \in \mathcal{L}(X)$. An irreducible sofic subshift has a finite strongly connected graph presentation. For each ordered pair (u, v) of vertices, choose a directed path of positive length from u to v. Since there are only finitely many such pairs, the lengths of these paths have a finite maximum; denote it by p. Given $I, J \in \mathcal{L}(X)$, choose paths whose labels are I and J. If u is the terminal vertex of the first path and v is the initial vertex of the second, let W be the word formed by the labels of the vertices between u and v on the chosen path. Then $|W| \le p$ and $IWJ \in \mathcal{L}(X)$. Thus every irreducible sofic subshift satisfies weak specification and is covered by Theorem 1.1. We therefore focus on reducible sofic subshifts, which need not satisfy weak specification.

Theorem 1.2. *For every sofic subshift X, the box dimension of $K = R(X)$ exists. Write $\gamma = \dim_{\mathrm{H}} K$. Moreover:*

(i) *If $\dim_{\mathrm{H}} K = \dim_{\mathrm{B}} K$, then $\mathcal{H}^{\gamma}(K) > 0$ and $\mathcal{H}^{\gamma}$ is σ-finite on K, but $\mathcal{H}^{\gamma}(K)$ need not be finite.*

(ii) *If $\dim_{\mathrm{H}} K < \dim_{\mathrm{B}} K$, then $\mathcal{H}^{\gamma}(K)$ may be positive and finite, or it may be infinite and σ-finite.*

Remark 1.3. For a sofic subshift, Proposition 7.2 not only proves that the box dimension exists, but also gives its formula in terms of the topological entropies of the strongly connected components and their factor subshifts.

Together, these results compare the weak-specification setting with the reducible sofic setting. Under weak specification, Theorem 1.1 gives a complete dichotomy for the critical Hausdorff measure according to whether the dimensions are equal. For sofic subshifts, dimension equality implies positivity and σ-finiteness but not finiteness. When the dimensions differ, both finite and infinite critical Hausdorff measure can occur.

Structure of the manuscript. Section 2 introduces the notation and reduces the second implication in Theorem 1.1 to Propositions 2.2 and 2.3. Section 3 defines approximate cubes and proves the required density estimates. Section 4 proves the partition lemma for finite admissible word sets. Section 5 uses this partition to construct a probability measure and prove Proposition 2.2. Section 6 proves Proposition 2.3. Finally, Section 7 proves Theorem 1.2 and examines the dimensions and critical Hausdorff measure of reducible sofic subshifts.

2. Preliminaries and reduction of the main theorem

Throughout, log denotes the natural logarithm. For $x \in \mathbb{R}$, $\lfloor x \rfloor$ is the greatest integer less than or equal to x, and $\lceil x \rceil$ is the least integer greater than or equal to x. For positive quantities A and B, the notation $A \lesssim B$ means that $A \le CB$ for a constant $C > 0$, $A \gtrsim B$ means that $B \lesssim A$, and $A \asymp B$ means that both $A \lesssim B$ and $B \lesssim A$. For a real quantity F and a positive quantity B, the notation $F = O(B)$ means $|F| \lesssim B$, while $F = o(B)$ means that $F/B \to 0$ in the stated limit. A subscript on $\lesssim$, $\gtrsim$, $\asymp$, or O lists the parameters on which the corresponding constant may depend. For a finite set E, the symbol $\#E$ denotes its cardinality.

2.1. Factor maps and weighted pressure.

Throughout Sections 2–6, $X \subset \mathcal{A}^{\mathbb{N}}$ denotes a subshift satisfying weak specification, $K = R(X)$, and $\mathcal{S} = \{x_1 : x \in X\}$ is the set of digits that occur in X. For $x = (x_j)_{j \ge 1}$, write $x|N = x_1 \cdots x_N$. If $I = i_1 \cdots i_N$ and $0 \le a \le b \le N$,

write $I|_a^b = i_{a+1}\cdots i_b$, abbreviate $I|b = I|_0^b$, and write $|I| = N$. We use the convention $I|_a^a = \varnothing$. If Y is a subshift over a finite alphabet, let $\mathcal{L}_0(Y) = \{\varnothing\}$ and

$$\mathcal{L}_N(Y) = \{y_1\cdots y_N : y = (y_j)_{j\geq 1} \in Y\} \quad (N \geq 1), \qquad \mathcal{L}(Y) = \bigcup_{N\geq 0} \mathcal{L}_N(Y).$$

For $N \geq 1$ and $I \in \mathcal{L}_N(Y)$, the cylinder generated by I is $[I] = \{y \in Y : y|N = I\}$.

We write $\mathcal{M}_\sigma(Y)$ for the set of invariant Borel probability measures on Y. For a Borel probability measure ν on Y, set

$$H(\nu|\mathcal{L}_N(Y)) = -\sum_{I\in\mathcal{L}_N(Y)} \nu[I]\log\nu[I], \tag{2.1}$$

where $0\log 0 = 0$. For $\mu \in \mathcal{M}_\sigma(Y)$, its measure-theoretic entropy is

$$h(\mu) = \lim_{N\to\infty} \frac{1}{N} H(\mu|\mathcal{L}_N(Y)),$$

whereas the topological entropy of the whole subshift is

$$h_{\mathrm{top}}(Y) = \lim_{N\to\infty} \frac{1}{N}\log \#\mathcal{L}_N(Y).$$

Both limits exist by subadditivity and Fekete's lemma, since $H(\mu|\mathcal{L}_{N+M}(Y)) \leq H(\mu|\mathcal{L}_N(Y)) + H(\mu|\mathcal{L}_M(Y))$ and $\#\mathcal{L}_{N+M}(Y) \leq \#\mathcal{L}_N(Y)\,\#\mathcal{L}_M(Y)$. In particular, $h(\mu) \leq h_{\mathrm{top}}(Y)$.

Recall that there are integers $n_1 > \cdots > n_s \geq 2$, where $s \geq 2$, and $0 = d_0 < d_1 < \cdots < d_s = d$ such that

$$m_r = n_i \qquad d_{i-1} < r \leq d_i.$$

Set $n_0 = \infty$, $\mathcal{A}_1 = \mathcal{A}$, and

$$\mathcal{A}_i = \prod_{r>d_{i-1}} \{0,\dots,m_r - 1\} \qquad 2 \leq i \leq s.$$

Let $\pi_1 : \mathcal{A} \to \mathcal{A}_1$ be the identity. For $2 \leq i \leq s$, let $\pi_i : \mathcal{A}_{i-1} \to \mathcal{A}_i$ delete the coordinates numbered at most d_{i-1}. The maps are applied coordinatewise to sequences and words. Write

$$\tau_i = \pi_i \circ \cdots \circ \pi_1, \qquad X_i = \tau_i(X), \qquad X = X_1 \xrightarrow{\pi_2} X_2 \xrightarrow{\pi_3} \cdots \xrightarrow{\pi_s} X_s. \tag{2.2}$$

We call the sequence of factor maps in (2.2) the *factor chain*.

If X satisfies weak specification with bound p, then every factor X_i satisfies weak specification with the same bound. To see this, let $U, V \in \mathcal{L}(X_i)$, choose $I, J \in \mathcal{L}(X)$ such that $\tau_i(I) = U$ and $\tau_i(J) = V$. There is $W \in \mathcal{L}(X)$ such that $|W| \leq p$ and $IWJ \in \mathcal{L}(X)$. Hence

$$U\,\tau_i(W)\,V = \tau_i(IWJ) \in \mathcal{L}(X_i), \qquad |\tau_i(W)| = |W| \leq p.$$

If μ is a Borel probability measure on X, its push-forward under τ_i is the probability measure $\tau_i\mu = (\tau_i)_*\mu$ on X_i given by

$$(\tau_i\mu)(B) = \mu(\tau_i^{-1}(B)) \qquad \text{for every Borel set } B \subset X_i.$$

If μ is invariant under σ, then $\tau_i\mu$ is also invariant, as $\tau_i \circ \sigma = \sigma \circ \tau_i$.

Set

$$\theta_0 = 0, \qquad \theta_i = \frac{\log n_s}{\log n_i}, \qquad a_i = \theta_i - \theta_{i-1}, \qquad \mathbf{a} = (a_1,\dots,a_s). \tag{2.3}$$

Then $0 < \theta_1 < \cdots < \theta_s = 1$.

If $\mathbf{b} = (b_1, \dots, b_s)$, where $b_i > 0$ and $\sum_i b_i = 1$, write

$$P^{\mathbf{b}}(\sigma) = \sup_{\mu \in \mathcal{M}_\sigma(X)} \sum_{i=1}^{s} b_i h(\tau_i \mu). \tag{2.4}$$

$P^{\mathbf{b}}(\sigma)$ is called the $\mathbf{b}$-*weighted pressure* of σ. In the notation of D.-J. Feng [Fen11], this is $P^{\mathbf{b}}(\sigma, 0)$; we omit the zero potential from the notation. An invariant measure attaining the supremum in (2.4) is called a $\mathbf{b}$-*weighted equilibrium state* [Fen11].

We shall use the following results from Kenyon and Peres [KP96b], D.-J. Feng [Fen11], and Z. Feng [Fen26].

Proposition 2.1. *Let $X \subset \mathcal{A}^{\mathbb{N}}$ be a subshift satisfying weak specification and set $K = R(X)$. Then:*

(i)
$$\dim_{\mathrm{H}} K = \frac{P^{\mathbf{a}}(\sigma)}{\log n_s}, \qquad \dim_{\mathrm{B}} K = \frac{1}{\log n_s} \sum_{i=1}^{s} a_i h_{\mathrm{top}}(X_i). \tag{2.5}$$

(ii) *Every positive weight vector $\mathbf{b}$ has a unique $\mathbf{b}$-weighted equilibrium state $\eta_{\mathbf{b}}$.*

(iii) *For each fixed $\mathbf{b}$ and $1 \le i \le s$,*
$$H\big((\tau_i \eta_{\mathbf{b}}) | \mathcal{L}_N(X_i)\big) = N h(\tau_i \eta_{\mathbf{b}}) + O_{\mathbf{b}}(1). \tag{2.6}$$

(iv) *If $\gamma = \dim_{\mathrm{H}} K$ and $\dim_{\mathrm{H}} K = \dim_{\mathrm{B}} K$, then*
$$0 < \mathcal{H}^{\gamma}(K) < \infty. \tag{2.7}$$

Proof. The Ledrappier–Young formula and the full-dimension variational principle in [KP96b, Theorem 1.1 and Lemma 4.3] give

$$\dim_{\mathrm{H}} K = \sup_{\mu \in \mathcal{M}_\sigma(X)} \sum_{i=1}^{s} \left(\frac{1}{\log n_i} - \frac{1}{\log n_{i-1}} \right) h(\tau_i \mu) = \frac{1}{\log n_s} \sup_{\mu \in \mathcal{M}_\sigma(X)} \sum_{i=1}^{s} a_i h(\tau_i \mu) = \frac{P^{\mathbf{a}}(\sigma)}{\log n_s}.$$

Since X satisfies weak specification, [Fen26, Proposition 3.3] shows that the box dimension of K exists and gives

$$\dim_{\mathrm{B}} K = \frac{1}{\log n_s} \sum_{i=1}^{s} a_i h_{\mathrm{top}}(X_i).$$

This proves (i).

Conclusion (ii) follows directly from [Fen11, Theorem 1.2], applied to the zero potential.

We next prove (iii). For $i = 1$, equation (2.6) is precisely the entropy estimate in [Fen11, Theorem 7.3(iii)]. Let $2 \le i \le s$. The factor X_i satisfies weak specification. Applying [Fen11, Lemma 7.2(ii)] successively to $\pi_2, \dots, \pi_i$ shows that $\tau_i \eta_{\mathbf{b}}$ is the weighted equilibrium state of the induced potential on the factor chain $X_i \to \cdots \to X_s$. By [Fen11, Lemma 5.7], each recursively induced word function remains in the class required by [Fen11, Theorems 5.5 and 7.3]. This induced potential need not be zero. If $i < s$, its weight vector is $\left(\sum_{\ell=1}^{i} b_\ell, b_{i+1}, \dots, b_s\right)$, so [Fen11, Theorem 7.3(iii)] applies. For $i = s$, [Fen11, Theorem 5.5] applies. In either case,

$$\sum_{J \in \mathcal{L}_N(X_i)} (\tau_i \eta_{\mathbf{b}})[J] \log (\tau_i \eta_{\mathbf{b}})[J] = -N h(\tau_i \eta_{\mathbf{b}}) + O_{\mathbf{b}}(1).$$

Multiplying by -1 and using (2.1) gives (2.6).

Finally, under the assumption $\dim_{\mathrm{H}} K = \dim_{\mathrm{B}} K$, conclusion (i) of [Fen26, Theorem 1.1] gives $0 < \mathcal{H}^{\dim_{\mathrm{H}} K}(K) < \infty$. Taking $\gamma = \dim_{\mathrm{H}} K$ proves (iv). □

2.2. **A reduction of Theorem 1.1.** Sections 3–6 prove the second implication in Theorem 1.1. For $0 < \beta \le 1$, write

$$a_i(\beta) = \theta_i^\beta - \theta_{i-1}^\beta, \qquad \mathbf{a}(\beta) = (a_1(\beta), \dots, a_s(\beta)). \tag{2.8}$$

Note that $\mathbf{a}(1) = \mathbf{a}$, where $\mathbf{a}$ is defined in (2.3). Since $0 = \theta_0 < \theta_1 < \cdots < \theta_s = 1$,

$$a_i(\beta) > 0, \qquad \sum_{i=1}^{s} a_i(\beta) = 1.$$

Thus $\mathbf{a}(\beta)$ is a positive weight vector. Let $\eta = \eta_{\mathbf{a}}$, and write

$$c(\beta) = P^{\mathbf{a}(\beta)}(\sigma) - \sum_{i=1}^{s} a_i(\beta) h(\tau_i \eta) \ge 0. \tag{2.9}$$

The last inequality follows directly from the variational definition of $P^{\mathbf{a}(\beta)}(\sigma)$.

Given a continuous increasing function $\phi : [0, \infty) \to [0, \infty)$ with $\phi(0) = 0$, called a *gauge function*, the associated Hausdorff measure is

$$\mathcal{H}^\phi(E) = \liminf_{\delta \downarrow 0} \left\{ \sum_j \phi(\operatorname{diam} U_j) : E \subset \bigcup_j U_j,\ \operatorname{diam} U_j \le \delta \right\}.$$

For $t > 0$, taking $\phi(r) = r^t$ gives the usual Hausdorff measure $\mathcal{H}^t$.

We now reduce the second implication in Theorem 1.1 to two propositions. The first shows that $c(\beta) > 0$ gives the required Hausdorff-measure conclusion, and the second shows that the dimension gap $\dim_{\mathrm{H}} K < \dim_{\mathrm{B}} K$ gives $c(\beta) > 0$. Their proofs are given in Sections 5 and 6, respectively. Together with Proposition 2.1(iv), these two propositions prove Theorem 1.1.

Proposition 2.2. *Let $X \subset \mathcal{A}^{\mathbb{N}}$ be a subshift satisfying weak specification, set $K = R(X)$, and let $\gamma = \dim_{\mathrm{H}} K$. If $c(\beta) > 0$ for some $\beta \in (1/2, 1)$, then there is $\varepsilon > 0$ such that the gauge function g_β, satisfying $g_\beta(r) = r^\gamma \exp\{-\varepsilon(\log(1/r))^\beta\}$ for $0 < r \le 1$, has*

$$\mathcal{H}^{g_\beta}(K) = \infty, \qquad \mathcal{H}^\gamma(K) = \infty.$$

Neither $\mathcal{H}^{g_\beta}$ nor $\mathcal{H}^\gamma$ is σ-finite on K.

Proposition 2.3. *Let $X \subset \mathcal{A}^{\mathbb{N}}$ be a subshift satisfying weak specification and set $K = R(X)$. If $\dim_{\mathrm{H}} K < \dim_{\mathrm{B}} K$, then $c(\beta) > 0$ for some $\beta \in (1/2, 1)$.*

Proof of Theorem 1.1 based on Propositions 2.2 and 2.3. Suppose first that $\dim_{\mathrm{H}} K = \dim_{\mathrm{B}} K$. Then (2.7) gives

$$0 < \mathcal{H}^\gamma(K) < \infty.$$

Suppose next that $\dim_{\mathrm{H}} K < \dim_{\mathrm{B}} K$. Proposition 2.3 gives $\beta \in (1/2, 1)$ such that $c(\beta) > 0$. Proposition 2.2 then gives

$$\mathcal{H}^\gamma(K) = \infty,$$

and shows that $\mathcal{H}^\gamma$ is not σ-finite on K.

The general inequality $\dim_{\mathrm{H}} K \le \dim_{\mathrm{B}} K$ shows that these are the only two possible cases. The finite-measure conclusion in the first case and the infinite-measure conclusion in the second case prove both converse implications in (1.2) and (1.3). □

Remark 2.4. The proof of Proposition 2.3 gives a stronger conclusion. If $\dim_{\mathrm{H}} K < \dim_{\mathrm{B}} K$, then every nonempty open interval $I \subset (0, 1)$ contains some β for which $c(\beta) > 0$. Otherwise, the function c would vanish throughout I, and the argument in Section 6 would imply $\dim_{\mathrm{H}} K = \dim_{\mathrm{B}} K$. Consequently, for every $\beta_0 \in (1/2, 1)$, there is some $\beta \in (\beta_0, 1)$ with

$c(\beta) > 0$. Proposition 2.2 then shows that the exponent β in the gauge function g_β may be chosen arbitrarily close to 1.

3. Approximate cubes and the density estimates

This section introduces the approximate cubes used throughout the proof. We describe both their geometric form and their symbolic representation. We then use a density estimate to deduce infinite and non-σ-finite Hausdorff measure from sufficiently small approximate-cube masses.

For a word $I = i_1 \cdots i_N$ and $0 \le k \le N$, write

$$R_k(I) = \sum_{\ell=1}^{k} \Lambda^{-\ell} i_\ell, \qquad R_0(I) = 0.$$

Thus

$$R(x) = R_k(x|k) + \Lambda^{-k} R(\sigma^k x).$$

For $I \in \mathcal{A}^k$, the *level-k approximate cube* is

$$Q_k(I) = \prod_{i=1}^{s} \prod_{r=d_{i-1}+1}^{d_i} \left[R_{\lfloor \theta_i k \rfloor}(I)_r, R_{\lfloor \theta_i k \rfloor}(I)_r + n_i^{-\lfloor \theta_i k \rfloor} \right),$$

with the adjustment that a factor is closed on the right whenever its right endpoint is 1. Here $R_{\lfloor \theta_i k \rfloor}(I)_r$ denotes the r-th coordinate of $R_{\lfloor \theta_i k \rfloor}(I)$. Note that the level-$k$ approximate cubes form a partition of $[0,1]^d$. These partitions are nested: every level-$(k+1)$ approximate cube is contained in a unique level-k approximate cube. Every side length of $Q_k(I)$ is comparable to n_s^{-k}; hence there are constants $c_0, C_0 > 0$, independent of k and I, such that

$$c_0 n_s^{-k} \le \operatorname{diam} Q_k(I) \le C_0 n_s^{-k}. \tag{3.1}$$

For $x \in \mathcal{A}^{\mathbb{N}}$, write $Q_k(x) = Q_k(x|k)$.

