

# ESCAPE OF ENTROPY FOR COUNTABLE MARKOV SHIFTS

GODOFREDO IOMMI, MIKE TODD, AND ANIBAL VELOZO

ABSTRACT. In this paper we study ergodic theory of countable Markov shifts. These are dynamical systems defined over non-compact spaces. Our main result relates the escape of mass, the measure theoretic entropy, and the entropy at infinity of the system. This relation has several consequences. For example we obtain that the entropy map is upper semi-continuous and that the ergodic measures form an entropy dense subset. Our results also provide new proofs of results describing the existence and stability of the measure of maximal entropy. We relate the entropy at infinity with the Hausdorff dimension of the set of recurrent points that escape on average. Of independent interest, we prove a version of Katok's entropy formula in this non-compact setting.

## 1. INTRODUCTION

Many problems in ergodic theory and dynamical systems involve properties of limits of sequences of invariant probability measures. If the phase space is compact then the space of invariant probability measures is also compact in the weak\* topology, which is partly a consequence of convergence in this topology preserving mass. However, when the phase space is non-compact, the space of invariant probability measures might also be non-compact, and thus mass, as well as other quantities of interest, may escape in the limit. In this paper, we are principally interested in how the entropy of sequences of measures behaves in this setting.

More specifically, we consider countable Markov shifts (CMS)  $(\Sigma, \sigma)$ , which in general are not even locally compact. We discuss the difficulties with the various classical topologies in this context in the next section, where we also give details of the space of invariant sub-probability measures endowed with the so-called cylinder topology, introduced in [IV]. This topology generalises the vague topology to a non-locally compact setting (see Section 2.2.2). If  $(\mu_n)_n$  is a sequence of  $\sigma$ -invariant probability measures that converges in the cylinder topology to the measure  $\mu$  then the total mass  $|\mu| := \mu(\Sigma) \in [0, 1]$ . In particular, this topology captures the escape of mass. Moreover,  $\mu$  is an invariant measure and the normalisation  $\mu/|\mu|$  is an invariant probability measure (whenever  $\mu$  is not the zero measure). Denote by  $h_\nu(\sigma)$  the entropy of the invariant probability measure  $\nu$  (see Section 2.3 for details). Our first main result answers one of the classical questions about sequences of measures: how does entropy change in the limit?

**Theorem 1.1.** *Let  $(\Sigma, \sigma)$  be a transitive CMS with finite topological entropy. Let  $(\mu_n)_n$  be a sequence of  $\sigma$ -invariant probability measures converging on cylinders to*

---

*Date:* August 29, 2019.

GI was partially supported by CONICYT PIA ACT172001 and by Proyecto Fondecyt 1190194. MT and AV thank UC for their hospitality. AV thanks Paul Apisa for interesting discussions about the set of points that escape on average.

$\mu$ . Then

$$\limsup_{n \rightarrow \infty} h_{\mu_n}(\sigma) \leq |\mu| h_{\mu/|\mu|}(\sigma) + (1 - |\mu|) \delta_\infty. \quad (1.1)$$

If the sequence converges on cylinders to the zero measure then the right hand side is understood as  $\delta_\infty$ .

The number  $\delta_\infty$  is the *topological entropy at infinity* of the system. Since it plays a crucial role in this article, we define it here, leaving details of the background to this to Section 2.6. The idea is to measure how complicated the dynamics is near infinity. Of course, such a notion only makes sense for dynamical systems defined on non-compact phase spaces. As in the classical entropy theory, we will study two ways of measuring the complexity of the system near infinity. One topological in nature and the other measure theoretic.

**Definition 1.2.** Let  $(\Sigma, \sigma)$  be a CMS. Let  $M, q \in \mathbb{N}$ . For  $n \in \mathbb{N}$  let  $z_n(M, q)$  be the number of cylinders of the form  $[x_0, \dots, x_{n+1}]$ , where  $x_0 \leq q$ ,  $x_{n+1} \leq q$ , and

$$\#\{i \in \{0, 1, \dots, n+1\} : x_i \leq q\} \leq \frac{n+2}{M}.$$

Define

$$\delta_\infty(M, q) := \limsup_{n \rightarrow \infty} \frac{1}{n} \log z_n(M, q),$$

and

$$\delta_\infty(q) := \liminf_{M \rightarrow \infty} \delta_\infty(M, q).$$

The *topological entropy at infinity* of  $(\Sigma, \sigma)$  is defined by  $\delta_\infty := \liminf_{q \rightarrow \infty} \delta_\infty(q)$ .

The measure theoretic counterpart is given by:

**Definition 1.3.** Let  $(\Sigma, \sigma)$  be a finite entropy CMS. The *measure theoretic entropy at infinity* of  $(\Sigma, \sigma)$  is defined by

$$h_\infty := \sup_{(\mu_n)_{n \rightarrow 0}} \limsup_{n \rightarrow \infty} h_{\mu_n}(\sigma), \quad (1.2)$$

where  $(\mu_n)_n \rightarrow 0$  means that the sequence  $(\mu_n)_n$  converges on cylinders to the zero measure.

Other authors have considered related concepts. Most notably, Buzzi [B1, Definition 1.13] proposed a notion of entropy at infinity for CMS. His definition is given in terms of the graph  $G$  which defines the CMS  $(\Sigma, \sigma)$ :

$$b_\infty := \inf_F \inf_{\lambda > 0} \sup \{h_\mu(\sigma) : \mu([F]) < \lambda\},$$

where  $F$  ranges over the finite sub-graphs of  $G$  and  $[F] := \{x \in \Sigma : x_0 \in \mathcal{A}_F\}$ , where  $\mathcal{A}_F$  denotes the symbols appearing as vertex of  $F$ . It turns out that Buzzi's notion coincides with ours. Indeed, our next result states that all these three notions coincide.

**Theorem 1.4.** *Let  $(\Sigma, \sigma)$  be a CMS of finite topological entropy. Then*

$$\delta_\infty = h_\infty = b_\infty.$$

The equality  $\delta_\infty = h_\infty$  can be understood as a variational principle at infinity.

Einsiedler, Lindenstrauss, Michel and Venkatesh [ELMV, Lemma 4.4] were the first to obtain an inequality similar to (1.1). It appeared in their ergodic theoretic proof of Duke's theorem on equidistribution of closed geodesics on the modular

surface. After that, similar results in the context of homogeneous dynamics were obtained in [EK, Theorem 1.2] and [EKP, Theorem A]. For different classes of geodesic flows defined on non-compact manifolds of negative sectional curvature related results were obtained in [IRV, Theorem 1.2] and [RV, Theorem 1.1]. In this context the most general result was obtained in [Vel1, Theorems 1.4 and 1.6] where an inequality like (1.1) was proved for the geodesic flow defined on an arbitrary complete Riemannian manifolds with pinched negative sectional curvature. The manifolds studied are locally compact, thus the topology considered in the space of invariant measures is the vague topology. A more interesting and subtle point is the quantity playing the role of the entropy at infinity. Due to the geometric nature of the examples studied, the entropy at infinity is related to the critical exponent of the Poincaré series associated to the non-compact parts of the space (in the geometrically finite case this reduces to the critical exponent of the parabolic subgroups of the fundamental group). Let us mention that the topological entropy at infinity of the geodesic flow was also studied by Schapira and Tapie [ST] in their work about the rate of change of the topological entropy under perturbations of the metric.

A major difference with previous works is that in the context of CMS the behaviour of the orbits approaching infinity can be very complicated and that we do not assume the phase space to be locally compact. These are major difficulties that have to be overcome making the analysis more technical. As a general principle we follow the method employed in [Vel1] with appropriate modifications. Loosely speaking the entropy at infinity of the geodesic flow counts geodesics that start and end at a given base point, but do not return near this point for intermediate times. In our setup the entropy at infinity counts orbits that might return near a base point many times, but the number of returns become negligible on average, which can occur due to the lack of local compactness.

There are several interesting consequences of Theorem 1.1, some of them are discussed in Section 8. For example, in Theorem 8.1 it is proved that the entropy map is upper semi-continuous for every transitive finite entropy CMS. The continuity properties of the entropy map have been studied for a long time. Major results in the area are that for expansive systems defined on compact metric spaces the entropy map is upper semi-continuous [Wa, Theorem 8.2]. Another fundamental result is that if  $f$  is a  $C^\infty$  diffeomorphism defined on a smooth compact manifold then again the entropy map is upper semi-continuous [N, Theorem 4.1]. As explained in Remark 8.2, for infinite entropy CMS the entropy map is not upper semi-continuous. In a recent article we proved [ITV, Corollary 1.2] that if  $(\Sigma, \sigma)$  is a finite entropy transitive CMS then the entropy map is upper semi-continuous when restricted to ergodic measures. A complete solution to the problem can be obtained as a consequence of Theorem 1.1. In Section 8 we also prove that the set of ergodic measures is ‘entropy dense’ in the space of invariant probability measures. This result not only provides a fine description of the structure of the space of invariant probability measures but also provides an important tool to study Large Deviations in this setting.

There is a classification of transitive CMS in terms of their recurrence properties: they can be transient, null recurrent or positive recurrent (see Definition 2.10 for  $\varphi = 0$ ). Positive recurrent CMS are precisely those with a measure of maximal entropy. A particularly important role is played by strongly positive recurrent CMS

(SPR); which are a sub-class of positive recurrent Markov shifts. The dynamical properties of this class of systems are similar to that of sub-shifts of finite type. Buzzi gave a characterisation of SPR shifts using  $b_\infty$  in [B1, Proposition 6.1], based on the work of Gurevich-Zargaryan, Gurevich-Savchenko and Ruette. We note in Proposition 2.20 that we can now restate this result as saying that  $(\Sigma, \sigma)$  is SPR if and only if  $\delta_\infty < h_{top}(\sigma)$ , where  $h_{top}(\sigma)$  is the Gurevich entropy of  $(\Sigma, \sigma)$  (for precise definitions see Section 2.4). In Section 8.5 we use Theorem 1.1 to obtain stability properties of the measure of maximal entropy for SPR CMS (recovering results from [GS]). Similar arguments are used to prove the existence of equilibrium states for potentials in  $C_0(\Sigma)$ , the space of test functions for the cylinder topology (see Section 8.6). To the author's knowledge, this is the first result on the existence of equilibrium states for CMS that goes beyond regular potentials (e.g. with summable variations or the Walters property). Finally, in Theorem 8.16 we prove that for SPR systems it is possible to bound the amount of mass that escapes the system in terms of the entropy of the measures. Sequences of measures with large entropy can not lose much mass.

The entropy at infinity has yet another important appearance in dynamics: it is related to the Hausdorff dimension of the set of points that escape on average (see [AAEKMU, EKP, KKLM, KP]). These are points for which the frequency of visits to every cylinder equals to zero. In particular, no invariant measure is supported on that set. This notion has been studied recently in contexts of homogeneous and Teichmüller dynamics. The motivation comes from work of Dani [D] in the mid 1980s who proved that singular matrices are in one-to-one correspondence with certain divergent orbits of one parameter diagonal groups on the space of lattices. In Theorem 8.9 we prove that the Hausdorff dimension of the set of recurrent points that escape on average is bounded above by  $\delta_\infty/\log 2$ , where the factor  $\log 2$  comes from the metric in the symbolic space.

While our interest in this paper lies in the realm of Markov shifts, to provide context we mention some applications of this theory. Symbolic methods have been used to describe dynamical properties of a variety of systems since the 1898 work of Hadamard on closed geodesics on surfaces of negative curvature, at the latest. Compact Markov shifts have been used to study uniformly hyperbolic dynamical systems defined on compact spaces, see for example the work of Bowen in [Bo3]. Many deep results can be obtained with this coding. Mostly after the work of Sarig [S3], countable Markov partitions have been constructed for a wide range of dynamical systems. This gives a semiconjugacy between a relevant part of the dynamics, albeit not all of it, and a CMS. Examples of systems for which Markov partitions have been constructed include positive entropy diffeomorphisms defined on compact manifolds [B2, Ov, S3] and Sinai and Bunimovich billiards [LM]. Remarkable results have been proved making use of these codings, for example in [BCS, Main Theorem] it is shown that a positive entropy  $C^\infty$  diffeomorphism of a closed surface admits at most finitely many ergodic measures of maximal entropy. Results in this paper apply to all the symbolic codings mentioned above. However, due to topologies possibly not being preserved by the coding, it is not clear that the results pass to the original systems.

In 1980 Katok [Ka, Theorem 1.1] established a formula for the entropy of an invariant probability measure in analogy to the definition of topological entropy of a dynamical system [Bo2, Di]. This formula is now known as *Katok's entropy*

*formula*. An important assumption in [Ka, Theorem 1.1] is the compactness of the phase space. In Section 3 we prove that Katok's entropy formula holds in the non-compact setting of CMS. We require this formula in the proof of Theorem 1.1, but it is also of independent interest.

## 2. PRELIMINARIES

**2.1. Basic definitions for CMS.** Let  $\mathcal{S}$  be an alphabet of countably many symbols and  $M$  an  $\mathcal{S} \times \mathcal{S}$  matrix with entries 0 or 1. The symbolic space associated to  $M$  is defined by

$$\Sigma := \{(x_0, x_1, \dots) \in \mathcal{S}^{\mathbb{N}_0} : M(x_i, x_{i+1}) = 1 \text{ for every } i \in \mathbb{N}_0\},$$

where  $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ . We endow  $\mathcal{S}$  with the discrete topology and  $\mathcal{S}^{\mathbb{N}_0}$  with the product topology. On  $\Sigma$  we consider the induced topology given by the natural inclusion  $\Sigma \subset \mathcal{S}^{\mathbb{N}_0}$ . We stress that, in general, this is a non-compact space. The space  $\Sigma$  is locally compact if and only if for every  $i \in \mathcal{S}$  we have  $\sum_{j \in \mathcal{S}} M(i, j) < \infty$  (see [Ki, Observation 7.2.3]).

The *shift map*  $\sigma : \Sigma \rightarrow \Sigma$  is defined by  $(\sigma(x))_i = x_{i+1}$ , where  $x = (x_0, x_1, \dots) \in \Sigma$ . Note that  $\sigma$  is a continuous map. The pair  $(\Sigma, \sigma)$  is called a one sided *countable Markov shift* (CMS). The matrix  $M$  can be identified with a directed graph  $G$  with no multiple edges (but allowing edges connecting a vertex to itself).

An *admissible word* of length  $N$  is a string  $\mathbf{w} = a_0 a_1 \dots a_{N-1}$  of letters in  $\mathcal{S}$  such that  $M(a_i, a_{i+1}) = 1$ , for every  $i \in \{0, \dots, N-2\}$ . We use bold letters to denote admissible words. The length of an admissible word  $\mathbf{w}$  is  $\ell(\mathbf{w})$ .

