Ergod. Th. & Dynam. Sys. (2021), **41**, 2542–2560
doi:10.1017/etds.2020.47

Symbolic dynamics in mean dimension theory

MAO SHINODA† and MASAKI TSUKAMOTO‡

† *Department of Human Coexistence, Graduate School of Human and Environmental Studies, Kyoto University, Yoshida-Nihonmaths-cho, Sakyo-ku, Kyoto 606-8501, Japan*
(e-mail: shinoda-mao@keio.jp)

‡ *Department of Mathematics, Kyushu University, Moto-oka 744, Nishi-ku, Fukuoka 819-0395, Japan*
(e-mail: masaki.tsukamoto@gmail.com)

(*Received* 7 *October* 2019 *and accepted in revised form* 16 *April* 2020)

Abstract. Furstenberg [Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation. *Math. Syst. Theory* **1** (1967), 1–49] calculated the Hausdorff and Minkowski dimensions of one-sided subshifts in terms of topological entropy. We generalize this to $\mathbb{Z}^2$-subshifts. Our generalization involves mean dimension theory. We calculate the metric mean dimension and the mean Hausdorff dimension of $\mathbb{Z}^2$-subshifts with respect to a subaction of $\mathbb{Z}$. The resulting formula is quite analogous to Furstenberg's theorem. We also calculate the rate distortion dimension of $\mathbb{Z}^2$-subshifts in terms of Kolmogorov–Sinai entropy.

Key words: subshift, metric mean dimension, mean Hausdorff dimension, rate distortion dimension

2020 Mathematics Subject Classification: 37B10, 37C45 (Primary); 37A05, 94A34 (Secondary)

1. *Introduction*

1.1. *Hausdorff and Minkowski dimensions of subshifts.* Let A be a finite set (alphabet). We consider the one-sided infinite product $A^{\mathbb{N}} = A \times A \times A \times \cdots$ with the shift map $\sigma : A^{\mathbb{N}} \to A^{\mathbb{N}}$ defined by

$$\sigma((x_n)_{n\in\mathbb{N}}) = (x_{n+1})_{n\in\mathbb{N}}.$$

Take $\alpha > 1$. We define a distance d on $A^{\mathbb{N}}$ by

$$d(x, y) = \alpha^{-\min\{n \mid x_n \neq y_n\}}.$$

https://doi.org/10.1017/etds.2020.47 Published online by Cambridge University Press

Let $\mathcal{X} \subset A^{\mathbb{N}}$ be a σ-invariant closed subset. Furstenberg [**Fur67**, Proposition III.1] calculated the Hausdorff and Minkowski dimensions of $\mathcal{X}$ with respect to d:

$$\dim_{\mathrm{H}}(\mathcal{X}, d) = \dim_{\mathrm{M}}(\mathcal{X}, d) = \frac{h_{\mathrm{top}}(\mathcal{X}, \sigma)}{\log \alpha}. \tag{1.1}$$

Here $h_{\mathrm{top}}(\mathcal{X}, \sigma)$ is the topological entropy of $(\mathcal{X}, \sigma)$. The purpose of the paper is to extend this result to *higher rank actions*.

1.2. *Mean dimension theory.* Mean dimension theory provides a meaningful framework for extending (1.1) to higher rank actions. This is the theory first introduced by Gromov [**Gro99**] and further developed by Lindenstrauss and Weiss [**LW00**], Lindenstrauss [**Lin99**] and more recently Lindenstrauss and the second named author [**LT19**]. We review the basic ingredients here. (The precise definitions will be given in §2.)

A pair $(\mathcal{X}, T)$ is called a *dynamical system* if $\mathcal{X}$ is a compact metric space and $T : \mathcal{X} \to \mathcal{X}$ is a homeomorphism†. Gromov [**Gro99**] defined *mean topological dimension* $\mathrm{mdim}(\mathcal{X}, T)$. This is a dynamical analogue of topological dimension and it evaluates the number of parameters per iterate for describing the orbits of $(\mathcal{X}, T)$. As the name suggested, the mean topological dimension is a *topological invariant* of dynamical systems. There are many important works around this quantity [**LW00**, **Lin99**, **Gut15**, **GLT16**, **GQT19**, **GT20**, **LL18**, **Tsu18**, **MT19**, **LT19**]. However, mean topological dimension is *not* the right notion for the purpose of this paper because Furstenberg's theorem (1.1) concerns Hausdorff and Minkowski dimensions, not the topological one. (The topological dimension of a subshift $\mathcal{X} \subset A^{\mathbb{N}}$ is simply zero.)

Let d be a metric (i.e. a distance function) on $\mathcal{X}$. Lindenstrauss and Weiss [**LW00**] defined *metric mean dimension* $\mathrm{mdim}_{\mathrm{M}}(\mathcal{X}, T, d)$. This is a dynamical analogue of Minkowski dimension. Lindenstrauss and the second named author [**LT19**] defined *mean Hausdorff dimension* $\mathrm{mdim}_{\mathrm{H}}(\mathcal{X}, T, d)$. This is a dynamical analogue of Hausdorff dimension. Metric mean dimension and mean Hausdorff dimension are metric-dependent quantities. They provide a good framework for the purpose of the paper.

It is well known in geometric measure theory [**Mat95**] that metrical dimensions are deeply connected to measure theory. In particular, we can introduce the concept of (metric-dependent) dimension for measure (see e.g. [**Rén59**, **Youn82**, **KD94**]). Similarly, we can introduce a mean dimensional quantity for invariant measures of dynamical systems. Let μ be a T-invariant Borel probability measure on $\mathcal{X}$. Let X be a random variable taking values in $\mathcal{X}$ according to the law μ; we consider the stochastic process $\{T^n X\}_{n\in\mathbb{Z}}$. We denote by $R(d, \mu, \varepsilon)$ ($\varepsilon > 0$) the *rate distortion function* of this stochastic process. This is the key quantity of Shannon's rate distortion theory [**Sh48**, **Sh59**]. It evaluates how many bits per iterate we need for describing the process within the distortion (with respect to d) bound by ε. Following Kawabata and Dembo [**KD94**], we define the *upper and lower rate*

† We can also consider a non-invertible map T as in §1.1. But we consider only invertible T here for simplicity.

distortion dimensions by†

$$\overline{\mathrm{rdim}}(\mathcal{X}, T, d, \mu) = \limsup_{\varepsilon \to 0} \frac{R(d, \mu, \varepsilon)}{\log(1/\varepsilon)}, \quad \underline{\mathrm{rdim}}(\mathcal{X}, T, d, \mu) = \liminf_{\varepsilon \to 0} \frac{R(d, \mu, \varepsilon)}{\log(1/\varepsilon)}. \tag{1.2}$$

When the upper and lower limits coincide, we denote the common value by $\mathrm{rdim}(\mathcal{X}, T, d, \mu)$.

Metric mean dimension, mean Hausdorff dimension and rate distortion dimension are related to each other. See Proposition 2.1 and Theorem 2.4 below.

1.3. *Statement of the main result.* Let A be a finite set as in §1.1. We consider the infinite product $A^{\mathbb{Z}^2}$ indexed by $\mathbb{Z}^2$. We define the shifts σ_1 and σ_2 on $A^{\mathbb{Z}^2}$ by

$$\sigma_1((x_{m,n})_{m,n\in\mathbb{Z}}) = (x_{m+1,n})_{m,n\in\mathbb{Z}}, \quad \sigma_2((x_{m,n})_{m,n\in\mathbb{Z}}) = (x_{m,n+1})_{m,n\in\mathbb{Z}}.$$

Fix $\alpha > 1$ and define a distance d on $A^{\mathbb{Z}^2}$ by

$$d(x, y) = \alpha^{-\min\{|u|_\infty \mid x_u \neq y_u\}}, \tag{1.3}$$

where $|u|_\infty = \max(|m|, |n|)$ for $u = (m, n) \in \mathbb{Z}^2$. We call a closed subset $\mathcal{X} \subset A^{\mathbb{Z}^2}$ a *subshift* if it is invariant under both σ_1 and σ_2.

The following is our main result.

THEOREM 1.1. *Let $\mathcal{X} \subset A^{\mathbb{Z}^2}$ be a subshift. Then*

$$\mathrm{mdim}_{\mathrm{H}}(\mathcal{X}, \sigma_1, d) = \mathrm{mdim}_{\mathrm{M}}(\mathcal{X}, \sigma_1, d) = \frac{2h_{\mathrm{top}}(\mathcal{X}, \sigma_1, \sigma_2)}{\log \alpha}. \tag{1.4}$$

Here $h_{\mathrm{top}}(\mathcal{X}, \sigma_1, \sigma_2)$ is the topological entropy of $(\mathcal{X}, \sigma_1, \sigma_2)$. Moreover, if μ is a Borel probability measure on $\mathcal{X}$ invariant under both σ_1 and σ_2, then

$$\mathrm{rdim}(\mathcal{X}, \sigma_1, d, \mu) = \frac{2h_\mu(\mathcal{X}, \sigma_1, \sigma_2)}{\log \alpha}.$$

Here $h_\mu(\mathcal{X}, \sigma_1, \sigma_2)$ is the Kolmogorov–Sinai entropy of $(\mathcal{X}, \sigma_1, \sigma_2)$ with respect to the measure μ.

In particular, if μ is a maximal entropy measure (i.e. $h_\mu(\mathcal{X}, \sigma_1, \sigma_2) = h_{\mathrm{top}}(\mathcal{X}, \sigma_1, \sigma_2)$), then $\mathrm{rdim}(\mathcal{X}, \sigma_1, d, \mu)$ coincides with the mean Hausdorff dimension and metric mean dimension.