For $x \in X$, let the corresponding word set be

$$\Gamma_k(x) = \{ I \in \mathcal{L}_k(X) : \tau_i(I|\lfloor \theta_i k \rfloor) = \tau_i(x|\lfloor \theta_i k \rfloor) \text{ for } 1 \le i \le s \}. \tag{3.2}$$

The corresponding *cylinder set* is

$$[\Gamma_k(x)] = \bigcup_{I \in \Gamma_k(x)} [I].$$

The following elementary observation from Euclidean geometry will be used: for every level-k approximate cube Q with $R^{-1}(Q) \cap X \neq \varnothing$, there are sequences $x^{(1)}, \dots, x^{(L)} \in X$ such that

$$R^{-1}(Q) \cap X \subseteq \bigcup_{\ell=1}^{L} [\Gamma_k(x^{(\ell)})], \qquad L \le 2^d. \tag{3.3}$$

Set $\mathcal{Q}_k(X) = \{Q_k(x) : x \in X\}$. For every $x \in X$, the point $R(x)$ belongs to the closure of $Q_k(x)$. Consequently, the closures of the approximate cubes in $\mathcal{Q}_k(X)$ cover K. For $z \in [0,1]^d$, let $Q_k(z)$ be the unique level-k approximate cube containing z. If $z = R(x)$ is not on a grid boundary, then $Q_k(z) = Q_k(x)$. Let

$$\mathcal{D}_k(X) = \Bigg\{ (J_1, \dots, J_s) \in \prod_{i=1}^{s} \mathcal{L}_{\lfloor \theta_i k \rfloor - \lfloor \theta_{i-1} k \rfloor}(X_i) : \text{there is } I \in \mathcal{L}_k(X) \text{ such that } \tau_i(I|_{\lfloor \theta_{i-1} k \rfloor}^{\lfloor \theta_i k \rfloor}) = J_i \text{ for every } i \Bigg\}. \tag{3.4}$$

By [Fen26, Lemma 2.7],

$$\#\mathcal{Q}_k(X) = \#\mathcal{D}_k(X). \tag{3.5}$$

Lemma 3.1. *Let λ be a Borel probability measure supported on K, and let g be a gauge function. If*

$$\limsup_{k\to\infty} \frac{\lambda(Q_k(z))}{g(n_s^{-k})} = 0 \quad \text{for } \lambda\text{-almost every } z,$$

then $\mathcal{H}^g(K) = \infty$. Moreover, $\mathcal{H}^g$ is not σ-finite on K.

Proof. For every $C > 0$, the assumed upper density is at most C almost everywhere. Hence [Fen26, Lemma 2.9(i)] gives $\mathcal{H}^g(K) \gtrsim C^{-1}$, where the implicit constant is independent of C. Letting $C \downarrow 0$ proves $\mathcal{H}^g(K) = \infty$. Let $E \subset K$ be Borel with $\lambda(E) > 0$ and define $\lambda_E(A) = \lambda(A \cap E)/\lambda(E)$. For λ_E-almost every z,

$$\frac{\lambda_E(Q_k(z))}{g(n_s^{-k})} \le \frac{1}{\lambda(E)} \frac{\lambda(Q_k(z))}{g(n_s^{-k})} \longrightarrow 0.$$

Applying the same argument to λ_E gives $\mathcal{H}^g(E) = \infty$. Thus every Borel set of finite $\mathcal{H}^g$ measure has zero λ-mass. By Borel regularity of Hausdorff measure, every $\mathcal{H}^g$-measurable set of finite measure is contained in a Borel set of the same finite measure. Therefore no countable family of finite-$\mathcal{H}^g$-measure sets can cover K, and $\mathcal{H}^g$ is not σ-finite on K. □

4. Partition of admissible words

This section proves the finite-word decomposition used in the proof of Proposition 2.2. For each N, we partition $\mathcal{L}_N(X)$ into $O(N^{s-1})$ sets having uniform fibres at every factor level.

The idea comes from a standard approximation method for planar self-affine carpets. For a Bedford–McMullen carpet, one can pass to a sufficiently high iterate and select words whose digit frequencies are close to those of a measure of full Hausdorff dimension. The selected words define a Bedford–McMullen subcarpet with uniform fibres and Hausdorff dimension arbitrarily close to that of the original carpet. This approximation from within was used in [FJS10, Section 4] and [FS16, Section 6]. Here a single subsystem is not enough. Instead, for each fixed word length, we divide all admissible words into finitely many sets, each having uniform fibres at every factor level.

Fix $N \in \mathbb{N}$ and write $\mathcal{W}_i = \mathcal{L}_N(X_i)$. When no confusion can arise, we use the same symbols for the restrictions

$$\pi_{i+1} : \mathcal{W}_i \to \mathcal{W}_{i+1}, \qquad \tau_i : \mathcal{W}_1 \to \mathcal{W}_i.$$

For $C \subset \mathcal{W}_1$ and $J \in \tau_i(C)$, write

$$m_i(C, J) = \#\{I \in C : \tau_i(I) = J\}.$$

We consider sets $C \subset \mathcal{W}_1$ for which, for every $1 \le i \le s$, the number $m_i(C, J)$ is independent of $J \in \tau_i(C)$. In other words, all fibres of the map $\tau_i : C \to \tau_i(C)$ have the same cardinality. We say that such a set C has *uniform fibres at every level*. Denote the common fibre cardinality by $M_i(C)$, and write $B_i(C) = \#\tau_i(C)$. Then

$$\#C = B_i(C)M_i(C) \qquad 1 \le i \le s. \tag{4.1}$$

Recall that $\mathcal{S} = \{x_1 : x \in X\}$, and write $\mathcal{S}_i = \tau_i(\mathcal{S})$. A partition of $\mathcal{W}_1$ is a family of pairwise disjoint subsets whose union is $\mathcal{W}_1$.

Lemma 4.1. *The set $\mathcal{W}_1$ has a partition $\{C_t\}_{t=1}^{q_N}$ such that every C_t has uniform fibres at every level. Moreover,*

$$q_N \le \prod_{i=1}^{s-1} \left(1 + N \left\lceil \frac{\log \#\mathcal{S}_i}{\log 2} \right\rceil\right) = O(N^{s-1}). \tag{4.2}$$

Proof. We construct the partition recursively along the factor chain. The construction will use sets of the form

$$G_i(r_1, \dots, r_{i-1}; J_i) \subset \mathcal{W}_1, \qquad J_i \in \mathcal{W}_i, \quad 2 \le i \le s.$$

For each i, we construct the nonempty sets of this form so that they satisfy the following three properties.

(P1) They are pairwise disjoint and their union is $\mathcal{W}_1$.
(P2) Every word in $G_i(r_1, \dots, r_{i-1}; J_i)$ projects to J_i under τ_i.
(P3) For $1 \le \ell \le i$ and $L \in \tau_\ell(G_i(r_1, \dots, r_{i-1}; J_i))$,

$$\#\{I \in G_i(r_1, \dots, r_{i-1}; J_i) : \tau_\ell(I) = L\} = 2^{r_1 + \cdots + r_{\ell-1}}. \tag{4.3}$$

For $\ell = 1$, the exponent is the empty sum 0, so the right-hand side is 1.

In particular, taking $\ell = i$ in (4.3) and using (P2) gives

$$\#G_i(r_1, \dots, r_{i-1}; J_i) = 2^{r_1 + \cdots + r_{i-1}}. \tag{4.4}$$

We first construct the level-2 sets. For $J_2 \in \mathcal{W}_2$, let

$$q_1(J_2) = \#\{I \in \mathcal{W}_1 : \tau_2(I) = J_2\} \ge 1.$$

Write its binary expansion as

$$q_1(J_2) = \sum_{r_1 \ge 0} \varepsilon_{r_1}(J_2) 2^{r_1}, \qquad \varepsilon_{r_1}(J_2) \in \{0, 1\}.$$

Using this binary expansion, split the fibre $\tau_2^{-1}(J_2)$ into sets

$$G_2(r_1; J_2), \qquad \#G_2(r_1; J_2) = 2^{r_1},$$

for each r_1 with $\varepsilon_{r_1}(J_2) = 1$. The cardinalities of these sets add up to $q_1(J_2)$. As J_2 varies, these sets partition $\mathcal{W}_1$. Every word $I \in G_2(r_1; J_2)$ satisfies $\tau_2(I) = J_2$, and τ_1 is the identity. Consequently, properties (P1)–(P3) hold for $i = 2$.

Suppose now that the level-i sets have been constructed for some $2 \le i < s$. Fix $(r_1, \dots, r_{i-1})$ and $J_{i+1} \in \mathcal{W}_{i+1}$, and consider the finite family

$$\mathscr{F}_i(r_1, \dots, r_{i-1}; J_{i+1}) = \left\{ G_i(r_1, \dots, r_{i-1}; J_i) : \begin{array}{l} J_i \in \mathcal{W}_i,\ \pi_{i+1}(J_i) = J_{i+1}, \\ G_i(r_1, \dots, r_{i-1}; J_i) \ne \varnothing \end{array} \right\}. \tag{4.5}$$

Let $q_i(r_1, \dots, r_{i-1}; J_{i+1})$ be the number of sets in this family. If this number is zero, there is nothing to do. Otherwise, write its binary expansion as

$$q_i(r_1, \dots, r_{i-1}; J_{i+1}) = \sum_{r_i \ge 0} \varepsilon_{r_i} 2^{r_i}, \qquad \varepsilon_{r_i} \in \{0, 1\}.$$

Partition the family in (4.5) into disjoint subfamilies $\mathscr{F}_{i,r_i}$ with $\varepsilon_{r_i} = 1$, so that $\#\mathscr{F}_{i,r_i} = 2^{r_i}$. Let

$$G_{i+1}(r_1, \dots, r_i; J_{i+1}) = \bigcup_{G \in \mathscr{F}_{i,r_i}} G. \tag{4.6}$$

The new sets are unions of disjoint level-i sets, and every level-i set is used exactly once. Hence they are pairwise disjoint and cover $\mathcal{W}_1$, which proves (P1). Every word in a new set

projects to J_{i+1}, as $\tau_{i+1}(I) = \pi_{i+1}(\tau_i(I)) = J_{i+1}$ for every word I in that set. This proves (P2).

To prove (P3), first take $1 \le \ell \le i$ and $L \in \tau_\ell(G_{i+1}(r_1, \dots, r_i; J_{i+1}))$. Among the level-$i$ sets whose union defines the new set in (4.6), exactly one contains a word whose τ_ℓ-image is L. Indeed, suppose that two sets $G_i(r_1, \dots, r_{i-1}; J_i)$ and $G_i(r_1, \dots, r_{i-1}; J_i')$ have τ_ℓ-images containing L. If $\ell < i$, then

$$J_i = (\pi_i \circ \cdots \circ \pi_{\ell+1})(L) = J_i'.$$

If $\ell = i$, then $L = J_i = J_i'$. Thus the two sets are the same. Therefore the fibre over L is contained in one level-i set, and the induction hypothesis gives

$$\#\{I \in G_{i+1}(r_1, \dots, r_i; J_{i+1}) : \tau_\ell(I) = L\} = 2^{r_1 + \cdots + r_{\ell-1}}.$$

For $\ell = i+1$, the τ_ℓ-image of $G_{i+1}(r_1, \dots, r_i; J_{i+1})$ consists only of J_{i+1}. The union in (4.6) contains 2^{r_i} level-i sets, and each has $2^{r_1+\cdots+r_{i-1}}$ words by (4.4). Hence

$$\#G_{i+1}(r_1, \dots, r_i; J_{i+1}) = 2^{r_i} 2^{r_1+\cdots+r_{i-1}} = 2^{r_1+\cdots+r_i}.$$

This proves (P3) and completes the induction.

After the level-s construction, fix a complete index vector $\mathbf{r} = (r_1, \dots, r_{s-1})$ and let

$$C_{\mathbf{r}} = \bigcup_{\substack{J_s \in \mathcal{W}_s \\ G_s(r_1,\dots,r_{s-1};J_s) \neq \varnothing}} G_s(r_1, \dots, r_{s-1}; J_s).$$

Retain only the nonempty sets $C_{\mathbf{r}}$. Since the level-s sets form a partition of $\mathcal{W}_1$, the sets $C_{\mathbf{r}}$ also form a partition of $\mathcal{W}_1$. Fix $1 \le i \le s$ and $J_i \in \tau_i(C_{\mathbf{r}})$. For $1 \le i < s$, the word J_i has the uniquely determined image $J_s = (\pi_s \circ \cdots \circ \pi_{i+1})(J_i) \in \mathcal{W}_s$; when $i = s$, set $J_s = J_i$. Applying (4.3) at level s gives

$$m_i(C_{\mathbf{r}}, J_i) = 2^{r_1+\cdots+r_{i-1}}, \qquad 1 \le i \le s.$$

The right-hand side is independent of J_i. Therefore $C_{\mathbf{r}}$ has uniform fibres at every level, with $M_1(C_{\mathbf{r}}) = 1$ and $M_i(C_{\mathbf{r}}) = 2^{r_1+\cdots+r_{i-1}}$ for $2 \le i \le s$.

It remains to count the possible vectors $\mathbf{r}$. At the first step, $q_1(J_2)$ is the number of words in a fibre of $\pi_2 : \mathcal{W}_1 \to \mathcal{W}_2$, so

$$q_1(J_2) \le (\#\mathcal{S}_1)^N.$$

For $2 \le i \le s-1$, each member of (4.5) corresponds to a different J_i in the fibre of $\pi_{i+1} : \mathcal{W}_i \to \mathcal{W}_{i+1}$. Every word in $\mathcal{W}_i$ has length N over the alphabet $\mathcal{S}_i$, and hence

$$q_i(r_1, \dots, r_{i-1}; J_{i+1}) \le \#\pi_{i+1}^{-1}(J_{i+1}) \le (\#\mathcal{S}_i)^N.$$

Thus, whenever an exponent r_i occurs in one of the binary expansions, $q_i(r_1, \dots, r_{i-1}; J_{i+1})$ is at least 2^{r_i} and at most $(\#\mathcal{S}_i)^N$. Consequently,

$$0 \le r_i \le N \frac{\log \#\mathcal{S}_i}{\log 2} \le N \left\lceil \frac{\log \#\mathcal{S}_i}{\log 2} \right\rceil.$$

Thus there are at most

$$\prod_{i=1}^{s-1} \left(1 + N \left\lceil \frac{\log \#\mathcal{S}_i}{\log 2} \right\rceil\right)$$

possible complete index vectors, which proves (4.2). □

Remark 4.2. Let $C \subset \mathcal{W}_1$ have uniform fibres at every level, and let $\mathbb{P}_C$ be the uniform probability distribution on C. Then (4.1) gives that $(\tau_i)_*\mathbb{P}_C$ is the uniform probability distribution on $\tau_i(C)$, i.e.

$$(\tau_i)_*\mathbb{P}_C(\{J\}) = \frac{M_i(C)}{\#C} = \frac{1}{B_i(C)} \qquad J \in \tau_i(C). \tag{4.7}$$

Lemma 4.3. *Let ζ_N be a probability distribution on $\mathcal{W}_1$, and let $\{C_t\}_{t=1}^{q_N}$ be the partition given by Lemma 4.1. Write $\vartheta_t = \zeta_N(C_t), B_{t,i} = B_i(C_t) = \#\tau_i(C_t)$. For $1 \le i \le s$, let $\zeta_{N,i} = \tau_i\zeta_N$ be the distribution on $\mathcal{W}_i$. Then, for every $1 \le i \le s$,*

$$\sum_t \vartheta_t \log B_{t,i} \ge H(\zeta_{N,i}) - \log q_N, \tag{4.8}$$

where $H(\zeta_{N,i}) = -\sum_{J\in\mathcal{W}_i} \zeta_{N,i}(J)\log\zeta_{N,i}(J)$. Moreover, there exist $m \in \{1,\dots,s+1\}$, pairwise distinct indices $t_1,\dots,t_m$, and positive numbers $\rho_1,\dots,\rho_m$ with $\sum_{j=1}^m \rho_j = 1$ such that

$$\sum_t \vartheta_t \log B_{t,i} = \sum_{j=1}^m \rho_j \log B_{t_j,i} \qquad 1 \le i \le s.$$

Proof. For $\vartheta_t > 0$, let $p_{t,i}$ be the probability distribution on $\mathcal{W}_i$ given by

$$p_{t,i}(J) = \frac{1}{\vartheta_t} \sum_{\substack{I\in C_t \\ \tau_i(I)=J}} \zeta_N(I).$$

This distribution is supported on $\tau_i(C_t)$, which has $B_{t,i}$ elements, and hence $H(p_{t,i}) \le \log B_{t,i}$. For $J \in \mathcal{W}_i$, write $a_{t,J} = \vartheta_t p_{t,i}(J)$. Then $\zeta_{N,i}(J) = \sum_{t:\vartheta_t>0} a_{t,J}$ and $a_{t,J} \le \zeta_{N,i}(J)$. Therefore,

$$\begin{aligned} H(\zeta_{N,i}) &= -\sum_{t:\vartheta_t>0}\sum_{J\in\mathcal{W}_i} a_{t,J}\log\zeta_{N,i}(J) \le -\sum_{t:\vartheta_t>0}\sum_{J\in\mathcal{W}_i} a_{t,J}\log a_{t,J} \\ &= -\sum_{t:\vartheta_t>0}\vartheta_t\log\vartheta_t + \sum_{t:\vartheta_t>0}\vartheta_t H(p_{t,i}) \le \log q_N + \sum_t \vartheta_t \log B_{t,i}. \end{aligned}$$

This proves (4.8).

For the last assertion, write

$$v_t = (\log B_{t,1},\dots,\log B_{t,s}) \in \mathbb{R}^s.$$

The vector $v = \sum_t \vartheta_t v_t$ belongs to the convex hull of the vectors v_t in $\mathbb{R}^s$. By Carathéodory's theorem, every point in the convex hull of a subset of $\mathbb{R}^s$ is a convex combination of at most $s+1$ points of that subset. There are $m \le s+1$ pairwise distinct indices $t_1,\dots,t_m$ and coefficients $\rho_1,\dots,\rho_m > 0$, with $\sum_{j=1}^m \rho_j = 1$, such that

$$v = \sum_{j=1}^m \rho_j v_{t_j}.$$

Equality of the i-th coordinates gives the last assertion of the lemma. □

Corollary 4.4. *Suppose $\omega \in \mathcal{M}_\sigma(X)$ satisfies, for every $N \ge 1$,*

$$H((\tau_i\omega)|\mathcal{L}_N(X_i)) = Nh(\tau_i\omega) + O(1) \qquad 1 \le i \le s.$$

For every $N \ge 2$, let $\{C_t\}_{t=1}^{q_N}$ be a partition supplied by Lemma 4.1. Then there exist $m \in \{1,\dots,s+1\}$, pairwise distinct sets $C_{t_1},\dots,C_{t_m} \subset \mathcal{L}_N(X)$, and positive numbers

$\rho_1, \dots, \rho_m$ *with* $\sum_{j=1}^m \rho_j = 1$. *If* $B_{t_j,i} = \#\tau_i(C_{t_j})$, *then*

$$\sum_{j=1}^m \rho_j \log B_{t_j,i} \geq Nh(\tau_i\omega) - O(\log N) \qquad 1 \leq i \leq s.$$

Proof. Let ζ_N be the probability distribution on $\mathcal{L}_N(X)$ given by

$$\zeta_N(I) = \omega([I]) \qquad I \in \mathcal{L}_N(X).$$

For every i, write $\zeta_{N,i} = \tau_i\zeta_N$. The hypothesis gives $H(\zeta_{N,i}) = Nh(\tau_i\omega) + O(1)$ for every i. Applying the two conclusions of Lemma 4.3 to ζ_N, and using $q_N = O(N^{s-1})$ from Lemma 4.1, yields $m \leq s+1$, pairwise distinct indices $t_1, \dots, t_m$, and positive coefficients $\rho_1, \dots, \rho_m$ summing to 1 such that, for every $1 \leq i \leq s$,

$$\sum_{j=1}^m \rho_j \log B_{t_j,i} \geq H(\zeta_{N,i}) - \log q_N = Nh(\tau_i\omega) - O(\log N).$$

This proves the corollary. □

Corollary 4.4 gives finitely many admissible word sets together with weights describing the proportions in which they should be used. In the measure construction in Section 5, we need a sequence of these sets whose occurrence frequencies follow the prescribed weights. The next lemma constructs such a sequence with a uniformly bounded error.