A *cylinder* of length  $N$  is the set

$$[a_0, \dots, a_{N-1}] := \{x = (x_0, x_1, \dots) \in \Sigma : x_i = a_i \text{ for } 0 \leq i \leq N-1\}.$$

If  $a_0 \dots a_{N-1}$  is an admissible word then  $[a_0, \dots, a_{N-1}] \neq \emptyset$ . We use the notation  $C_n(x)$  to denote the cylinder of length  $n$  containing  $x$ . Since a cylinder can be identified with an admissible word, we also denote the length of a cylinder  $C$  by  $\ell(C)$ . Note that the topology generated by the cylinder sets coincides with that induced by the product topology.

The space  $\Sigma$  is metrisable. Indeed, let  $d : \Sigma \times \Sigma \rightarrow \mathbb{R}$  be the function defined by

$$d(x, y) := \begin{cases} 1 & \text{if } x_0 \neq y_0; \\ 2^{-k} & \text{if } x_i = y_i \text{ for } i \in \{0, \dots, k-1\} \text{ and } x_k \neq y_k; \\ 0 & \text{if } x = y. \end{cases} \quad (2.1)$$

The function  $d$  is a metric and it generates the same topology as that of the cylinders sets. Moreover, the ball  $B(x, 2^{-N})$  is precisely  $C_N(x)$ . Given  $\varphi : \Sigma \rightarrow \mathbb{R}$ , we define

$$\text{var}_n(\varphi) := \sup \{|\varphi(x) - \varphi(y)| : \forall x, y \in \Sigma \text{ such that } d(x, y) \leq 2^{-n}\}.$$

A function  $\varphi : \Sigma \rightarrow \mathbb{R}$  is said to have *summable variations* if  $\sum_{n \geq 2} \text{var}_n(\varphi) < \infty$ . A function  $\varphi$  is called *weakly Hölder* if there exist  $\theta \in (0, 1)$  and a positive constant  $O \in \mathbb{R}$  such that  $\text{var}_n(\varphi) \leq O\theta^n$ , for every  $n \geq 2$ . A weakly Hölder continuous function is Hölder if and only if it is bounded. The  $C^0$ -norm of  $\varphi$  is  $\|\varphi\|_0 := \sup_{x \in \Sigma} |\varphi(x)|$ . We denote by

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

the *Birkhoff sum* of  $\varphi$  at the point  $x$ .

We say that  $(\Sigma, \sigma)$  is *topologically transitive* if its associated directed graph  $G$  is connected. We say that  $(\Sigma, \sigma)$  is *topologically mixing* if for each pair  $a, b \in \mathcal{S}$ , there exists a number  $N(a, b)$  such that for every  $n \geq N(a, b)$  there is an admissible word of length  $n$  connecting  $a$  and  $b$ . There is a particular class of CMS that will be of interest to us,

**Definition 2.1.** A CMS  $(\Sigma, \sigma)$  is said to satisfy the  $\mathcal{F}$ -*property* if for every element of the alphabet  $a$  and natural number  $n$ , there are only finitely many admissible words of length  $n$  starting and ending at  $a$ .

*Remark 2.2.* A CMS  $(\Sigma, \sigma)$  satisfies the  $\mathcal{F}$ -*property* if and only if there are only finitely many periodic orbits of length  $n$  intersecting  $[a]$ . Note that every locally compact CMS satisfies the  $\mathcal{F}$ -*property*.

*Remark 2.3.* Equivalent definitions and properties as those discussed in this section can be given for *two sided CMS*. In this case the acting group is  $\mathbb{Z}$ . It turns out that, in general, thermodynamic formalism for the two sided CMS can be reduced to the one sided case (see [S4, Section 2.3]).

**2.2. Topologies in the space of invariant measures.** The space of invariant measures can be endowed with different topologies, some of which can account for the escape of mass phenomenon whereas others can not. In this section we not only fix notation for later use, but we also recall definitions and properties of several topologies in the space of measures. First note that in this article a measure is always a countably additive non-negative Borel measure defined in the symbolic space  $\Sigma$ . The mass of a measure  $\mu$  is defined as  $|\mu| := \mu(\Sigma)$ .

Denote by  $\mathcal{M}(\Sigma, \sigma)$  the space of  $\sigma$ -invariant probability measures on  $\Sigma$  and by  $\mathcal{M}_{\leq 1}(\Sigma, \sigma)$  the space of  $\sigma$ -invariant measures on  $\Sigma$  with mass in  $[0, 1]$ . In other words,  $\mathcal{M}_{\leq 1}(\Sigma, \sigma)$  is the space of  $\sigma$ -invariant sub-probability measures on  $\Sigma$ . Note that  $\mathcal{M}(\Sigma, \sigma) \subset \mathcal{M}_{\leq 1}(\Sigma, \sigma)$ . The set of ergodic probability measures is denoted by  $\mathcal{E}(\Sigma, \sigma)$ .

**2.2.1. The weak\* topology.** Denote by  $C_b(\Sigma)$  the space of real valued bounded continuous function on  $\Sigma$ . A sequence of measures  $(\mu_n)_n$  in  $\mathcal{M}(\Sigma, \sigma)$  converges to a measure  $\mu$  in the weak\* topology if for every  $f \in C_b(\Sigma)$  we have

$$\lim_{n \rightarrow \infty} \int f d\mu_n = \int f d\mu.$$

Note that since the constant function equal to one belongs to  $C_b(\Sigma)$  the measure  $\mu$  is also a probability measure. A basis for this topology is given by the collection of sets of the form

$$V(f_1, \dots, f_k, \mu, \epsilon) := \left\{ \nu \in \mathcal{M}(\Sigma, \sigma) : \left| \int f_i d\nu - \int f_i d\mu \right| < \epsilon, \text{ for } i \in \{1, \dots, k\} \right\}, \quad (2.2)$$

where  $\mu \in \mathcal{M}(\Sigma, \sigma)$ ,  $(f_i)_i$  are elements from  $C_b(\Sigma)$  and  $\epsilon > 0$ . Note that in this notion of convergence we can replace the set of test functions (bounded and continuous) by the space of bounded uniformly continuous functions (see [B, 8.3.1 Remark]). Convergence with respect to the weak\* topology can be characterised as follows, see [Bi, Theorem 2.1].

**Proposition 2.4** (Portmanteau Theorem). *Let  $(\mu_n)_n, \mu$  be probability measures on  $\Sigma$ . The following statements are equivalent.*

- (a) The sequence  $(\mu_n)_n$  converges to  $\mu$  in the weak\* topology.
- (b) If  $O \subset \Sigma$  is an open set, then  $\mu(O) \leq \liminf_{n \rightarrow \infty} \mu_n(O)$ .
- (c) If  $C \subset \Sigma$  is a closed set, then  $\mu(C) \geq \limsup_{n \rightarrow \infty} \mu_n(C)$ .
- (d) If  $A \subset \Sigma$  has  $\mu(\partial A) = 0$ , where  $\partial A$  is the boundary of  $A$ , then  $\lim_{n \rightarrow \infty} \mu_n(A) = \mu(A)$ .

Note that the space  $\mathcal{M}(\Sigma, \sigma)$  is closed in the weak\* topology ([Wa, Theorem 6.10]). If  $\Sigma$  is compact then so is  $\mathcal{M}(\Sigma, \sigma)$  with respect to the weak\* topology (see [Wa, Theorem 6.10]). If  $\Sigma$  is not compact then, in general (e.g., whenever the  $\mathcal{F}$ -property holds),  $\mathcal{M}(\Sigma, \sigma)$  is not compact with respect to the weak\* topology. Finally, the space  $\mathcal{M}(\Sigma, \sigma)$  is a convex set whose extreme points are ergodic measures (see [Wa, Theorem 6.10]).

**2.2.2. The topology of convergence on cylinders.** In this section we recall the definition and properties of the topology of convergence on cylinders. This topology was introduced and studied in [IV] as a way to compactify  $\mathcal{M}(\Sigma, \sigma)$  under suitable assumptions on  $\Sigma$ .

Let  $(C^n)_n$  be an enumeration of the cylinders of  $\Sigma$ . Given  $\mu, \nu \in \mathcal{M}_{\leq 1}(\Sigma, \sigma)$  we define

$$\rho(\mu, \nu) = \sum_{n=1}^{\infty} \frac{1}{2^n} |\mu(C^n) - \nu(C^n)|.$$

It follows from the outer regularity of Borel measures on metric spaces that  $\rho(\mu, \nu) = 0$ , if and only if  $\mu = \nu$ . Moreover, the function  $\rho$  defines a metric on  $\mathcal{M}_{\leq 1}(\Sigma, \sigma)$ . The topology induced by this metric is called the *topology of convergence on cylinders*. We say that a sequence  $(\mu_n)_n$  in  $\mathcal{M}_{\leq 1}(\Sigma, \sigma)$  *converges on cylinders* to  $\mu$  if

$$\lim_{n \rightarrow \infty} \mu_n(C) = \mu(C),$$

for every cylinder  $C \subset \Sigma$ . Of course,  $(\mu_n)_n$  converges on cylinders to  $\mu$  if and only if  $(\mu_n)_n$  converges to  $\mu$  in the topology of convergence on cylinders. In the next lemma we see that in the case that mass does not escape then weak\* and convergence on cylinders coincide.

**Lemma 2.5.** [IV, Lemma 3.17] *Let  $(\Sigma, \sigma)$  be a CMS,  $\mu$  and  $(\mu_n)_n$  be invariant probability measures on  $\Sigma$ . The following assertions are equivalent.*

- (a) The sequence  $(\mu_n)_n$  converges in the weak\* topology to  $\mu$ .
- (b) The sequence  $(\mu_n)_n$  converges on cylinders to  $\mu$ .

Let  $\Sigma$  be a locally compact space and  $(\mu_n)_n, \mu$  in  $\mathcal{M}_{\leq 1}(\Sigma, \sigma)$ . The sequence  $(\mu_n)_n$  converges to  $\mu$  in the *vague topology* if  $\lim_{n \rightarrow \infty} \int f d\mu_n = \int f d\mu$ , for every function  $f$  continuous and of compact support (note that the set of test functions can be replaced by the set of continuous functions vanishing at infinity). If  $(\Sigma, \sigma)$  is locally compact then the topology of convergence on cylinders coincides with the vague topology (see [IV, Lemma 3.18]). It is important to note that if  $\Sigma$  is transitive and not locally compact then the space of continuous functions of compact support is empty, so the vague topology is of no use and the topology of convergence on cylinders is a suitable generalisation (see [IV, Remark 3.13]).

If  $C$  is a cylinder of length  $m$ , denote by

$$C(\geq n) := \left\{ x \in C : \sigma^m(x) \in \bigcup_{k \geq n} [k] \right\}.$$

For a non-empty set  $A \subset \Sigma$  we define

$$\text{var}^A(f) := \sup \{|f(x) - f(y)| : x, y \in A\}.$$

We declare  $\text{var}^A(f) = 0$  if  $A$  is the empty set.

**Definition 2.6.** We say that  $f$  belongs to  $C_0(\Sigma)$  if the following three conditions hold:

- (a)  $f$  is uniformly continuous.
- (b)  $\lim_{n \rightarrow \infty} \sup_{x \in [n]} |f(x)| = 0$ .
- (c)  $\lim_{n \rightarrow \infty} \text{var}^{C(\geq n)}(f) = 0$ , for every cylinder  $C \subset \Sigma$ .

In this case we say that  $f$  *vanishes at infinity*.

The set  $C_0(\Sigma)$  is the space of test functions for the cylinder topology (see [IV, Lemma 3.19]). In other words,  $(\mu_n)_n$  is a sequence in  $\mathcal{M}_{\leq 1}(\Sigma, \sigma)$  that converges in the cylinder topology to  $\mu \in \mathcal{M}_{\leq 1}(\Sigma, \sigma)$  if and only if  $\lim_{n \rightarrow \infty} \int f d\mu_n = \int f d\mu$ , for every  $f \in C_0(\Sigma)$ . Since cylinder topology generalises the vague topology for non-locally compact CMS, the space  $C_0(\Sigma)$  is the natural substitute to the set of continuous functions that vanish at infinity.

The following result was proved in [IV, Theorem 1.2], and is an important ingredient for many of our applications.

**Theorem 2.7.** *Let  $(\Sigma, \sigma)$  be a transitive CMS with the  $\mathcal{F}$ -property. Then  $\mathcal{M}_{\leq 1}(\Sigma, \sigma)$  is a compact metrisable space with respect to the cylinder topology. Moreover,  $\mathcal{M}_{\leq 1}(\Sigma, \sigma)$  is affine homeomorphic to the Poulsen simplex.*

We remark that, as shown in [IV, Proposition 4.19], Theorem 2.7 is sharp in a strong sense: if  $(\Sigma, \sigma)$  does not have the  $\mathcal{F}$ -property, then  $\mathcal{M}_{\leq 1}(\Sigma, \sigma)$  is not compact. More precisely, there exists a sequence of periodic measures that converges on cylinders to a finitely additive measure which is not countably additive.

**2.3. Entropy of a measure.** In this section we recall the definition of the entropy of an invariant measure  $\mu \in \mathcal{M}(\Sigma, \sigma)$  (see [Wa, Chapter 4] for more details). We take the opportunity to fix the notation that will be used in what follows. A partition  $\beta$  of a probability space  $(\Sigma, \mu)$  is a countable (finite or infinite) collection of pairwise disjoint subsets of  $\Sigma$  whose union has full measure. The *entropy* of the partition  $\beta$  is defined by

$$H_\mu(\beta) := - \sum_{P \in \beta} \mu(P) \log \mu(P),$$

where  $0 \log 0 := 0$ . It is possible that  $H_\mu(\beta) = \infty$ . Given two partitions  $\beta$  and  $\beta'$  of  $\Sigma$  we define the new partition

$$\beta \vee \beta' := \{P \cap Q : P \in \beta, Q \in \beta'\}.$$

Let  $\beta$  be a partition of  $\Sigma$ . We define the partition  $\sigma^{-1}\beta := \{\sigma^{-1}P : P \in \beta\}$  and for  $n \in \mathbb{N}$  we set  $\beta^n := \bigvee_{i=0}^{n-1} \sigma^{-i}\beta$ . Since the measure  $\mu$  is  $\sigma$ -invariant, the sequence  $H_\mu(\beta^n)$  is sub-additive. The *entropy of  $\mu$  with respect to  $\beta$*  is defined by

$$h_\mu(\beta) := \lim_{n \rightarrow \infty} \frac{1}{n} H_\mu(\beta^n).$$

Finally, the *entropy of  $\mu$*  is defined by

$$h_\mu(\sigma) := \sup \{h_\mu(\beta) : \beta \text{ a partition with } H_\mu(\beta) < \infty\}.$$

*Remark 2.8.* Krengel [Kr, Remark p.166] observes that since the entropy of a finite invariant measure  $\mu$  is usually defined as the entropy of the normalised measure  $\mu/|\mu|$ , the linearity (in the standard sense) of the entropy map is destroyed. Following Krengel's line of thought, the number  $|\mu|h_{\mu/|\mu|}(\sigma)$  appearing in Theorem 1.1 can be understood, as the entropy of the finite measure  $\mu$  (see also [Wa, Theorem 8.1] for example).