The point of the statement is that we consider various mean dimensional quantities for the action of σ_1, not the total $\mathbb{Z}^2$-action generated by σ_1 and σ_2. In other words, we consider only σ_1 and disregard σ_2. Nevertheless, we can recover the entropy of the total $\mathbb{Z}^2$-action. This might look a bit strange at first sight. But in fact it has the same spirit as Furstenberg's theorem (see 1.1). In (1.1), we consider the Hausdorff and Minkowski dimensions of one-sided subshifts. Hausdorff and Minkowski dimensions are purely metric invariants and do not involve dynamics. So here we disregard the action at all. However, we can recover the topological entropy. See Remark 1.2(4) below for more background behind the formulation of the theorem.

† Throughout the paper, we assume that the base of the logarithm is two.

https://doi.org/10.1017/etds.2020.47 Published online by Cambridge University Press

Remark 1.2.

(1) Subshifts $\mathcal{X} \subset A^{\mathbb{Z}^2}$ are totally disconnected. So the mean topological dimension of $(\mathcal{X}, \sigma_1)$ is zero.

(2) Probably some readers might notice a slight difference between Furstenberg's theorem (1.1) and our (1.4): our formula involves the coefficient '2' where Furstenberg's theorem does not. This difference comes from the point that Furstenberg's theorem considers one-sided shifts (i.e. actions of $\mathbb{N}$, not $\mathbb{Z}$). If we consider two-sided shifts, then we get a result completely analogous to (1.4).

(3) Theorem 1.1 can be generalized to $\mathbb{Z}^k$-shifts and, probably, some non-commutative group actions. But we stick to $\mathbb{Z}^2$ for simplicity of the exposition.

(4) A guiding principle behind our theorem is as follows: let $T : \mathbb{Z}^k \times \mathcal{X} \to \mathcal{X}$ be a continuous action of $\mathbb{Z}^k$ on a compact metric space $\mathcal{X}$. If T has some hyperbolicity-like property, then we can control the mean dimensional quantities of the restriction of T to subgroups $G \subset \mathbb{Z}^k$ with rank $G = k - 1$.
When $k = 1$, the subgroup G must be trivial. So, in particular, this principle claims that we can control the dimensions of $\mathcal{X}$ if $\mathcal{X}$ admits an action of $\mathbb{Z}$ with some 'hyperbolicity'. Furstenberg's theorem (1.1) is a typical example of such results because symbolic dynamics can be seen as an extreme case of hyperbolic dynamics. Theorem (1.1) corresponds to the case of $k = 2$ in this principle.
Another manifestation of the above principle was given by the work of Meyerovitch and the second named author [**MT19**]. They proved that if $T : \mathbb{Z}^k \times \mathcal{X} \to \mathcal{X}$ is expansive, then the mean topological dimension of $T|_G$ is finite for any rank-$(k - 1)$ subgroups $G \subset \mathbb{Z}^k$. In particular, when $k = 1$, a compact metric space has finite topological dimension if it admits an expansive action of $\mathbb{Z}$. This is a classical theorem of Mañé [**Ma79**].

(5) We consider the action of σ_1 in Theorem 1.1. This corresponds to a study of the action of the subgroup $\{(n, 0) \mid n \in \mathbb{Z}\} \subset \mathbb{Z}^2$. According to the principle in (4) above, it is also natural to consider other rank-one subgroups. Namely, we should study various mean dimensional quantities for the action of $\sigma_1^a \sigma_2^b$ for any non-zero $(a, b) \in \mathbb{Z}^2$, which corresponds to the subgroup $\{(an, bn) \mid n \in \mathbb{Z}\}$.
Indeed we can calculate them. Take a non-zero $(a, b) \in \mathbb{Z}^2$. Then, for a subshift $\mathcal{X} \subset A^{\mathbb{Z}^2}$, we have

$$\begin{aligned} \mathrm{mdim}_{\mathrm{M}}(\mathcal{X}, \sigma_1^a\sigma_2^b, d) = \mathrm{mdim}_{\mathrm{H}}(\mathcal{X}, \sigma_1^a\sigma_2^b, d) &= 2(|a| + |b|)\frac{h_{\mathrm{top}}(\mathcal{X}, \sigma_1, \sigma_2)}{\log \alpha}, \\ \mathrm{rdim}(\mathcal{X}, \sigma_1^a\sigma_2^b, d, \mu) &= 2(|a| + |b|)\frac{h_\mu(\mathcal{X}, \sigma_1, \sigma_2)}{\log \alpha}. \end{aligned} \tag{1.5}$$

Here d is the metric defined by (1.3) and μ is a Borel probability measure on $\mathcal{X}$ invariant under σ_1 and σ_2.
The factor $2(|a| + |b|)$ in (1.5) has the following geometric meaning. For natural numbers M and N, we define $\Lambda_{a,b}(M, N) \subset \mathbb{Z}^2$ as the set of points

$$(an + x, bn + y) \quad (0 \le n < N,\ |(x, y)|_\infty < M).$$

https://doi.org/10.1017/etds.2020.47 Published online by Cambridge University Press

Here n, x, y are integers. (Namely, we consider the parallel translations of $(-M, M)^2$ along the segment $\{(an, bn)|\, 0 \le n < N\}$.) Then we have

$$2(|a| + |b|) = \lim_{M\to\infty}\left(\lim_{N\to\infty}\frac{|\Lambda_{a,b}(M, N)|}{MN}\right) \quad (|\cdot| \text{ denotes the cardinality}).$$

The square $(-M, M)^2$ is the disk of radius M in the ℓ^∞-norm $|u|_\infty$. The relevance of the ℓ^∞-norm here comes from the point that the metric (1.3) uses it. If we use a different metric, then we get a different result. For example, consider the following metric ρ on $A^{\mathbb{Z}^2}$:

$$\rho(x, y) = \alpha^{-\min\{\sqrt{m^2+n^2}|\, x_{m,n}\neq y_{m,n}\}}. \tag{1.6}$$

This metric uses the ℓ^2-norm $\sqrt{m^2+n^2}$ instead of the ℓ^∞-norm. For this metric, we have

$$\begin{aligned} \mathrm{mdim_M}(\mathcal{X}, \sigma_1^a\sigma_2^b, \rho) = \mathrm{mdim_H}(\mathcal{X}, \sigma_1^a\sigma_2^b, \rho) = 2\sqrt{a^2+b^2}\cdot\frac{h_{\mathrm{top}}(\mathcal{X}, \sigma_1, \sigma_2)}{\log\alpha}, \\ \mathrm{rdim}(\mathcal{X}, \sigma_1^a\sigma_2^b, \rho, \mu) = 2\sqrt{a^2+b^2}\cdot\frac{h_\mu(\mathcal{X}, \sigma_1, \sigma_2)}{\log\alpha}. \end{aligned} \tag{1.7}$$

The proofs of (1.5) and (1.7) are conceptually the same as the proof of Theorem 1.1. However, they become notationally more messy. So we decided to concentrate on the statement of Theorem 1.1. It clarifies the ideas in the simplest form.

2. *Preliminaries*

The purpose of this section is to define the three dynamical dimensions (metric mean dimension, mean Hausdorff dimension and rate distortion dimension)† and explain some of their basic properties.

2.1. *Metric mean dimension and mean Hausdorff dimension.*

Let $(\mathcal{X}, d)$ be a compact metric space. For $\varepsilon > 0$, we define $\#(\mathcal{X}, d, \varepsilon)$ as the minimum natural number n such that $\mathcal{X}$ can be covered by open sets $U_1, \dots, U_n$ with diam $U_i < \varepsilon$ for all $1 \le i \le n$. The upper and lower Minkowski dimensions of $(\mathcal{X}, d)$ are given by

$$\overline{\dim}_{\mathrm{M}}(\mathcal{X}, d) = \limsup_{\varepsilon\to 0}\frac{\log\#(\mathcal{X}, d, \varepsilon)}{\log(1/\varepsilon)}, \quad \underline{\dim}_{\mathrm{M}}(\mathcal{X}, d) = \liminf_{\varepsilon\to 0}\frac{\log\#(\mathcal{X}, d, \varepsilon)}{\log(1/\varepsilon)}.$$

For $s \ge 0$ and $\varepsilon > 0$, we define

$$\mathcal{H}^s_\varepsilon(\mathcal{X}, d) = \inf\left\{\sum_{i=1}^\infty(\mathrm{diam}\, E_i)^s \,\middle|\, \mathcal{X} = \bigcup_{n=1}^\infty E_i \text{ with diam } E_i < \varepsilon \text{ for all } i \ge 1\right\}.$$

Here we use the convention that $0^0 = 1$ and $\mathrm{diam}(\emptyset)^s = 0$. Since $\mathcal{X}$ is compact, this is equal to the infimum of

$$\sum_{i=1}^n(\mathrm{diam}\, U_i)^s$$

over all finite open covers $\{U_1, \dots, U_n\}$ of $\mathcal{X}$ with diam $U_i < \varepsilon$ for all $1 \le i \le n$. We set

$$\dim_{\mathrm{H}}(\mathcal{X}, d, \varepsilon) = \sup\{s \ge 0|\, \mathcal{H}^s_\varepsilon(\mathcal{X}, d) \ge 1\}.$$

† We do not use mean topological dimension in the paper. So we skip defining it.

https://doi.org/10.1017/etds.2020.47 Published online by Cambridge University Press