Lemma 4.5. *Let* $q \in \mathbb{N}$, *let* $\rho_1, \dots, \rho_q > 0$, *and suppose* $\sum_{t=1}^q \rho_t = 1$. *There is a sequence* $u_1, u_2, \dots \in \{1, \dots, q\}$ *with the following property. For* $r \in \mathbb{N}$, *let* $N_t(r) = \#\{1 \leq j \leq r : u_j = t\}$. *Then, for every* $1 \leq t \leq q$ *and* $r \in \mathbb{N}$,

$$|N_t(r) - \rho_t r| \leq q - 1. \tag{4.9}$$

Consequently, for every $1 \leq t \leq q$ *and all integers* $0 \leq a < b$,

$$|\#\{a < j \leq b : u_j = t\} - \rho_t(b-a)| \leq 2(q-1).$$

Proof. If $q = 1$, take $u_r = 1$ for every r. Hence assume $q \geq 2$.

Choose u_1 arbitrarily. For $1 \leq t \leq q$ and $r \in \mathbb{N}$, write $e_t(r) = N_t(r) - r\rho_t$. Having chosen the first r terms, where $r \geq 1$, write $p_t(r) = (r+1)\rho_t - N_t(r) = \rho_t - e_t(r)$, and choose u_{r+1} to be an index for which $p_t(r)$ is largest. Since

$$\sum_{t=1}^q p_t(r) = (r+1)\sum_{t=1}^q \rho_t - \sum_{t=1}^q N_t(r) = (r+1) - r = 1,$$

the largest number $p_t(r)$ is positive.

We prove by induction that for $1 \leq t \leq q$ and $r \in \mathbb{N}$,

$$e_t(r) < 1. \tag{4.10}$$

For $r = 1$, this is immediate. Assume it holds at time r, and let $t_* = u_{r+1}$ be the chosen index. For this index,

$$e_{t_*}(r+1) = e_{t_*}(r) + 1 - \rho_{t_*} = 1 - p_{t_*}(r) < 1,$$

since $p_{t_*}(r) > 0$. If $t \neq t_*$, then

$$e_t(r+1) = e_t(r) - \rho_t < 1,$$

by the induction hypothesis and $\rho_t > 0$. This proves (4.10).

For every $r \in \mathbb{N}$ we also have

$$\sum_{t=1}^{q} e_t(r) = \sum_{t=1}^{q} N_t(r) - r\sum_{t=1}^{q} \rho_t = r - r = 0.$$

Consequently, for each $1 \le t \le q$,

$$e_t(r) = -\sum_{\substack{1\le v\le q\\ v\neq t}} e_v(r) > -(q-1).$$

Together with (4.10), this gives $|e_t(r)| \le q-1$, and hence proves (4.9).

Finally, for $1 \le a < b$,

$$\#\{a < j \le b : u_j = t\} - \rho_t(b-a) = N_t(b) - N_t(a) - \rho_t(b-a) = e_t(b) - e_t(a).$$

This gives the stated bound when $a \ge 1$; when $a = 0$, it follows directly from (4.9). □

5. Proof of Proposition 2.2

The preceding section produced finite sets of admissible words with uniform fibres and factor cardinalities that give a required entropy lower bound. We now use these sets to choose independent random source blocks, join the chosen blocks by weak specification, and push the resulting product measure forward to K. The following proposition gives the quantitative estimate needed for Proposition 2.2.

Proposition 5.1. *Let X be a subshift satisfying weak specification and set $K = R(X)$. Suppose that $\eta, \xi \in \mathcal{M}_\sigma(X)$ satisfy the following conditions.*

(i) *for every $\omega \in \{\eta, \xi\}$, $1 \le i \le s$, and $N \ge 1$, $H((\tau_i\omega)|\mathcal{L}_N(X_i)) = Nh(\tau_i\omega) + O_\omega(1)$;*
(ii) *η is an $\mathbf{a}$-weighted equilibrium state, i.e., $\sum_{i=1}^{s} a_i h(\tau_i\eta) = P^{\mathbf{a}}(\sigma)$;*
(iii) *there is $\beta \in (1/2, 1)$ such that $\sum_{i=1}^{s} \Delta_i(\theta_i^\beta - \theta_{i-1}^\beta) > 0$, where $\Delta_i = h(\tau_i\xi) - h(\tau_i\eta)$ for $1 \le i \le s$.*

Then there exist a Borel probability measure λ supported on K, a constant $c_0 > 0$, and an integer k_0 such that

$$\log \lambda(Q_k(z)) + kP^{\mathbf{a}}(\sigma) \le -c_0 k^\beta \tag{5.1}$$

for every $z \in K$ and every $k \ge k_0$.

We first derive Proposition 2.2 from Proposition 5.1.

Proof of Proposition 2.2 based on Proposition 5.1. Let $\eta = \eta_{\mathbf{a}}$ and $\xi = \eta_{\mathbf{a}(\beta)}$. Then $P^{\mathbf{a}(\beta)}(\sigma) = \sum_{i=1}^{s} a_i(\beta) h(\tau_i\xi)$. Proposition 2.1(iii) shows that η and ξ satisfy condition (i) of Proposition 5.1. Since $\eta = \eta_{\mathbf{a}}$, Proposition 2.1(ii) gives condition (ii). With Δ_i as in condition (iii), the definition of $c(\beta)$ in (2.9) gives

$$\sum_{i=1}^{s} \Delta_i(\theta_i^\beta - \theta_{i-1}^\beta) = \sum_{i=1}^{s} a_i(\beta)\big(h(\tau_i\xi) - h(\tau_i\eta)\big) = c(\beta) > 0.$$

Proposition 5.1 gives a Borel probability measure λ on K and constants $c_0 > 0$ and k_0 such that, for every $z \in K$ and $k \ge k_0$,

$$\lambda(Q_k(z)) \le \exp\{-kP^{\mathbf{a}}(\sigma) - c_0 k^\beta\}. \tag{5.2}$$

Write $L = \log n_s$ and $\rho_k = n_s^{-k} = e^{-kL}$ for $k \geq 1$. By (2.5), $P^{\mathbf{a}}(\sigma) = \gamma \log n_s = \gamma L$. Choose $0 < \varepsilon < c_0/L^\beta$ and write $\delta = c_0 - \varepsilon L^\beta > 0$. With this value of ε, let g_β be a gauge function satisfying $g_\beta(r) = r^\gamma \exp\{-\varepsilon(\log(1/r))^\beta\}$ for $0 < r \leq 1$. Then

$$g_\beta(\rho_k) = \exp\{-kP^{\mathbf{a}}(\sigma) - \varepsilon L^\beta k^\beta\}. \tag{5.3}$$

For every $z \in K$, equations (5.2) and (5.3) give

$$0 \leq \frac{\lambda(Q_k(z))}{g_\beta(\rho_k)} \leq e^{-\delta k^\beta} \longrightarrow 0 \qquad k \to \infty. \tag{5.4}$$

Lemma 3.1, applied to (5.4), shows that $\mathcal{H}^{g_\beta}(K) = \infty$ and that $\mathcal{H}^{g_\beta}$ is not σ-finite on K. Since $g_\beta(r) \leq r^\gamma$, we have $\mathcal{H}^{g_\beta}(E) \leq \mathcal{H}^\gamma(E)$ for every $E \subset K$. Therefore $\mathcal{H}^\gamma(K) = \infty$, and $\mathcal{H}^\gamma$ is not σ-finite on K. □

The proof of Proposition 5.1 has four parts. We first choose the source-block coordinate sets C_r in Lemma 4.1, use Lemma 4.5 to schedule them, and construct the product measure $\mathbb{P}$. We then use weak specification to insert joining words and define the push-forward measure λ. Next, we estimate the λ-mass of approximate cubes. Finally, the entropy estimates from Corollary 4.4, together with the block schedule, prove Proposition 5.1.

5.1. **Measure construction.** We shall construct a product space whose coordinates are admissible words, called *source blocks*. We first prescribe the length of every source block and then choose the finite set of words allowed at each coordinate. Fix a value $\beta \in (1/2, 1)$ satisfying condition (iii) of Proposition 5.1.

For $j \geq 1$, let $R_j = L_j = 2^j$ and $Q_j = \sum_{h=1}^{j} R_h = 2^{j+1} - 2$, with $Q_0 = 0$. For $j \geq 1$, the source blocks with indices $Q_{j-1} < r \leq Q_j$ form stage j. For every such r, let the r-th source block have length $L(r) = L_j$. For $r \geq 1$, place the first r source blocks consecutively and denote their right endpoint by $\overline{s}_r = \sum_{u=1}^{r} L(u)$; set $\overline{s}_0 = 0$. For $j \geq 1$, denote the right endpoint of stage j in these source coordinates by $\overline{S}_j = \overline{s}_{Q_j} = \sum_{h=1}^{j} R_h L_h$; set $\overline{S}_0 = 0$. For $n \in \mathbb{N} \cup \{0\}$, let $f(n) = (n+1)^{\beta-1}$, so that $0 < f(n) \leq 1$.

Lemma 5.2. *Under the assumptions of Proposition 5.1, there exist a sequence $(\varepsilon_r)_{r\geq 1} \in \{0,1\}^{\mathbb{N}}$ and finite nonempty sets*

$$C_r \subset \mathcal{L}_{L(r)}(X), \qquad r \geq 1,$$

each having uniform fibres at every level. Set

$$\Omega = \prod_{r=1}^{\infty} C_r.$$

If ν_r is the uniform probability measure on C_r, then $\mathbb{P} = \bigotimes_{r=1}^{\infty} \nu_r$ is a Borel probability measure on Ω.

In particular, for every $m \geq 1$ and every choice $J_r \in C_r$, $1 \leq r \leq m$,

$$\mathbb{P}\{(I_r)_{r\geq 1} \in \Omega : I_1 = J_1, \ldots, I_m = J_m\} = \prod_{r=1}^{m} \frac{1}{\#C_r}. \tag{5.5}$$

The set C_r is selected from the block construction for η when $\varepsilon_r = 0$, and from the block construction for ξ when $\varepsilon_r = 1$.

The proof is straightforward. The sequence $(\varepsilon_r)_{r\geq 1}$ and the sets C_r will be chosen in a particular way for later estimates.

Choice of $(\varepsilon_r)_{r\ge1}$ *in Lemma 5.2.* For $r \ge 1$, let the interval of the r-th source block be $\overline{\mathcal{I}}_r = (\overline{s}_{r-1}, \overline{s}_r] \cap \mathbb{N}$:

$$x_r = \frac{1}{L(r)} \sum_{n\in\overline{\mathcal{I}}_r} f(n) \in [0,1].$$

For each $j \ge 1$ and $1 \le u \le R_j$, write

$$r(j,u) = Q_{j-1} + u, \qquad A_{j,u} = \sum_{v=1}^{u} x_{r(j,v)}, \tag{5.6}$$

and set $A_{j,0} = 0$. Since $A_{j,u} - A_{j,u-1} = x_{r(j,u)} \in [0,1]$, we choose

$$\varepsilon_{r(j,u)} = \lfloor A_{j,u} \rfloor - \lfloor A_{j,u-1} \rfloor \in \{0,1\}. \tag{5.7}$$

We assign the r-th source block to η when $\varepsilon_r = 0$ and to ξ when $\varepsilon_r = 1$.

Choice of $(C_r)_{r\ge1}$ *in Lemma 5.2.* Condition (i) of Proposition 5.1 allows us to apply Corollary 4.4 to both η and ξ. Fix j and $\omega \in \{\eta, \xi\}$. Applying Corollary 4.4 with $N = L_j$ gives $m_{j,\omega} \le s+1$, sets $C_{j,\omega,1}, \dots, C_{j,\omega,m_{j,\omega}}$, and weights $\rho_{j,\omega,1}, \dots, \rho_{j,\omega,m_{j,\omega}}$. Let $N_{j,\omega}$ be the number of source blocks in stage j assigned to ω. When $N_{j,\omega} > 0$, list them as $r_{j,\omega}(1) < \cdots < r_{j,\omega}(N_{j,\omega})$. Lemma 4.5 gives indices $u_{j,\omega}(v) \in \{1, \dots, m_{j,\omega}\}$, $1 \le v \le N_{j,\omega}$, such that, for every $1 \le q \le m_{j,\omega}$ and $0 \le a < b \le N_{j,\omega}$,

$$\left|\#\{a < v \le b : u_{j,\omega}(v) = q\} - \rho_{j,\omega,q}(b-a)\right| \le 2s. \tag{5.8}$$

For $1 \le v \le N_{j,\omega}$, we choose the $r_{j,\omega}(v)$-th source block set to be

$$C_{r_{j,\omega}(v)} = C_{j,\omega,u_{j,\omega}(v)}.$$

For $t \in \mathbb{N}$, write $E_\beta(t) = \sum_{n=1}^{t} f(n)$, and set $E_\beta(0) = 0$. To describe the source blocks before the joining words are inserted, define, for every integer $t \ge 1$,

$$r_*(t) = \min\{r \ge 1 : t \le \overline{s}_r\}, \qquad L_*(t) = L(r_*(t)).$$

Thus $r_*(t)$ and $L_*(t)$ give, respectively, the number and the maximum length of the source blocks that meet the first t digit positions before the joining words are inserted.

We now introduce the joining words. For each source-block sequence $\mathbf{I} = (I_r)_{r\ge1} \in \Omega$, we shall insert joining words and obtain

$$I_1 W_1(\mathbf{I}) I_2 W_2(\mathbf{I}) I_3 W_3(\mathbf{I}) \cdots \in X, \tag{5.9}$$

Here $W_r(\mathbf{I})$ is the *joining word* placed between I_r and I_{r+1}, and $|W_r(\mathbf{I})| \le p$ by weak specification. The sequence in (5.9) is called the *concatenated sequence*. Let $S_j(\mathbf{I})$ be the position of the last digit of the final source block in stage j, including all joining words preceding that block, and let $\mathcal{I}_r(\mathbf{I}) \subset \mathbb{N}$ be the interval occupied by I_r in the concatenated sequence in (5.9). For every integer $t \ge 1$, let

$$M_{\mathbf{I}}(t) = \sum_{r\ge1} \varepsilon_r \#\big(\mathcal{I}_r(\mathbf{I}) \cap [1,t]\big).$$

Set $M_{\mathbf{I}}(0) = 0$. Thus $M_{\mathbf{I}}(t)$ counts the digits belonging to ξ-blocks among the first t positions of the concatenated sequence in (5.9). It does not count digits in joining words.

Lemma 5.3. *Let $(\varepsilon_r)_{r\ge1}$, $(C_r)_{r\ge1}$, Ω, and $\mathbb{P}$ be given by Lemma 5.2. There is a continuous map $\Pi : \Omega \to X$. Define the push-forward measure by*

$$\lambda = (R \circ \Pi)_* \mathbb{P}. \tag{5.10}$$

Then λ is a Borel probability measure on K. The following estimates hold for $t \ge 1$.

$$S_j(\mathbf{I}) \asymp 4^j, \qquad r_*(t) = O(t^{1/2}), \qquad L_*(t) = O(t^{1/2}). \tag{5.11}$$

Moreover, for every $\mathbf{I} \in \Omega$ and $t \ge 1$,

$$|M_{\mathbf{I}}(t) - E_\beta(t)| \lesssim L_*(t) + p\, r_*(t). \tag{5.12}$$

Proof. We first construct the map Π and the probability measure λ on K. Let $\mathcal{L}_{\le p}(X)$ be the finite set of admissible words of length at most p. Weak specification implies that, for every $U, V \in \mathcal{L}(X)$, the set

$$\{W \in \mathcal{L}_{\le p}(X) : UWV \in \mathcal{L}(X)\}$$

is nonempty. For each pair (U, V), choose one word from this set and denote it by $\mathscr{W}(U, V)$. This choice is fixed throughout the construction.

Fix $\mathbf{I} = (I_r)_{r\ge1} \in \Omega$. For $r \ge 1$, denote

$$W_r(\mathbf{I}) = \mathscr{W}(I_r, I_{r+1}). \tag{5.13}$$

Every finite prefix of the following infinite word is admissible, and therefore

$$\Pi(\mathbf{I}) = I_1 W_1(\mathbf{I}) I_2 W_2(\mathbf{I}) I_3 W_3(\mathbf{I}) \cdots \in X.$$

For every n, the first n digits of $\Pi(\mathbf{I})$ depend on only finitely many coordinates I_r of $\mathbf{I}$. Hence the inverse image under Π of every cylinder in X is open in Ω. Thus Π is continuous and, in particular, Borel measurable. The composition $R \circ \Pi : \Omega \to K$ is Borel measurable. With λ given by (5.10), λ is therefore a Borel probability measure on K.

For the estimates, fix $\mathbf{I} \in \Omega$ and abbreviate $W_r = W_r(\mathbf{I})$, $S_j = S_j(\mathbf{I})$, $\mathcal{I}_r = \mathcal{I}_r(\mathbf{I})$, and $M(t) = M_{\mathbf{I}}(t)$. Since $R_h = L_h = 2^h$,

$$\overline{S}_j = \sum_{h=1}^{j} R_h L_h = \sum_{h=1}^{j} 4^h \asymp 4^j.$$

For $t \ge 1$, choose $j \ge 1$ so that $\overline{S}_{j-1} \le t < \overline{S}_j$. Then

$$r_*(t) \le Q_j \le 2^{j+1}, \qquad L_*(t) \le L_j = 2^j.$$

If $j \ge 2$, then $t \ge \overline{S}_{j-1} \ge 4^{j-1}$, and hence

$$r_*(t) = O(t^{1/2}), \qquad L_*(t) = O(t^{1/2}). \tag{5.14}$$

By its definition,

$$S_j = \overline{S}_j + \sum_{r=1}^{Q_j - 1} |W_r|.$$

Consequently,

$$\overline{S}_j \le S_j \le \overline{S}_j + pQ_j, \qquad S_j \asymp 4^j. \tag{5.15}$$

Equations (5.14) and (5.15) prove (5.11).