**2.4. Thermodynamic formalism for CMS.** Throughout this section we assume that  $(\Sigma, \sigma)$  is topologically transitive and that every function  $\varphi : \Sigma \rightarrow \mathbb{R}$  has summable variations. Let  $A \subset \Sigma$  and  $1_A(x)$  be the characteristic function of the set  $A$ . In this setting we define,

$$Z_n(\varphi, a) := \sum_{\sigma^n x = x} e^{S_n \varphi(x)} 1_{[a]}(x),$$

where  $a \in \mathcal{S}$ . The *Gurevich pressure* of  $\varphi$  is defined by

$$P_G(\varphi) := \limsup_{n \rightarrow \infty} \frac{1}{n} \log Z_n(\varphi, a).$$

This definition was introduced by Sarig [S1], based on the work of Gurevich [Gu2]. We remark that the right hand side in the definition of  $P_G(\varphi)$  is independent of  $a \in \mathcal{S}$ , and that if  $(\Sigma, \sigma)$  is topologically mixing, then the limsup can be replaced by a limit (see [S1, Theorem 1] and [S4, Theorem 4.3]). This definition of pressure satisfies the variational principle (see [S1, Theorem 3] and [IJT, Theorem 2.10]) and can be approximated by the pressure of compact invariant subsets [S1, Theorem 2 and Corollary 1]. Indeed,

$$\begin{aligned} P_G(\varphi) &= \sup \{P_{\text{top}}(\varphi|K) : K \subset \Sigma \text{ compact and } \sigma^{-1}K = K\} \\ &= \sup_{\mu \in \mathcal{M}(\Sigma, \sigma)} \left\{ h_\mu(\sigma) + \int \varphi d\mu : \int \varphi d\mu > -\infty \right\}, \end{aligned}$$

where  $P_{\text{top}}(\cdot)$  is the classical pressure on compact spaces [Wa, Chapter 9]. A measure  $\mu \in \mathcal{M}(\Sigma, \sigma)$  is an *equilibrium state* for  $\varphi$  if  $\int \varphi d\mu > -\infty$  and

$$P_G(\varphi) = h_\mu(\sigma) + \int \varphi d\mu.$$

The following function will be of importance in this article.

**Definition 2.9.** Let  $A \subset \Sigma$ . Denote by  $R_A(x) := 1_A(x) \inf\{n \geq 1 : \sigma^n x \in A\}$  the first return time map to the set  $A$ . In the particular case in which the set  $A$  is a cylinder  $[a]$  we denote  $R_{[a]}(x) := R_a(x)$ .

Sarig [S1, Section 4.2] introduced the following:

$$Z_n^*(\varphi, a) := \sum_{\sigma^n(x)=x} e^{S_n \varphi(x)} 1_{[R_a=n]}(x),$$

where  $[R_a = n] := \{x \in \Sigma : R_a(x) = n\}$ . Extending notions of Markov chains, Sarig [S1] classified potentials according to its recurrence properties.

**Definition 2.10** (Classification of potentials). Let  $(\Sigma, \sigma)$  be a topologically transitive CMS and  $\varphi$  a summable variation potential with finite Gurevich pressure. Define  $\lambda = \exp(P_G(\varphi))$  and fix  $a \in \mathcal{S}$ .

- (a) If  $\sum_{n \geq 1} \lambda^{-n} Z_n(\varphi, a)$  diverges we say that  $\varphi$  is *recurrent*.
- (b) If  $\sum_{n \geq 1} \lambda^{-n} Z_n(\varphi, a)$  converges we say that  $\varphi$  is *transient*.

- (c) If  $\varphi$  is recurrent and  $\sum_{n \geq 1} n\lambda^{-n} Z_n^*(\varphi, a)$  converges we say that  $\varphi$  is *positive recurrent*.
- (d) If  $\varphi$  is recurrent but  $\sum_{n \geq 1} n\lambda^{-n} Z_n^*(\varphi, a)$  diverges we say that  $\varphi$  is *null recurrent*.

Topological transitivity implies that above definitions do not depend on the choice of the symbol  $a$ .

*Remark 2.11.* The classification in Definition 2.10 is invariant under the addition of coboundaries and constants. That is, if  $\psi : \Sigma \rightarrow \mathbb{R}$  is of summable variations and  $C \in \mathbb{R}$  we have that: the potential  $\varphi$  is recurrent (resp. transient) if and only if the potential  $\varphi + \psi - \psi \circ \sigma + C$  is recurrent (resp. transient). Moreover, the potential  $\varphi$  is positive recurrent (resp. null recurrent) if and only if the potential  $\varphi + \psi - \psi \circ \sigma + C$  is positive recurrent (resp. null recurrent).

The following result describes existence and uniqueness of equilibrium states. Parts (a) and (b) follow from Theorems 1.1 and Theorem 1.2 of [BS], respectively.

**Theorem 2.12.** *Let  $(\Sigma, \sigma)$  be a topologically transitive CMS and  $\varphi$  a summable variation potential with finite Gurevich pressure. Then*

- (a) *There exists at most one equilibrium state for  $\varphi$ .*
- (b) *If  $\varphi$  has an equilibrium state then  $\varphi$  is positive recurrent.*

In this article the potential  $\varphi = 0$  will play a particularly important role. The *topological entropy* of  $(\Sigma, \sigma)$ , that we denote by  $h_{top}(\sigma)$ , is defined as the Gurevich pressure of the potential  $\varphi = 0$ , that is

$$h_{top}(\sigma) := P_G(0).$$

We say that  $(\Sigma, \sigma)$  is recurrent, transient, null recurrent or positive recurrent according to the corresponding properties of  $\varphi = 0$ . If  $(\Sigma, \sigma)$  is positive recurrent, then Theorem 2.12 implies that  $(\Sigma, \sigma)$  admits a unique measure of maximal entropy. This was first proved by Gurevich [Gu1].

*Remark 2.13.* Note that every finite entropy, transitive CMS satisfies the  $\mathcal{F}$ -property (see Definition 2.1).

**2.5. Strongly positive recurrent CMS.** Properties of CMS may be significantly different from those of sub-shifts of finite type defined on finite alphabets. In this section we describe a special class of CMS with properties analogous to those of compact sub-shifts. This study is based on work of Vere-Jones [Ve1, Ve2] developed during the 1960s, where he first defined an equivalent class in the setting of stochastic matrices. Several people have contributed to the understanding of this class, for example, Salama [Sal], Gurevich and Savchenko [GS], Sarig [S2], Ruelle [R1], Boyle, Buzzi and Gómez [BBG] and Cyr and Sarig [CS]. In these works the following quantities, or related ones, have been defined and studied.

**Definition 2.14.** Let  $(\Sigma, \sigma)$  be topologically transitive CMS and  $a \in \mathcal{S}$ . Let

$$\Delta_\infty([a]) := \limsup_{n \rightarrow \infty} \frac{1}{n} \log Z_n^*(0, a),$$

and

$$\Delta_\infty := \inf_{a \in \mathcal{S}} \Delta_\infty([a]).$$

*Remark 2.15.* The number  $\Delta_\infty([a])$  can depend on the symbol  $a$ , see [R1, Remark 2,1].

**Definition 2.16** (Strongly positive recurrent CMS). Let  $(\Sigma, \sigma)$  be a topologically transitive CMS with finite topological entropy. We say that  $(\Sigma, \sigma)$  is *strongly positive recurrent* (SPR) if  $\Delta_\infty < h_{top}(\sigma)$ .

*Remark 2.17.* A strongly positive recurrent CMS is positive recurrent. In particular it admits a unique measure of maximal entropy. Moreover, with respect to this measure the system  $(\Sigma, \sigma)$  is exponentially recurrent (see [BBG, Proposition 2.3] for precise statements). The class of strongly positive recurrent CMS was intensively studied by Gurevich and Savchenko in [GS]. Note, however, that in [GS] these are called *stable-positive recurrent*. We also remark that there exists CMS that are positive recurrent but not strongly positive recurrent (see [R1, Example 2.9]).

*Remark 2.18.* Strongly positive recurrent CMS have the property that the entropy is concentrated inside the system and not near infinity. Indeed, let  $(\Sigma, \sigma)$  be a CMS and  $G$  its associated graph. Gurevich and Zargaryan [GZ] (see also [GS]) showed that a condition equivalent to SPR is the existence of a finite connected subgraph  $H \subset G$  such that there are more paths inside than outside  $H$  (in term of exponential growth). See [R1, Section 3.1] for precise statements. On the other hand, for graphs that are not strongly positive recurrent the entropy is supported by the infinite paths that spend most of the time outside a finite subgraph (see [R1, Proposition 3.3]).

Along the lines of the observations made in Remark 2.18, in the next section (see Proposition 2.20) we characterise SPR for CMS as those having entropy at infinity strictly smaller than the topological entropy.

Sarig [S2] generalised the notion of strong positive recurrence to potentials  $\varphi$ . Using his definition, we recover the topological notion in Definition 2.16 for the potential  $\varphi \equiv 0$ , i.e., this potential is strongly positive recurrent if and only if  $(\Sigma, \sigma)$  is SPR (see [R1, Remark 2.11]). For Sarig's class of potentials the associated thermodynamic formalism enjoys most of the properties of the thermodynamics for Hölder potentials on sub-shifts of finite type. In particular, the transfer operator corresponding to a strongly positive recurrent potential has a spectral gap (see [CS, Theorem 2.1]). This readily implies that the pressure function is analytic and there exist formulas for its derivatives ([S2, Theorem 3 and 4] and [CS, Theorem 1.1]), there exists a unique equilibrium state and it has exponential decay of correlations and satisfies the Central Limit Theorem ([CS, Theorem 1.1]). Moreover, in the space of potentials strongly positive recurrence is a robust property. Indeed, it was proved by Cyr and Sarig [CS, Theorem 2.2] that the space of strongly positive recurrent potentials is open and dense (with respect to the uniform metric) in the space of weakly Hölder potentials with finite pressure.

**2.6. Entropy at infinity.** A fundamental consequence of Theorem 1.1 is that a great deal of dynamical information of the system is captured by its complexity at infinity. As discussed in the introduction, we have defined two different ways of quantifying this complexity. Namely, the topological entropy at infinity (Definition 1.2) and the measure theoretic one (Definition 1.3). In this section we will elaborate on these notions and put our results into context.

We first discuss the topological entropy at infinity of  $(\Sigma, \sigma)$ , given in Definition 1.2. Observe that if  $M_1 < M_2$ , then  $z_n(M_2, q) \leq z_n(M_1, q)$ , for every  $n, q \in \mathbb{N}$ , so

$$\delta_\infty(q) = \inf_M \delta_\infty(M, q) = \lim_{M \rightarrow \infty} \delta_\infty(M, q).$$

If  $(\Sigma, \sigma)$  is a transitive CMS then for every  $a, b \in \mathcal{S}$ ,

$$\delta_\infty(M, q) = \limsup_{n \rightarrow \infty} \frac{1}{n} \log z_n(M, q, a, b), \quad (2.3)$$

where  $z_n(M, q, a, b)$  is the number of cylinders of the form  $[x_0, \dots, x_{n+1}]$ , where  $x_0 = a$ ,  $x_{n+1} = b$ , and

$$\#\{i \in \{0, \dots, n+1\} : x_i \leq q\} \leq \frac{n+2}{M}.$$

Note that  $q < q'$  implies the inequality  $z_n(M, q', a, b) \leq z_n(M, q, a, b)$ . In particular  $(\delta_\infty(M, q))_q$  is decreasing in  $q$ . We conclude that

$$\delta_\infty = \inf_q \delta_\infty(q) = \inf_{M, q} \delta_\infty(M, q). \quad (2.4)$$

Since in our results we will usually assume that the symbolic space is transitive, we can consider (2.4) as the definition of the topological entropy at infinity.

We now consider the measure theoretic entropy at infinity, defined for finite entropy CMS as

$$h_\infty := \sup_{(\mu_n)_n \rightarrow 0} \limsup_{n \rightarrow \infty} h_{\mu_n}(\sigma),$$

where  $(\mu_n)_n \rightarrow 0$  means that the sequence  $(\mu_n)_n$  converges on cylinders to the zero measure. Note that the finite entropy assumption, and more generally the  $\mathcal{F}$ -property, ensures the existence of sequences of measures converging on cylinders to the zero measure (see [IV, Lemma 4.16]). In particular,  $h_\infty$  is well defined. In [IV, Example 4.17], an example of a CMS made of infinitely many loops of length two based at a common vertex is considered. The entropy is infinite and there are no sequences of measures converging to zero. Every measure gives weight at least  $1/2$  to the base cylinder.

In Section 7 we will prove that both, the topological and the measure theoretic entropies at infinity coincide. This has several consequences, in particular we obtain that Theorem 1.1 is sharp. Indeed,  $\delta_\infty$  is the smallest number for which inequality (1.1) holds.

In the context of CMS the entropy at infinity was already investigated by Gurevich and Zargaryan [GZ], Ruelle [R1] and Buzzi [B1]. It is important to mention that they also had two flavours of entropy at infinity, a topological and a measure theoretic version. It is proven by Ruelle [R1] that both notions coincide (for a precise statement see [B1, Proposition 6.1]). It turned out that the notions of entropy at infinity presented in this work coincide with theirs. Recall that if  $G$  is the graph which determines  $(\Sigma, \sigma)$ , then

$$b_\infty = \inf_F \inf_{\lambda > 0} \sup \{h_\mu(\sigma) : \mu([F]) < \lambda\},$$

where  $F$  ranges over the finite sub-graphs of  $G$  and  $[F] := \{x \in \Sigma : x_0 \in \mathcal{A}_F\}$ , where  $\mathcal{A}_F$  denotes the symbols appearing as vertex of  $F$ . We first show the relation between  $h_\infty$  and  $b_\infty$ .

**Lemma 2.19.** *For a sequence  $(\mu_n)_n$  in  $\mathcal{M}(\Sigma, \sigma)$ , the following are equivalent:*

- (a) for any collection of cylinders  $C^1, \dots, C^N$ , and  $\varepsilon > 0$ , there is  $n_0 \in \mathbb{N}$  such that  $\mu_n(\bigcup_{i=1}^N C^i) < \varepsilon$  for all  $n \geq n_0$ ;
- (b) for any finite subgraph  $F$  of  $G$  and any  $\varepsilon > 0$ , there is  $n_1 \in \mathbb{N}$  such that  $\mu_n([F]) < \varepsilon$  for all  $n \geq n_1$ .

An easy consequence of the lemma is that convergence on cylinders in this setting corresponds to the type of limits featuring in the definition of  $b_\infty$  and thus  $b_\infty = h_\infty$ .