The Hausdorff dimension of $(\mathcal{X}, d)$ is given by

$$\dim_{\mathrm{H}}(\mathcal{X}, d) = \lim_{\varepsilon \to 0} \dim_{\mathrm{H}}(\mathcal{X}, d, \varepsilon).$$

Given a homeomorphism $T : \mathcal{X} \to \mathcal{X}$, we define metrics d_N^T $(N \geq 1)$ on $\mathcal{X}$ by

$$d_N^T(x, y) = \max_{0 \leq n < N} d(T^n x, T^n y).$$

We define the *entropy at the resolution* $\varepsilon > 0$ by

$$S(\mathcal{X}, T, d, \varepsilon) = \lim_{N \to \infty} \frac{\log \#(\mathcal{X}, d_N^T, \varepsilon)}{N}.$$

This limit exists because $\log \#(\mathcal{X}, d_N^T, \varepsilon)$ is subadditive in N. We define the *upper and lower metric mean dimensions* by

$$\overline{\mathrm{mdim}}_{\mathrm{M}}(\mathcal{X}, T, d) = \limsup_{\varepsilon \to 0} \frac{S(\mathcal{X}, T, d, \varepsilon)}{\log(1/\varepsilon)}, \quad \underline{\mathrm{mdim}}_{\mathrm{M}}(\mathcal{X}, T, d) = \liminf_{\varepsilon \to 0} \frac{S(\mathcal{X}, T, d, \varepsilon)}{\log(1/\varepsilon)}.$$

When the upper and lower limits coincide, we denote the common value by $\mathrm{mdim}_{\mathrm{M}}(\mathcal{X}, T, d)$.

We define the *upper and lower mean Hausdorff dimensions* by

$$\overline{\mathrm{mdim}}_{\mathrm{H}}(\mathcal{X}, T, d) = \lim_{\varepsilon \to 0} \left(\limsup_{N \to \infty} \frac{\dim_{\mathrm{H}}(\mathcal{X}, d_N, \varepsilon)}{N} \right),$$
$$\underline{\mathrm{mdim}}_{\mathrm{H}}(\mathcal{X}, T, d) = \lim_{\varepsilon \to 0} \left(\liminf_{N \to \infty} \frac{\dim_{\mathrm{H}}(\mathcal{X}, d_N, \varepsilon)}{N} \right).$$

When these two quantities are equal to one other, we denote the common value by $\mathrm{mdim}_{\mathrm{H}}(\mathcal{X}, T, d)$.

The following is the dynamical analogue of the fact that Minkowski dimension bounds Hausdorff dimension. It was proved in [**LT19**, Proposition 3.2].

PROPOSITION 2.1. *For a dynamical system* $(\mathcal{X}, T)$ *with a metric* d*,*

$$\underline{\mathrm{mdim}}_{\mathrm{H}}(\mathcal{X}, T, d) \leq \overline{\mathrm{mdim}}_{\mathrm{H}}(\mathcal{X}, T, d) \leq \underline{\mathrm{mdim}}_{\mathrm{M}}(\mathcal{X}, T, d) \leq \overline{\mathrm{mdim}}_{\mathrm{M}}(\mathcal{X}, T, d).$$

Remark 2.2. Here is one remark about the notation. In the paper [**LT19**], the lower mean Hausdorff dimension played no role. So the upper mean Hausdorff dimension was simply denoted by $\mathrm{mdim}_{\mathrm{H}}(\mathcal{X}, T, d)$ in [**LT19**].

2.2. *Mutual information.* Let $(\Omega, \mathbb{P})$ be a probability space. Let $\mathcal{X}$ and $\mathcal{Y}$ be measurable spaces, and let $X : \Omega \to \mathcal{X}$ and $Y : \Omega \to \mathcal{Y}$ be measurable maps. We want to define their *mutual information* $I(X; Y)$ as the measure of the amount of information X and Y share. (This will be used in the definition of rate distortion function in the next subsection.) The basic reference is [**CT06**].

Case 1: When $\mathcal{X}$ *and* $\mathcal{Y}$ *are finite sets.* In this case† we set

$$\begin{aligned} I(X; Y) &= H(X) + H(Y) - H(X, Y) = H(Y) - H(Y|X) \\ &= \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \mathbb{P}(X = x, Y = y) \log \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(X = x)\mathbb{P}(Y = y)}. \end{aligned} \tag{2.1}$$

† We always assume that the σ-algebras of finite sets are the largest ones, i.e. the sets of all subsets.

https://doi.org/10.1017/etds.2020.47 Published online by Cambridge University Press

Here we have used the convention that $0 \log(0/a) = 0$ for all $a \geq 0$.

Case 2: General case. Let $f : \mathcal{X} \to A$ and $g : \mathcal{Y} \to B$ be measurable maps such that A and B are finite sets. Then we can define $I(f \circ X; g \circ Y)$ by (2.1). We define $I(X; Y)$ as the supremum of $I(f \circ X; g \circ Y)$ over all finite range measurable maps f on $\mathcal{X}$ and g on $\mathcal{Y}$. When $\mathcal{X}$ and $\mathcal{Y}$ are finite sets, this definition is compatible with (2.1). (Namely, the supremum is attained when f and g are the identity maps.)

The mutual information is symmetric and non-negative: $I(X; Y) = I(Y; X) \geq 0$. The following basic result immediately follows from the above definition.

LEMMA 2.3. (Data-processing inequality) *Let $\mathcal{Z}$ and $\mathcal{W}$ be measurable spaces. If $f : \mathcal{X} \to \mathcal{Z}$ and $g : \mathcal{Y} \to \mathcal{W}$ are measurable maps, then $I(f(X); g(Y)) \leq I(X; Y)$.*

2.3. *Rate distortion theory.* Here we introduce the rate distortion function. As Shannon entropy is the fundamental limit of *lossless* data compression, the rate distortion function is the fundamental limit of *lossy* data compression†. A friendly introduction can be found in [**CT06**, Ch. 10].

Let $(\mathcal{X}, T)$ be a dynamical system with a metric d and an invariant Borel probability measure μ. We define the *rate distortion function* $R(d, \mu, \varepsilon)$ ($\varepsilon > 0$) as the infimum of

$$\frac{I(X; Y)}{N},$$

where N runs over natural numbers and X and $Y = (Y_0, \ldots, Y_{N-1})$ are random variables defined on some probability space $(\Omega, \mathbb{P})$ such that:

- X takes values in $\mathcal{X}$ according to the law μ;
- $Y_0, \ldots, Y_{N-1}$ take values in $\mathcal{X}$ and satisfy

$$\mathbb{E}\left(\frac{1}{N}\sum_{n=0}^{N-1} d(T^n X, Y_n)\right) < \varepsilon. \tag{2.2}$$

The condition (2.2) means that $Y = (Y_0, \ldots, Y_{N-1})$ approximates the stochastic process $X, TX, \ldots, T^{N-1}X$ within the averaged distortion bound by ε. We define the upper and lower rate distortion dimensions $\overline{\mathrm{rdim}}(\mathcal{X}, T, d, \mu)$ and $\underline{\mathrm{rdim}}(\mathcal{X}, T, d, \mu)$ by (1.2) in §1.2.

The rate distortion function $R(d, \mu, \varepsilon)$ is the minimum rate when we try to quantize the process $\{T^n X\}_{n\in\mathbb{Z}}$ within the averaged distortion bound by ε. See [**CT06**, Ch. 10], [**Gra90**, Ch. 11] and [**ECG94**, **LDN79**] for the precise meaning of this statement.

The rest of this subsection is not used in the proof of Theorem 1.1. We include this for providing the reader with a wider view of the subject. A metric d is said to have the *tame growth of covering numbers* if, for any $\delta > 0$,

$$\lim_{\varepsilon \to 0} \varepsilon^{\delta} \log \#(\mathcal{X}, d, \varepsilon) = 0. \tag{2.3}$$

Note that this is purely a condition on the metric structure and does not involve dynamics. The condition (2.3) is a mild condition. It is known [**LT19**, Lemma 3.10] that every compact metrizable space admits a metric satisfying (2.3). For example, the metrics (1.3)

† For example, expanding a given signal in a wavelet basis and discarding small terms.

https://doi.org/10.1017/etds.2020.47 Published online by Cambridge University Press

and (1.6) on the shift space $A^{\mathbb{Z}^2}$ satisfy (2.3). The following theorem [**LT19**, Proposition 3.2 and Theorem 3.11] provides a link between rate distortion dimension and various mean dimensions. Here we denote by $\mathscr{M}^T(\mathcal{X})$ the set of all invariant Borel probability measures on $\mathcal{X}$.

THEOREM 2.4. *In the above setting,*

$$\overline{\mathrm{rdim}}(\mathcal{X}, T, d, \mu) \le \overline{\mathrm{mdim}}_{\mathrm{M}}(\mathcal{X}, T, d), \quad \underline{\mathrm{rdim}}(\mathcal{X}, T, d, \mu) \le \underline{\mathrm{mdim}}_{\mathrm{M}}(\mathcal{X}, T, d).$$

If d has the tame growth of covering numbers, then

$$\overline{\mathrm{mdim}}_{\mathrm{H}}(\mathcal{X}, T, d) \le \sup_{\mu \in \mathscr{M}^T(\mathcal{X})} \underline{\mathrm{rdim}}(\mathcal{X}, T, d, \mu).$$

3. *Proof of Theorem 1.1*

First we recall the notation of §1.3. The shift $A^{\mathbb{Z}^2}$ is the $\mathbb{Z}^2$-full shift on the alphabet (finite set) A with the shifts σ_1 and σ_2. Fix $\alpha > 1$; we define the metric d on $A^{\mathbb{Z}^2}$ by

$$d(x, y) = \alpha^{-\min\{|u|_\infty \mid x_u \ne y_u\}}.$$

Let $\mathcal{X} \subset A^{\mathbb{Z}^2}$ be a subshift (closed shift-invariant set) with a Borel probability measure μ invariant under both σ_1 and σ_2.