Inserting joining words only moves a source block to the right. Therefore, if I_r meets one of the first t positions of the concatenated sequence, then $r \le r_*(t)$. By the definition of $\mathcal{I}_r$,

$$0 \le \min \mathcal{I}_r - (\overline{s}_{r-1} + 1) = \sum_{u<r} |W_u| \le p(r-1).$$

For $t \in \mathbb{N}$, let $\overline{M}(t)$ be the number of digits belonging to ξ-blocks among the first t positions before the joining words are inserted, and set $\overline{M}(0) = 0$. For $t \geq 1$, choose the unique integers $j \geq 1$, $0 \leq u < R_j$, and $0 \leq q < L_j$ such that

$$t = \overline{S}_{j-1} + uL_j + q.$$

With $T = \overline{S}_{j-1} + uL_j$, and using $r(h, v) = Q_{h-1} + v$ from (5.6), we have

$$\overline{M}(t) = \sum_{h<j} L_h \sum_{v=1}^{R_h} \varepsilon_{r(h,v)} + L_j \sum_{v=1}^{u} \varepsilon_{r(j,v)} + q\varepsilon_{r(j,u+1)},$$
$$E_\beta(t) = \sum_{h<j} L_h \sum_{v=1}^{R_h} x_{r(h,v)} + L_j \sum_{v=1}^{u} x_{r(j,v)} + \sum_{n=T+1}^{T+q} f(n).$$

For every $h \geq 1$ and $0 \leq w \leq R_h$, summing (5.7) gives

$$\sum_{v=1}^{w} \varepsilon_{r(h,v)} = \lfloor A_{h,w} \rfloor, \qquad \left| \sum_{v=1}^{w} \varepsilon_{r(h,v)} - \sum_{v=1}^{w} x_{r(h,v)} \right| = \left| \lfloor A_{h,w} \rfloor - A_{h,w} \right| < 1.$$

Apply this estimate with $w = R_h$ for every $h < j$, and then with $(h, w) = (j, u)$. Since $0 \leq f \leq 1$,

$$\left| \overline{M}(t) - E_\beta(t) \right| \leq \sum_{h<j} L_h + L_j + q \leq \sum_{h<j} L_h + 2L_j \lesssim L_j \lesssim L_*(t). \tag{5.16}$$

For an integer $t \geq 1$, let $V(t)$ be the number of joining-word digits among the first t positions of the concatenated sequence, and write $U(t) = t - V(t)$. Since the first position belongs to I_1, we have $U(t) \geq 1$. By the definitions of $M(t)$, $\overline{M}(t)$, and $U(t)$,

$$M(t) = \overline{M}(U(t)), \qquad 1 \leq U(t) \leq t.$$

Moreover,

$$V(t) \leq \sum_{r=1}^{r_*(U(t))} |W_r| \leq p\, r_*(U(t)).$$

Since $0 \leq f(n) \leq 1$,

$$|E_\beta(t) - E_\beta(U(t))| = \sum_{n=U(t)+1}^{t} f(n) \leq t - U(t) = V(t).$$

Applying (5.16) and the monotonicity of L_* and r_* gives

$$\begin{aligned} |M(t) - E_\beta(t)| &\leq \left| \overline{M}(U(t)) - E_\beta(U(t)) \right| + |E_\beta(U(t)) - E_\beta(t)| \\ &\lesssim L_*(U(t)) + p\, r_*(U(t)) \\ &\lesssim L_*(t) + p\, r_*(t), \end{aligned}$$

which proves (5.12) uniformly for $\mathbf{I} \in \Omega$. □

5.2. **The mass for approximate cubes.** Lemma 5.2 makes the source blocks independent, whereas each joining word inserted in Lemma 5.3 depends on the two source blocks it joins. Their lengths, however, take only $p+1$ possible values. We therefore first fix the entire joining-length pattern and then drop the requirement that the joining words have the prescribed lengths, which can only enlarge the event. The remaining conditions concern distinct source blocks and, by their independence under the original product measure, the probability of the enlarged event is the product of the corresponding probabilities.

For the sets C_r constructed in Lemma 5.2, write $B_{r,i} = \#\tau_i(C_r)$ for $r \geq 1$ and $1 \leq i \leq s$.

Lemma 5.4. *Let λ be the probability measure constructed in Lemma 5.3 from the sets C_r given by Lemma 5.2. Fix an integer $k \geq 1$. For $1 \leq i \leq s$, let $I_i(k) = (\lfloor \theta_{i-1} k \rfloor, \lfloor \theta_i k \rfloor] \cap \mathbb{N}$. Write also $r_k = r_*(k)$. Let*

$$\mathfrak{L}_k = \{(|W_1(\mathbf{I})|, \dots, |W_{r_k}(\mathbf{I})|) : \mathbf{I} \in \Omega\},$$

where $W_r(\mathbf{I})$ is defined in (5.13). For $\boldsymbol{\ell} = (\ell_1, \dots, \ell_{r_k}) \in \mathfrak{L}_k$ and $1 \leq r \leq r_k$, define the left and right endpoints of the r-th source block by

$$a_r(\boldsymbol{\ell}) = \overline{s}_{r-1} + \sum_{u=1}^{r-1} \ell_u, \qquad b_r(\boldsymbol{\ell}) = a_r(\boldsymbol{\ell}) + L(r), \tag{5.17}$$

and let $\mathcal{I}_r(\boldsymbol{\ell}) = (a_r(\boldsymbol{\ell}), b_r(\boldsymbol{\ell})] \cap \mathbb{N}$. For $1 \leq i \leq s$, let $\mathcal{B}_{i,k}(\boldsymbol{\ell}) = \{1 \leq r \leq r_k : \mathcal{I}_r(\boldsymbol{\ell}) \subset I_i(k)\}$. Then

$$\#\mathfrak{L}_k \leq (p+1)^{r_k}. \tag{5.18}$$

Moreover, for every $z \in K$,

$$\lambda(Q_k(z)) \leq 2^d (p+1)^{r_k} \max_{\boldsymbol{\ell} \in \mathfrak{L}_k} \prod_{i=1}^{s} \prod_{r \in \mathcal{B}_{i,k}(\boldsymbol{\ell})} B_{r,i}^{-1}. \tag{5.19}$$

Proof. Fix $\mathbf{I} \in \Omega$. If the source block I_r or the joining word $W_r(\mathbf{I})$ meets the first k positions of the concatenated sequence, then $r \leq r_k$. Consequently, every source block contained in one of the intervals $I_i(k)$ has index at most r_k, and its location is determined by $|W_1(\mathbf{I})|, \dots, |W_{r_k}(\mathbf{I})|$. Since this holds for every $\mathbf{I} \in \Omega$ and every joining-word length belongs to $\{0, \dots, p\}$, there are at most $(p+1)^{r_k}$ possible length patterns. This proves (5.18).

For $\boldsymbol{\ell} \in \mathfrak{L}_k$, let

$$\Omega_{\boldsymbol{\ell}} = \{\mathbf{I} = (I_u)_{u \geq 1} \in \Omega : (|W_1(\mathbf{I})|, \dots, |W_{r_k}(\mathbf{I})|) = \boldsymbol{\ell}\}.$$

The word $W_r(\mathbf{I})$ depends only on $I_1, \dots, I_{r+1}$, so every $\Omega_{\boldsymbol{\ell}}$ is a Borel subset of Ω. These sets are pairwise disjoint and their union is the whole product probability space:

$$\Omega = \bigsqcup_{\boldsymbol{\ell} \in \mathfrak{L}_k} \Omega_{\boldsymbol{\ell}}. \tag{5.20}$$

By (5.17), for every $\boldsymbol{\ell} \in \mathfrak{L}_k$, $\mathbf{I} \in \Omega_{\boldsymbol{\ell}}$, and $1 \leq r \leq r_k$, the r-th source block occupies the interval $\mathcal{I}_r(\boldsymbol{\ell})$ in the concatenated sequence. For $r \geq 1$, let $\mathrm{pr}_r : \Omega \to C_r$ be the r-th coordinate projection, so that $\mathrm{pr}_r((I_u)_{u \geq 1}) = I_r$. Fix $x \in X$ and $\boldsymbol{\ell} \in \mathfrak{L}_k$. For $1 \leq i \leq s$ and $r \in \mathcal{B}_{i,k}(\boldsymbol{\ell})$, let

$$A_{r,i}(x, \boldsymbol{\ell}) = \left\{\mathbf{I} = (I_u)_{u \geq 1} \in \Omega : \tau_i(\mathrm{pr}_r(\mathbf{I})) = \tau_i\big(x|_{a_r(\boldsymbol{\ell})}^{b_r(\boldsymbol{\ell})}\big)\right\}.$$

Thus $A_{r,i}(x,\boldsymbol{\ell})$ is the set of source-block sequences for which the i-th factor of the r-th source block agrees with the corresponding factor word of x. The definition of $\Gamma_k(x)$ in (3.2) gives

$$\Pi^{-1}([\Gamma_k(x)])\cap\Omega_{\boldsymbol{\ell}}\subseteq\bigcap_{i=1}^{s}\bigcap_{r\in\mathcal{B}_{i,k}(\boldsymbol{\ell})}A_{r,i}(x,\boldsymbol{\ell}).$$

Indeed, let $\mathbf{I}$ belong to the set on the left, and fix $1\le i\le s$ and $r\in\mathcal{B}_{i,k}(\boldsymbol{\ell})$. Since $\mathbf{I}\in\Omega_{\boldsymbol{\ell}}$, the r-th source block occupies the interval $(a_r(\boldsymbol{\ell}),b_r(\boldsymbol{\ell})]$. This interval is contained in $I_i(k)$ by the definition of $\mathcal{B}_{i,k}(\boldsymbol{\ell})$. On the other hand, since $\Pi(\mathbf{I})\in[\Gamma_k(x)]$, we have

$$\tau_i\big(\Pi(\mathbf{I})|_{\lfloor\theta_i k\rfloor}\big)=\tau_i\big(x|_{\lfloor\theta_i k\rfloor}\big).$$

Restricting this equality to $(a_r(\boldsymbol{\ell}),b_r(\boldsymbol{\ell})]$ yields

$$\tau_i(I_r)=\tau_i\big(x|_{a_r(\boldsymbol{\ell})}^{b_r(\boldsymbol{\ell})}\big).$$

Thus $\mathbf{I}\in A_{r,i}(x,\boldsymbol{\ell})$, which proves the inclusion.

The intervals $I_i(k)$ are disjoint, so each block index r belongs to at most one set $\mathcal{B}_{i,k}(\boldsymbol{\ell})$. Consequently, the sets $A_{r,i}(x,\boldsymbol{\ell})$ in the intersection are determined by pairwise distinct coordinates of Ω, and, by (5.5), their probabilities multiply under $\mathbb{P}$. The r-th marginal of $\mathbb{P}$ is uniform on C_r, so (4.7) gives

$$\mathbb{P}(A_{r,i}(x,\boldsymbol{\ell}))\le B_{r,i}^{-1};$$

the left-hand side is 0 if the required factor word is not in $\tau_i(C_r)$, and otherwise it equals $B_{r,i}^{-1}$. Hence

$$\begin{aligned}\mathbb{P}\big(\Pi^{-1}([\Gamma_k(x)])\cap\Omega_{\boldsymbol{\ell}}\big)&\le\mathbb{P}\left(\bigcap_{i=1}^{s}\bigcap_{r\in\mathcal{B}_{i,k}(\boldsymbol{\ell})}A_{r,i}(x,\boldsymbol{\ell})\right)\\&=\prod_{i=1}^{s}\prod_{r\in\mathcal{B}_{i,k}(\boldsymbol{\ell})}\mathbb{P}(A_{r,i}(x,\boldsymbol{\ell}))\\&\le\prod_{i=1}^{s}\prod_{r\in\mathcal{B}_{i,k}(\boldsymbol{\ell})}B_{r,i}^{-1}.\end{aligned}$$

Summing over the partition in (5.20) and using (5.18) gives

$$\begin{aligned}\mathbb{P}\big(\Pi^{-1}([\Gamma_k(x)])\big)&=\sum_{\boldsymbol{\ell}\in\mathfrak{L}_k}\mathbb{P}\big(\Pi^{-1}([\Gamma_k(x)])\cap\Omega_{\boldsymbol{\ell}}\big)\\&\le\sum_{\boldsymbol{\ell}\in\mathfrak{L}_k}\prod_{i=1}^{s}\prod_{r\in\mathcal{B}_{i,k}(\boldsymbol{\ell})}B_{r,i}^{-1}\\&\le(p+1)^{r_k}\max_{\boldsymbol{\ell}\in\mathfrak{L}_k}\prod_{i=1}^{s}\prod_{r\in\mathcal{B}_{i,k}(\boldsymbol{\ell})}B_{r,i}^{-1}.\end{aligned}\tag{5.21}$$

Fix $z\in K$ and write $Q=Q_k(z)$. By (3.3), there are $x^{(1)},\dots,x^{(L)}\in X$, with $L\le 2^d$, such that

$$R^{-1}(Q)\cap X\subseteq\bigcup_{h=1}^{L}[\Gamma_k(x^{(h)})].$$

By (5.10), $\lambda(Q) = \mathbb{P}((R \circ \Pi)^{-1}(Q))$. Therefore

$$\lambda(Q) \le \sum_{h=1}^{L} \mathbb{P}\Big(\Pi^{-1}([\Gamma_k(x^{(h)})])\Big).$$

Applying (5.21) to each $x^{(h)}$ and using $L \le 2^d$ proves (5.19). □

We now turn the product estimate in (5.19) into an entropy lower bound. Applying Corollary 4.4 block by block yields the estimate in the following lemma.

Lemma 5.5. *For the measure λ constructed in Lemma 5.3, the following estimate holds uniformly for $z \in K$ as $k \to \infty$:*

$$\begin{aligned} -\log \lambda(Q_k(z)) \ge \sum_{i=1}^{s} (\lfloor \theta_i k \rfloor - \lfloor \theta_{i-1} k \rfloor) h(\tau_i \eta) \\ + \sum_{i=1}^{s} \Delta_i \big(E_\beta(\lfloor \theta_i k \rfloor) - E_\beta(\lfloor \theta_{i-1} k \rfloor)\big) - o(k^\beta). \end{aligned} \tag{5.22}$$

Here $\Delta_i = h(\tau_i \xi) - h(\tau_i \eta)$.

Proof. Fix $k \ge 2$. Write $k_i = \lfloor \theta_i k \rfloor$ for $0 \le i \le s$. For $1 \le i \le s$, write $D_i = k_i - k_{i-1}$ and $I_i(k) = (k_{i-1}, k_i] \cap \mathbb{N}$. Let $r_k = r_*(k)$ and set

$$\mathcal{E}_k = r_k \log k + L_*(k) + p\, r_k + r_k \log(p+1).$$

By (5.11) and $\beta > 1/2$,

$$\frac{\mathcal{E}_k}{k^\beta} = O\big(k^{1/2-\beta} \log k\big) \longrightarrow 0 \qquad k \to \infty. \tag{5.23}$$

Fix $1 \le i \le s$, a stage $j \ge 1$, and a measure $\omega \in \{\eta, \xi\}$. Let $\mathcal{J}$ be the set of block indices in a consecutive segment of the sequence $(r_{j,\omega}(v))_{1 \le v \le N_{j,\omega}}$. Thus, for some $0 \le a < b \le N_{j,\omega}$, $\mathcal{J} = \{r_{j,\omega}(v) : a < v \le b\}$. Let $n = b - a = \#\mathcal{J}$ and, for $1 \le q \le m_{j,\omega}$, write

$$n_q = \#\{a < v \le b : u_{j,\omega}(v) = q\}, \qquad B_{j,\omega,q,i} = \#\tau_i(C_{j,\omega,q}).$$

Recall that $\mathcal{S}_i = \tau_i(\mathcal{S})$ is the set of digits that occur in X_i. Corollary 4.4 and (5.8) give

$$\begin{aligned} \sum_{q=1}^{m_{j,\omega}} \rho_{j,\omega,q} \log B_{j,\omega,q,i} \ge L_j h(\tau_i \omega) - O(\log L_j), \\ |n_q - \rho_{j,\omega,q} n| \le 2s, \\ \log B_{j,\omega,q,i} \le L_j \log \#\mathcal{S}_i. \end{aligned}$$

Therefore

$$\begin{aligned} \sum_{r \in \mathcal{J}} \log B_{r,i} &= \sum_{q=1}^{m_{j,\omega}} n_q \log B_{j,\omega,q,i} \\ &= n \sum_{q=1}^{m_{j,\omega}} \rho_{j,\omega,q} \log B_{j,\omega,q,i} + \sum_{q=1}^{m_{j,\omega}} (n_q - \rho_{j,\omega,q} n) \log B_{j,\omega,q,i} \\ &\ge n L_j h(\tau_i \omega) - O\big(n \log L_j\big) - 2s \sum_{q=1}^{m_{j,\omega}} \log B_{j,\omega,q,i} \\ &\ge n L_j h(\tau_i \omega) - O\big(n \log L_j + L_j\big). \end{aligned} \tag{5.24}$$

We now return to all factor levels. Fix $\boldsymbol{\ell} \in \mathfrak{L}_k$. For each $1 \le i \le s$, write

$$\ell_{\eta,i,k}(\boldsymbol{\ell}) = \sum_{\substack{r \in \mathcal{B}_{i,k}(\boldsymbol{\ell}) \\ \varepsilon_r = 0}} L(r), \qquad \ell_{\xi,i,k}(\boldsymbol{\ell}) = \sum_{\substack{r \in \mathcal{B}_{i,k}(\boldsymbol{\ell}) \\ \varepsilon_r = 1}} L(r).$$

For $j \ge 1$, $1 \le i \le s$, and $\omega \in \{\eta, \xi\}$, let

$$\mathcal{J}_{j,\omega,i}(\boldsymbol{\ell}) = \begin{cases} \{r \in \mathcal{B}_{i,k}(\boldsymbol{\ell}) : Q_{j-1} < r \le Q_j,\ \varepsilon_r = 0\}, & \omega = \eta, \\ \{r \in \mathcal{B}_{i,k}(\boldsymbol{\ell}) : Q_{j-1} < r \le Q_j,\ \varepsilon_r = 1\}, & \omega = \xi. \end{cases}$$

If this set is nonempty, there are integers $a < b$ such that

$$\mathcal{J}_{j,\omega,i}(\boldsymbol{\ell}) = \{r_{j,\omega}(v) : a < v \le b\}.$$

Hence, for each fixed $\boldsymbol{\ell} \in \mathfrak{L}_k$, the set $\mathcal{J}_{j,\omega,i}(\boldsymbol{\ell})$ is a consecutive segment of $(r_{j,\omega}(v))_{1 \le v \le N_{j,\omega}}$. We may therefore take $\mathcal{J} = \mathcal{J}_{j,\omega,i}(\boldsymbol{\ell})$ in (5.24). The index ranges $Q_{j-1} < r \le Q_j$ are disjoint for different stages j, so summing over j counts each block at most once. For the fixed length pattern $\boldsymbol{\ell}$, let $j_k(\boldsymbol{\ell})$ be the largest j for which a stage-j block meets one of the first k positions of the concatenated sequence. For $1 \le j \le j_k(\boldsymbol{\ell})$ and $\omega \in \{\eta, \xi\}$, write

$$n_{j,\omega,i} = \#\mathcal{J}_{j,\omega,i}(\boldsymbol{\ell}).$$

If $n_{j,\omega,i} > 0$, then a complete stage-j block lies in $I_i(k)$, and hence $L_j \le D_i \le k$. Therefore

$$\sum_{j=1}^{j_k(\boldsymbol{\ell})} \sum_{\omega \in \{\eta,\xi\}} n_{j,\omega,i} \le r_k, \qquad \log L_j \le \log k, \qquad \sum_{j=1}^{j_k(\boldsymbol{\ell})} L_j \lesssim L_*(k),$$

where the last inequality follows from $\sum_{j=1}^{j_k(\boldsymbol{\ell})} L_j \lesssim L_{j_k(\boldsymbol{\ell})}$ and $L_{j_k(\boldsymbol{\ell})} \le L_*(k)$. Moreover,

$$\sum_{j=1}^{j_k(\boldsymbol{\ell})} n_{j,\eta,i} L_j = \ell_{\eta,i,k}(\boldsymbol{\ell}), \qquad \sum_{j=1}^{j_k(\boldsymbol{\ell})} n_{j,\xi,i} L_j = \ell_{\xi,i,k}(\boldsymbol{\ell}).$$

For fixed i and $\boldsymbol{\ell}$, the sets $\mathcal{J}_{j,\omega,i}(\boldsymbol{\ell})$, over all j and ω, form a partition of $\mathcal{B}_{i,k}(\boldsymbol{\ell})$. Summing (5.24) over j and ω therefore gives

$$\begin{aligned} \sum_{r \in \mathcal{B}_{i,k}(\boldsymbol{\ell})} \log B_{r,i} &= \sum_{j=1}^{j_k(\boldsymbol{\ell})} \sum_{\omega \in \{\eta,\xi\}} \sum_{r \in \mathcal{J}_{j,\omega,i}(\boldsymbol{\ell})} \log B_{r,i} \\ &\ge \sum_{j=1}^{j_k(\boldsymbol{\ell})} \big(n_{j,\eta,i} L_j h(\tau_i \eta) + n_{j,\xi,i} L_j h(\tau_i \xi)\big) - O\left(\sum_{j=1}^{j_k(\boldsymbol{\ell})} \sum_{\omega \in \{\eta,\xi\}} n_{j,\omega,i} \log L_j + \sum_{j=1}^{j_k(\boldsymbol{\ell})} L_j\right) \\ &\ge h(\tau_i \eta) \ell_{\eta,i,k}(\boldsymbol{\ell}) + h(\tau_i \xi) \ell_{\xi,i,k}(\boldsymbol{\ell}) - O\big(r_k \log k + L_*(k)\big). \end{aligned} \tag{5.25}$$

The positions in $I_i(k)$ not counted by $\ell_{\eta,i,k} + \ell_{\xi,i,k}$ lie in joining words or in the parts of at most two source blocks cut by the endpoints. Thus

$$0 \le D_i - \ell_{\eta,i,k}(\boldsymbol{\ell}) - \ell_{\xi,i,k}(\boldsymbol{\ell}) \le p\, r_k + 2L_*(k). \tag{5.26}$$

For a fixed length pattern $\boldsymbol{\ell}$, the values $M_{\mathbf{I}}(t)$, $1 \le t \le k$, are the same for all $\mathbf{I} \in \Omega_{\boldsymbol{\ell}}$; denote them by $M_{\boldsymbol{\ell}}(t)$, and set $M_{\boldsymbol{\ell}}(0) = 0$.