*Proof of Lemma 2.19.* Since (b) concerns 1-cylinders, the fact that (a) implies (b) is clear. To prove the reverse implication, we observe that if  $C^1, \dots, C^N$  is a collection of cylinders then we can take the subgraph defined by the first coordinate of each  $C^i$  as our subgraph.  $\square$

As previously mentioned, in Section 7 we will prove that  $h_\infty = \delta_\infty$ . This implies that the entropy at infinity defined in this section coincides with the previously defined one. One consequence is that, since [B1, Proposition 6.1] implies  $b_\infty < h_{top}(\sigma)$  is a characterisation of SPR, we thus have the following alternative characterisation:

**Proposition 2.20.** *A topologically transitive CMS  $(\Sigma, \sigma)$  is SPR if and only if  $h_\infty < h_{top}(\sigma)$  if and only if  $\delta_\infty < h_{top}(\sigma)$ .*

This result is consistent with the comments in Remark 2.18. Indeed, SPR systems are those for which the entropy is not concentrated at infinity; the inequality  $\delta_\infty < h_{top}(\sigma)$  has a wealth of dynamical consequences (see Remark 2.17).

From a slightly different point of view, it was not realised until recently that the entropy at infinity has a particularly important role in the regularity of the entropy map. In the context of homogeneous dynamics, for the diagonal action on  $G/\Gamma$ , where  $G$  is a  $\mathbb{R}$ -rank 1 semisimple Lie group with finite centre and  $\Gamma \leq G$  a lattice, a formula like Theorem 1.1 was obtained in [EKP]. In that context the constant playing the role of the entropy at infinity is half the topological entropy of the flow. It was later proved in [KP] that half the topological entropy is in fact sharp and equal to the measure theoretic entropy at infinity in that setup. The method employed in [EKP] was used in [RV] to prove that a similar result holds for the geodesic flow on a geometrically finite manifold. Unfortunately, an obstruction to run the method from [EKP] is the existence of periodic orbits that escape to infinity. This issue was overcome in [Vel1], where the results in [RV] were generalised to all complete negatively curved manifolds. For CMS the existence of periodic orbits that escape phase space is quite common so our approach is similar to the one in [Vel1]. Additional complications arise from the possible lack of locally compactness of  $\Sigma$ . In Section 5 and Section 6 we will address these issues and prove Theorem 1.1.

The entropy at infinity has further applications to suspension flows, entropy density, the dimension of points which escape on average, existence of equilibrium states and bounds on mass escape, all of which we give in Section 8.

### 3. KATOK'S ENTROPY FORMULA

In the early 1970s Bowen [Bo2] and Dinaburg [Di] provided a new definition of topological entropy of a dynamical system. Inspired by these results, Katok [Ka] established a formula for the measure theoretic entropy in analogy to the definition of topological entropy by Bowen and Dinaburg. We now recall his result in a particular context.

Let  $(\Sigma, \sigma)$  be a CMS and let  $d$  be the metric on  $\Sigma$  defined in (2.1). The dynamical metric  $d_n$  is defined by the formula

$$d_n(x, y) := \max_{k \in \{0, \dots, n-1\}} d(\sigma^k x, \sigma^k y).$$

The open ball of radius  $r$  centred at  $x$  with respect to the metric  $d_n$  is denoted by  $B_n(x, r)$ . By the definition of the metric  $d$  we know that  $B_n(x, 2^{-N}) = C_{n+N}(x)$ . A ball of the form  $B_n(x, r)$  is called a  $(n, r)$ -dynamical ball. The following theorem was proved in [Ka, Theorem 1.1].

**Theorem 3.1.** *Let  $(\Sigma, \sigma)$  be a sub-shift of finite type defined on a finite alphabet and  $\mu$  an ergodic  $\sigma$ -invariant probability measure. Then*

$$h_\mu(\sigma) = \lim_{\epsilon \rightarrow 0} \liminf_{n \rightarrow \infty} \frac{1}{n} \log N_\mu(n, \epsilon, \delta), \quad (3.1)$$

where  $N_\mu(n, \epsilon, \delta)$  is the minimum number of  $(n, \epsilon)$ -dynamical balls needed to cover a set of  $\mu$ -measure strictly bigger than  $1 - \delta$ . In particular the limit above does not depend on  $\delta \in (0, 1)$ .

The relation established in (3.1) is known as Katok's entropy formula. It turns out that Katok's proof is rather flexible. It was observed by Gurevich and S. Katok [GK, Section 4] and also by Riquelme [Ri, Theorem 2.6] that the proof in [Ka, Theorem 1.1] yields that if  $(X, d)$  is a metric space (not necessarily compact) and  $T : X \rightarrow X$  a continuous map then

$$h_\mu(T) \leq \lim_{\epsilon \rightarrow 0} \liminf_{n \rightarrow \infty} \frac{\log N_\mu(n, \epsilon, \delta)}{n}.$$

The compactness assumption on  $X$  is used in the proof of the other inequality. It is routine to check that compactness assumption can be replaced by the existence of a totally bounded metric.

This section is devoted to proving that formula (3.1) holds for CMS of finite topological entropy. Moreover, we will prove the limit is independent of  $\epsilon$ . We prove:

**Theorem 3.2.** *Let  $(\Sigma, \sigma)$  be a CMS and  $\mu$  an ergodic  $\sigma$ -invariant probability measure. Then for every  $\delta \in (0, 1)$  we have*

$$h_\mu(\sigma) \leq \lim_{n \rightarrow \infty} \frac{1}{n} \log N_\mu(n, 1, \delta).$$

If  $(\Sigma, \sigma)$  has finite topological entropy, then

$$h_\mu(\sigma) = \lim_{n \rightarrow \infty} \frac{1}{n} \log N_\mu(n, 1, \delta).$$

For simplicity we fix once and for all an identification of the alphabet  $\mathcal{S}$  with  $\mathbb{N}$  and we use  $\mathbb{N}$  instead. Define the following collection of sets, for every  $m \in \mathbb{N}$  let

$$K_m := \bigcup_{s=1}^m [s]. \quad (3.2)$$

Note that if  $\Sigma$  is locally compact, then  $K_m$  is compact for every  $m \in \mathbb{N}$ . To every sequence of natural numbers  $(a_i)_{i=0}^\infty$  we associate the set

$$K((a_i)_i) = \prod_{i \geq 0} \{1, \dots, a_i\}. \quad (3.3)$$

Observe that  $K((a_i)_i)$  is the intersection of a closed set with a compact set and is thus a compact subset of  $\Sigma$ . Moreover, every compact set  $K \subset \Sigma$  is contained in a set of the form  $K((a_i)_i)$ . The following lemma follows directly from [P, Theorem 3.2]. For concreteness we provide a simple proof of this general fact.

**Lemma 3.3.** *Let  $\mu$  a Borel measure on  $\Sigma$ . For every  $\epsilon > 0$ , there exists a sequence of natural numbers  $(a_i)_i$  such that  $\mu(K((a_i)_i)) > 1 - \epsilon$ .*

*Proof.* Fix a sequence  $(b_i)_i$  satisfying

$$\left(1 - \frac{\epsilon}{2}\right) \prod_{i=1}^{\infty} b_i > 1 - \epsilon,$$

where  $b_i \in (0, 1)$  for every  $i \in \mathbb{N}$ . We will construct the sequence  $(a_i)_i$  inductively. Choose  $a_0$  such that  $\mu(\bigcup_{i=1}^{a_0} [i]) > 1 - \frac{\epsilon}{2}$ . For every  $i \in \{1, \dots, a_0\}$  we choose  $c(i) \in \mathbb{N}$  such that

$$\mu\left(\bigcup_{k=1}^{c(i)} [ik]\right) \geq \mu([i])b_1.$$

Let  $a_1 := \max_{i \in \{1, \dots, a_0\}} c(i)$ . For  $(i_1, i_2) \in \prod_{i=0}^1 \{1, \dots, a_i\}$  we define  $c(i_1, i_2)$  such that

$$\mu\left(\bigcup_{k=1}^{c(i_1, i_2)} [i_1 i_2 k]\right) \geq \mu([i_1 i_2])b_2.$$

Define  $a_2 = \max_{(i, j) \in \prod_{i=0}^1 \{1, \dots, a_i\}} c(i, j)$ . We continue this procedure inductively. It follows from the construction that

$$\mu(K((a_i)_i)) = \mu\left(\prod_{i=0}^{\infty} \{1, \dots, a_i\}\right) \geq \left(1 - \frac{\epsilon}{2}\right) \prod_{i=1}^{\infty} b_i > 1 - \epsilon,$$

as desired.  $\square$

*Remark 3.4.* Katok proved [Ka, Theorem 1.1] that if  $\mathcal{P}$  is any finite partition of  $\Sigma$  satisfying  $\mu(\partial\mathcal{P}) = 0$ , then for any  $\delta \in (0, 1)$

$$h_\mu(\mathcal{P}) \leq \lim_{r \rightarrow 0} \liminf_{n \rightarrow \infty} \frac{1}{n} \log N_\mu(n, r, \delta).$$

For a CMS it is easy to check that the partitions

$$\mathcal{P}_n = \left\{ [1], \dots, [n], \bigcup_{s>n} [s] \right\}$$

are such that  $\partial\mathcal{P}_n = \emptyset$ , and  $\lim_{n \rightarrow \infty} h_\mu(\sigma, \mathcal{P}_n) = h_\mu(\sigma)$ . From this we conclude that

$$h_\mu(\sigma) \leq \lim_{r \rightarrow 0} \liminf_{n \rightarrow \infty} \frac{1}{n} \log N_\mu(n, r, \delta).$$

Our next result is inspired by the proof of [Ri, Theorem 2.10 and Theorem 2.11]. In our context we do not have local compactness of  $\Sigma$ : the finite entropy assumption is important in overcoming this issue.

**Lemma 3.5.** *Let  $(\Sigma, \sigma)$  be a CMS with finite topological entropy. If  $\mu$  is an ergodic  $\sigma$ -invariant probability measure, then for every  $\delta \in (0, 1)$  we have*

$$h_\mu(\sigma) = \lim_{N \rightarrow \infty} \liminf_{n \rightarrow \infty} \frac{1}{n} \log N_\mu(n, 2^{-N}, \delta).$$

*Proof.* As observed in Remark 3.4 the inequality

$$h_\mu(\sigma) \leq \lim_{N \rightarrow \infty} \liminf_{n \rightarrow \infty} \frac{1}{n} \log N_\mu(n, 2^{-N}, \delta)$$

is known to hold. For the converse inequality it suffices to prove that for every  $\ell \in \mathbb{N}$  there exists a partition  $\mathcal{P} = \mathcal{P}(\ell)$  of  $\Sigma$  and a subset  $K \subset \Sigma$  satisfying:

- (a) The partition  $\mathcal{P}(\ell)$  has finite entropy with respect to  $\mu$ .
- (b)  $\mu(K) > 1 - \frac{\delta}{6}$ .
- (c) For every  $x \in K \cap \sigma^{-n}K$  we have  $\mathcal{P}^n(x) \subset B_n(x, 2^{-\ell})$ .

In this situation a slight modification of the first part of the proof in [Ka, Theorem 1.1] yields the desired inequality, as we show here. Suppose that the partition  $\mathcal{P} = \mathcal{P}(\ell)$  has been constructed. Let  $\epsilon > 0$ . Since the measure  $\mu$  is ergodic by the Shannon-McMillan-Breiman theorem there exists  $N_0 \in \mathbb{N}$  such that the set

$$A_{\epsilon, N_0} := \{x \in \Sigma : \mu(\mathcal{P}^n(x)) \geq \exp(-n(h_\mu(\mathcal{P}) + \epsilon)), \text{ for all } n \geq N_0\}.$$

satisfies  $\mu(A_{\epsilon, N_0}) > 1 - \frac{\delta}{6}$ . Let  $n \geq N_0$  and  $B_n := A_{\epsilon, N_0} \cap K \cap \sigma^{-n}K$ . Observe that  $\mu(B_n) \geq 1 - \frac{\delta}{2}$  and that if  $x \in B_n$  then  $x \in K \cap \sigma^{-n}K$ , and therefore  $\mathcal{P}^n(x) \subset B_n(x, 2^{-\ell})$ . The set  $A_{\epsilon, N_0}$  requires at most  $\exp(n(h_\mu(\mathcal{P}) + \epsilon))$  elements of the partition  $\mathcal{P}^n$  to cover it. Therefore,  $B_n$  requires at most  $\exp(n(h_\mu(\mathcal{P}) + \epsilon))$   $(n, 2^{-\ell})$ -dynamical balls to cover it, where  $\mu(B_n) > 1 - \frac{\delta}{2}$ . We conclude that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log N_\mu(n, 2^{-\ell}, \delta) \leq h_\mu(\mathcal{P}) + \epsilon \leq h_\mu(\sigma) + \epsilon.$$

Since  $\epsilon > 0$  was arbitrary we obtain

$$\lim_{\ell \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n} \log N_\mu(n, 2^{-\ell}, \delta) \leq h_\mu(\sigma),$$

concluding the proof of the lemma.

We now prove the existence of such a partition  $\mathcal{P} = \mathcal{P}(\ell)$ . By Lemma 3.3 there exists a sequence  $(a_i)_i$  such that the compact set  $K_0 := K((a_i)_i)$  satisfies  $\mu(K_0) \geq 1 - \frac{\delta}{6}$ . Denote by  $S$  the set of points in  $\Sigma$  that enter  $K_0$  infinitely many times under iterates of  $\sigma$ . It is a consequence of Birkhoff's Ergodic Theorem that  $\mu(S) = 1$ . Define  $K := K_0 \cap S$ , and observe that  $\mu(K) \geq 1 - \frac{\delta}{6}$ . For every  $k \geq 1$ , let

$$R_K(x) := \inf \{k \geq 1 : \sigma^k(x) \in K_0\} \text{ for } x \in K, \text{ and } A_k := \{x \in K : R_K(x) = k\}.$$

Partition  $A_k$  using cylinders of length  $k + \ell + 1$  and denote such partition by  $\mathcal{Q}_k$ . It is important to observe that  $\#\mathcal{Q}_k$  is finite for all  $k$ . This follows from the definition of  $K_0$  and the finite topological entropy of  $(\Sigma, \sigma)$ . Indeed, if  $x = (x_0, x_1, \dots, x_k \dots) \in A_k$ , then  $x_0, x_k \in \{1, \dots, a_0\}$ . Moreover, there are at most  $C = \prod_{i=0}^{\ell} a_i$  cylinders of the form  $[y_0 y_1 \dots y_\ell]$  intersecting  $K$ , so for  $k$  large enough,

$$\#\mathcal{Q}_k \leq C e^{k(h_{\text{top}}(\sigma) + 1)}.$$

Finally, consider the partition of  $\Sigma$  defined by  $\mathcal{P} = \mathcal{Q} \cup (\bigcup_{k=1}^{\infty} \mathcal{Q}_k)$ , where  $\mathcal{Q} := \Sigma \setminus \bigcup_{k=1}^{\infty} \mathcal{Q}_k$ . We claim that this countable partition satisfies the remaining required properties, that is:

- (a) The partition  $\mathcal{P} = \mathcal{P}(\ell)$  has finite entropy with respect to  $\mu$ .
- (b) For every  $x \in K \cap \sigma^{-n}K$  we have  $\mathcal{P}^n(x) \subset B_n(x, 2^{-\ell})$ .