The proof of Theorem 1.1 is divided into four steps.

(1) Prove the upper bound on the upper metric mean dimension

$$\overline{\mathrm{mdim}}_{\mathrm{M}}(\mathcal{X}, \sigma_1, d) \le \frac{2h_{\mathrm{top}}(\mathcal{X}, \sigma_1, \sigma_2)}{\log \alpha}.$$

(2) Prove the lower bound on the lower mean Hausdorff dimension

$$\underline{\mathrm{mdim}}_{\mathrm{H}}(\mathcal{X}, \sigma_1, d) \ge \frac{2h_{\mathrm{top}}(\mathcal{X}, \sigma_1, \sigma_2)}{\log \alpha}.$$

(3) Prove the upper bound on the upper rate distortion dimension

$$\overline{\mathrm{rdim}}(\mathcal{X}, \sigma_1, d, \mu) \le \frac{2h_{\mu}(\mathcal{X}, \sigma_1, \sigma_2)}{\log \alpha}.$$

(4) Prove the lower bound on the lower rate distortion dimension

$$\underline{\mathrm{rdim}}(\mathcal{X}, \sigma_1, d, \mu) \ge \frac{2h_{\mu}(\mathcal{X}, \sigma_1, \sigma_2)}{\log \alpha}.$$

Since we know that $\underline{\mathrm{mdim}}_{\mathrm{H}}(\mathcal{X}, \sigma_1, d) \le \overline{\mathrm{mdim}}_{\mathrm{M}}(\mathcal{X}, \sigma_1, d)$ by Proposition 2.1, the steps (1) and (2) show that

$$\mathrm{mdim}_{\mathrm{H}}(\mathcal{X}, \sigma_1, d) = \mathrm{mdim}_{\mathrm{M}}(\mathcal{X}, \sigma_1, d) = \frac{2h_{\mathrm{top}}(\mathcal{X}, \sigma_1, \sigma_2)}{\log \alpha}.$$

The steps (3) and (4) show that

$$\mathrm{rdim}(\mathcal{X}, \sigma_1, d, \mu) = \frac{2h_{\mu}(\mathcal{X}, \sigma_1, \sigma_2)}{\log \alpha}.$$

The steps (1) and (3) are easy. The step (2) is the most involved. The four steps are independent of each other.

For $\Omega \subset \mathbb{Z}^2$, we denote by $\pi_\Omega : \mathcal{X} \to A^\Omega$ the natural projection. As in §2.1, we set $d_N^{\sigma_1}(x, y) = \max_{0 \le n < N} d(\sigma_1^n x, \sigma_1^n y)$ for $N > 0$. In this section, intervals mean *discrete* intervals. Namely, for example, $[a, b] = \{a, a+1, \ldots, b-1, b\}$ and $(a, b) = \{a + 1, a+2, \ldots, b-1\}$ for integers $a \le b$.

https://doi.org/10.1017/etds.2020.47 Published online by Cambridge University Press

3.1. *Step 1: Proof of* $\overline{\mathrm{mdim}}_{\mathrm{M}}(\mathcal{X}, \sigma_1, d) \le 2h_{\mathrm{top}}(\mathcal{X}, \sigma_1, \sigma_2)/\log\alpha$. Let $0 < \varepsilon < 1$ and take a natural number M with $\alpha^{-M} < \varepsilon \le \alpha^{-M+1}$. Then

$$\#(\mathcal{X}, d_N^{\sigma_1}, \varepsilon) \le |\pi_{(-M,N+M)\times(-M,M)}(\mathcal{X})|.$$

(Here $|\cdot|$ denotes the cardinality.) Since $(M-1)\log\alpha \le \log(1/\varepsilon) < M\log\alpha$,

$$\begin{aligned}\overline{\mathrm{mdim}}_{\mathrm{M}}(\mathcal{X}, \sigma_1, d) &= \limsup_{\varepsilon\to 0}\left(\lim_{N\to\infty}\frac{\log\#(\mathcal{X}, d_N^{\sigma_1}, \varepsilon)}{N\log(1/\varepsilon)}\right)\\ &\le \lim_{M\to\infty}\left(\lim_{N\to\infty}\frac{\log|\pi_{(-M,N+M)\times(-M,M)}(\mathcal{X})|}{N(M-1)\log\alpha}\right)\\ &= \frac{2h_{\mathrm{top}}(\mathcal{X}, \sigma_1, \sigma_2)}{\log\alpha}.\end{aligned}$$

3.2. *Step 2: Proof of* $\underline{\mathrm{mdim}}_{\mathrm{H}}(\mathcal{X}, \sigma_1, d) \ge 2h_{\mathrm{top}}(\mathcal{X}, \sigma_1, \sigma_2)/\log\alpha$. First we prepare some terminology about the geometry of $\mathbb{Z}^2$. In this subsection, *rectangles* mean sets of the form $[a, b] \times [c, d]$ in $\mathbb{Z}^2$ for integers $a \le b$ and $c \le d$. For a rectangle $R = [a, b] \times [c, d]$, we define a new rectangle $3R$ by

$$3R = [2a - b, 2b - a] \times [2c - d, 2d - c].$$

We have $|3R| = (3b - 3a + 1)(3d - 3c + 1) \le 9|R|$.

For two rectangles $R = [a, b] \times [c, d]$ and $R' = [a', b'] \times [c', d']$, we denote $R \le R'$ if $b - a \le b' - a'$ and $d - c \le d' - c'$. This defines an order among rectangles. (Strictly speaking, this is a 'pre-order' because $R \le R'$ and $R' \le R$ do not imply that $R = R'$.) A set of rectangles $\{R_1, \ldots, R_n\}$ is said to be *totally ordered* if any two elements are comparable, i.e. for any R_i and R_j we have either $R_i \le R_j$ or $R_j \le R_i$.

The following trivial fact will be used later: suppose that $\{R_1, \ldots, R_n\}$ is totally ordered. If a set of rectangles $\{R'_1, \ldots, R'_{n'}\}$ has the property that each R'_i is a parallel translation of some R_j (namely, $R'_i = u + R_j$ for some $u \in \mathbb{Z}^2$), then $\{R'_1, \ldots, R'_{n'}\}$ is also totally ordered.

The next lemma is a kind of finite Vitali covering lemma [**EW11**, Lemma 2.27] adapted to our situation.

LEMMA 3.1. *Suppose that a set of rectangles* $\{R_1, \ldots, R_n\}$ *is totally ordered. Then we can find a disjoint subfamily* $\{R_{i_1}, \ldots, R_{i_m}\}$ *satisfying*

$$R_1 \cup \cdots \cup R_n \subset 3R_{i_1} \cup 3R_{i_2} \cup \cdots \cup 3R_{i_m}.$$

Note that this implies that

$$|R_{i_1} \cup \cdots \cup R_{i_m}| \ge \tfrac{1}{9}|R_1 \cup \cdots \cup R_n|.$$

Proof. We use a simple greedy algorithm. We first choose (one of) the largest rectangle, say R_{i_1}. Next, suppose that we have chosen $R_{i_1}, \ldots, R_{i_k}$. We choose as $R_{i_{k+1}}$ the largest rectangle disjoint to $R_{i_1} \cup \cdots \cup R_{i_k}$. If there is no such a rectangle, the algorithm stops.

Suppose that the algorithm stops after m steps. For any R_j, there exists R_{i_k} with $R_{i_k} \ge R_j$ and $R_{i_k} \cap R_j \ne \emptyset$. This implies that $R_j \subset 3R_{i_k}$. □

https://doi.org/10.1017/etds.2020.47 Published online by Cambridge University Press

For two sets $\Omega, \Lambda \subset \mathbb{Z}^2$, we define $\partial_\Lambda \Omega$ as the set of $u \in \mathbb{Z}^2$ such that $u + \Lambda$ has non-empty intersections with both Ω and $\mathbb{Z}^2 \setminus \Omega$. We set $\mathrm{Int}_\Lambda \Omega = \Omega \setminus \partial_\Lambda \Omega$. This is the set of $u \in \Omega$ with $u + \Lambda \subset \Omega$.

Let $R \subset \mathbb{Z}^2$ be a rectangle. A subset $C \subset \mathcal{X}$ is called a *cylinder over* R if there is $x \in \mathcal{X}$ such that C is equal to the set of $y \in \mathcal{X}$ satisfying $\pi_R(y) = \pi_R(x)$.

Set

$$s = \frac{2h_{\mathrm{top}}(\mathcal{X}, \sigma_1, \sigma_2)}{\log \alpha}.$$

Suppose that $\underline{\mathrm{mdim}}_{\mathrm{H}}(\mathcal{X}, \sigma_1, d) < s$. We would like to get a contradiction. We fix $\varepsilon > 0$ satisfying $\underline{\mathrm{mdim}}_{\mathrm{H}}(\mathcal{X}, \sigma_1, d) < s - 2\varepsilon$.