The difference $M_{\boldsymbol{\ell}}(k_i) - M_{\boldsymbol{\ell}}(k_{i-1})$ counts the ξ-block digits in $I_i(k) = (k_{i-1}, k_i] \cap \mathbb{N}$, including the parts of at most two blocks cut by the endpoints. Hence

$$0 \le M_{\boldsymbol{\ell}}(k_i) - M_{\boldsymbol{\ell}}(k_{i-1}) - \ell_{\xi,i,k}(\boldsymbol{\ell}) \le 2L_*(k). \tag{5.27}$$

For $1 \le i \le s$, write

$$e_{0,i} = D_i - \ell_{\eta,i,k}(\boldsymbol{\ell}) - \ell_{\xi,i,k}(\boldsymbol{\ell}), \qquad e_{1,i} = M_{\boldsymbol{\ell}}(k_i) - M_{\boldsymbol{\ell}}(k_{i-1}) - \ell_{\xi,i,k}(\boldsymbol{\ell}).$$

Equations (5.26) and (5.27) give

$$0 \le e_{0,i} \le p\, r_k + 2L_*(k), \qquad 0 \le e_{1,i} \le 2L_*(k).$$

Since $h(\tau_i \xi) = h(\tau_i \eta) + \Delta_i$, the two main terms on the right-hand side of (5.25) are exactly

$$\begin{aligned}
&h(\tau_i\eta)\ell_{\eta,i,k}(\boldsymbol{\ell}) + h(\tau_i\xi)\ell_{\xi,i,k}(\boldsymbol{\ell}) \\
&= h(\tau_i\eta)\big(\ell_{\eta,i,k}(\boldsymbol{\ell}) + \ell_{\xi,i,k}(\boldsymbol{\ell})\big) + \Delta_i \ell_{\xi,i,k}(\boldsymbol{\ell}) \\
&= h(\tau_i\eta)(D_i - e_{0,i}) + \Delta_i\big(M_{\boldsymbol{\ell}}(k_i) - M_{\boldsymbol{\ell}}(k_{i-1}) - e_{1,i}\big) \\
&= D_i h(\tau_i\eta) + \Delta_i\big(M_{\boldsymbol{\ell}}(k_i) - M_{\boldsymbol{\ell}}(k_{i-1})\big) - h(\tau_i\eta)e_{0,i} - \Delta_i e_{1,i}.
\end{aligned}$$

Since the entropies and the numbers Δ_i are fixed, (5.25) and the preceding bounds give

$$\begin{aligned}
\sum_{r \in \mathcal{B}_{i,k}(\boldsymbol{\ell})} \log B_{r,i} &\ge D_i h(\tau_i\eta) \\
&\quad + \Delta_i\big(M_{\boldsymbol{\ell}}(k_i) - M_{\boldsymbol{\ell}}(k_{i-1})\big) - O\big(r_k \log k + L_*(k) + p\, r_k\big).
\end{aligned} \tag{5.28}$$

Applying (5.12) at both endpoints and using $0 \le k_{i-1} \le k_i \le k$ gives

$$\big|\big(M_{\boldsymbol{\ell}}(k_i) - M_{\boldsymbol{\ell}}(k_{i-1})\big) - \big(E_\beta(k_i) - E_\beta(k_{i-1})\big)\big| \lesssim L_*(k) + p\, r_k.$$

Substitution in (5.28) yields, uniformly in $\boldsymbol{\ell}$,

$$\begin{aligned}
\sum_{r \in \mathcal{B}_{i,k}(\boldsymbol{\ell})} \log B_{r,i} &\ge D_i h(\tau_i\eta) + \Delta_i\big(E_\beta(k_i) - E_\beta(k_{i-1})\big) \\
&\quad - O\big(r_k \log k + L_*(k) + p\, r_k\big).
\end{aligned} \tag{5.29}$$

If $\lambda(Q_k(z)) = 0$, the conclusion is immediate, with $-\log 0 = +\infty$. Otherwise, (5.19) gives

$$-\log \lambda(Q_k(z)) \ge \min_{\boldsymbol{\ell} \in \mathfrak{L}_k} \sum_{i=1}^{s} \sum_{r \in \mathcal{B}_{i,k}(\boldsymbol{\ell})} \log B_{r,i} - r_k \log(p+1) - d \log 2.$$

For every $\boldsymbol{\ell} \in \mathfrak{L}_k$, summing (5.29) over i yields

$$\begin{aligned}
\sum_{i=1}^{s} \sum_{r \in \mathcal{B}_{i,k}(\boldsymbol{\ell})} \log B_{r,i} &\ge \sum_{i=1}^{s} D_i h(\tau_i\eta) \\
&\quad + \sum_{i=1}^{s} \Delta_i\big(E_\beta(k_i) - E_\beta(k_{i-1})\big) - O\big(r_k \log k + L_*(k) + p\, r_k\big).
\end{aligned}$$

The right-hand side is independent of $\boldsymbol{\ell}$. Taking the minimum over $\boldsymbol{\ell}$ and adding the loss $r_k \log(p+1) + d \log 2$ gives a total error of order $O(\mathcal{E}_k)$. Now $\mathcal{E}_k = o(k^\beta)$ by (5.23), which proves (5.22). □

5.3. **Completion of the proof.**

Proof of Proposition 5.1. Let λ be the probability measure constructed in Lemma 5.3. Set

$$A = \sum_{i=1}^{s} \Delta_i(\theta_i^\beta - \theta_{i-1}^\beta) > 0.$$

The function $x \mapsto (x+1)^{\beta-1}$ is decreasing. Hence, for every integer $N \geq 1$,

$$\int_1^{N+1} (x+1)^{\beta-1}\,dx \leq E_\beta(N) \leq \int_0^N (x+1)^{\beta-1}\,dx.$$

Evaluating the two integrals gives

$$E_\beta(N) = \frac{N^\beta}{\beta} + O(1). \tag{5.30}$$

By (5.30), the identities $\lfloor \theta_i k \rfloor = \theta_i k + O(1)$ and $E_\beta(0) = 0$ give

$$\sum_{i=1}^{s} \Delta_i\big(E_\beta(\lfloor \theta_i k \rfloor) - E_\beta(\lfloor \theta_{i-1} k \rfloor)\big) = \frac{k^\beta}{\beta} \sum_{i=1}^{s} \Delta_i(\theta_i^\beta - \theta_{i-1}^\beta) + O(1) = \frac{A}{\beta} k^\beta + O(1). \tag{5.31}$$

Also, by (2.3) and condition (ii),

$$\sum_{i=1}^{s} (\lfloor \theta_i k \rfloor - \lfloor \theta_{i-1} k \rfloor) h(\tau_i \eta) = k \sum_{i=1}^{s} (\theta_i - \theta_{i-1}) h(\tau_i \eta) + O(1) = k P^{\mathbf{a}}(\sigma) + O(1). \tag{5.32}$$

Equations (5.22), (5.31), and (5.32) give

$$-\log \lambda(Q_k(z)) \geq k P^{\mathbf{a}}(\sigma) + \frac{A}{\beta} k^\beta - o(k^\beta).$$

Here the $o(k^\beta)$ term is uniform for $z \in K$. Hence there is k_0 such that, for every $z \in K$ and $k \geq k_0$, the absolute value of this error is at most $Ak^\beta/(2\beta)$. Then

$$\log \lambda(Q_k(z)) + k P^{\mathbf{a}}(\sigma) \leq -\frac{A}{2\beta} k^\beta.$$

Thus (5.1) holds with $c_0 = A/(2\beta)$. □

6. Proof of Proposition 2.3

Recall that $\eta = \eta_{\mathbf{a}}$ is the unique $\mathbf{a}$-weighted equilibrium state from Proposition 2.1(ii). Thus $P^{\mathbf{a}}(\sigma) = \sum_{i=1}^{s} a_i h(\tau_i \eta)$. We prove that $\dim_{\mathrm{H}} K < \dim_{\mathrm{B}} K$ implies $c(\beta) > 0$ for some $\beta \in (1/2, 1)$. Suppose instead that c vanishes on an open interval. We show that η is then the equilibrium state for an open family of weights. The resulting uniform fibre estimates imply $\dim_{\mathrm{H}} K = \dim_{\mathrm{B}} K$, a contradiction.

6.1. **Common equilibrium states.** We first show that if c vanishes on an open interval, then η is the equilibrium state for an open family of weights.

For a positive probability weight vector $\mathbf{b} = (b_1, \ldots, b_s)$, write

$$A_i(\mathbf{b}) = \sum_{\ell=1}^{i} b_\ell, \qquad 0 < A_1(\mathbf{b}) < \cdots < A_s(\mathbf{b}) = 1.$$

For $\mathbf{b} = \mathbf{a}(\beta)$, where $\mathbf{a}(\beta)$ is defined in (2.8),

$$A_i(\mathbf{a}(\beta)) = \theta_i^\beta. \tag{6.1}$$

Lemma 6.1. *Let $I \subset (0,1)$ be a nonempty open interval. There are distinct $\beta_1, \dots, \beta_s \in I$ such that $\mathbf{a}(\beta_1), \dots, \mathbf{a}(\beta_s)$ are affinely independent. More explicitly, if real numbers $t_1, \dots, t_s$ satisfy*

$$\sum_{j=1}^{s} t_j = 0, \qquad \sum_{j=1}^{s} t_j \mathbf{a}(\beta_j) = 0,$$

then $t_1 = \cdots = t_s = 0$.

Proof. Write $m = s - 1$. We regard vectors in $\mathbb{R}^m$ as column vectors. The map from $\mathbf{b}$ to $\big(A_1(\mathbf{b}), \dots, A_m(\mathbf{b})\big)^{\mathsf{T}}$ is affine and one-to-one. Consider the column vector

$$F(\beta) = \big(\theta_1^{\beta}, \dots, \theta_m^{\beta}\big)^{\mathsf{T}}.$$

Fix $\beta_0 \in I$, and write $x_i = \log \theta_i$ for $1 \le i \le m$. For $1 \le r \le m$, the r-th derivative of F at β_0 is

$$F^{(r)}(\beta_0) = \big(\theta_1^{\beta_0} x_1^r, \dots, \theta_m^{\beta_0} x_m^r\big)^{\mathsf{T}}.$$

Consequently, factoring $\theta_i^{\beta_0} x_i$ from row i gives

$$\begin{aligned} \det\big(F'(\beta_0), \dots, F^{(m)}(\beta_0)\big) &= \prod_{i=1}^{m} \theta_i^{\beta_0} x_i \cdot \det(x_i^{r-1})_{1 \le i,r \le m} \\ &= \prod_{i=1}^{m} \theta_i^{\beta_0} \log \theta_i \prod_{1 \le i < j \le m} (\log \theta_j - \log \theta_i). \end{aligned} \tag{6.2}$$

Since $0 < \theta_1 < \cdots < \theta_m < 1$, (6.2) is nonzero.

Choose distinct real numbers $u_1, \dots, u_s$. For $2 \le j \le s$, Taylor's formula at β_0 gives

$$F(\beta_0 + \varepsilon u_j) - F(\beta_0 + \varepsilon u_1) = \sum_{r=1}^{m} \frac{\varepsilon^r (u_j^r - u_1^r)}{r!} F^{(r)}(\beta_0) + O(\varepsilon^{m+1}). \tag{6.3}$$

Let C be the $m \times m$ matrix whose entry in row r and column $j-1$ is $(u_j^r - u_1^r)/r!$. Subtracting the first column from every other column in the usual Vandermonde determinant for $u_1, \dots, u_s$, and then expanding along its first row, gives

$$\det C = \frac{1}{\prod_{r=1}^{m} r!} \prod_{1 \le i < j \le s} (u_j - u_i) \ne 0. \tag{6.4}$$

By the multilinearity of the determinant, (6.3) therefore gives

$$\begin{aligned} &\det\big(F(\beta_0 + \varepsilon u_2) - F(\beta_0 + \varepsilon u_1), \dots, F(\beta_0 + \varepsilon u_s) - F(\beta_0 + \varepsilon u_1)\big) \\ &\quad = \varepsilon^{m(m+1)/2} \det\big(F'(\beta_0), \dots, F^{(m)}(\beta_0)\big) \det C + o\big(\varepsilon^{m(m+1)/2}\big). \end{aligned} \tag{6.5}$$

The coefficient of the leading term is nonzero by (6.2) and (6.4). Thus the determinant in (6.5) is nonzero for every sufficiently small positive ε. We may also choose ε so small that $\beta_j = \beta_0 + \varepsilon u_j$ belongs to I for every $1 \le j \le s$. Since this determinant is nonzero, the vectors $F(\beta_j) - F(\beta_1)$, $2 \le j \le s$, are linearly independent. Hence the points $F(\beta_1), \dots, F(\beta_s)$ are affinely independent.

Finally, (6.1) says that the affine one-to-one map at the beginning of the proof sends $\mathbf{a}(\beta_j)$ to $F(\beta_j)$. Hence $\mathbf{a}(\beta_1), \dots, \mathbf{a}(\beta_s)$ are affinely independent. □

Lemma 6.2. *Suppose $c(\beta) = 0$ for every β in a nonempty open interval I, and let $\eta = \eta_{\mathbf{a}}$. Then there is a relatively open subset U of the positive weight simplex such that η is the unique $\mathbf{b}$-weighted equilibrium state for every $\mathbf{b} \in U$.*

Proof. Choose $\beta_1, \dots, \beta_s$ as in Lemma 6.1. The equality $c(\beta_j) = 0$ means that η attains the supremum defining $P^{\mathbf{a}(\beta_j)}(\sigma)$. If

$$\mathbf{b} = \sum_{j=1}^{s} t_j \mathbf{a}(\beta_j), \qquad t_j > 0, \qquad \sum_j t_j = 1,$$

then for every $\mu \in \mathcal{M}_\sigma(X)$,

$$\begin{aligned}\sum_i b_i h(\tau_i \mu) &= \sum_j t_j \sum_i a_i(\beta_j) h(\tau_i \mu) \\ &\le \sum_j t_j \sum_i a_i(\beta_j) h(\tau_i \eta) = \sum_i b_i h(\tau_i \eta).\end{aligned}$$

Thus η is an equilibrium state for every weight vector in the relative interior of the convex hull of these vectors. Uniqueness follows from Proposition 2.1. □

6.2. **Recursive-function estimate.** Combining the common equilibrium state given by Lemma 6.2 with the Gibbs estimates of D.-J. Feng [Fen11], we will obtain a recursive-function estimate.

For a positive probability weight vector $\mathbf{b} = (b_1, \dots, b_s)$, set

$$\alpha_1(\mathbf{b}) = 0, \qquad \alpha_i(\mathbf{b}) = \frac{A_{i-1}(\mathbf{b})}{A_i(\mathbf{b})} \quad 2 \le i \le s. \tag{6.6}$$

When $\mathbf{b}$ is fixed, we write simply α_i for $\alpha_i(\mathbf{b})$. For the vector $\mathbf{a}$ in (2.3), this is exactly

$$\alpha_i(\mathbf{a}) = \frac{\theta_{i-1}}{\theta_i} = \frac{\log n_i}{\log n_{i-1}}.$$

The map $\mathbf{b} \mapsto (\alpha_2, \dots, \alpha_s)$ is a diffeomorphism from the positive weight simplex onto $(0,1)^{s-1}$. Let V be the image of the set U from Lemma 6.2 under this map. The open set V contains an open box

$$B = I_2 \times \cdots \times I_s, \qquad \overline{B} \subset V \subset (0,1)^{s-1}. \tag{6.7}$$

For $\boldsymbol{\alpha} \in B$, let $\mathbf{b}(\boldsymbol{\alpha}) \in U$ denote the unique weight vector given by the inverse diffeomorphism.

We now define recursive functions associated with the successive factor maps. Their definition agrees with the functions used in D.-J. Feng's weighted equilibrium construction, so the Gibbs estimates in [Fen11] can be applied. Fix $\mathbf{b}$, set $\phi_{\mathbf{b}}^{(1)} \equiv 1$ on $\mathcal{L}(X_1)$, and define the remaining functions recursively by

$$\begin{aligned}\phi_{\mathbf{b}}^{(2)}(J) &= \#\{I \in \mathcal{L}_{|J|}(X_1) : \pi_2(I) = J\}, \quad J \in \mathcal{L}(X_2), \\ \phi_{\mathbf{b}}^{(i)}(J) &= \sum_{\substack{I \in \mathcal{L}_{|J|}(X_{i-1}) \\ \pi_i(I) = J}} \phi_{\mathbf{b}}^{(i-1)}(I)^{\alpha_{i-1}(\mathbf{b})}, \qquad J \in \mathcal{L}(X_i), \quad 3 \le i \le s.\end{aligned} \tag{6.8}$$

For $N \ge 1$, write

$$Z_{\mathbf{b}}(N) = \sum_{J \in \mathcal{L}_N(X_s)} \phi_{\mathbf{b}}^{(s)}(J)^{\alpha_s(\mathbf{b})}. \tag{6.9}$$

The function $\phi_{\mathbf{b}}^{(i)}$ depends only on $\alpha_2(\mathbf{b}), \dots, \alpha_{i-1}(\mathbf{b})$.

The following proposition is based on [Fen11].