The second property follows from the construction of  $\mathcal{P}$ . Indeed, let  $z \in \mathcal{P}^n(x)$ , where  $x, \sigma^n(x) \in K$ . We claim that  $z \in B_n(x, 2^{-\ell})$ . For simplicity we will assume that  $x$  has its first return to  $K$  at time  $n$  (the general case is just an iteration of the argument in this setting). Since  $x \in A_n$  we have that  $\mathcal{P}(x)$  is a cylinder of length  $n + \ell + 1$ , which readily implies that  $z \in B_n(x, 2^{-\ell})$ .

We now verify that  $H_\mu(\mathcal{P}) < \infty$ . For  $r$  sufficiently large,

$$\begin{aligned} H_\mu(\mathcal{P}) + R &= \sum_{k \geq r} \sum_{P \in \mathcal{Q}_k} -\mu(P) \log \mu(P) \\ &= \sum_{k \geq r} \mu(A_k) \left( \sum_{P \in \mathcal{Q}_k} -\frac{\mu(P)}{\mu(A_k)} \log \frac{\mu(P)}{\mu(A_k)} - \frac{\mu(P)}{\mu(A_k)} \log \mu(A_k) \right) \\ &\leq \sum_{k \geq r} \mu(A_k) \log(|\mathcal{Q}_k|) - \sum_{k \geq r} \mu(A_k) \log \mu(A_k) \\ &\leq \sum_{k \geq r} k \mu(A_k) \log \left( e^{h_{\text{top}}(\sigma)+1} C^{1/k} \right) - \sum_{k \geq r} \mu(A_k) \log \mu(A_k) \\ &\leq C' \sum_{k \geq r} k \mu(A_k) - \sum_{k \geq r} \mu(A_k) \log \mu(A_k), \end{aligned}$$

where  $R = \mu(\mathcal{Q}) \log \mu(\mathcal{Q}) + \sum_{k=1}^{r-1} \sum_{P \in \mathcal{Q}_k} \mu(P) \log \mu(P) \in \mathbb{R}$ . It follows from Kac's lemma that  $\sum k \mu(A_k) = 1$ . This and the inequality

$$-\sum_{k \geq r} \mu(A_k) \log \mu(A_k) \leq \sum_{k \geq r} k \mu(A_k) + 2e^{-1} \sum_{k \geq r} e^{-k/2},$$

see [M, Lemma 1], imply the finiteness of  $H_\mu(\mathcal{P})$ . This concludes the proof.  $\square$

**Lemma 3.6.** *Let  $(\Sigma, \sigma)$  be a CMS and  $\mu$  an ergodic  $\sigma$ -invariant probability measure. Then for any  $\delta \in (0, 1)$ , we have*

$$h_\mu(\sigma) \leq \liminf_{n \rightarrow \infty} \frac{1}{n} \log N_\mu(n, 1, \delta).$$

*Proof.* Let  $A \subset \Sigma$  be a set such that  $\mu(A) \geq 1 - \delta$ . Denote by  $a_\mu(n, \delta)$  the minimum number of cylinders of length  $n$  that cover a  $A$ . Observe that

$$N_\mu(n, 2^{-t}, \delta) = a_\mu(n + t, \delta),$$

and that

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log a_\mu(n, \delta) = \liminf_{n \rightarrow \infty} \frac{1}{n} \log a_\mu(n + t, \delta),$$

for every  $t \in \mathbb{N}$ . In particular we have that  $\liminf_{n \rightarrow \infty} \frac{1}{n} \log N_\mu(n, 2^{-\ell}, \delta)$  is independent of  $\ell$ . From Remark 3.4 we conclude that

$$\begin{aligned} h_\mu(\sigma) &\leq \lim_{t \rightarrow \infty} \liminf_{n \rightarrow \infty} \frac{1}{n} \log N_\mu(n, 2^{-t}, \delta) \\ &= \lim_{t \rightarrow \infty} \liminf_{n \rightarrow \infty} \frac{1}{n} \log a_\mu(n + t, \delta) \\ &= \liminf_{n \rightarrow \infty} \frac{1}{n} \log a_\mu(n, \delta). \end{aligned}$$

$\square$

*Proof of Theorem 3.2.* The proof follows combining Lemma 3.5 and Lemma 3.6.  $\square$

We now prove a result related to Lemma 3.6. We say that two points  $x, y \in \Sigma$  are  $(n, r)$ -separated if  $d_n(x, y) \geq r$ . In particular  $x$  and  $y$  are  $(n, 1)$ -separated if they do not belong to the same cylinder of length  $n$ .

**Lemma 3.7.** *Let  $X$  be a  $\sigma$ -invariant compact subset of  $\Sigma$ . Then*

$$h_{top}(\sigma|X) = \limsup_{n \rightarrow \infty} \frac{1}{n} \log N(X, n),$$

where  $N(X, n)$  is the maximal number of  $(n, 1)$ -separated points in  $X$ , and  $h_{top}(\sigma|X)$  is the topological entropy of  $(X, \sigma)$ .

*Proof.* By definition of the topological entropy of a compact metric space we know that

$$h_{top}(\sigma|X) = \lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n} \log N(X, n, k),$$

where  $N(X, n, k)$  is the maximal number of  $(n, 2^{-k})$ -separated points in  $X$ . Observe that being  $(n, 2^{-k})$ -separated is the same as being  $(n+k, 1)$ -separated. This implies that  $N(X, n, k) = N(X, n+k)$ . Note that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log N(X, n, k) = \limsup_{n \rightarrow \infty} \frac{1}{n} \log N(X, n+k) = \limsup_{n \rightarrow \infty} \frac{1}{n} \log N(X, n).$$

Therefore

$$h_{top}(\sigma|X) = \lim_{k \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n} \log N(X, n, k) = \limsup_{n \rightarrow \infty} \frac{1}{n} \log N(X, n),$$

as desired.  $\square$

#### 4. WEAK ENTROPY DENSITY

In this section we describe the inclusion  $\mathcal{E}(\Sigma, \sigma) \subset \mathcal{M}(\Sigma, \sigma)$ , where  $\mathcal{E}(\Sigma, \sigma)$  is the subset of ergodic measures. It is well known that, even in this non-compact setting, the set  $\mathcal{E}(\Sigma, \sigma)$  is dense in  $\mathcal{M}(\Sigma, \sigma)$  with respect to the weak\* topology (see [CSc, Section 6]). We prove that any finite entropy measure can be approximated by an ergodic measure with entropy sufficiently large, see Proposition 4.1. This result can be thought of as a weak form of entropy density. In Section 8.3 we will make use of this result to prove that any invariant measure  $\mu$  can be approximated by ergodic measures with entropy converging to  $h_\mu(\sigma)$  (see Theorem 8.7). Moreover, Proposition 4.1 will be used in the proof of our main result (see Theorem 1.1). Both the statement and the proof of Proposition 4.1 closely follow that of [EKW, Theorem B], but modifications are required to deal with the non-compactness of the space  $\Sigma$ .

**Proposition 4.1.** *Let  $(\Sigma, \sigma)$  be a transitive CMS. Then for every  $\mu \in \mathcal{M}(\Sigma, \sigma)$  with  $h_\mu(\sigma) < \infty$ ,  $\epsilon > 0$ ,  $\eta > 0$ , and  $f_1, \dots, f_l \in C_b(\Sigma)$ , there exists an ergodic measure  $\mu_\epsilon \in V(f_1, \dots, f_l, \mu, \epsilon)$  (see equation (2.2)) such that  $h_{\mu_\epsilon}(\sigma) > h_\mu(\sigma) - \eta$ . We can moreover assume that  $\text{supp}(\mu_\epsilon)$  is compact.*

Analogously to the proof of [EKW, Theorem B] we will use the following fact.

**Lemma 4.2.** *Let  $\mu \in \mathcal{E}(\Sigma, \sigma)$ ,  $\alpha > 0$ ,  $\beta > 0$ ,  $f_1, \dots, f_l \in C_b(\Sigma)$ , and a set  $K \subset \Sigma$  satisfying  $\mu(K) > 3/4$ . Assume that  $h_\mu(\sigma) < \infty$ . Then there exists  $n_0 \in \mathbb{N}$  such that for all  $n \geq n_0$  there is a finite set  $\mathcal{G} = \mathcal{G}(n) \subset \Sigma$  satisfying the following properties:*

- (a)  $\mathcal{G} \subset K \cap \sigma^{-n}K$

- (b)  $d(x, y) > 2^{-n}$ , for every pair of distinct points  $x, y \in \mathcal{G}$ .
- (c)  $\#\mathcal{G} \geq \exp(n(h_\mu(\sigma) - \alpha))$ .
- (d)  $|\frac{1}{n} \sum_{k=0}^{n-1} f_j(\sigma^k x) - \int f_j d\mu| < \beta$ , for all  $x \in \mathcal{G}$  and  $j \in \{1, \dots, \ell\}$ .

*Proof.* Let

$$A_{k,\beta} := \left\{ x \in \Sigma : \left| \sum_{i=0}^{n-1} f_j(\sigma^i x) - \int f_j d\mu \right| < \beta, \forall j \in \{1, \dots, \ell\} \text{ and } n \geq k \right\}.$$

By Birkhoff's Ergodic Theorem there exists  $s_0 \in \mathbb{N}$  such that  $\mu(A_{s_0,\beta}) > 3/4$ . From Lemma 3.6 we have that

$$h_\mu(\sigma) \leq \liminf_{n \rightarrow \infty} \frac{1}{n} \log N_\mu(n, 1, 1/4).$$

There exists  $s_1 \in \mathbb{N}$  such that if  $n \geq s_1$ , then

$$\exp(n(h_\mu(\sigma) - \alpha)) \leq N_\mu(n, 1, 1/4).$$

Let  $B_n := K \cap \sigma^{-n}K \cap A_{s_0,\beta}$  and observe that  $\mu(B_n) > 1/4$ . In what follows we assume that  $n \geq n_0 := \max\{s_0, s_1\}$ . From the definition of  $N_\mu(n, 1, 1/4)$  the minimal number of cylinders of length  $n$  needed to cover  $B_n$  is at least  $N_\mu(n, 1, 1/4)$ . More precisely, let  $(C_i)_{i \in I}$  be a minimal collection of cylinders of length  $n$  covering  $B_n$ . In particular for every  $i \in I$  we have  $C_i \cap B_n \neq \emptyset$ . For every  $i \in I$  choose a point  $x_i \in C_i \cap B_n$ . We claim that the set  $(x_i)_{i \in I}$  satisfies the properties required on  $\mathcal{G}$ . Conditions (a) and (d) follow from the definition of  $B_n$ . Condition (b) follows from the fact that if  $i \neq j$ , then  $x_i$  and  $x_j$  are in different cylinders of length  $n$ . Condition (c) follows from the inequality

$$\#I \geq N_\mu(n, 1, 1/4) \geq \exp(n(h_\mu(\sigma) - \alpha)).$$

□

*Proof of Proposition 4.1.* Recall that we want to prove that given  $\mu \in \mathcal{M}(\Sigma, \sigma)$ ,  $\epsilon > 0$ ,  $\eta > 0$ , and  $f_1, \dots, f_\ell \in C_b(\Sigma)$ , there exists an ergodic measure  $\mu_\epsilon \in V(f_1, \dots, f_\ell, \mu, \epsilon)$  such that  $h_{\mu_\epsilon}(\sigma) > h_\mu(\sigma) - \eta$ . In the following remarks we observe that this general situation can be simplified.

As observed in Section 2.2.1 or in [B, 8.3.1 Remark] it suffices to consider the case in which the functions  $(f_i)_i$  in Proposition 4.1 are uniformly continuous. Therefore, under this assumption, there exists  $A = A(f_1, \dots, f_\ell) \in \mathbb{N}$ , such that if  $d(x, y) < 2^{-A}$ , then  $|f_i(x) - f_i(y)| < \frac{\epsilon}{4}$ . Also define  $W = \max_{i \in \{1, \dots, \ell\}} |f_i|_0$ .

Since periodic measures are dense in  $\mathcal{M}(\Sigma, \sigma)$ , see [CSc, Section 6], we will assume that  $h_\mu(\sigma) - \eta > 0$ , otherwise we can approximate  $\mu$  by a periodic measure. By the affinity of the entropy map [Wa, Theorem 8.1] and [IV, Lemma 6.13] we can reduce the problem to the case in which  $\mu = \frac{1}{N} \sum_{i=1}^N \mu_i$ , where  $\{\mu_i\}_{i=1}^N$  is a collection of ergodic measures.

Let  $m \in \mathbb{N}$  be such that the set  $K = K_m$ , defined as in (3.2), satisfies  $\mu_i(K) > 3/4$  for every  $i \in \{1, \dots, N\}$ . Since  $(\Sigma, \sigma)$  is transitive, there exists a constant  $L = L(m)$  such that for each pair  $(a, b) \in \{1, \dots, m\}^2$ , there exists an admissible word  $arb$ , where  $\ell(\mathbf{r}) \leq L$ . It follows from Lemma 4.2, setting  $\beta = \epsilon/4$  and  $\alpha = \eta/2$ , that there exists  $n' \in \mathbb{N}$  such that for every  $n > n'$  and every measure  $\mu_i$ , with  $i \in \{1, \dots, N\}$ , there exists  $(n, 1)$ -separated sets  $\mathcal{G}_i \subset K \cap \sigma^{-n}K$  satisfying properties (a), (b), (c) and (d) of Lemma 4.2.

Denote by  $\mathbf{a}(x)$  the word defined concatenating the first  $(n+1)$ -coordinates of  $x \in \Sigma$ . Given  $\hat{x} = (x^1, x^2, \dots, x^{MN}) \in (\prod_{i=1}^N \mathcal{G}_i)^M$ , we define an admissible word  $w_0(\hat{x}) = \mathbf{a}(x^1)\mathbf{r}_1\mathbf{a}(x^2)\mathbf{r}_2 \dots \mathbf{a}(x^{MN})\mathbf{r}_{MN}\mathbf{a}(x^1)$ , where  $\mathbf{r}_k$ s are words chosen so that  $w_0(\hat{x})$  is an admissible word and  $\ell(\mathbf{r}_k) \leq L$  (note that this is possible since  $(\mathbf{a}(x^i))_0$  and  $(\mathbf{a}(x^i))_n$  are in  $\{1, \dots, m\}$  by definition of  $\mathcal{G}_i$ ). The word  $w_0(\hat{x})$  defines a periodic point in  $\Sigma$  that we denote by  $w(\hat{x})$ . We have that

$$w(\hat{x}) = \overline{\mathbf{a}(x^1)\mathbf{r}_1\mathbf{a}(x^2)\mathbf{r}_2 \dots \mathbf{a}(x^{MN})\mathbf{r}_{MN}}.$$

Let  $\mathcal{G} := \prod_{M \geq 1} (\prod_{i=1}^N \mathcal{G}_i)^M$ . Following the same procedure of concatenation described above, for every  $\hat{x} \in \mathcal{G}$  we define a point  $w(\hat{x}) \in \Sigma$ . Define

$$\Psi = \bigcup_{\hat{x} \in \mathcal{G}} O(w(\hat{x})),$$

where  $O(w(\hat{x}))$  is the orbit of  $w(\hat{x})$  and define  $\Psi_0$  to be the topological closure of  $\Psi$ .