LEMMA 3.2. *For any finite subset $\Lambda \subset \mathbb{Z}^2$ and any positive number L, we can find rectangles $R_1, \ldots, R_M \subset \mathbb{Z}^2$ and subsets $C_1, \ldots, C_M \subset \mathcal{X}$ such that:*

- *each C_m is a cylinder over R_m and they satisfy $\mathcal{X} = \bigcup_{m=1}^M C_m$;*
- *all the rectangles R_m contain the origin and they are all sufficiently large so that*

$$|\partial_\Lambda R_m| < \frac{|R_m|}{L}, \quad |R_m| > L;$$

- *the rectangles $R_1, \ldots, R_M$ are totally ordered and satisfy*

$$\sum_{m=1}^M \alpha^{-1/2(s-\varepsilon)|R_m|} < 1.$$

Proof. We choose a natural number r_0 such that:

- every $r \geq r_0$ satisfies $(s - 2\varepsilon)r < (s - \varepsilon)(r - 1)$;
- if a rectangle $R = [a, b] \times [c, d] \subset \mathbb{Z}^2$ satisfies $b - a \geq r_0$ and $d - c \geq r_0$, then

$$|\partial_\Lambda R| < \frac{|R|}{L}, \quad |R| > L.$$

From $\underline{\mathrm{mdim}}_{\mathrm{H}}(\mathcal{X}, \sigma_1, d) < s - 2\varepsilon$, we can find $N > 0$ satisfying

$$\frac{1}{N} \dim_{\mathrm{H}}(\mathcal{X}, d_N^{\sigma_1}, \alpha^{-r_0}) < s - 2\varepsilon.$$

This implies that there exists a covering $\mathcal{X} = E_1 \cup \cdots \cup E_M$ satisfying

$$\mathrm{diam}(E_m, d_N^{\sigma_1}) < \alpha^{-r_0} \text{ (for all} 1 \leq m \leq M), \quad \sum_{m=1}^M (\mathrm{diam}(E_m, d_N^{\sigma_1}))^{(s-2\varepsilon)N} < 1.$$

Set $\alpha^{-r_m} := \mathrm{diam}(E_m, d_N^{\sigma_1})$. Then r_m is a natural number with $r_m > r_0$. Choose a point x_m from each E_m and let $C_m \subset \mathcal{X}$ be a cylinder over the rectangle

$$R_m := [-r_m + 1, N + r_m - 2] \times [-r_m + 1, r_m - 1]$$

defined by $C_m = \pi_{R_m}^{-1}(\pi_{R_m}(x_m))$. Then $E_m \subset C_m$ and hence $\mathcal{X} = C_1 \cup \cdots \cup C_m$. The rectangles R_m are totally ordered ($R_m \leq R_{m'}$ if and only if $r_m \leq r_{m'}$).

Recall that $r_m > r_0$ for all $1 \leq m \leq M$. From the choice of r_0,

$$|\partial_\Lambda R_m| < \frac{|R_m|}{L}, \quad |R_m| > L.$$

From $|R_m| = (N + 2r_m - 2)(2r_m - 1) \geq N(2r_m - 1)$,

$$\begin{aligned}\tfrac{1}{2}(s-\varepsilon)|R_m| &\geq \tfrac{1}{2}(s-\varepsilon)N(2r_m - 1)\\ &> \tfrac{1}{2}(s-2\varepsilon)N(2r_m) \quad \text{by the choice of } r_0\\ &= (s-2\varepsilon)Nr_m.\end{aligned}$$

Hence,

$$\alpha^{-1/2(s-\varepsilon)|R_m|} < \alpha^{-(s-2\varepsilon)Nr_m} = (\operatorname{diam}(E_m, d_N^{\sigma_1}))^{(s-2\varepsilon)N}.$$

Therefore,

$$\sum_{m=1}^{M} \alpha^{-1/2(s-\varepsilon)|R_m|} < \sum_{m=1}^{M} (\operatorname{diam}(E_m, d_N^{\sigma_1}))^{(s-2\varepsilon)N} < 1. \qquad \square$$

We choose a real number $0 < \delta < 1/2$ and a natural number p satisfying the following conditions.

$$\left(\tfrac{17}{18}\right)^p < \delta, \quad H(\delta) + \delta \log p < \frac{\varepsilon}{8}\log\alpha, \quad |A|^\delta < \alpha^{\varepsilon/8}. \tag{3.1}$$

Here $H(\delta) = -\delta\log\delta - (1-\delta)\log(1-\delta)$. (Recall that the base of the logarithm is two.) The first condition is satisfied for $p \approx \log(1/\delta)$. Then we choose a sufficiently small δ satisfying the second and third conditions.

By using Lemma 3.2 iteratively, we find rectangles $R_{i,m}$ and subsets $C_{i,m} \subset \mathcal{X}$ for $i = 1, \ldots, p$ and $m = 1, \ldots, M_i$ (where M_i is a natural number depending on i) satisfying the following conditions.

(a) Each $C_{i,m}$ is a cylinder over $R_{i,m}$. For each $1 \leq i \leq p$, we have $\mathcal{X} = \bigcup_{m=1}^{M_i} C_{i,m}$.

(b) For each $1 \leq i \leq p$, the rectangles $R_{i,1}, R_{i,2}, \ldots, R_{i,M_i}$ are totally ordered and satisfy

$$\sum_{m=1}^{M_i} \alpha^{-1/2(s-\varepsilon)|R_{i,m}|} < 1. \tag{3.2}$$

(c) All the rectangles $R_{i,m}$ contain the origin and they satisfy $|R_{i,m}| > 1/\delta$.

(d) Set $\hat{R}_i = \bigcup_{m=1}^{M_i} R_{i,m}$. Then, for all $j < i$ and $m = 1, \ldots, M_j$, we have

$$|\partial_{\hat{R}_i} R_{j,m}| < \frac{\delta}{4}|R_{j,m}|.$$

Roughly speaking, the condition (d) means that the rectangles in one level (say, j) are much larger than the rectangles in higher levels (say, $i > j$). The construction goes from the level p to the bottom. First, by Lemma 3.2, we construct $R_{p,m}$ and $C_{p,m}$. Next, by using the lemma again, we construct $R_{p-1,m}$ and $C_{p-1,m}$. We continue this process until we come to the first level ($R_{1,m}$ and $C_{1,m}$). The condition (d) connects the constructions in different levels.

LEMMA 3.3. *If $N > 0$ is sufficiently large, then the following statement holds. For each $x \in \mathcal{X}$, we can choose a subset*

$$D(x) \subset \{(u, i, m) \mid u \in [0, N-1]^2, 1 \leq i \leq p, 1 \leq m \leq M_i\}$$

such that:

https://doi.org/10.1017/etds.2020.47 Published online by Cambridge University Press

(1) *for* $(u, i, m) \in D(x)$*, we have* $\sigma^u(x) \in C_{i,m}$ *and* $u + R_{i,m} \subset [0, N-1]^2$*;*
(2) *if* (u, i, m) *and* (u', i', m') *are two different elements of* $D(x)$*, then* $(u + R_{i,m}) \cap (u' + R_{i',m'}) = \emptyset$*. In particular (recall that* $R_{i,m}$ *contain the origin),* $u \neq u'$*;*
(3) *we have*

$$\left|[0, N-1]^2 \setminus \bigcup_{(u,i,m)\in D(x)} (u + R_{i,m})\right| < \delta N^2.$$

Proof. Let N be sufficiently large so that

$$|\partial_{\hat{R}_i}[0, N-1]^2| < \frac{\delta}{4} N^2 \quad \text{for all } 1 \le i \le p. \tag{3.3}$$

Here recall that $\hat{R}_i = \bigcup_{m=1}^{M_i} R_{i,m}$. Fix $x \in \mathcal{X}$. Set $E_0 = [0, N-1]^2$. We will inductively construct $E_0 \supset E_1 \supset E_2 \supset \cdots \supset E_p$.

Suppose that we have defined $E_0, E_1, \ldots, E_{i-1}$. Consider the following set of rectangles:

$$\{u + R_{i,m} \mid u \in E_{i-1}, m \in [1, M_i] \text{ with } \sigma^u(x) \in C_{i,m} \text{ and } u + R_{i,m} \subset E_{i-1}\}. \tag{3.4}$$

Since $R_{i,1}, \ldots, R_{i,M_i}$ are totally ordered, so is (3.4). (Here the point is that i is fixed.) The rectangles (3.4) cover $\mathrm{Int}_{\hat{R}_i} E_{i-1}$. Then, by Lemma 3.1, we can find a subset

$$D_i(x) \subset \{(u, m) \mid u \in E_{i-1}, 1 \le m \le M_i\}$$

such that:

- for $(u, m) \in D_i(x)$, we have $\sigma^u(x) \in C_{i,m}$ and $u + R_{i,m} \subset E_{i-1}$;
- if (u, m) and (u', m') are two different elements of $D_i(x)$, then $(u + R_{i,m}) \cap (u' + R_{i,m'}) = \emptyset$;
- the rectangles $u + R_{i,m}$, $(u, m) \in D_i(x)$, cover at least one-ninth of $\mathrm{Int}_{\hat{R}_i} E_{i-1}$:

$$\left|\bigcup_{(u,m)\in D_i(x)} (u + R_{i,m})\right| \ge \frac{1}{9} |\mathrm{Int}_{\hat{R}_i} E_{i-1}|. \tag{3.5}$$

We set

$$E_i = E_{i-1} \setminus \bigcup_{(u,m)\in D_i(x)} (u + R_{i,m}).$$

We define $D(x)$ by

$$D(x) = \{(u, i, m) \mid 1 \le i \le p, (u, m) \in D_i(x)\}.$$

The properties (1) and (2) of $D(x)$ immediately follow from the construction. The property (3) is equivalent to the claim that $|E_p| < \delta N^2$. We will prove this.