Proposition 6.3. *Let $\eta_{\mathbf{b}}$ be the $\mathbf{b}$-weighted equilibrium state. Then:*

(i) *for every* $N \geq 1$ *and* $J \in \mathcal{L}_N(X_s)$,

$$(\tau_s \eta_{\mathbf{b}})[J] \asymp_{\mathbf{b}} \frac{\phi_{\mathbf{b}}^{(s)}(J)^{\alpha_s(\mathbf{b})}}{Z_{\mathbf{b}}(N)}; \tag{6.10}$$

(ii) *if* $1 \leq j < s$, $N \geq 1$, *and* $I_j \in \mathcal{L}_N(X_j)$, *let* $I_k = \pi_k \circ \cdots \circ \pi_{j+1}(I_j)$ *for* $j < k \leq s$, *and write*

$$\widetilde{g}_{j,\mathbf{b},N}(I_j) = \left(\prod_{k=j}^{s-1} \frac{\phi_{\mathbf{b}}^{(k)}(I_k)^{\alpha_k(\mathbf{b})}}{\phi_{\mathbf{b}}^{(k+1)}(I_{k+1})} \right) \frac{\phi_{\mathbf{b}}^{(s)}(I_s)^{\alpha_s(\mathbf{b})}}{Z_{\mathbf{b}}(N)}. \tag{6.11}$$

Then

$$\sum_{I_j \in \mathcal{L}_N(X_j)} \widetilde{g}_{j,\mathbf{b},N}(I_j) = 1, \qquad \widetilde{g}_{j,\mathbf{b},N}(I_j) \lesssim_{\mathbf{b}} (\tau_j \eta_{\mathbf{b}})[I_j]. \tag{6.12}$$

All implicit constants in this proposition are independent of N *and of the words involved.*

Proof. For a subshift Y satisfying weak specification with bound p, recall from [Fen11, Section 5] that $\mathcal{D}_w(Y,p)$ is the collection of positive word functions $\psi : \mathcal{L}(Y) \to (0,\infty)$ satisfying the following two properties: for some $C \geq 1$,

$$\psi(IJ) \leq C\psi(I)\psi(J) \quad \text{whenever } IJ \in \mathcal{L}(Y),$$

and, for every $I, J \in \mathcal{L}(Y)$, there is a word $W \in \mathcal{L}(Y)$ with $|W| \leq p$ such that

$$IWJ \in \mathcal{L}(Y), \qquad \psi(IWJ) \geq C^{-1}\psi(I)\psi(J).$$

The function $\phi_{\mathbf{b}}^{(1)} \equiv 1$ on $\mathcal{L}(X_1)$ belongs to $\mathcal{D}_w(X_1,p)$. By [Fen11, Lemma 5.7], every $\phi_{\mathbf{b}}^{(i)}$, and every positive power of it, belongs to $\mathcal{D}_w(X_i,p)$, since positive powers and fibre sums under a factor map preserve the required two inequalities.

For part (i), [Fen11, Lemma 7.2] identifies $\tau_s\eta_{\mathbf{b}}$ as the equilibrium state on X_s generated by $\big(\phi_{\mathbf{b}}^{(s)}\big)^{\alpha_s(\mathbf{b})}$. Thus [Fen11, Theorem 5.5], applied to this word function, gives (6.10).

For part (ii), [Fen11, Theorem 7.3(i)] gives the required inequality when $j = 1$. Now fix $2 \leq j < s$. The recursion (6.8) says that, for every $1 \leq k < s$ and $I_{k+1} \in \mathcal{L}_N(X_{k+1})$,

$$\sum_{\pi_{k+1}(I_k) = I_{k+1}} \frac{\phi_{\mathbf{b}}^{(k)}(I_k)^{\alpha_k(\mathbf{b})}}{\phi_{\mathbf{b}}^{(k+1)}(I_{k+1})} = 1.$$

Consequently, successively summing the preceding expression for $\widetilde{g}_{1,\mathbf{b},N}$ over the fibres of $\pi_2, \ldots, \pi_j$ gives

$$\widetilde{g}_{j,\mathbf{b},N}(I_j) = \sum_{\substack{I_1 \in \mathcal{L}_N(X_1) \\ (\pi_j \circ \cdots \circ \pi_2)(I_1) = I_j}} \widetilde{g}_{1,\mathbf{b},N}(I_1). \tag{6.13}$$

Using the definition (6.9), the same fibre-sum calculation, carried all the way to level s, gives

$$\sum_{I_1 \in \mathcal{L}_N(X_1)} \widetilde{g}_{1,\mathbf{b},N}(I_1) = \frac{1}{Z_{\mathbf{b}}(N)} \sum_{I_s \in \mathcal{L}_N(X_s)} \phi_{\mathbf{b}}^{(s)}(I_s)^{\alpha_s(\mathbf{b})} = 1.$$

Summing (6.13) over $I_j \in \mathcal{L}_N(X_j)$ gives

$$\sum_{I_j \in \mathcal{L}_N(X_j)} \widetilde{g}_{j,\mathbf{b},N}(I_j) = \sum_{I_1 \in \mathcal{L}_N(X_1)} \widetilde{g}_{1,\mathbf{b},N}(I_1) = 1,$$

which proves the first assertion in (6.12). Finally, using the $j = 1$ inequality and (6.13) gives

$$\widetilde{g}_{j,\mathbf{b},N}(I_j) \lesssim_{\mathbf{b}} \sum_{(\pi_j \circ \cdots \circ \pi_2)(I_1) = I_j} \eta_{\mathbf{b}}[I_1] = (\tau_j \eta_{\mathbf{b}})[I_j],$$

which proves the second assertion in (6.12). □

For a subshift Y, we call a positive function ϕ on $\mathcal{L}(Y)$ *comparable* if there is a constant $C \geq 1$ such that

$$C^{-1} \leq \frac{\phi(I)}{\phi(J)} \leq C \quad \text{whenever } I, J \in \mathcal{L}_N(Y),\ N \geq 1, \tag{6.14}$$

where the constant C is independent of N, I, J.

Lemma 6.4. *Fix $1 \leq j < s$ and $\mathbf{b}$. Suppose $\phi_{\mathbf{b}}^{(j+1)}, \ldots, \phi_{\mathbf{b}}^{(s)}$ satisfy (6.14). Then, for every $N \geq 1$ and $I \in \mathcal{L}_N(X_j)$,*

$$(\tau_j \eta_{\mathbf{b}})[I] \asymp_{\mathbf{b}} \frac{\phi_{\mathbf{b}}^{(j)}(I)^{\alpha_j(\mathbf{b})}}{\sum_{L \in \mathcal{L}_N(X_j)} \phi_{\mathbf{b}}^{(j)}(L)^{\alpha_j(\mathbf{b})}}. \tag{6.15}$$

Proof. Fix $N \geq 1$. For each $j + 1 \leq k \leq s$, choose a word $J_{k,N} \in \mathcal{L}_N(X_k)$. Condition (6.14) gives

$$\phi_{\mathbf{b}}^{(k)}(I_k) \asymp_{\mathbf{b}} \phi_{\mathbf{b}}^{(k)}(J_{k,N}) \qquad I_k \in \mathcal{L}_N(X_k).$$

Substitution in (6.11) gives

$$\widetilde{g}_{j,\mathbf{b},N}(I) \asymp_{\mathbf{b}} T_N \phi_{\mathbf{b}}^{(j)}(I)^{\alpha_j(\mathbf{b})}, \qquad T_N = \frac{1}{Z_{\mathbf{b}}(N)} \prod_{k=j+1}^{s} \phi_{\mathbf{b}}^{(k)}(J_{k,N})^{\alpha_k(\mathbf{b})-1}.$$

The quantity T_N may depend on N, but it does not depend on I. The normalisation in (6.12) now gives

$$1 = \sum_{I \in \mathcal{L}_N(X_j)} \widetilde{g}_{j,\mathbf{b},N}(I) \asymp_{\mathbf{b}} T_N \sum_{I \in \mathcal{L}_N(X_j)} \phi_{\mathbf{b}}^{(j)}(I)^{\alpha_j(\mathbf{b})}.$$

Solving this comparison for T_N and substituting it into the preceding estimate gives

$$\widetilde{g}_{j,\mathbf{b},N}(I) \asymp_{\mathbf{b}} \frac{\phi_{\mathbf{b}}^{(j)}(I)^{\alpha_j(\mathbf{b})}}{\sum_{L \in \mathcal{L}_N(X_j)} \phi_{\mathbf{b}}^{(j)}(L)^{\alpha_j(\mathbf{b})}}. \tag{6.16}$$

If $j = 1$, then $\left(\phi_{\mathbf{b}}^{(1)}\right)^{\alpha_1(\mathbf{b})} = 1$, which belongs to $\mathcal{D}_w(X_1, p)$, as noted in the proof of Proposition 6.3. If $2 \leq j < s$, the verification in the proof of Proposition 6.3 shows that $\left(\phi_{\mathbf{b}}^{(j)}\right)^{\alpha_j(\mathbf{b})}$ belongs to $\mathcal{D}_w(X_j, p)$. Thus [Fen11, Theorem 5.5] gives a unique equilibrium state ν on X_j. Moreover, ν is ergodic and satisfies

$$\nu[I] \asymp_{\mathbf{b}} \frac{\phi_{\mathbf{b}}^{(j)}(I)^{\alpha_j(\mathbf{b})}}{\sum_{L \in \mathcal{L}_N(X_j)} \phi_{\mathbf{b}}^{(j)}(L)^{\alpha_j(\mathbf{b})}}. \tag{6.17}$$

Equations (6.12), (6.16), and (6.17) imply that

$$\nu[I] \lesssim_{\mathbf{b}} (\tau_j \eta_{\mathbf{b}})[I]$$

for every cylinder $[I] \subset X_j$. Every open subset of X_j can be written as a countable disjoint union of cylinders, so the same inequality holds for open sets. By regularity of Borel probability measures, it follows that $\nu \ll \tau_j \eta_{\mathbf{b}}$.

By [Fen11, Theorem 7.3], the weighted equilibrium state $\eta_{\mathbf{b}}$ is ergodic, and hence its factor $\tau_j \eta_{\mathbf{b}}$ is ergodic. Since ν is also ergodic, distinct ergodic invariant probability measures are mutually singular, whereas $\nu \ll \tau_j \eta_{\mathbf{b}}$. Therefore $\nu = \tau_j \eta_{\mathbf{b}}$, and (6.17) proves (6.15). □

We now combine the preceding estimates with the fact that the equilibrium state is the same throughout the box B. By changing one coordinate α_i while keeping the others fixed, two Gibbs estimates for the same measure can be compared.

Lemma 6.5. *Suppose that, for every $\boldsymbol{\alpha} = (\alpha_2, \dots, \alpha_s)$ in the box $B = I_2 \times \cdots \times I_s$ defined in (6.7), the same measure η is the $\mathbf{b}(\boldsymbol{\alpha})$-weighted equilibrium state. Then, for every fixed $\boldsymbol{\alpha} \in B$ and $1 \le i \le s$, the function $\phi^{(i)}_{\mathbf{b}(\boldsymbol{\alpha})}$ satisfies (6.14).*

Proof. The case $i = 1$ is immediate since $\phi^{(1)}_{\mathbf{b}(\boldsymbol{\alpha})} \equiv 1$. We argue from $i = s$ down to $i = 2$. First fix $\alpha_2, \dots, \alpha_{s-1}$ and choose distinct $u, v \in I_s$. For $t \in \{u, v\}$, let $\mathbf{b}_t$ be the weight vector corresponding to $(\alpha_2, \dots, \alpha_{s-1}, t)$. The function $\phi^{(s)}_{\mathbf{b}_t}$ does not depend on t; denote this common function by Φ_s. The measure $\tau_s \eta$ is also unchanged. Write

$$S_{s,t}(N) = \sum_{L \in \mathcal{L}_N(X_s)} \Phi_s(L)^t, \qquad t \in \{u, v\}$$

and apply (6.10) to $\mathbf{b}_u$ and $\mathbf{b}_v$. This gives

$$\frac{\Phi_s(J)^u}{S_{s,u}(N)} \asymp_{\mathbf{b}_u} (\tau_s \eta)[J] \asymp_{\mathbf{b}_v} \frac{\Phi_s(J)^v}{S_{s,v}(N)}.$$

Therefore

$$\Phi_s(J)^{u-v} \asymp_{\mathbf{b}_u, \mathbf{b}_v} \frac{S_{s,u}(N)}{S_{s,v}(N)}.$$

The comparison constants are independent of N and J. Hence, for $J, L \in \mathcal{L}_N(X_s)$,

$$\left(\frac{\Phi_s(J)}{\Phi_s(L)}\right)^{u-v} \asymp_{\mathbf{b}_u, \mathbf{b}_v} 1,$$

with an implicit constant independent of N, J, L. Since $u \neq v$, this proves (6.14) for $\phi^{(s)}_{\mathbf{b}(\boldsymbol{\alpha})}$ at every $\boldsymbol{\alpha} \in B$.

Now suppose the assertion has been proved for $\phi^{(j+1)}_{\mathbf{b}}, \dots, \phi^{(s)}_{\mathbf{b}}$ at every point of B. Fix all coordinates of $\boldsymbol{\alpha}$ except α_j, and choose distinct $u, v \in I_j$. For $t \in \{u, v\}$, let $\mathbf{b}_t$ be the weight vector obtained by setting $\alpha_j = t$. The function $\phi^{(j)}_{\mathbf{b}_t}$ depends only on $\alpha_2, \dots, \alpha_{j-1}$ and is therefore independent of t; denote it by Φ_j. Lemma 6.4, applied to $\mathbf{b}_u$ and $\mathbf{b}_v$, gives

$$(\tau_j \eta)[I] \asymp_{\mathbf{b}_u} \frac{\Phi_j(I)^u}{S_{j,u}(N)}, \qquad (\tau_j \eta)[I] \asymp_{\mathbf{b}_v} \frac{\Phi_j(I)^v}{S_{j,v}(N)},$$

where $S_{j,t}(N) = \sum_{L \in \mathcal{L}_N(X_j)} \Phi_j(L)^t$. Since $\tau_j \eta$ is unchanged,

$$\Phi_j(I)^{u-v} \asymp_{\mathbf{b}_u, \mathbf{b}_v} \frac{S_{j,u}(N)}{S_{j,v}(N)}.$$

Again the comparison constants are independent of N and I. Thus, for $I, L \in \mathcal{L}_N(X_j)$,

$$\left(\frac{\Phi_j(I)}{\Phi_j(L)}\right)^{u-v} \asymp_{\mathbf{b}_u,\mathbf{b}_v} 1,$$

with an implicit constant independent of N, I, L. Since $u \neq v$, this proves (6.14) for $\phi^{(j)}_{\mathbf{b}(\boldsymbol{\alpha})}$ at every $\boldsymbol{\alpha} \in B$. □

6.3. **Fibre estimates.** We now pass from the recursive functions to the fibres of the factor maps. We first treat each adjacent map π_i and then each composite map τ_i. The resulting estimates imply dimension equality and complete the contradiction argument.

For $N \geq 1$, $1 \leq i \leq s$, and $J \in \mathcal{L}_N(X_i)$, write

$$f_1(J) = 1, \qquad f_i(J) = \#\{I \in \mathcal{L}_N(X_{i-1}) : \pi_i(I) = J\} \quad (2 \leq i \leq s).$$

Lemma 6.6. *For $2 \leq i \leq s$, if $\phi^{(i-1)}_{\mathbf{b}}$ and $\phi^{(i)}_{\mathbf{b}}$ satisfy (6.14), then so does f_i. The function f_1 also satisfies (6.14).*

Proof. The constant function $f_1 \equiv 1$ on $\mathcal{L}(X_1)$ trivially satisfies (6.14). Let $2 \leq i \leq s$, and write

$$m_N = \min_{I \in \mathcal{L}_N(X_{i-1})} \phi^{(i-1)}_{\mathbf{b}}(I), \qquad M_N = \max_{I \in \mathcal{L}_N(X_{i-1})} \phi^{(i-1)}_{\mathbf{b}}(I).$$

Condition (6.14) gives $M_N/m_N \leq C_{i-1}$, where $C_{i-1} \geq 1$ is independent of N. By (6.8),

$$f_i(J) m_N^{\alpha_{i-1}} \leq \phi^{(i)}_{\mathbf{b}}(J) \leq f_i(J) M_N^{\alpha_{i-1}}.$$

Comparing this estimate for two words $J, J' \in \mathcal{L}_N(X_i)$ proves that $f_i(J)/f_i(J')$ is bounded above and below independently of N, J, J'. □

For $N \geq 1$, $1 \leq i \leq s$, and $J \in \mathcal{L}_N(X_i)$, set

$$F_i(J) = \#\{I \in \mathcal{L}_N(X) : \tau_i(I) = J\}.$$

Lemma 6.7. *If every f_i, $1 \leq i \leq s$, satisfies (6.14), then every F_i satisfies the same condition.*

Proof. The assertion is clear for $F_1 \equiv 1$. If it holds for F_i, then

$$F_{i+1}(J) = \sum_{\substack{L \in \mathcal{L}_N(X_i) \\ \pi_{i+1}(L) = J}} F_i(L).$$

This lies between $f_{i+1}(J) \min_L F_i(L)$ and $f_{i+1}(J) \max_L F_i(L)$. Since both F_i and f_{i+1} have uniformly comparable values on words of the same length, the same is true of F_{i+1}. □

Lemma 6.8. *If all the functions F_i satisfy (6.14), then $\dim_{\mathrm{H}} K = \dim_{\mathrm{B}} K$.*

Proof. The subshifts X and X_i satisfy weak specification. Hence [Fen26, Proposition 2.2] gives

$$\#\mathcal{L}_N(X) \asymp e^{N h_{\mathrm{top}}(X)}, \qquad \#\mathcal{L}_N(X_i) \asymp e^{N h_{\mathrm{top}}(X_i)}.$$

For $J \in \mathcal{L}_N(X_i)$, condition (6.14) and the identity $\sum_{L \in \mathcal{L}_N(X_i)} F_i(L) = \#\mathcal{L}_N(X)$ give

$$F_i(J) \asymp \frac{\#\mathcal{L}_N(X)}{\#\mathcal{L}_N(X_i)} \asymp e^{N(h_{\mathrm{top}}(X) - h_{\mathrm{top}}(X_i))}.$$

Thus condition (c) of [Fen26, Theorem 3.1] holds, and that theorem gives $\dim_{\mathrm{H}} K = \dim_{\mathrm{B}} K$. □

Proof of Proposition 2.3. Recall from (2.9) that $c(\beta) \geq 0$ for every $\beta \in (0,1)$. Suppose, to the contrary, that there is no $\beta \in (1/2,1)$ for which $c(\beta) > 0$. Then

$$c(\beta) = 0 \qquad \text{for every } \beta \in (1/2,1). \tag{6.18}$$

Apply Lemma 6.2 with $I = (1/2,1)$. There is a relatively open set U of positive weight vectors such that the fixed measure $\eta = \eta_{\mathbf{a}}$ is the unique $\mathbf{b}$-weighted equilibrium state for every $\mathbf{b} \in U$. Under the ratio coordinates in (6.6), the image of U contains the box $B = I_2 \times \cdots \times I_s$ from (6.7). Fix $\boldsymbol{\alpha} \in B$ and set $\mathbf{b} = \mathbf{b}(\boldsymbol{\alpha})$. Lemma 6.5 shows that $\phi_{\mathbf{b}}^{(i)}$ satisfies (6.14) for every $1 \leq i \leq s$. Lemmas 6.6 and 6.7 then show that every F_i satisfies the same condition. Lemma 6.8 gives $\dim_{\mathrm{H}} K = \dim_{\mathrm{B}} K$, contrary to the hypothesis $\dim_{\mathrm{H}} K < \dim_{\mathrm{B}} K$. Hence (6.18) is impossible, and $c(\beta) > 0$ for some $\beta \in (1/2,1)$. □

7. Proof of Theorem 1.2

We prove Theorem 1.2 in three steps. First, for a sofic subshift X, we establish the existence and formula for the box dimension of its corresponding self-affine set $K = R(X)$. We then treat separately the cases $\dim_{\mathrm{H}} K = \dim_{\mathrm{B}} K$ and $\dim_{\mathrm{H}} K < \dim_{\mathrm{B}} K$. In the first case, the critical Hausdorff measure is positive and σ-finite but need not be finite. In the second case, examples show that it may be positive and finite or infinite and σ-finite.