Note that the space  $\Psi_0$  is a compact  $\sigma$ -invariant subset of  $\Sigma$ . By definition the set  $\Psi$  is closed and invariant. Observe that the number of symbols appearing in elements belonging to  $\Psi$  is finite: there are finitely many admissible words  $\mathbf{a}(x^i)$  (recall that each  $\mathcal{G}_i$  is a finite set) and we could use finitely many connecting words  $\mathbf{r}_i$ . Therefore there exists  $J \in \mathbb{N}$  such that  $\Psi \subset \{1, \dots, J\}^{\mathbb{N}}$ . Thus,  $\Psi_0$  is a closed subset of a compact set.

By property (d) of Lemma 4.2, and assuming that  $n'$ , which also depends on  $A$ ,  $W$  and  $L$ , is sufficiently large,

$$\Psi \subset \left\{ x \in \Sigma : \left| \frac{1}{n} S_n f_j(x) - \int f_j d\mu \right| \leq \epsilon, \forall j \in \{1, \dots, \ell\} \right\}. \quad (4.1)$$

Since the set in right hand side of (4.1) is closed, the same inclusion holds if  $\Psi$  is replaced by  $\Psi_0$ . Also, since  $\Psi_0$  is  $\sigma$ -invariant we have

$$\begin{aligned} \Psi_0 &\subset \left\{ x \in \Sigma : \left| \frac{1}{n} S_n f_j(\sigma^s x) - \int f_j d\mu \right| \leq \epsilon, \forall j \in \{1, \dots, \ell\} \text{ and } s \geq 0 \right\} \\ &\subset \left\{ x \in \Sigma : \left| \frac{1}{nk} S_{nk} f_j(x) - \int f_j d\mu \right| \leq \epsilon, \forall j \in \{1, \dots, \ell\} \text{ and } k \in \mathbb{N} \right\}. \end{aligned}$$

This implies that every ergodic measure supported in  $\Psi_0$  belongs to  $V(f_1, \dots, f_\ell, \mu, \epsilon)$ . Indeed, consider a generic point for the ergodic measure and use the inclusion above.

By construction, if  $x, y \in \left( \prod_{i=1}^N \mathcal{G}_i \right)^M$  and  $x \neq y$ , then

$$d_{NM(n+1+L)}(w(x), w(y)) = 1.$$

In other words  $\Psi_0$  contains a  $(NM(n+1+L), 1)$ -separated set of cardinality at least

$$\exp \left( nNM \left( \frac{1}{N} \sum_{k=1}^N h_{\mu_k}(\sigma) - \frac{\eta}{2} \right) \right).$$

Here we used property (c) of Lemma 4.2 for our sets  $\mathcal{G}_i$ . It follows from Lemma 3.7 that

$$h_{top}(\Psi_0) \geq \limsup_{M \rightarrow \infty} \frac{nNM(h_\mu(\sigma) - \frac{\eta}{2})}{NM(n+1+L)} = \frac{n(h_\mu(\sigma) - \frac{\eta}{2})}{(n+1+L)} > h_\mu(\sigma) - \eta.$$

Finally, let  $\mu_e$  be an ergodic measure supported in  $\Psi_0$  with entropy at least  $h_\mu(\sigma) - \eta$  (which exists by the standard variational principle in the compact setting), since we already proved that  $\mu_e \in V(f_1, \dots, f_l, \mu, \epsilon)$  this finishes the proof.  $\square$

## 5. MAIN ENTROPY INEQUALITY

This section is devoted to the proof of the main entropy inequality. This is stated in Theorem 5.1 and relates the entropy of a sequence of ergodic measures with the amount of mass lost and the topological entropy at infinity.

Recall that, as explained in (3.3), to every sequence of natural numbers  $(a_i)_i$  we assign a compact set  $K = K((a_i)_i) \subset \Sigma$ . The definition of  $K$  implies that if  $x \in K^c$ , then  $x_i > a_i$ , for some  $i \in \mathbb{N}_0$ . For  $x \in K^c$  we define  $i : K^c \rightarrow \mathbb{N}_0$  by

$$i(x) := \min \{n \in \mathbb{N}_0 : x_n > a_n\}. \quad (5.1)$$

For  $n \in \mathbb{N}$  we define

$$T_n(K) := K_{a_0} \cap \sigma^{-1}K^c \cap \dots \cap \sigma^{-n}K^c \cap \sigma^{-(n+1)}K_{a_0}, \quad (5.2)$$

where  $K_{a_0} = \bigcup_{i=1}^{a_0} [i]$  (as defined in (3.2)). Let

$$\hat{T}_n(K) := \{x \in T_n(K) : i(\sigma^k(x)) \leq n - k, \text{ for every } k \in \{1, \dots, n\}\}. \quad (5.3)$$

Let  $\hat{z}_n(K)$  be the minimal number of cylinders of length  $(n+2)$  needed to cover  $\hat{T}_n(K)$  and define

$$\hat{\delta}_\infty(K) := \limsup_{n \rightarrow \infty} \frac{1}{n} \log \hat{z}_n(K). \quad (5.4)$$

The reason why we define  $\hat{\delta}_\infty(K)$  covering the sets  $\hat{T}_n(K)$ , and not  $T_n(K)$ , is to ensure Lemma 6.2. This allows us to relate  $\hat{\delta}_\infty(K)$  with the topological entropy at infinity of  $(\Sigma, \sigma)$ .

Our next result is fundamental in this paper.

**Theorem 5.1.** *Let  $(\Sigma, \sigma)$  be a finite entropy CMS. Let  $(\mu_n)_n$  be a sequence of ergodic probability measures converging on cylinders to an invariant measure  $\mu$ . Let  $(a_i)_i$  be an increasing sequence of natural numbers such that the corresponding compact set  $K = K((a_i)_i)$  satisfies that  $\mu_n(K) > 0$ , for all  $n \in \mathbb{N}$ . Then*

$$\limsup_{n \rightarrow \infty} h_{\mu_n}(\sigma) \leq |\mu| h_{\mu/|\mu|}(\sigma) + (1 - \mu(Y)) \hat{\delta}_\infty(K),$$

where  $Y = \bigcup_{s=0}^{\infty} \sigma^{-s}K$ .

The proof of this theorem requires some propositions and lemmas, which we will prove first before completing the proof of the theorem at the end of this section.

The fact that  $K \subset \sigma(K)$ , which follows since  $(a_i)_i$  is an increasing sequence, will be used several times here. Let  $A_k := \{x \in K : R_K(x) = k\}$ , where  $R_K(x)$  is the first return time function to the set  $K$  (see Definition 2.9). For  $x \in Y$  we define the following:

$$\begin{aligned} n_1(x) &:= \min \{n \in \mathbb{N}_0 : \text{there exists } y \in K \text{ such that } \sigma^n(y) = x\}, \\ n_2(x) &:= \min \{n \in \mathbb{N}_0 : \sigma^n x \in K\}. \end{aligned}$$

We emphasise that the function  $n_1(x)$  is well defined. Indeed, observe that if  $x \in Y$  then  $\sigma^{n_2(x)}x \in K$ . Let  $r \in \mathbb{N}$  be such that  $r > n_2(x) - 1$ . Since the sequence  $(a_i)_i$  is increasing we have that  $a_r \geq \max\{a_i : i \in \{0, \dots, n_2(x) - 1\}\}$ .

Since  $x \in \sigma^r(K) = \prod_{k=r}^{\infty} \{1, \dots, a_k\} \cap \Sigma$  we have that  $n_1(x)$  is finite for every  $x \in Y$ . Let

$$n(x) := \begin{cases} n_1(x) + n_2(x) & \text{if } x \in Y, \\ \infty & \text{if } x \in \Sigma \setminus Y. \end{cases}$$

For  $n \in \mathbb{N}_0 \cup \{\infty\}$  define

$$\mathcal{C}_n := \{x \in \Sigma : n(x) = n\}.$$

Note that  $\mathcal{C}_0 = K$  and  $\mathcal{C}_1 = \emptyset$ . For  $n \geq 2$  observe that  $x \in \mathcal{C}_n$  if it belongs to the orbit of a point in  $A_n$ . More precisely, for every  $n \geq 2$  we have that  $\mathcal{C}_n \subset \bigcup_{k=1}^{n-1} \sigma^k(A_n)$ . We define the following sets,

$$\alpha_{\leq N} := \left( \bigcup_{n=2}^N \mathcal{C}_n \right), \alpha_{N,M} := \left( \bigcup_{n>N}^M \mathcal{C}_n \right) \text{ and } \alpha_{>M} := \left( \bigcup_{n>M} \mathcal{C}_n \right) \cup \mathcal{C}_{\infty}.$$

*Remark 5.2.* The set  $\alpha_{\leq M}$  can be covered with finitely many cylinders of length  $L$ . Indeed, observe that for every  $n \geq 2$  we have

$$\mathcal{C}_n \subset \bigcup_{s=1}^{n-1} \sigma^s(A_n) \subset \bigcup_{s=1}^{n-1} \sigma^s(K) \subset \sigma^{n-1}(K).$$

Therefore,

$$\alpha_{\leq M} = \bigcup_{n=2}^M \mathcal{C}_n \subset \sigma^{M-1}(K) = \prod_{s=M-1}^{\infty} \{1, \dots, a_s\} \cap \Sigma.$$

Since the set  $\prod_{s=M-1}^{\infty} \{1, \dots, a_s\} \cap \Sigma$  can be covered with at most  $\prod_{s=M-1}^{M-2+L} a_s$  cylinders of length  $L$ , the same holds for  $\alpha_{\leq M}$ .

Observe that it follows directly from the definition of  $\hat{\delta}_{\infty}(K)$  (see (5.4)) that for every  $\epsilon > 0$ , there exists  $N_0 = N_0(\epsilon) \in \mathbb{N}$  such that for every  $n \geq N_0$  we have

$$\hat{z}_n(K) \leq e^{n(\hat{\delta}_{\infty}(K) + \epsilon)}.$$

At this point we fix  $\epsilon > 0$  and  $k, N \in \mathbb{N}$  large enough so that  $kN \geq N_0(\epsilon)$ : these will appear explicitly in the proof of Theorem 5.1.

Given  $A \subset \Sigma$  and  $t \in \mathbb{N}$  we define

$$U_t(A) := \{x \in \Sigma : d(x, A) \leq 2^{-t}\}.$$

Now let

$$\begin{aligned} K(k, N) &:= U_{kN+2}(K), \\ \gamma_{\leq N} &:= U_{(k+1)N+2}(\alpha_{\leq N}) \setminus K(k, N), \\ G_{k,N} &:= K(k, N) \cup \gamma_{\leq N}, \end{aligned}$$

and

$$\begin{aligned} \gamma_{N,kN} &:= U_{2(k+1)N+2}(\alpha_{N,kN}) \setminus G_{k,N}, \\ \gamma_{>kN} &:= \Sigma \setminus (G_{k,N} \cup \gamma_{N,kN}), \\ B_{k,N} &:= \gamma_{N,kN} \cup \gamma_{>kN}. \end{aligned}$$

Denote by  $\mathcal{Q}_1(k, N)$  the minimal cover of  $K(k, N)$  with cylinders of length  $kN + 2$ . Similarly, denote by  $\mathcal{Q}'_2(k, N)$  the minimal cover of  $\alpha_{\leq N}$  with cylinders of length  $(k+1)N + 2$ . Observe that every element in  $\mathcal{Q}'_2(k, N)$  is disjoint or contained in

an element of  $\mathcal{Q}_1(k, N)$ . In particular  $\gamma_{\leq N}$  is a finite union of cylinders of length  $(k+1)N+2$ ; this collection of cylinders is denoted by  $\mathcal{Q}_2(k, N)$ . Define

$$\beta'_{k,N} := \mathcal{Q}_1(k, N) \cup \mathcal{Q}_2(k, N) \quad (5.5)$$

and observe that  $\beta'_{k,N}$  is a partition of the set  $G_{k,N}$ . Define the following partition of  $\Sigma$ ,

$$\beta_{k,N} := \{\gamma_{>kN}, \gamma_{N,kN}\} \cup \beta'_{k,N}. \quad (5.6)$$

Recall that the refinement  $\beta_{k,N}^n$  follows as in Section 2.3.

*Notation:* We use the following notation for an interval of integers  $[a, b) := \{n \in \mathbb{N} : a \leq n < b\}$  and  $|[a, b) = b - a$ .

**Definition 5.3.** Let  $Q \in \beta_{k,N}^n$  be such that  $(Q \cup \sigma^{n-1}Q) \subset G_{k,N}$ . An interval  $[r, s) \subset [0, n)$  is called an *excursion* of  $Q$  into  $\gamma_{>kN}$  (resp.  $B_{k,N}$ ) if  $\sigma^t Q \subset \gamma_{>kN}$  (resp.  $\sigma^t Q \subset B_{k,N}$ ) for every  $t \in [r, s)$  and  $(\sigma^{r-1}Q \cup \sigma^s Q) \subset G_{k,N}$ .

An excursion  $[r, s)$  of  $Q$  into  $B_{k,N}$  is said to *enter*  $\gamma_{>kN}$  if there exists  $i \in [r, s)$  such that  $\sigma^i Q \subset \gamma_{>kN}$ .

The next three lemmas are preparation for the proof of Proposition 5.8. These give us control on the return times to  $K(k, N)$  and the length of excursions into  $B_{k,N}$ .