Suppose that $|E_p| \ge \delta N^2$. Then we also have $|E_{i-1}| \ge \delta N^2$ for all $1 \le i \le p$. We estimate $|\partial_{\hat{R}_i} E_{i-1}|$ for $1 \le i \le p$. We have

$$\partial_{\hat{R}_i} E_{i-1} \subset \partial_{\hat{R}_i}[0, N-1]^2 \cup \bigcup_{j=1}^{i-1} \bigcup_{(u,m)\in D_j(x)} \partial_{\hat{R}_i}(u + R_{j,m}).$$

https://doi.org/10.1017/etds.2020.47 Published online by Cambridge University Press

Recall (3.3) and that $|\partial_{\hat{R}_i} R_{j,m}| < (\delta/4)|R_{j,m}|$ for $j < i$ by the condition (d) of the choice of $R_{i,m}$. Then

$$|\partial_{\hat{R}_i} E_{i-1}| \le |\partial_{\hat{R}_i}[0, N-1]^2| + \sum_{j=1}^{i-1} \sum_{(u,m)\in D_j(x)} |\partial_{\hat{R}_i}(u + R_{j,m})|$$
$$< \frac{\delta}{4}N^2 + \frac{\delta}{4}\sum_{j=1}^{i-1} \sum_{(u,m)\in D_j(x)} |u + R_{j,m}|.$$

The rectangles $u + R_{j,m}$, $1 \le j \le i-1$ and $(u, m) \in D_j(x)$, are disjoint and contained in $[0, N-1]^2$. Therefore,

$$\sum_{j=1}^{i-1} \sum_{(u,m)\in D_j(x)} |u + R_{j,m}| \le N^2.$$

Thus, $|\partial_{\hat{R}_i} E_{i-1}| < (\delta/2)N^2$. Since we assumed that $|E_{i-1}| \ge \delta N^2$, we have $|\partial_{\hat{R}_i} E_{i-1}| < (1/2)|E_{i-1}|$. Namely,

$$|\mathrm{Int}_{\hat{R}_i} E_{i-1}| > \tfrac{1}{2}|E_{i-1}|.$$

From (3.5),

$$\left| \bigcup_{(u,m)\in D_i(x)} (u + R_{i,m}) \right| \ge \frac{1}{9}|\mathrm{Int}_{\hat{R}_i} E_{i-1}| > \frac{1}{18}|E_{i-1}|.$$

So we get

$$|E_i| = \left| E_{i-1} \setminus \bigcup_{(u,m)\in D_i(x)} (u + R_{i,m}) \right| < \frac{17}{18}|E_{i-1}|.$$

This holds for all $1 \le i \le p$. Therefore,

$$|E_p| < \left(\frac{17}{18}\right)^p |E_0| = \left(\frac{17}{18}\right)^p N^2.$$

Recall that p satisfies $(17/18)^p < \delta$ by (3.1). So $|E_p| < \delta N^2$. This is a contradiction. □

In the rest of this subsection, N is assumed to be so large that the statement of Lemma 3.3 holds. For each $x \in \mathcal{X}$, we define $\underline{D}(x) \subset [0, N-1]^2 \times [1, p]$ as the set of $(u, i) \in [0, N-1]^2 \times [1, p]$ such that there exists $m \in [1, M_i]$ with $(u, i, m) \in D(x)$. (Notice that the sets $D(x)$ and $\underline{D}(x)$ depend on N. So it might be better to use the notation $D^{(N)}(x)$ and $\underline{D}^{(N)}(x)$. But we prefer the simpler one here.)

LEMMA 3.4. *If N is sufficiently large, then the number of possibilities of $\underline{D}(x)$ is bounded as follows:*

$$|\{\underline{D}(x) \mid x \in \mathcal{X}\}| < \alpha^{(\varepsilon/8)N^2}.$$

Proof. We use the well-known bound on the binomial coefficient:

$$\binom{n}{k} \le 2^{nH(k/n)}. \tag{3.6}$$

https://doi.org/10.1017/etds.2020.47 Published online by Cambridge University Press

This follows from

$$1 = \left\{\frac{k}{n} + \left(1 - \frac{k}{n}\right)\right\}^n \geq \binom{n}{k}\left(\frac{k}{n}\right)^k\left(1 - \frac{k}{n}\right)^{n-k} = \binom{n}{k}2^{-nH(k/n)}.$$

Let $x \in \mathcal{X}$ and set $D(x) = \{(u_1, i_1, m_1), \ldots, (u_k, i_k, m_k)\}$. (Then we have $\underline{D}(x) = \{(u_1, i_1), \ldots, (u_k, i_k)\}$.) By (1) and (2) of Lemma 3.3, $u_1, \ldots, u_k$ are different from each other and the rectangles $u_1 + R_{i_1,m_1}, \ldots, u_k + R_{i_k,m_k}$ are disjoint and contained in $[0, N-1]^2$. Since $|R_{i,m}| > 1/\delta$ (the condition (c) of the choice of $R_{i,m}$), we have $k < \delta N^2$.

Then the number of possibilities of $\underline{D}(x)$ is bounded by

$$\begin{aligned}
&\underbrace{\left\{\binom{N^2}{1} + \binom{N^2}{2} + \cdots + \binom{N^2}{\lfloor \delta N^2 \rfloor}\right\}}_{\text{choices of } u_1, \ldots, u_k} \times \underbrace{p^{\delta N^2}}_{\text{choices of } i_1, \ldots, i_k} \\
&\leq N^2 \cdot 2^{N^2 H(\delta)} \times p^{\delta N^2} \quad \text{by (3.6)} \\
&= N^2 \cdot 2^{N^2(H(\delta) + \delta \log p)}.
\end{aligned}$$

We assumed that $H(\delta) + \delta \log p < (\varepsilon/8) \log \alpha$ in (3.1). Hence, if N is sufficiently large, then

$$N^2 \cdot 2^{N^2(H(\delta)+\delta \log p)} < 2^{N^2(\varepsilon/8)\log \alpha} = \alpha^{(\varepsilon/8)N^2}. \qquad \square$$

Take a subset $E \subset [0, N-1]^2 \times [1, p]$ such that there exists $x \in \mathcal{X}$ with $\underline{D}(x) = E$. We denote by $\mathcal{X}_E$ the set of $x \in \mathcal{X}$ with $\underline{D}(x) = E$. Let $E = \{(u_1, i_1), (u_2, i_2), \ldots, (u_k, i_k)\}$.

LEMMA 3.5. *We have the following estimate:*

$$\begin{aligned}
&|\pi_{[0,N-1]^2}(\mathcal{X}_E)| \cdot \alpha^{-1/2(s-\varepsilon)N^2} \\
&\leq |A|^{\delta N^2}\left(\sum_{m=1}^{M_{i_1}} \alpha^{-1/2(s-\varepsilon)|R_{i_1,m}|}\right) \times \cdots \times \left(\sum_{m=1}^{M_{i_k}} \alpha^{-1/2(s-\varepsilon)|R_{i_k,m}|}\right). \qquad (3.7)
\end{aligned}$$

Proof. For $\mathbf{m} = (m_1, \ldots, m_k) \in [1, M_{i_1}] \times \cdots \times [1, M_{i_k}]$, we denote by $\mathcal{X}_{E,\mathbf{m}} \subset \mathcal{X}_E$ the set of $x \in \mathcal{X}_E$ with $D(x) = \{(u_1, i_1, m_1), (u_2, i_2, m_2), \ldots, (u_k, i_k, m_k)\}$. We have $\sigma^{u_j}(x) \in C_{i_j,m_j}$ for $x \in \mathcal{X}_{E,\mathbf{m}}$. Hence, over each rectangle $u_j + R_{i_j,m_j}$, the value of $\pi_{u_j+R_{i_j,m_j}}(x)$ ($x \in \mathcal{X}_{E,\mathbf{m}}$) is fixed. (Namely, we have $\pi_{u_j+R_{i_j,m_j}}(x) = \pi_{u_j+R_{i_j,m_j}}(x')$ for any two $x, x' \in \mathcal{X}_{E,\mathbf{m}}$.) Therefore, we have

$$|\pi_{[0,N-1]^2}(\mathcal{X}_{E,\mathbf{m}})| \leq |A|^{|[0,N-1]^2 \setminus \bigcup_{j=1}^k (u_j + R_{i_j,m_j})|} < |A|^{\delta N^2}.$$

Here the second inequality follows from the condition (3) of Lemma 3.3. We decompose the left-hand side of (3.7) as

$$\begin{aligned}
|\pi_{[0,N-1]^2}(\mathcal{X}_E)| \cdot \alpha^{-1/2(s-\varepsilon)N^2} &= \sum_{\mathbf{m}} |\pi_{[0,N-1]^2}(\mathcal{X}_{E,\mathbf{m}})| \cdot \alpha^{-1/2(s-\varepsilon)N^2} \\
&\leq \sum_{\mathbf{m} \text{ with } \mathcal{X}_{E,\mathbf{m}} \neq \emptyset} |A|^{\delta N^2} \cdot \alpha^{-1/2(s-\varepsilon)N^2}. \qquad (3.8)
\end{aligned}$$

https://doi.org/10.1017/etds.2020.47 Published online by Cambridge University Press