Fix a finite vertex-labelled directed graph $G = (V, E, \ell)$ presenting X, where V is the vertex set, $E \subset V \times V$ is the edge set, and $\ell : V \to \mathcal{S} \subset \mathcal{A}$ is the label map. Write

$$\Sigma_G = \{v = (v_k)_{k \geq 1} \in V^{\mathbb{N}} : (v_k, v_{k+1}) \in E \text{ for every } k \geq 1\}.$$

Applying ℓ coordinatewise, the graph presentation gives

$$X = \ell(\Sigma_G) = \{(\ell(v_k))_{k \geq 1} : v \in \Sigma_G\}.$$

For a finite vertex path $v_1 \cdots v_N$, we also write $\ell(v_1 \cdots v_N) = \ell(v_1) \cdots \ell(v_N)$. We use the following convention for directed paths and strongly connected components; compare [MW88].

A *directed path* from u to v is a finite sequence of vertices $u = v_0, v_1, \ldots, v_r = v$, where $r \geq 1$, such that $(v_{\ell-1}, v_\ell) \in E$ for every $1 \leq \ell \leq r$. A *directed cycle* is a directed path whose first and last vertices are the same. We say that v is *reachable* from u if there is a directed path from u to v.

A nonempty set $C \subset V$ is *strongly connected* if, for every $u, v \in C$, there is a directed path from u to v. A *strongly connected component* is a maximal strongly connected subset in V. A vertex belongs to one of these components exactly when it lies on a directed cycle. Distinct components are disjoint.

An ordered tuple $(D_1, \ldots, D_r)$ of strongly connected components is called a *component chain* if, for every $1 \leq j < r$, either $D_j = D_{j+1}$ or a vertex of D_{j+1} is reachable from a vertex of D_j.

Every sequence in X is the label sequence of at least one infinite path in G. The vertices that occur infinitely often along such a path lie in one strongly connected component, and the path eventually remains in that component. Let $C_1, \ldots, C_M$ be all the strongly connected components of G. For a component C, define

$$\Sigma_C = \{v \in \Sigma_G : v_k \in C \text{ for every } k \geq 1\}, \qquad X_C = \ell(\Sigma_C), \qquad K_C = R(X_C).$$

The path shift Σ_C is irreducible, so its factor X_C satisfies weak specification.

For a strongly connected component C and $1 \leq i \leq s$, set

$$h_i(C) = h_{\mathrm{top}}\big(\tau_i(X_C)\big). \tag{7.1}$$

For $1 \le j \le M$, let

$$\mathcal{U}_j = \{u \in \mathcal{L}(X) : uy \in X \text{ for some } y \in X_{C_j}\}.$$

This is a countable set and contains the empty word. If $u = d_1 \cdots d_q \in \mathcal{U}_j$, define

$$f_u(z) = \sum_{\ell=1}^{q} \Lambda^{-\ell} d_\ell + \Lambda^{-q} z$$

for $z \in \mathbb{R}^d$. For every $1 \le j \le M$, we have $K_{C_j} \subset K$. Every $x \in X$ can be written as $x = uy$, where $u \in \mathcal{U}_j$ and $y \in X_{C_j}$ for some j, and (1.1) gives $R(uy) = f_u(R(y))$. Thus

$$\bigcup_{j=1}^{M} K_{C_j} \subseteq K \subseteq \bigcup_{j=1}^{M} \bigcup_{u \in \mathcal{U}_j} f_u(K_{C_j}). \tag{7.2}$$

Every f_u is a bi-Lipschitz affine embedding, so countable stability of Hausdorff dimension gives

$$\dim_{\mathrm{H}} K = \max_j \dim_{\mathrm{H}} K_{C_j}. \tag{7.3}$$

7.1. **Box dimension.** The finite graph presentation will be used to prove that the box dimension exists and to derive its formula. This finiteness assumption is essential: for general subshifts, even the box dimension may fail to exist [Jur23].

We now estimate the number of level-k approximate cubes. By (3.5), this number is $\#\mathcal{D}_k(X)$. Recall from (3.4) that

$$\mathcal{D}_k(X) = \Bigg\{ (J_1, \ldots, J_s) \in \prod_{i=1}^{s} \mathcal{L}_{\lfloor \theta_i k \rfloor - \lfloor \theta_{i-1} k \rfloor}(X_i) : \text{there is } I \in \mathcal{L}_k(X) \text{ such that } \tau_i(I|_{\lfloor \theta_{i-1} k \rfloor}^{\lfloor \theta_i k \rfloor}) = J_i \text{ for every } i \Bigg\}.$$

For a component C, $1 \le i \le s$, and $N \ge 1$, a vertex path $\widehat{J} = v_1 \cdots v_N \in \mathcal{L}_N(\Sigma_C)$ is called a *lift* of $J \in \mathcal{L}_N(\tau_i(X_C))$ if $\tau_i(\ell(\widehat{J})) = J$.

Lemma 7.1. *Let C be a strongly connected component of G and let $1 \le i \le s$. For every $\varepsilon > 0$ and all sufficiently large N, there are vertices $u_{i,N}, v_{i,N} \in C$ and a set $\mathcal{V}_{i,N} \subset \mathcal{L}_N(\tau_i(X_C))$ such that*

$$\#\mathcal{V}_{i,N} \ge \exp\big(N(h_i(C) - \varepsilon)\big),$$

and every word in $\mathcal{V}_{i,N}$ has a lift that begins at $u_{i,N}$ and ends at $v_{i,N}$.

Proof. Choose one lift for each $J \in \mathcal{L}_N(\tau_i(X_C))$ and partition $\mathcal{L}_N(\tau_i(X_C))$ according to the initial and terminal vertices of the chosen lifts. There are at most $(\#C)^2$ classes, so one class $\mathcal{V}_{i,N}$, with common endpoint pair $(u_{i,N}, v_{i,N}) \in C \times C$, satisfies

$$\#\mathcal{V}_{i,N} \ge \frac{\#\mathcal{L}_N(\tau_i(X_C))}{(\#C)^2}.$$

By the definition of $h_i(C)$,

$$\lim_{N \to \infty} \frac{1}{N} \log \#\mathcal{L}_N(\tau_i(X_C)) = h_i(C).$$

Therefore, for all sufficiently large N, we have

$$\begin{aligned}\log \#\mathcal{V}_{i,N} &\geq \log \#\mathcal{L}_N(\tau_i(X_C)) - 2\log \#C \\ &\geq N(h_i(C) - \varepsilon/2) - N\varepsilon/2 \\ &= N(h_i(C) - \varepsilon).\end{aligned}$$

This completes the proof. □

Proposition 7.2. *Let $X \subset \mathcal{A}^{\mathbb{N}}$ be a sofic subshift with the finite vertex-labelled graph presentation fixed above, let $C_1, \ldots, C_M$ be all its strongly connected components, and set $K = R(X)$. Then the box dimension of K exists and*

$$\dim_{\mathrm{B}} K = \frac{1}{\log n_s} \max_{(D_1,\ldots,D_s)} \sum_{i=1}^{s} a_i h_i(D_i), \tag{7.4}$$

where the maximum is over component chains $(D_1, \ldots, D_s)$ with $D_i \in \{C_1, \ldots, C_M\}$, and h_i is defined in (7.1).

Proof. For $k \geq 1$ and $0 \leq i \leq s$, write $k_i = \lfloor \theta_i k \rfloor$, and write $m_i = k_i - k_{i-1}$ for $1 \leq i \leq s$. Then $\sum_{i=1}^{s} m_i = k$ and $m_i/k \to a_i = \theta_i - \theta_{i-1}$.

Each level-k approximate cube has diameter comparable to n_s^{-k}, while every set of diameter at most n_s^{-k} meets at most a bounded number of level-k approximate cubes. Thus (3.1), (3.3), and (3.5) show that the logarithms of the covering number of K at scale n_s^{-k} and of $\#\mathcal{D}_k(X)$ differ by $O(1)$. Consequently, the lower and upper box dimensions are obtained from the corresponding lower and upper limits of $k^{-1}\log \#\mathcal{D}_k(X)$ by division by $\log n_s$. It is therefore enough to estimate $\log \#\mathcal{D}_k(X)$.

Fix $\varepsilon > 0$. We first prove the upper bound. Since there are finitely many pairs (i, C), we have, uniformly for all $N \geq 1$, i, and C,

$$\#\mathcal{L}_N(\tau_i(X_C)) \lesssim_\varepsilon \exp\bigl(N(h_i(C) + \varepsilon)\bigr). \tag{7.5}$$

For every $\mathbf{J} = (J_1, \ldots, J_s) \in \mathcal{D}_k(X)$, choose a word $I(\mathbf{J}) \in \mathcal{L}_k(X)$ such that

$$\tau_i\bigl(I(\mathbf{J})|_{k_{i-1}}^{k_i}\bigr) = J_i \qquad 1 \leq i \leq s.$$

Since $I(\mathbf{J})$ determines $\mathbf{J}$, the words $I(\mathbf{J})$ chosen for distinct $\mathbf{J}$ are distinct. For each $I(\mathbf{J})$, choose a vertex path $\widehat{I}(\mathbf{J}) \in \mathcal{L}_k(\Sigma_G)$ with $\ell(\widehat{I}(\mathbf{J})) = I(\mathbf{J})$. Decompose this path as

$$\widehat{I}(\mathbf{J}) = U_0 P^{(1)} U_1 \cdots U_{q-1} P^{(q)} U_q,$$

where each $P^{(j)}$ is a maximal vertex subpath contained in a component $C^{(j)}$, and every vertex in the subpath U_j lies outside all components. A vertex outside the components cannot occur twice in a directed path, since the segment between two occurrences would be a directed cycle. A path also cannot leave and later return to the same component, since the intervening vertices would then be strongly connected to that component. Hence $C^{(1)}, \ldots, C^{(q)}$ are distinct and form a component chain, $q \leq M$, and $\sum_{j=0}^{q} |U_j| \leq \#V$. For $1 \leq i \leq s$ and $1 \leq j \leq q$, let $n_{i,j}$ be the number of positions in $(k_{i-1}, k_i] \cap \mathbb{N}$ occupied by $P^{(j)}$. For each $\mathbf{J}$, form the tuple

$$T(\widehat{I}(\mathbf{J})) = \left(q, (C^{(j)})_{j=1}^{q}, (U_j)_{j=0}^{q}, (n_{i,j})_{\substack{1\leq i\leq s \\ 1\leq j\leq q}}\right).$$

Let $\mathfrak{T}_k = \{T(\widehat{I}(\mathbf{J})) : \mathbf{J} \in \mathcal{D}_k(X)\}$. The number of possibilities for q, $(C^{(j)})_{j=1}^q$, and $(U_j)_{j=0}^q$ is independent of k. Since $0 \le n_{i,j} \le k$,

$$\#\mathfrak{T}_k \lesssim \sum_{q=0}^{M}(k+1)^{sq} = \frac{(k+1)^{s(M+1)}-1}{(k+1)^s-1} \lesssim (k+1)^{sM}.$$

Fix $T \in \mathfrak{T}_k$ and define

$$\mathcal{J}_i(T) = \{1 \le j \le q : n_{i,j} > 0\}.$$

For all sufficiently large k, $m_i > \#V$ for every i, so $\mathcal{J}_i(T)$ is nonempty. Choose $j_i \in \mathcal{J}_i(T)$ and set $D_i = C^{(j_i)}$ so that

$$h_i(D_i) = \max_{j \in \mathcal{J}_i(T)} h_i(C^{(j)}).$$

The component subpaths $P^{(1)}, \dots, P^{(q)}$ occur in this order along $\widehat{I}(\mathbf{J})$, and the index intervals $(k_{i-1}, k_i] \cap \mathbb{N}$ are ordered from left to right. Hence $j_i \le j_{i+1}$ for $1 \le i < s$, and therefore $(D_1, \dots, D_s)$ is a component chain. Moreover,

$$\sum_{j=1}^{q} n_{i,j} \le m_i, \qquad \sum_{j=1}^{q} n_{i,j} h_i(C^{(j)}) \le h_i(D_i) \sum_{j=1}^{q} n_{i,j} \le m_i h_i(D_i).$$

For fixed T, the vertex words U_j and hence their labels are fixed. The part of J_i corresponding to $P^{(j)}$ is a word in $\mathcal{L}_{n_{i,j}}(\tau_i(X_{C^{(j)}}))$. Thus (7.5) gives

$$\begin{aligned}
\#\{\mathbf{J} \in \mathcal{D}_k(X) : T(\widehat{I}(\mathbf{J})) = T\} &\le \prod_{i=1}^{s} \prod_{\substack{1 \le j \le q \\ n_{i,j} > 0}} \#\mathcal{L}_{n_{i,j}}\big(\tau_i(X_{C^{(j)}})\big) \\
&\lesssim_\varepsilon \exp\left\{ \sum_{i=1}^{s} \sum_{j=1}^{q} n_{i,j}\big(h_i(C^{(j)}) + \varepsilon\big) \right\} \\
&\lesssim_\varepsilon \exp\left\{ \sum_{i=1}^{s} m_i h_i(D_i) + \varepsilon k \right\}.
\end{aligned}$$

Summing over $T \in \mathfrak{T}_k$ gives, for a constant $C_\varepsilon > 0$ independent of k,

$$\#\mathcal{D}_k(X) \le C_\varepsilon (k+1)^{sM} \times \exp\left\{ \max_{\substack{(D_1,\dots,D_s) \text{ is} \\ \text{a component chain}}} \sum_{i=1}^{s} m_i h_i(D_i) + \varepsilon k \right\}.$$

Taking logarithms gives

$$\log \#\mathcal{D}_k(X) \le \max_{\substack{(D_1,\dots,D_s) \text{ is} \\ \text{a component chain}}} \sum_{i=1}^{s} m_i h_i(D_i) + \varepsilon k + sM \log(k+1) + \log C_\varepsilon. \tag{7.6}$$

For the lower bound, fix a component chain $(D_1, \dots, D_s)$ and set $B = \#V$. Since (D_i, D_{i+1}) is a component chain, any vertex in D_i can be joined to any vertex in D_{i+1} by a positive path of length at most B.

For all sufficiently large k, Lemma 7.1, applied for each i with $N = m_i - B$, gives vertices $u_i := u_{i,m_i-B}$ and $v_i := v_{i,m_i-B}$, and a set $\mathcal{V}_i := \mathcal{V}_{i,m_i-B} \subset \mathcal{L}_{m_i-B}(\tau_i(X_{D_i}))$ such that

$$\log \#\mathcal{V}_i \ge (m_i - B)(h_i(D_i) - \varepsilon),$$

and, for every $J_i \in \mathcal{V}_i$, a chosen vertex-path lift $\widehat{J}_i \in \mathcal{L}_{m_i-B}(\Sigma_{D_i})$ begins at u_i and ends at v_i.

For $1 \le i < s$, choose a vertex path $v_i \widehat{W}_i u_{i+1}$ of length at most B, and write $q_i = |\widehat{W}_i| \le B$. This is independent of J_i, J_{i+1}. Hence every tuple $(J_1, \dots, J_s) \in \prod_{i=1}^{s} \mathcal{V}_i$ gives a vertex path

$$\widehat{J}_1 \widehat{W}_1 \widehat{J}_2 \widehat{W}_2 \cdots \widehat{W}_{s-1} \widehat{J}_s \in \mathcal{L}(\Sigma_G)$$

of length

$$\sum_{i=1}^{s} (m_i - B) + \sum_{i=1}^{s-1} q_i = k - sB + \sum_{i=1}^{s-1} q_i < k.$$

The terminal vertex of $\widehat{J}_s$ is fixed and lies in D_s. Starting from this vertex, extend every such path by the same vertex path until the resulting path has length k, and let $I(J_1, \dots, J_s) \in \mathcal{L}_k(X)$ be the word obtained by applying ℓ to the resulting path.

For $1 \le i \le s$, let $d_i = (i-1)B - \sum_{r=1}^{i-1} q_r$. The number of labels preceding $\ell(\widehat{J}_i)$ is

$$\sum_{r=1}^{i-1} (m_r - B) + \sum_{r=1}^{i-1} q_r = k_{i-1} - d_i.$$

Therefore $\ell(\widehat{J}_i)$ occupies the positions

$$(k_{i-1} - d_i, k_i - B - d_i] \cap \mathbb{N}, \qquad 0 \le d_i \le (i-1)B.$$

For sufficiently large k, we have $d_i < m_i - B$. The digits of X_{D_i} belong to $\ell(D_i)$, and the map τ_i cannot increase their cardinality. Hence $\tau_i(X_{D_i})$ has at most $\#D_i \le B$ digits, so there are at most B^{d_i} possible length-d_i prefixes among the words in $\mathcal{V}_i$. We may therefore choose $\mathcal{V}_i' \subset \mathcal{V}_i$ whose words have the same first d_i digits and such that

$$\#\mathcal{V}_i' \ge B^{-d_i} \#\mathcal{V}_i \ge B^{-(i-1)B} \#\mathcal{V}_i.$$

Map each tuple $(J_1', \dots, J_s') \in \prod_i \mathcal{V}_i'$ to the vector in $\mathcal{D}_k(X)$ determined by $I(J_1', \dots, J_s')$. For $1 \le i \le s$, the i-th coordinate J_i' gives

$$\tau_i\big(I(J_1', \dots, J_s')|_{k_{i-1}}^{k_i}\big)|_0^{m_i - B - d_i} = J_i'|_{d_i}^{m_i - B}.$$

The deleted prefix is fixed on $\mathcal{V}_i'$, so different choices of J_i' give different i-th coordinates. The map is therefore injective. It follows that

$$\begin{aligned}
\log \#\mathcal{D}_k(X) &\ge \sum_{i=1}^{s} \log \#\mathcal{V}_i' \\
&\ge \sum_{i=1}^{s} (m_i - B)(h_i(D_i) - \varepsilon) - B \log B \sum_{i=1}^{s} (i-1) \\
&\ge \sum_{i=1}^{s} m_i h_i(D_i) - \varepsilon k - O(1).
\end{aligned}$$

Taking the maximum over all component chains gives

$$\log \#\mathcal{D}_k(X) \ge \max_{\substack{(D_1, \dots, D_s) \text{ is} \\ \text{a component chain}}} \sum_{i=1}^{s} m_i h_i(D_i) - \varepsilon k - O(1). \tag{7.7}$$

Since the number of component chains is finite and $m_i/k \to a_i$, (7.6) and (7.7) give, after $\varepsilon \downarrow 0$,

$$\lim_{k\to\infty} \frac{1}{k} \log \#\mathcal{D}_k(X) = \lim_{k\to\infty} \frac{1}{k} \max_{\substack{(D_1,\dots,D_s) \text{ is} \\ \text{a component chain}}} \sum_{i=1}^{s} m_i h_i(D_i)$$

$$= \max_{\substack{(D_1,\dots,D_s) \text{ is} \\ \text{a component chain}}} \sum_{i=1}^{s} a_i h_i(D_i).$$

Therefore the box dimension of K exists and is given by (7.4). □

When $s = 2$ and $\Lambda = \operatorname{diag}(n, m)$ with $n > m$, Proposition 7.2 becomes

$$\dim_{\mathrm{B}} R(X) = \max_{C} \left\{ \frac{h_{\mathrm{top}}(X_C)}{\log n} + \max_{\substack{D:\ (C,D) \text{ is} \\ \text{a component chain}}} h_{\mathrm{top}}\big(\tau_2(X_D)\big) \left(\frac{1}{\log m} - \frac{1}{\log n} \right) \right\}. \tag{7.8}$$

This agrees with the planar formula of Fraser and Jurga [FJ23, Theorem 1.2].