**Lemma 5.4.** *If  $[r, r+s)$  is an excursion of  $Q$  into  $B_{k,N}$  that does not enter  $\gamma_{>kN}$  then  $s < kN$ .*

*Proof.* Since the excursion does not enter  $\gamma_{>kN}$  we have that  $\sigma^r Q \subset \gamma_{N,kN}$ . Fix  $x \in \sigma^r Q$ . By the definition of  $\gamma_{N,kN}$  there exists  $x_0 \in \alpha_{N,kN}$  such that  $d(x, x_0) \leq 2^{-(2(k+1)N+2)}$ . Since  $x_0 \in \alpha_{N,kN}$  we have that  $n(x_0) \leq kN$  and therefore  $n_2(x_0) < kN$ . In particular  $\sigma^t(x_0) \in \alpha_{\leq N}$ , for some  $t \in [0, kN)$ . Observe that

$$d(\sigma^t(x), \sigma^t(x_0)) \leq 2^{-(2(k+1)N+2)+t} \leq 2^{-((k+1)N+2)}.$$

This readily implies that  $\sigma^t(x) \in U_{(k+1)N+2}(\alpha_{\leq N}) \subset G_{k,N}$ . We conclude that  $\sigma^{r+t} Q \subset G_{k,N}$ , and therefore  $s < kN$ .  $\square$

**Lemma 5.5.** *If  $Q \subset G_{k,N}$  then there exists  $t \in [0, N)$  such that  $\sigma^t Q \subset K(k, N)$ .*

*Proof.* If  $Q \subset K(k, N)$  there is nothing to prove. Assume that  $Q \subset \gamma_{\leq N}$ . Let  $x \in Q$  and  $y \in \alpha_{\leq N}$  such that  $d(x, y) \leq 2^{-((k+1)N+2)}$ . Since  $y \in \alpha_{\leq N}$  we have that  $\sigma^t(y) \in K$ , for some  $t < N$ . Observe that

$$d(\sigma^t(x), \sigma^t(y)) \leq 2^{-((k+1)N+2)+t} < 2^{-(kN+2)}.$$

We conclude that there exists  $t \in [0, N)$  such that  $\sigma^t(x) \in U_{kN+2}(K) = K(k, N)$ . This implies that for some  $t < N$  we have  $\sigma^t Q \subset K(k, N)$ .  $\square$

**Lemma 5.6.** *If  $[r, r+s)$  is an excursion of  $Q$  into  $\gamma_{>kN}$  such that  $s \geq N$  then  $\sigma^{r-1}Q \subset K(k, N)$ .*

*Proof.* From the definition of an excursion, the set  $Q_0 := \sigma^{r-1}Q$  must lie in  $G_{k,N}$ , so to derive a contradiction we will assume that  $Q_0 \subset \gamma_{\leq N}$ . Let  $x \in Q_0$ . By the construction of  $\gamma_{\leq N}$  there exists  $y \in \alpha_{\leq N}$  such that  $d(x, y) \leq 2^{-((k+1)N+2)}$ . Since  $y \in \alpha_{\leq N}$  there exists  $t \leq N$  such that  $\sigma^t(y) \in K$ . Therefore

$$d(\sigma^t(x), \sigma^t(y)) \leq 2^{-((k+1)N+2)+t} \leq 2^{-(kN+2)}.$$

We conclude that  $\sigma^t(x) \in U_{kN+2}(K) = K(k, N)$ . This contradicts the fact the length of the excursion is larger than  $N$ .  $\square$

**Definition 5.7.** Denote by  $m_{n,k,N}(Q)$  the number of excursions of length greater or equal to  $kN$  into  $B_{k,N}$  that enter  $\gamma_{>kN}$  and let

$$E_{n,k,N} := \# \{i \in [0, n) : \sigma^i Q \subset B_{k,N}\}.$$

The following result shows that an atom  $Q \in \beta_{k,N}^n$  such that  $Q \subset K(k, N) \cap \sigma^{-(n-1)}K(k, N)$  can be covered by cylinders of length  $n$  in a controlled way. This is an estimate closely related to [EKP, Lemma 7.4] (see also [Vel1, Proposition 4.5]). The constant  $\delta_\infty(K)$  naturally appears when we try to control the time spent in the ‘bad’ part  $B_{k,N}$ .

**Proposition 5.8.** *Let  $\beta_{k,N}$  be the partition defined in (5.6). Then an atom  $Q \in \beta_{k,N}^n$  such that  $Q \subset K(k, N) \cap \sigma^{-(n-1)}K(k, N)$ , can be covered by at most*

$$e^{E_{n,k,N}(Q)(\delta_\infty(K)+\epsilon)} e^{m_{n,k,N}(Q)N(\delta_\infty(K)+\epsilon)}$$

*cylinders of length  $n$ .*

*Proof.* To simplify notation we drop the sub-indices  $N$  and  $k$ . The proof of Proposition 5.8 is by induction on  $n$ . First decompose  $[0, n-1]$  into

$$[0, n-1] = W_1 \cup V_1 \cup W_2 \cup \dots \cup V_s \cup W_{s+1},$$

according to the excursions into  $B_{k,N}$  that contain at least one excursion into  $\gamma_{>kN}$ . More precisely, let  $V_i = [m_i, m_i + h_i)$  and  $W_i = [l_i, l_i + L_i)$  with  $l_i + L_i = m_i$  and  $m_i + h_i = l_{i+1}$ . The segment  $V_i$  denotes an excursion into  $B_{k,N}$  that contains an excursion into  $\gamma_{>kN}$ . Given  $i \in \mathbb{N}$  define  $J_i := \sum_{j=1}^i |V_j| 1_{[kN, \infty)}(|V_j|)$ , where  $1_{[kN, \infty)}$  is the characteristic function of the interval  $[kN, \infty)$ . Similarly define  $H_i := \sum_{j=1}^i 1_{[kN, \infty)}(|V_j|)$ . Observe that  $Q \subset K(k, N)$  implies that  $Q$  is already contained in a cylinder of length  $kN + 2$ .

*Step 1:* Assume that  $Q$  has been covered with  $c_i$  cylinders of length  $l_i$ , where

$$c_i \leq e^{J_i(\delta_\infty(K)+\epsilon)} e^{NH_i(\delta_\infty(K)+\epsilon)}.$$

(As mentioned above, the set  $Q$  is covered by one cylinder of length 1, therefore take  $c_1 = 1$ .) We claim that the same number of cylinders of length  $(l_i + L_i)$  cover  $Q$ . Observe that by hypothesis  $\sigma^{l_i}Q$  is contained in an element of  $\beta'$ , therefore  $\text{diam}(\sigma^{l_i}Q) \leq 2^{-(kN+2)}$ . Since the elements of  $\beta'$  all have diameter smaller than  $2^{-(kN+2)}$ , the same holds if  $Q$  spends some extra time in  $\beta'$ . By Lemma 5.4, if  $Q$  has an excursion into  $B_{k,N}$  that does not enter  $\gamma_{>kN}$ , then it must come back to  $\beta'$  before  $kN$  iterates. In particular if the excursion into  $B_{k,N}$  is  $[p_i, p_i + q_i)$ , then  $q_i < kN$ . Observe that  $\text{diam}(\sigma^{p_i-1}Q) \leq 2^{-(kN+2)}$ , implies that  $\text{diam}(\sigma^{p_i+t}Q) \leq 2^{-2}$ , for every  $t \in [0, kN)$ . In particular the same holds for  $t \in [0, q_i]$ . Repeating this process we conclude that  $\text{diam}(\sigma^tQ) \leq 2^{-2}$ , for every  $t \in [l_i, l_i + L_i)$ . This immediately implies that  $\sigma^{l_i}Q$  is contained in a cylinder of length  $L_i$ , which implies our claim. We go next to Step 2.

*Step 2:* Assume we have covered  $Q$  with  $c_i$  cylinders of length  $m_i$ , where

$$c_i \leq e^{J_i(\delta_\infty(K)+\epsilon)} e^{NH_i(\delta_\infty(K)+\epsilon)}.$$

We want to estimate the number of cylinders of length  $(m_i + h_i)$  needed to cover  $Q$ . Define  $Q_i := \sigma^{m_i-1}Q$ . If we are able to cover  $Q_i$  with  $R$  cylinders of length  $(h_i + 1)$ , then we will be able to cover  $Q$  with  $Rc_i$  cylinders of length  $(m_i + h_i)$ . We will separate into two cases:

*Case 1:  $h_i < kN$ .*

Observe that  $Q_i \subset G_{k,N}$  and is therefore contained in an element of  $\beta'$ , which implies  $\text{diam}(Q_i) \leq 2^{-(kN+2)}$ . This implies that  $Q_i$  is contained in a cylinder of length  $(kN + 2)$ . Since  $h_i < kN$ , this implies that  $Q_i$  can be covered with one cylinder of length  $(h_i + 1)$ . We conclude that

$$c_{i+1} = c_i \leq e^{J_i(\hat{\delta}_\infty(K)+\epsilon)} e^{NH_i(\hat{\delta}_\infty(K)+\epsilon)} = e^{J_{i+1}(\hat{\delta}_\infty(K)+\epsilon)} e^{NH_{i+1}(\hat{\delta}_\infty(K)+\epsilon)}.$$

*Case 2:  $h_i \geq kN$ .*

By Lemma 5.6,  $Q_i = \sigma^{m_i-1}Q \subset K(k, N)$ . Observe that by assumption  $\sigma^{h_i+1}Q_i \subset \gamma_{\leq N}$ . By Lemma 5.5 there exists  $0 \leq t_i < N$ , such that  $\sigma^{h_i+1+t_i}Q_i \subset K(k, N)$  (we assume  $t_i$  is the smallest such number). We conclude that every  $x \in Q_i$  satisfies  $x \in K_{a_0}$ ,  $\sigma^{h_i+1+t_i}(x) \in K_{a_0}$ , and  $\sigma^s x \in K^c$ , for every  $s \in \{1, \dots, h_i + t_i\}$ . In other words  $Q_i \subset T_{h_i+t_i}(K)$ . We now claim that  $Q_i \subset \hat{T}_{h_i+t_i}(K)$ . Observe that if  $x \in Q_i$ , then  $\sigma^{h_i+t_i+1}(x) \in K(k, N)$ , and  $\sigma^k(x) \in K(k, N)^c$ , for every  $k \in \{1, \dots, h_i + t_i\}$ . We argue by contradiction and suppose that  $i(\sigma^k(x)) > (h_i + t_i) - k$  for some  $k \in \{1, \dots, h_i + t_i\}$ . This implies that  $(\sigma^k(x))_j \leq a_j$ , for  $j \in \{0, \dots, h_i + t_i - k\}$ . Observe that  $(\sigma^k(x))_{h_i+t_i-k+j+1} = (\sigma^{h_i+t_i+1}(x))_j$ , and for  $j \in \{0, \dots, kN + 1\}$  we have  $(\sigma^{h_i+t_i+1}(x))_j \leq a_j$ . We conclude that  $(\sigma^k(x))_{h_i+t_i-k+j+1} \leq a_j$ , for  $j \in \{0, \dots, kN\}$ . In particular we have that  $(\sigma^k(x))_j \leq a_j$ , for every  $j \in \{0, \dots, kN + 1\}$ , which contradicts that  $\sigma^k(x) \in K(k, N)^c$ , completing the proof of our claim. This implies, from the definition of  $\hat{\delta}_\infty(K)$ , that  $Q_i$  can be covered by at most  $e^{(h_i+t_i)(\hat{\delta}_\infty(K)+\epsilon)}$  cylinders of length  $(h_i + 1 + t_i)$ ; and by at most  $e^{(h_i+N)(\hat{\delta}_\infty(K)+\epsilon)}$  cylinders of length  $(h_i + 1)$ . We conclude that  $Q$  can be covered by at most  $c_{i+1}$  cylinders of length  $(n_i + h_i)$ , where

$$\begin{aligned} c_{i+1} &\leq e^{(h_i+N)(\hat{\delta}_\infty(K)+\epsilon)} (e^{J_i(\hat{\delta}_\infty(K)+\epsilon)} e^{NH_i(\hat{\delta}_\infty(K)+\epsilon)}) \\ &= e^{J_{i+1}(\hat{\delta}_\infty(K)+\epsilon)} e^{NH_{i+1}(\hat{\delta}_\infty(K)+\epsilon)}. \end{aligned}$$

Adding these steps together and noting that  $J_s = E_{n,k,N}(Q)$  and  $H_s = m_{n,k,N}(Q)$  completes the proof of the proposition.  $\square$

The idea now is to use Proposition 5.8 to compare the entropy of a measure with the corresponding entropy of our partition  $\beta_{k,N}$ . This is a natural idea: the map  $\mu \mapsto h_\mu(\beta_{k,N})$  is typically better behaved under sequences of measures; at this point we crucially use that the partition  $\beta_{k,N}$  is finite.

**Proposition 5.9.** *Let  $\beta_{k,N}$  be the partition defined in (5.6) and  $\mu$  an ergodic  $\sigma$ -invariant probability measure satisfying  $\mu(K(k, N)) > 0$ . Then*

$$h_\mu(\sigma) \leq h_\mu(\beta_{k,N}) + \left( \mu(B_{k,N}) + \frac{1}{k} \right) (\hat{\delta}_\infty(K) + \epsilon).$$

*Proof.* To simplify notation we denote the partition  $\beta_{k,N}$  by  $\beta$ . We will apply Theorem 3.2, so the main task is to estimate  $N_\mu(n, 1, \delta)$  for some  $\delta \in (0, 1)$ . Since  $\mu$  is an ergodic measure such that  $\mu(K(k, N)) > 0$  there exists  $\delta_1 > 0$  and an increasing sequence  $(n_i)_i$  satisfying

$$\mu(K(k, N) \cap \sigma^{-n_i}K(k, N)) > \delta_1,$$

for every  $i \in \mathbb{N}$ . Given  $\epsilon_1 > 0$ , by the Shannon-McMillan-Breiman theorem the set

$$\mathcal{D}_{\epsilon_1, N} = \{x \in X : \forall n \geq N, \mu(\beta^n(x)) \geq \exp(-n(h_\mu(\beta) + \epsilon_1))\},$$

satisfies

$$\lim_{N \rightarrow \infty} \mu(\mathcal{D}_{\epsilon_1, N}) = 1.$$

By Birkhoff's Ergodic Theorem there exists a set  $W_{\epsilon_1} \subset \Sigma$  satisfying  $\mu(W_{\epsilon_1}) > 1 - \frac{\delta_1}{4}$  and  $n(\epsilon_1) \in \mathbb{N}$  such that for every  $x \in W_{\epsilon_1}$  and  $n \geq n(\epsilon_1)$ ,

$$\frac{1}{n} \sum_{i=0}^{n-1} 1_{B_{k, N}}(\sigma^i x) < \mu(B_{k, N}) + \epsilon_1.$$

Define

$$X_i := W_{\epsilon_1} \cap \mathcal{D}_{\epsilon_1, n_i} \cap K(k, N) \cap \sigma^{-n_i} K(k, N).$$

So for sufficiently large values of  $i \in \mathbb{N}$ , by construction we have that  $\mu(X_i) > \frac{\delta_1}{2}$ . In what follows we will assume that  $i \in \mathbb{N}$  is large enough that it satisfies this condition.

By definition of  $\mathcal{D}_{\epsilon_1, n_i}$  the set  $X_i$  can be covered by  $\exp(n_i(h_\mu(\beta) + \epsilon_1))$  many elements of  $\beta^{n_i}$ . We will make use of Proposition 5.8 to efficiently cover each of those atoms by cylinders. Let  $Q \in \beta^{n_i}$  be an atom intersecting  $X_i$ . In particular  $Q \in K \cap \sigma^{-(n-1)}K$ . It follows from the definition of  $W_{\epsilon_1}$  that

$$E_{n_i, k, N}(Q) < (\mu(B_{k, N}) + \epsilon_1) n_i.$$

Moreover,

$$m_{n_i, k, N}(Q) \leq \frac{1}{kN} n_i.$$

Indeed, each of the excursions counted in  $m_{n_i, k, N}$  has length at least  $kN$ , which implies that the number of excursions can not be larger than  $\frac{1}{kN} n_i$ . Therefore Proposition 5.8 implies that

$$N_\mu \left( n_i, 1, 1 - \frac{\delta_1}{2} \right) \leq e^{n_i(h_\mu(\beta) + \epsilon_1)} e^{n_i(\hat{\delta}_\infty(K) + \epsilon)} (\mu(B_{k, N}) + \epsilon_1) e^{\frac{1}{kN} n_i N (\hat{\delta}_\infty(K) + \epsilon)}.$$

It now follows from Katok's entropy formula (see Theorem 3.2) that

$$h_\mu(\sigma) \leq h_\mu(\beta_{k, N}) + \epsilon_1 + (\hat{\delta}_\infty(K) + \epsilon)(\mu(B_{k, N}) + \epsilon_1) + \frac{1}{k}(\hat{\delta}_\infty(K) + \epsilon).$$

Since  $\epsilon_1 > 0$  was arbitrary the proof is complete.  $\square$

As in Proposition 5.9 we denote the partition  $\beta_{k, N}$  by  $\beta$ . We may assume, possibly after refining the partition, that

$$\beta = \{C^1, \dots, C^q, R\},$$

where each  $C^i$  is a cylinder for the original partition and  $R = \gamma_{>kN}$  is the complement of a finite collection of cylinders. For simplicity we still denote this partition by  $\beta$ . We emphasise that Proposition 5.9 still holds for this new partition.