Take $\mathbf{m} = (m_1, \ldots, m_k) \in [1, M_{i_1}] \times \cdots \times [1, M_{i_k}]$ with $\mathcal{X}_{E,\mathbf{m}} \neq \emptyset$. The rectangles $u_j + R_{i_j,m_j}$ $(1 \le j \le k)$ are disjoint and contained in $[0, N-1]^2$ by the conditions (1) and (2) of Lemma 3.3. Hence,

$$N^2 \ge \sum_{j=1}^{k} |R_{i_j,m_j}|.$$

So

$$\alpha^{-1/2(s-\varepsilon)N^2} \le \prod_{j=1}^{k} \alpha^{-1/2(s-\varepsilon)|R_{i_j,m_j}|}.$$

Inserting this into (3.8), we get

$$|\pi_{[0,N-1]^2}(\mathcal{X}_E)| \cdot \alpha^{-1/2(s-\varepsilon)N^2} \le \sum_{\mathbf{m}} |A|^{\delta N^2} \prod_{j=1}^{k} \alpha^{-1/2(s-\varepsilon)|R_{i_j,m_j}|}.$$

The right-hand side is equal to

$$|A|^{\delta N^2} \left(\sum_{m=1}^{M_{i_1}} \alpha^{-1/2(s-\varepsilon)|R_{i_1,m}|} \right) \times \cdots \times \left(\sum_{m=1}^{M_{i_k}} \alpha^{-1/2(s-\varepsilon)|R_{i_k,m}|} \right).$$

□

We continue the estimates as follows.

$$\begin{aligned} &|\pi_{[0,N-1]^2}(\mathcal{X}_E)| \cdot \alpha^{-1/2(s-\varepsilon)N^2} \\ &\quad \le |A|^{\delta N^2} \left(\sum_{m=1}^{M_{i_1}} \alpha^{-1/2(s-\varepsilon)|R_{i_1,m}|} \right) \times \cdots \times \left(\sum_{m=1}^{M_{i_k}} \alpha^{-1/2(s-\varepsilon)|R_{i_k,m}|} \right) \\ &\quad < |A|^{\delta N^2} \quad \text{by (3.2)} \\ &\quad < \alpha^{(\varepsilon/8)N^2}, \quad \text{since we assumed } |A|^{\delta} < \alpha^{\varepsilon/8} \text{ in (3.1).} \end{aligned}$$

The number of choices of $E \subset [0, N-1]^2 \times [1, p]$ with $\mathcal{X}_E \neq \emptyset$ is bounded by $\alpha^{(\varepsilon/8)N^2}$ if N is sufficiently large (Lemma 3.4). Then

$$\begin{aligned} |\pi_{[0,N-1]^2}(\mathcal{X})| \cdot \alpha^{-1/2(s-\varepsilon)N^2} &= \sum_{E \text{ with } \mathcal{X}_E \neq \emptyset} |\pi_{[0,N-1]^2}(\mathcal{X}_E)| \cdot \alpha^{-1/2(s-\varepsilon)N^2} \\ &< \alpha^{(\varepsilon/8)N^2} \times \alpha^{(\varepsilon/8)N^2} = \alpha^{(\varepsilon/4)N^2}. \end{aligned}$$

Therefore,

$$|\pi_{[0,N-1]^2}(\mathcal{X})| < \alpha^{1/2(s-\frac{\varepsilon}{2})N^2}.$$

Namely,

$$\frac{\log |\pi_{[0,N-1]^2}(\mathcal{X})|}{N^2} < \frac{1}{2}\left(s - \frac{\varepsilon}{2}\right) \log \alpha.$$

Letting $N \to \infty$,

$$h_{\text{top}}(\mathcal{X}, \sigma_1, \sigma_2) \le \frac{1}{2}\left(s - \frac{\varepsilon}{2}\right) \log \alpha < \frac{1}{2} s \log \alpha = h_{\text{top}}(\mathcal{X}, \sigma_1, \sigma_2).$$

This is a contradiction.

Remark 3.6.

(1) The above proof (in particular, see the proof of Lemma 3.2) also shows a (seemingly) slightly stronger statement that

$$\lim_{\varepsilon\to 0}\left(\inf_{N\geq 1}\frac{\dim_{\mathrm{H}}(\mathcal{X}, d_N^{\sigma_1}, \varepsilon)}{N}\right) \geq \frac{2h_{\mathrm{top}}(\mathcal{X}, \sigma_1, \sigma_2)}{\log \alpha}.$$

Combined with Step 1, both sides actually coincide. However, we do not know whether the left-hand side is an important quantity or not.

(2) The above proof (in particular, the use of the covering argument) is motivated by the proof of the Shannon–McMillan–Breiman theorem (see e.g. [**OW83**, **Rud90**, **Lin01**]). We expect that there is a proof more directly using the Shannon–McMillan–Breiman theorem (or related measure-theoretic ideas) although we have not found it so far.

3.3. *Step 3: Proof of* $\overline{\mathrm{rdim}}(\mathcal{X}, \sigma_1, d, \mu) \leq 2h_\mu(\mathcal{X}, \sigma_1, \sigma_2)/\log\alpha$. Let X be a random variable taking values in $\mathcal{X}$ and obeying μ. Let $0 < \varepsilon < 1$ and take $M > 0$ with $\alpha^{-M} < \varepsilon \leq \alpha^{-M+1}$ as in Step 1. Let $N > 0$. For each point $x \in \pi_{(-M,N+M)\times(-M,M)}(\mathcal{X})$, we choose $q(x) \in \mathcal{X}$ with $\pi_{(-M,N+M)\times(-M,M)}(q(x)) = x$. Set $X' = q(\pi_{(-M,N+M)\times(-M,M)}(X))$ and $Y = (X', \sigma_1 X', \sigma_1^2 X', \ldots, \sigma_1^{N-1}X')$. Then

$$\frac{1}{N}\sum_{n=0}^{N-1} d(\sigma_1^n X, Y_n) = \frac{1}{N}\sum_{n=0}^{N-1} d(\sigma_1^n X, \sigma_1^n X') \leq \alpha^{-M} < \varepsilon,$$

$$I(X; Y) \leq H(Y) = H(X') = H\{(X_u)_{u\in(-M,N+M)\times(-M,M)}\}.$$

So

$$R(d, \mu, \varepsilon) \leq \frac{I(X;Y)}{N} \leq \frac{1}{N} H\{(X_u)_{u\in(-M,N+M)\times(-M,M)}\},$$

$$\frac{R(d,\mu,\varepsilon)}{\log(1/\varepsilon)} \leq \frac{2M}{\log(1/\varepsilon)}\cdot\frac{1}{2NM} H\{(X_u)_{u\in(-M,N+M)\times(-M,M)}\}.$$

We first take the limit with respect to N and next the limit with respect to ε. Noting that $M/\log(1/\varepsilon) \to 1/\log\alpha$, we get

$$\overline{\mathrm{rdim}}(\mathcal{X}, \sigma_1, d, \mu) \leq \frac{2h_\mu(\mathcal{X}, \sigma_1, \sigma_2)}{\log\alpha}.$$

3.4. *Step 4: Proof of* $\underline{\mathrm{rdim}}(\mathcal{X}, \sigma_1, d, \mu) \geq 2h_\mu(\mathcal{X}, \sigma_1, \sigma_2)/\log\alpha$. We need the following lemma.

LEMMA 3.7. *Let* $N \geq 1$ *and* B *be a finite set. Let* $X = (X_0, \ldots, X_{N-1})$ *and* $Y = (Y_0, \ldots, Y_{N-1})$ *be random variables taking values in* B^N *(namely, each* X_n *and* Y_n *takes values in* B*) such that for some* $0 < \delta < 1/2$,

$$\mathbb{E}(\textit{the number of } 0 \leq n < N \textit{ with } X_n \neq Y_n) < \delta N.$$

Then

$$I(X;Y) > H(X) - NH(\delta) - \delta N \log|B|,$$

where $H(\delta) = -\delta\log\delta - (1-\delta)\log(1-\delta)$ *as in Step 2.*

https://doi.org/10.1017/etds.2020.47 Published online by Cambridge University Press

Proof. The proof is close to [**LT18**, Lemma 17]. Let $Z_n = 1_{\{X_n \neq Y_n\}}$ and $Z = \{0 \le n < N \mid X_n \neq Y_n\}$. We can identify Z with $(Z_0, \ldots, Z_{N-1})$ and hence

$$\begin{aligned} H(Z) &\le H(Z_0) + \cdots + H(Z_{N-1}) \\ &= H(\mathbb{E}Z_0) + \cdots + H(\mathbb{E}Z_{N-1}) \\ &\le NH\left(\frac{1}{N}\sum_{n=0}^{N-1} \mathbb{E}Z_n\right) \quad \text{(by concavity of } H(\cdot)) \\ &< NH(\delta). \end{aligned}$$

So $H(Z) < NH(\delta)$. We decompose $H(X, Z|Y)$ in two ways:

$$H(X, Z|Y) = H(X|Y) + H(Z|X, Y) = H(Z|Y) + H(X|Y, Z).$$

Here $H(Z|X, Y) = 0$ because Z is determined by X and Y. Hence,

$$H(X|Y) = H(Z|Y) + H(X|Y, Z) < NH(\delta) + H(X|Y, Z).$$

We estimate

$$H(X|Y, Z) = \sum_{E \subset \{0,1,\ldots,N-1\}} \mathbb{P}(Z = E) H(X|Y, Z = E).$$

Given Y and the condition $Z = E$, the possibilities of X are at most $|B|^{|E|}$. Therefore, $H(X|Y, Z = E) \le |E| \log |B|$ and

$$\begin{aligned} H(X|Y, Z) &\le \sum_{E \subset \{0,1,\ldots,N-1\}} |E| \cdot \mathbb{P}(Z = E) \log |B| \\ &= \mathbb{E}|Z| \cdot \log |B| \\ &\le \delta N \log |B|. \end{aligned}$$