7.2. **The case $\dim_{\mathrm{H}} K = \dim_{\mathrm{B}} K$.** Proposition 7.2 shows that the box dimension exists. We now prove that dimension equality implies positivity and σ-finiteness. The critical Hausdorff measure, however, need not be finite.

Lemma 7.3. *Let X be a sofic subshift, set $K = R(X)$, and suppose that $\dim_{\mathrm{H}} K = \dim_{\mathrm{B}} K = \gamma$. Then $\mathcal{H}^\gamma(K) > 0$ and $\mathcal{H}^\gamma$ is σ-finite on K.*

Proof. Fix a finite vertex-labelled graph presentation of X, and let $C_1, \dots, C_M$ be all its strongly connected components. For every component C_j, write $\delta_j = \dim_{\mathrm{H}} K_{C_j}$. By (7.3),

$$\gamma = \max_{1\le j\le M} \delta_j.$$

If $\delta_j = \gamma$, then $K_{C_j} \subset K$ and monotonicity of box dimensions give

$$\gamma = \dim_{\mathrm{H}} K_{C_j} \le \underline{\dim}_{\mathrm{B}} K_{C_j} \le \overline{\dim}_{\mathrm{B}} K_{C_j} \le \dim_{\mathrm{B}} K = \gamma.$$

Hence the Hausdorff and box dimensions of K_{C_j} are both equal to γ. Theorem 1.1 therefore gives

$$0 < \mathcal{H}^\gamma(K_{C_j}) < \infty \qquad \text{whenever } \delta_j = \gamma. \tag{7.9}$$

If $\delta_j < \gamma$, the definition of Hausdorff dimension gives

$$\mathcal{H}^\gamma(K_{C_j}) = 0. \tag{7.10}$$

It follows from (7.9) and (7.10) that every component set has finite γ-dimensional Hausdorff measure. At least one component has $\delta_j = \gamma$, so (7.9) and $K_{C_j} \subset K$ also give $\mathcal{H}^\gamma(K) > 0$.

Each map f_u in (7.2) is a Lipschitz affine map. Hence (7.2) and the finiteness of the component measures proved above show that K is a countable union of Borel sets of finite $\mathcal{H}^\gamma$ measure. Thus $\mathcal{H}^\gamma$ is σ-finite on K. □

We now show that the conclusion of Lemma 7.3 cannot in general be strengthened to finiteness. For an integer $m \ge 2$ and a nonempty digit set $D \subseteq \{0, \dots, m-1\}$, write

$$E_m(D) = \left\{ \sum_{j=1}^{\infty} m^{-j} d_j : d_j \in D \right\}.$$

Example 7.4. Consider

$$\Lambda = \operatorname{diag}(6,3), \qquad \mathcal{S}_A = \{(3,2),(5,2)\}, \qquad \mathcal{S}_B = \{(0,0),(1,0)\}.$$

Let $\mathcal{S} = \mathcal{S}_A \cup \mathcal{S}_B$. Let $X \subset \mathcal{S}^{\mathbb{N}}$ be the subshift of finite type in which every transition inside $\mathcal{S}_A$, every transition inside $\mathcal{S}_B$, and every transition from $\mathcal{S}_A$ to $\mathcal{S}_B$ is allowed, while no transition from $\mathcal{S}_B$ to $\mathcal{S}_A$ is allowed. Set

$$X_A = \mathcal{S}_A^{\mathbb{N}}, \qquad X_B = \mathcal{S}_B^{\mathbb{N}}, \qquad K = R(X), \quad K_A = R(X_A), \quad K_B = R(X_B).$$

The coding map gives

$$K_A = E_6(\{3,5\}) \times \{1\}, \qquad K_B = E_6(\{0,1\}) \times \{0\}.$$

Consequently, for $\gamma = \frac{\log 2}{\log 6}$, we have

$$\dim_{\mathrm{H}} K_A = \dim_{\mathrm{B}} K_A = \dim_{\mathrm{H}} K_B = \dim_{\mathrm{B}} K_B = \gamma, \qquad 0 < \mathcal{H}^{\gamma}(K_A), \mathcal{H}^{\gamma}(K_B) < \infty.$$

Combining this with equation (7.3) gives $\dim_{\mathrm{H}} K = \gamma$. The component entropies are $h_{\mathrm{top}}(X_A) = h_{\mathrm{top}}(X_B) = \log 2$, while both projected component entropies are zero. Formula (7.8) therefore gives

$$\dim_{\mathrm{B}} K = \frac{\log 2}{\log 6} = \gamma = \dim_{\mathrm{H}} K.$$

It remains to compute the measure. For $r \geq 1$, let

$$Y_r = \{I\omega : I \in \mathcal{S}_A^r,\ \omega \in X_B\}.$$

Set $c_B = \mathcal{H}^{\gamma}(K_B)$. For each $I \in \mathcal{S}_A^r$, the set $R_r(I) + \Lambda^{-r} K_B$ is a copy of K_B on which distances are multiplied by 6^{-r}. Since $6^{-r\gamma} = 2^{-r}$,

$$\mathcal{H}^{\gamma}\big(R_r(I) + \Lambda^{-r} K_B\big) = 2^{-r} c_B. \tag{7.11}$$

Their first-coordinate cylinder sets corresponding to $I \in \mathcal{S}_A^r$ are separated, so for fixed r these 2^r copies are pairwise disjoint. It follows from (7.11) that

$$\mathcal{H}^{\gamma}(R(Y_r)) = 2^r 2^{-r} c_B = c_B. \tag{7.12}$$

The set $R(Y_r)$ lies on the horizontal line with second coordinate $1 - 3^{-r}$, so the Borel sets $R(Y_r)$ are pairwise disjoint as r varies. Therefore countable additivity and (7.12) give

$$\mathcal{H}^{\gamma}(K) \geq \sum_{r=1}^{\infty} \mathcal{H}^{\gamma}(R(Y_r)) = \sum_{r=1}^{\infty} c_B = \infty.$$

7.3. **The case** $\dim_{\mathrm{H}} K < \dim_{\mathrm{B}} K$**.** We give two examples with $\dim_{\mathrm{H}} K < \dim_{\mathrm{B}} K$. In the first, the critical Hausdorff measure is positive and finite; in the second, it is infinite but σ-finite.

Example 7.5. Consider

$$\Lambda = \operatorname{diag}(9,3), \qquad \theta_1 = \frac{\log 3}{\log 9} = \frac{1}{2}.$$

Take two disjoint full-shift digit sets

$$\mathcal{S}_A = \{0,1,\dots,6\} \times \{1,2\}, \qquad \mathcal{S}_B = (\{0,1,\dots,8\} \times \{0\}) \cup \{(7,1),(7,2)\}.$$

Let $\mathcal{S} = \mathcal{S}_A \cup \mathcal{S}_B$. Let $X \subset \mathcal{S}^{\mathbb{N}}$ be the subshift of finite type in which every transition inside $\mathcal{S}_A$, every transition inside $\mathcal{S}_B$, and every transition from $\mathcal{S}_A$ to $\mathcal{S}_B$ is allowed, while no transition from $\mathcal{S}_B$ to $\mathcal{S}_A$ is allowed. Set

$$X_A = \mathcal{S}_A^{\mathbb{N}}, \qquad X_B = \mathcal{S}_B^{\mathbb{N}}, \qquad K = R(X), \quad K_A = R(X_A), \quad K_B = R(X_B).$$

The Bedford-McMullen dimension formulas [Bed84, McM84] give

$$\dim_{\mathrm{H}} K_A = \dim_{\mathrm{B}} K_A = \frac{\log(2\sqrt{7})}{\log 3} \approx 1.51655,$$
$$\dim_{\mathrm{H}} K_B = \frac{\log 5}{\log 3} \approx 1.46497,$$
$$\dim_{\mathrm{B}} K_B = \frac{\log \sqrt{33}}{\log 3} \approx 1.59133.$$

Set $\gamma = \log(2\sqrt{7})/\log 3$. The identity $K_A = E_9(\{0,\dots,6\}) \times E_3(\{1,2\})$ shows that K_A is γ-Ahlfors regular, and hence $0 < \mathcal{H}^\gamma(K_A) < \infty$. Since $\dim_{\mathrm{H}} K_B < \gamma$, every affine image of K_B on the right-hand side of (7.2) has zero $\mathcal{H}^\gamma$ measure, while the corresponding images of K_A are contained in K_A. Therefore (7.2) gives

$$\mathcal{H}^\gamma(K) = \mathcal{H}^\gamma(K_A) \in (0,\infty).$$

The box dimension follows directly from (7.8). The component and projected entropies are

$$h_{\mathrm{top}}(X_A) = \log 14, \qquad h_{\mathrm{top}}(\tau_2(X_A)) = \log 2,$$
$$h_{\mathrm{top}}(X_B) = \log 11, \qquad h_{\mathrm{top}}(\tau_2(X_B)) = \log 3.$$

Hence (7.8) gives

$$\begin{aligned}\dim_{\mathrm{B}} K &= \max\left\{\frac{\log 14}{\log 9} + \log 3\left(\frac{1}{\log 3} - \frac{1}{\log 9}\right), \frac{\log 11}{\log 9} + \log 3\left(\frac{1}{\log 3} - \frac{1}{\log 9}\right)\right\} \\ &= \frac{\log 42}{\log 9}.\end{aligned}$$

Since $\gamma = \log 28/\log 9$, we obtain

$$\dim_{\mathrm{H}} K = \frac{\log 28}{\log 9} < \frac{\log 42}{\log 9} = \dim_{\mathrm{B}} K, \qquad 0 < \mathcal{H}^\gamma(K) < \infty.$$

Example 7.6. Consider

$$\Lambda = \mathrm{diag}(16,4), \qquad \mathcal{S}_A = \{(8,2),(10,2),(12,2),(14,2)\}, \qquad \mathcal{S}_B = \{(1,0),(1,1)\}.$$

Let $\mathcal{S} = \mathcal{S}_A \cup \mathcal{S}_B$. Let $X \subset \mathcal{S}^{\mathbb{N}}$ be the subshift of finite type in which every transition inside $\mathcal{S}_A$, every transition inside $\mathcal{S}_B$, and every transition from $\mathcal{S}_A$ to $\mathcal{S}_B$ is allowed, while no transition from $\mathcal{S}_B$ to $\mathcal{S}_A$ is allowed. Set

$$X_A = \mathcal{S}_A^{\mathbb{N}}, \qquad X_B = \mathcal{S}_B^{\mathbb{N}}, \qquad K = R(X), \qquad K_A = R(X_A), \quad K_B = R(X_B).$$

The definition of R gives

$$K_A = E_{16}(\{8,10,12,14\}) \times \left\{\frac{2}{3}\right\}, \qquad K_B = \left\{\frac{1}{15}\right\} \times E_4(\{0,1\}).$$

Both one-dimensional systems satisfy the strong separation condition. Consequently,

$$\dim_{\mathrm{H}} K_A = \dim_{\mathrm{B}} K_A = \frac{\log 4}{\log 16} = \frac{1}{2}, \qquad \dim_{\mathrm{H}} K_B = \dim_{\mathrm{B}} K_B = \frac{\log 2}{\log 4} = \frac{1}{2},$$

and

$$0 < \mathcal{H}^{1/2}(K_A) < \infty, \qquad 0 < \mathcal{H}^{1/2}(K_B) < \infty. \tag{7.13}$$

Equation (7.3) gives $\dim_{\mathrm{H}} K = 1/2$.

For $r \geq 1$, let

$$Y_r = \{I\omega : I \in \mathcal{L}_r(X_A),\ \omega \in X_B\}.$$

Set $c_B = \mathcal{H}^{1/2}(K_B)$. Since X_A is the full shift on four digits, $\#\mathcal{L}_r(X_A) = 4^r$. For $I \in \mathcal{L}_r(X_A)$, the set $R_r(I) + \Lambda^{-r}K_B$ varies only in the second coordinate, where its similarity ratio is 4^{-r}. Thus

$$\mathcal{H}^{1/2}\big(R_r(I) + \Lambda^{-r}K_B\big) = 2^{-r}c_B.$$

Their first-coordinate cylinder sets are separated, so for fixed r these 4^r copies are pairwise disjoint, and hence $\mathcal{H}^{1/2}(R(Y_r)) = 4^r 2^{-r} c_B = 2^r c_B$. Since $R(Y_r) \subset K$,

$$\mathcal{H}^{1/2}(K) \geq 2^r c_B \qquad r \geq 1.$$

Letting $r \to \infty$ gives $\mathcal{H}^{1/2}(K) = \infty$. On the other hand, (7.2) and (7.13) show that K is covered by countably many Borel sets of finite $\mathcal{H}^{1/2}$ measure. Thus $\mathcal{H}^{1/2}$ is σ-finite on K.

The box dimension follows from (7.8). We have

$$h_{\mathrm{top}}(X_A) = \log 4, \qquad h_{\mathrm{top}}(\tau_2(X_A)) = 0,$$
$$h_{\mathrm{top}}(X_B) = \log 2, \qquad h_{\mathrm{top}}(\tau_2(X_B)) = \log 2.$$

Hence (7.8) gives

$$\begin{aligned}\dim_{\mathrm{B}} K &= \max\left\{\frac{\log 4}{\log 16} + \log 2\left(\frac{1}{\log 4} - \frac{1}{\log 16}\right), \frac{\log 2}{\log 16} + \log 2\left(\frac{1}{\log 4} - \frac{1}{\log 16}\right)\right\} \\ &= \frac{3}{4}.\end{aligned}$$

Therefore

$$\dim_{\mathrm{H}} K = \frac{1}{2} < \frac{3}{4} = \dim_{\mathrm{B}} K, \qquad \mathcal{H}^{1/2}(K) = \infty,$$

and this measure is σ-finite on K.

Proof of Theorem 1.2. Proposition 7.2 proves that the box dimension exists and gives its formula. Suppose first that $\dim_{\mathrm{H}} K = \dim_{\mathrm{B}} K = \gamma$. Lemma 7.3 gives $\mathcal{H}^{\gamma}(K) > 0$ and shows that $\mathcal{H}^{\gamma}$ is σ-finite on K. Example 7.4 shows that $\mathcal{H}^{\gamma}(K)$ may nevertheless be infinite. This proves (i).

For (ii), Example 7.5 gives a dimension gap with positive and finite critical Hausdorff measure, whereas Example 7.6 gives a dimension gap with infinite but σ-finite critical Hausdorff measure. □

References

[Bar07] K. Barański, Hausdorff dimension of the limit sets of some planar geometric constructions, *Adv. Math.* **210** (2007), 215–245.

[Bar25] B. Bárány, Hausdorff measure of dominated planar self-affine sets with large dimension, preprint, arXiv:2504.04525, 2025.

[BKY26] B. Bárány, A. Käenmäki and H. Yu, Finer geometry of planar self-affine sets, *Adv. Math.* **493** (2026), Paper No. 110914.

[BB25] J. Barral and G. Brunet, Variational principles for Hausdorff and packing dimensions of fractal percolation on self-affine sponges, *Adv. Math.* **481** (2025), Paper No. 110549, 69 pp.

[Bed84] T. Bedford, Crinkly curves, Markov partitions and box dimension in self-similar sets, Ph.D. thesis, University of Warwick, 1984.

[DS17] T. Das and D. Simmons, The Hausdorff and dynamical dimensions of self-affine sponges: a dimension gap result, *Invent. Math.* **210** (2017), 85–134.

[Fal14] K. J. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*, 3rd ed., John Wiley & Sons, Chichester, 2014.

[Fen11] D.-J. Feng, Weighted equilibrium states for factor maps between subshifts, *Adv. Math.* **226** (2011), 2470–2502.

[Fen25] Z. Feng, Dimension of diagonal self-affine measures with exponentially separated projections, *Adv. Math.* **480** (2025), Paper No. 110511.

[Fen26] Z. Feng, On the coincidence of the Hausdorff and box dimensions for some affine-invariant sets, *Ergodic Theory Dynam. Systems* **46** (2026), 1–33.

[FJS10] A. Ferguson, T. Jordan and P. Shmerkin, The Hausdorff dimension of the projections of self-affine carpets, *Fund. Math.* **209** (2010), no. 3, 193–213.

[FH17] J. M. Fraser and D. Howroyd, Assouad type dimensions for self-affine sponges, *Ann. Acad. Sci. Fenn. Math.* **42** (2017), 149–174.

[FJ23] J. M. Fraser and N. Jurga, Box dimensions of $(\times m, \times n)$-invariant sets, *Indiana Univ. Math. J.* **72** (2023), 2341–2367.

[FS16] J. M. Fraser and P. Shmerkin, On the dimensions of a family of overlapping self-affine carpets, *Ergodic Theory Dynam. Systems* **36** (2016), no. 8, 2463–2481.

[Hut81] J. E. Hutchinson, Fractals and self-similarity, *Indiana Univ. Math. J.* **30** (1981), 713–747.

[Jur23] N. Jurga, Nonexistence of the box dimension for dynamically invariant sets, *Anal. PDE* **16** (2023), no. 10, 2385–2399.

[KP96a] R. Kenyon and Y. Peres, Hausdorff dimensions of sofic affine-invariant sets, *Israel J. Math.* **94** (1996), 157–178.

[KP96b] R. Kenyon and Y. Peres, Measures of full dimension on affine-invariant sets, *Ergodic Theory Dynam. Systems* **16** (1996), 307–323.

[Kol23] I. Kolossváry, The L^q spectrum of self-affine measures on sponges, *J. Lond. Math. Soc. (2)* **108** (2023), no. 2, 666–701.

[LG92] S. P. Lalley and D. Gatzouras, Hausdorff and box dimensions of certain self-affine fractals, *Indiana Univ. Math. J.* **41** (1992), 533–568.

[MW88] R. D. Mauldin and S. C. Williams, Hausdorff dimension in graph directed constructions, *Trans. Amer. Math. Soc.* **309** (1988), 811–829.

[McM84] C. McMullen, The Hausdorff dimension of general Sierpiński carpets, *Nagoya Math. J.* **96** (1984), 1–9.

[Per94] Y. Peres, The self-affine carpets of McMullen and Bedford have infinite Hausdorff measure, *Math. Proc. Cambridge Philos. Soc.* **116** (1994), 513–526.

[QW24] H. Qiu and Q. Wang, The Hausdorff measure and uniform fibre conditions for Barański carpets, preprint, 2024.

[Rap26] A. Rapaport, Dimension of diagonal self-affine sets and measures via nonconformal partitions, *Proc. Lond. Math. Soc. (3)* **132** (2026), no. 1, Article e70120.

School of Mathematics, Nanjing University, Nanjing, 210093, P. R. China.
Email address: huaqiu@nju.edu.cn

School of Mathematics, Nanjing University, Nanjing, 210093, P. R. China.
Email address: 602023210013@smail.nju.edu.cn