As a conclusion, $H(X|Y) < NH(\delta) + \delta N \log |B|$ and $I(X; Y) = H(X) - H(X|Y) > H(X) - NH(\delta) - \delta N \log |B|$. □

Let X be a random variable taking values in $\mathcal{X}$ with $\mathrm{Law}(X) = \mu$ as in Step 3. Let $0 < \varepsilon < \delta < 1/2$ and $N > 0$. Let $Y = (Y_0, \ldots, Y_{N-1})$ be a random variable taking values in $\mathcal{X}^N$ and satisfying

$$\mathbb{E}\left(\frac{1}{N}\sum_{n=0}^{N-1} d(\sigma_1^n X, Y_n)\right) < \varepsilon.$$

We estimate $I(X; Y)$ from below. Take $M \ge 0$ satisfying $\delta\alpha^{-M-1} < \varepsilon \le \delta\alpha^{-M}$. For $0 \le n < N$, we set

$$X_n' = \pi_{\{n\}\times[-M,M]}(X) = (X_{n,m})_{-M\le m\le M}, \quad Y_n' = \pi_{\{0\}\times[-M,M]}(Y_n) = ((Y_n)_{0,m})_{-M\le m\le M}.$$

If $X_n' \neq Y_n'$ for some n, then $d(\sigma_1^n X, Y_n) \ge \alpha^{-M}$. So $\mathbb{E}d(\sigma_1^n X, Y_n) \ge \alpha^{-M}\mathbb{P}(X_n' \neq Y_n')$ and hence

$$\begin{aligned} \mathbb{E}(\text{the number of } 0 \le n < N \text{ with } X_n' \neq Y_n') &= \sum_{n=0}^{N-1} \mathbb{P}(X_n' \neq Y_n') \\ &\le \alpha^M \mathbb{E}\left(\sum_{n=0}^{N-1} d(\sigma_1^n X, Y_n)\right) \\ &< \alpha^M \varepsilon N \le \delta N. \end{aligned}$$

https://doi.org/10.1017/etds.2020.47 Published online by Cambridge University Press

Apply Lemma 3.7 to X'_n and Y'_n with $B = A^{2M+1}$:

$$I(X'_0, \ldots, X'_{N-1}; Y'_0, \ldots, Y'_{N-1}) > H(X'_0, \ldots, X'_{N-1}) \\ - NH(\delta) - \delta N(2M+1)\log|A|.$$

By the data-processing inequality (Lemma 2.3),

$$I(X; Y) \geq I(X'_0, \ldots, X'_{N-1}; Y'_0, \ldots, Y'_{N-1}).$$

Therefore,

$$\frac{I(X;Y)}{N} \geq \frac{H\{(X_u)_{u\in[0,N)\times[-M,M]}\}}{N} - H(\delta) - \delta(2M+1)\log|A|.$$

This holds for any $N > 0$. So

$$R(d,\mu,\varepsilon) \geq \inf_{N>0} \frac{H\{(X_u)_{u\in[0,N)\times[-M,M]}\}}{N} - H(\delta) - \delta(2M+1)\log|A| \\ = \lim_{N\to\infty} \frac{H\{(X_u)_{u\in[0,N)\times[-M,M]}\}}{N} - H(\delta) - \delta(2M+1)\log|A|.$$

We divide this by $\log(1/\varepsilon)$ and take the limit $\varepsilon \to 0$. Noting that $\log(1/\varepsilon) < \log(1/\delta) + (M+1)\log\alpha$ (here δ has been fixed), we get

$$\underline{\mathrm{rdim}}(\mathcal{X}, \sigma_1, d, \mu) \geq \frac{2h_\mu(\mathcal{X}, \sigma_1, \sigma_2)}{\log\alpha} - \frac{2\delta\log|A|}{\log\alpha}.$$

Here we have used

$$h_\mu(\mathcal{X}, \sigma_1, \sigma_2) = \lim_{N,M\to\infty} \frac{H\{(X_u)_{u\in[0,N)\times[-M,M]}\}}{N(2M+1)}.$$

Take the limit $\delta \to 0$. We get $\underline{\mathrm{rdim}}(\mathcal{X}, \sigma_1, d, \mu) \geq 2h_\mu(\mathcal{X}, \sigma_1, \sigma_2)/\log\alpha$.

Acknowledgement. The first proof we gave to Theorem 1.1 contained a gap. Elon Lindenstrauss pointed this out and he also kindly explained to us how to fix the gap. We would like to thank him for the help.

M.S. was partially supported by a Grant-in-Aid for JSPS Research Fellows, JSPS KAKENHI Grant Number 17J03495. M.T. was partially supported by JSPS KAKENHI 18K03275.

REFERENCES

[CT06] T. M. Cover and J. A. Thomas. *Elements of Information Theory*, 2nd edn. Wiley, New York, 2006.

[ECG94] M. Effros, P. A. Chou and G. M. Gray. Variable-rate source coding theorems for stationary nonergodic sources. *IEEE Trans. Inform. Theory* **40** (1994), 1920–1925.

[EW11] M. Einsiedler and T. Ward. *Ergodic Theory with a View Towards Number Theory (Graduate Texts in Mathematics, 259)*. Springer, London, 2011.

[Fur67] H. Furstenberg. Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation. *Math. Systems Theory* **1** (1967), 1–49.

[GLT16] Y. Gutman, E. Lindenstrauss and M. Tsukamoto. Mean dimension of $\mathbb{Z}^k$-actions. *Geom. Funct. Anal.* **26**(3) (2016), 778–817.

[Gra90] R. M. Gray. *Entropy and Information Theory*. Springer, New York, 1990.

[**Gro99**] M. Gromov. Topological invariants of dynamical systems and spaces of holomorphic maps: I. *Math. Phys. Anal. Geom.* **2** (1999), 323–415.

[**Gut15**] Y. Gutman. Mean dimension and Jaworski-type theorems. *Proc. Lond. Math. Soc. (3)* **111**(4) (2015), 831–850.

[**GQT19**] Y. Gutman, Y. Qiao and M. Tsukamoto. Application of signal analysis to the embedding problem of $\mathbb{Z}^k$-actions. *Geom. Funct. Anal.* **29** (2019), 1440–1502.

[**GT20**] Y. Gutman and M. Tsukamoto. Embedding minimal dynamical systems into Hilbert cubes. *Invent. Math.* **221** (2020), 113–166.

[**KD94**] T. Kawabata and A. Dembo. The rate distortion dimension of sets and measures. *IEEE Trans. Inform. Theory* **40**(5) (1994), 1564–1572.

[**LDN79**] A. Leon-Garcia, L. D. Davisson and D. L. Neuhoff. New results on coding of stationary nonergodic sources. *IEEE Trans. Inform. Theory* **25** (1979), 137–144.

[**Lin01**] E. Lindenstrauss. Pointwise theorems for amenable groups. *Invent. Math.* **146** (2001), 259–296.

[**Lin99**] E. Lindenstrauss. Mean dimension, small entropy factors and an embedding theorem. *Publ. Math. Inst. Hautes Études Sci.* **89** (1999), 227–262.

[**LL18**] H. Li and B. Liang. Mean dimension, mean rank and von Neumann–Lück rank. *J. Reine Angew. Math.* **739** (2018), 207–240.

[**LT18**] E. Lindenstrauss and M. Tsukamoto. From rate distortion theory to metric mean dimension: variational principle. *IEEE Trans. Inform. Theory* **64**(5) (2018), 3590–3609.

[**LT19**] E. Lindenstrauss and M. Tsukamoto. Double variational principle for mean dimension. *Geom. Funct. Anal.* **29** (2019), 1048–1109.

[**LW00**] E. Lindenstrauss and B. Weiss. Mean topological dimension. *Israel J. Math.* **115** (2000), 1–24.

[**Ma79**] R. Mañé. Expansive homeomorphisms and topological dimension. *Trans. Amer. Math. Soc.* **252** (1979), 313–319.

[**Mat95**] P. Mattila. *Geometry of Sets and Measures in Euclidean Spaces, Fractals and Rectifiability (Cambridge Studies in Advanced Mathematics, 44)*. Cambridge University Press, Cambridge, 1995.

[**MT19**] T. Meyerovitch and M. Tsukamoto. Expansive multiparameter actions and mean dimension. *Trans. Amer. Math. Soc.* **371** (2019), 7275–7299.

[**OW83**] D. S. Ornstein and B. Weiss. The Shannon–McMillan–Breiman theorem for a class of amenable groups. *Israel J. Math.* **44** (1983), 53–60.

[**Rén59**] A. Rényi. On the dimension and entropy of probability distributions. *Acta Math. Acad. Sci. Hungar.* **10** (1959), 193–215.

[**Rud90**] D. J. Rudolph. *Fundamentals of Measurable Dynamics*. Clarendon Press, Oxford, 1990.

[**Sh48**] C. E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.* **27** (1948), 379–423, 623–656.

[**Sh59**] C. E. Shannon. Coding theorems for a discrete source with a fidelity criterion. *IRE National Convention Record* **7**(4) (1959), 142–163.

[**Tsu18**] M. Tsukamoto. Mean dimension of the dynamical system of Brody curves. *Invent. Math.* **211** (2018), 935–968.

[**Youn82**] L.-S. Young. Dimension, entropy and Lyapunov exponents. *Ergod. Th. & Dynam. Sys.* **2** (1982), 109–124.