Define, for large  $m$ ,  $F_m := \bigcap_{i=0}^{m-1} \sigma^{-i} R$ . We will require the following continuity result.

**Proposition 5.10.** *Suppose that  $(\mu_n)_n$  is a sequence of ergodic probability measures converging on cylinders to an invariant measure  $\mu$ , where  $\mu(\Sigma) > 0$ . For every  $P \in \beta^m \setminus \{F_m\}$ , we have*

$$\lim_{n \rightarrow \infty} \mu_n(P) = \mu(P).$$

*Proof.* In order to prove the proposition we will need the following fact.

**Claim 1.** *Let  $(H_i)_i$  be a collection of cylinders and  $(p_i)_i$  a sequence of natural numbers. Then  $H_0 \cap \sigma^{-p_1} H_1 \cap \dots \cap \sigma^{-p_k} H_k$ , is either a finite collection of cylinders, or the empty set.*

*Proof.* We begin with the case  $k = 2$ , in other words, we will prove that if  $C$  and  $D$  are cylinders, then for every  $p \in \mathbb{N}$  the set  $C \cap \sigma^{-p} D$  is a finite collection of cylinders or the empty set. If the length of  $C$  is larger than or equal to  $p$  then  $C \cap \sigma^{-p} D$  is empty or a cylinder. If  $p$  is larger than the length of  $C$ , then we use that there are only finitely many admissible words of given length connecting two fixed symbols. More precisely, if  $C = [x_0, \dots, x_{h-1}]$  and  $D = [y_0, \dots, y_{t-1}]$ , then there are finitely many admissible words of length  $p - h + 2$  connecting  $x_{h-1}$  and  $y_0$ . We conclude that  $C \cap \sigma^{-p} D$  is a finite collection of cylinder or the empty set. The same argument gives us the proof of the claim for arbitrary  $k$ .  $\square$

Let  $P = S_0 \cap \sigma^{-1} S_1 \cap \dots \cap \sigma^{-(m-1)} S_{m-1}$ , where  $S_i \in \beta$  and  $P_k := \bigcap_{i=k}^{m-1} \sigma^{-(i-k)} S_i$ . Define  $B = B(P) := \{i \in \{0, \dots, m-1\} : S_i = R\}$ ,  $G = G(P) := \{0, \dots, m-1\} \setminus B$ , and  $k = k(P) := (\min G) - 1$ . By assumption we know that  $G \neq \emptyset$ . Let  $Q_0 = Q_0(P) := \bigcup_{i=0}^k \sigma^{-i} R$ ,  $Q_1 = Q_1(P) := \bigcap_{i \in G} \sigma^{-i} S_i$ , and  $Q_2 = Q_2(P) := \bigcap_{i \in B \cap (k, \infty)} \sigma^{-i} S_i$ . We will first consider the case  $k = -1$ , where  $Q_0 = \emptyset$ .

**Claim 2.** *Let  $P = \bigcap_{i=0}^{m-1} \sigma^{-i} S_i$ , where  $S_0 \in \{C^1, \dots, C^q\}$ . Then*

$$\lim_{n \rightarrow \infty} \mu_n(P) = \mu(P).$$

*Proof.* Since  $Q_1$  is the disjoint union of  $P = (Q_1 \cap Q_2)$  and  $(Q_1 \cap Q_2^c)$ , for every  $n \in \mathbb{N}$  we obtain that

$$\mu_n(P) = \mu_n(Q_1) - \mu_n(Q_1 \cap Q_2^c).$$

Observe that

$$Q_1 \cap Q_2^c = \left( \bigcap_{j \in G} \sigma^{-j} S_j \right) \cap \left( \bigcup_{i \in B} \sigma^{-i} R^c \right) = \bigcup_{i \in B} \left( \sigma^{-i} R^c \cap \bigcap_{j \in G} \sigma^{-j} S_j \right).$$

From Claim 1 we conclude that for every  $i \in B$  the sets  $Q_1$  and  $(\sigma^{-i} R^c \cap \bigcap_{j \in G} \sigma^{-j} S_j)$  are a finite union of cylinders or the empty set. Therefore,  $Q_1$  and  $Q_1 \cap Q_2^c$  are a finite union of cylinders, or the empty set. From this we immediately obtain that

$$\lim_{n \rightarrow \infty} \mu_n(P) = \lim_{n \rightarrow \infty} \mu_n(Q_1) - \lim_{n \rightarrow \infty} \mu_n(Q_1 \cap Q_2^c) = \mu(Q_1) - \mu(Q_1 \cap Q_2^c) = \mu(P),$$

which proves the claim.  $\square$

We now explain how to reduce the case  $k \geq 0$  to Claim 2. Observe that  $P = R \cap \sigma^{-1} P_1$ , therefore  $\sigma^{-1} P_1$  is the disjoint union between  $P$  and  $S_1 := (R^c \cap \sigma^{-1} P_1) = \bigcup_{i=1}^q (C^i \cap \sigma^{-1} P_1)$ . Thus,

$$\mu_n(P) = \mu_n(\sigma^{-1} P_1) - \mu_n(R^c \cap \sigma^{-1} P_1) = \mu_n(P_1) - \sum_{i=1}^q \mu_n(C^i \cap \sigma^{-1} P_1).$$

By Claim 2 we know that  $\lim_{n \rightarrow \infty} \mu_n(C^i \cap \sigma^{-1} P_1) = \mu(C^i \cap \sigma^{-1} P_1)$ . Therefore it suffices to prove that  $\lim_{n \rightarrow \infty} \mu_n(P_1) = \mu(P_1)$ . Applying the above argument  $k$  times we obtain that the original problem is reduced to  $\lim_{n \rightarrow \infty} \mu_n(P_{k+1}) = \mu(P_{k+1})$ . Since  $P_{k+1} = S_{k+1} \cap \sigma^{-1} P_{k+2}$ , where  $S_{k+1} \in \{C^1, \dots, C^q\}$ , we conclude the proof of the proposition by applying Claim 2.  $\square$

*Proof of Theorem 5.1.* We first consider the case in which not all the mass escapes, that is, we assume that  $\mu(\Sigma) > 0$ . Let  $\varepsilon_0 > 0$ . Choose  $m \in \mathbb{N}$  sufficiently large such that

$$h_{\frac{\mu}{|\mu|}}(\sigma) + \varepsilon_0 > \frac{1}{m} H_{\frac{\mu}{|\mu|}}(\beta^m) \quad , \quad 2 \frac{e^{-1}}{m} < \frac{\varepsilon}{2} \quad \text{and} \quad - \left( \frac{1}{m} \right) \log \mu(\Sigma) < \varepsilon_0.$$

Then

$$|\mu| h_{\frac{\mu}{|\mu|}}(\sigma) + 2\varepsilon_0 > \frac{1}{m} \sum_{P \in \beta^m} \mu(P) \log \mu(P).$$

It follows from Proposition 5.10 that

$$\lim_{n \rightarrow \infty} \sum_{Q \in \beta^n \setminus \{F_m\}} \mu_n(Q) \log \mu_n(Q) = \sum_{Q \in \beta^m \setminus \{F_m\}} \mu(Q) \log \mu(Q).$$

For sufficiently large  $n \in \mathbb{N}$  we have the inequality

$$|\mu| h_{\frac{\mu}{|\mu|}}(\sigma) + 3\varepsilon_0 \geq \frac{1}{m} H_{\mu_n}(\beta^m).$$

By Proposition 5.9, we have that

$$\begin{aligned} |\mu| h_{\frac{\mu}{|\mu|}}(\sigma) + 3\varepsilon_0 &> \frac{1}{m} H_{\mu_n}(\beta^m) \geq h_{\mu_n}(\sigma, \beta) \\ &\geq h_{\mu_n}(\sigma) - (\hat{\delta}_\infty(K) + \epsilon) \mu_n(B_{k,N}) - \frac{1}{k} (\hat{\delta}_\infty(K) + \epsilon), \end{aligned}$$

and therefore

$$\limsup_{n \rightarrow \infty} h_{\mu_n}(\sigma) \leq |\mu| h_{\frac{\mu}{|\mu|}}(\sigma) + (\hat{\delta}_\infty(K) + \epsilon) (1 - \mu(G_{k,N})) + \frac{1}{k} (\hat{\delta}_\infty(K) + \epsilon). \quad (5.7)$$

We stress that Proposition 5.9 can be applied for arbitrary  $k, N \in \mathbb{N}$  since  $K \subset \text{supp}(\mu_n)$  and therefore  $\mu_n(K(k, N)) > 0$ . Finally, letting  $k \rightarrow \infty$  and  $\epsilon \rightarrow 0$  we obtain the inequality

$$\limsup_{n \rightarrow \infty} h_{\mu_n}(\sigma) \leq |\mu| h_{\mu/|\mu|}(\sigma) + \left( 1 - \sup_{k,N} \mu(G_{k,N}) \right) \hat{\delta}_\infty(K).$$

Observe that  $Y \subset \bigcup_{k,N} G_{k,N}$ , therefore  $\mu(Y) \leq \mu \left( \bigcup_{k,N} G_{k,N} \right) = \sup_{k,N} \mu(G_{k,N})$ .

We conclude that

$$\limsup_{n \rightarrow \infty} h_{\mu_n}(\sigma) \leq |\mu| h_{\mu/|\mu|}(\sigma) + (1 - \mu(Y)) \hat{\delta}_\infty(K).$$

The case  $\mu(\Sigma) = 0$  follows directly from Proposition 5.9 since  $h_{\mu_n}(\sigma, \beta) \rightarrow 0$  and  $\mu_n(B_{k,N}) \rightarrow 1$  as  $n \rightarrow \infty$ .  $\square$

## 6. PROOF OF THEOREM 1.1

In this section we prove our main result. We start with a simple result we will need later.

**Lemma 6.1.** *Let  $(a_j)_j$  and  $(b_j)_j$  be sequences of natural numbers such that for every  $i \in \mathbb{N}_0$  we have  $a_0 = b_0$  and  $a_j \leq b_j$ . Then  $\hat{\delta}_\infty(K((b_j)_j)) \leq \hat{\delta}_\infty(K((a_j)_j))$ .*

*Proof.* Denote by  $K_1 := K((a_j)_j)$  and  $K_2 := K((b_j)_j)$ . Recall that associated to each compact set defined in this way there is a function  $i$  (see (5.1) for the definition). Denote the function  $i$  associated to  $K_1$  (resp.  $K_2$ ) by  $i_1$  (resp.  $i_2$ ). It follows from the hypothesis that  $K_1 \subset K_2$ . In particular we have that  $K_2^c \subset K_1^c$  and therefore  $T_n(K_2) \subset T_n(K_1)$  (see (5.2) for the definition of  $T$ ). Moreover, we have that

$$\widehat{T}_n(K_2) \subset \widehat{T}_n(K_1),$$

(see (5.3) for the definition of  $\widehat{T}$ ). Indeed, let  $x \in \widehat{T}_n(K_2)$ , we have that  $i_2(\sigma^k(x)) \leq n - k$ . In particular

$$(\sigma^k(x))_{i_2(\sigma^k(x))} > b_{i_2(\sigma^k(x))} \geq a_{i_2(\sigma^k(x))}.$$

We conclude that  $i_1(\sigma^k(x)) \leq i_2(\sigma^k(x)) \leq n - k$ , and therefore  $x \in \widehat{T}_n(K_1)$ . Thus  $\widehat{T}_n(K_2) \subset \widehat{T}_n(K_1)$ , which readily implies that for every  $n \in \mathbb{N}$  we have  $\widehat{z}_n(K_2) \leq \widehat{z}_n(K_1)$ .  $\square$

In the next lemma we establish a relation between the quantities  $\widehat{\delta}_\infty(K)$  and  $\delta_\infty(q)$ , which in turn is necessary to relate Theorem 5.1 with Theorem 1.1. As mentioned before, in the definition of  $\widehat{\delta}_\infty(K)$  we covered the sets  $\widehat{T}_n(K)$  (and not  $T_n(K)$  which may seem more natural) in order to ensure this result.

**Lemma 6.2.** *Let  $(\Sigma, \sigma)$  be a CMS satisfying the  $\mathcal{F}$ -property, and  $M, q \in \mathbb{N}$ . Then there exists a sequence of natural numbers  $(a_i)_i$  such that  $a_0 = q$ , and*

$$\widehat{\delta}_\infty(K) \leq \delta_\infty(M, q),$$

where  $K = K((a_i)_i)$

*Proof.* Let  $i \in \mathbb{N}$ . Since  $\Sigma$  satisfies the  $\mathcal{F}$ -property, there are finitely many cylinders of the form  $[x_0, \dots, x_n]$ , where  $x_0 \leq q$ ,  $x_n \leq q$ , and  $n \leq iM$ . Thus, only a finite collection of symbols from the alphabet are used in this collection of cylinders. Denote by  $r_i \in \mathbb{N}$  the largest of this collection of symbols. Inductively define  $(a_i)_i \subset \mathbb{N}$  so that:

$$a_{i+1} > a_i \quad \text{and} \quad a_i > r_i.$$

We now prove that the set  $K = K((a_i)_i)$  is such that  $\widehat{z}_n(K) \leq \delta_\infty(M, q)(n)$ , for every  $n \in \mathbb{N}$ . Recall that  $\widehat{z}_n(K)$  is the minimal number of cylinders of length  $(n+2)$  needed to cover  $\widehat{T}_n$ . Let  $x = (x_0, x_1, \dots) \in \widehat{T}_n$ ,

$$E := \{k \in \{0, \dots, n+1\} : x_k \leq a_0\},$$

and  $B := \{0, \dots, n+1\} \setminus E$ . For  $k \in E$  we define  $p_k := i(\sigma^k(x))$ . We emphasise that since  $k \in E$  then  $x_k \leq a_0$ , thus  $p_k = i(\sigma^k(x)) \geq 1$ . Let  $r \in E$  and observe that  $x_{p_r+r} = (\sigma^r(x))_{p_r} > a_{p_r}$ , where  $p_r \leq n - r$ . Because of the choice of  $a_{p_r}$ , there is no admissible word of length less or equal to  $p_r M$  connecting  $x_{p_r+r}$  and a symbol in the set  $\{0, 1, \dots, q\}$ . Since  $x_{n+1} \leq q$ , this means that we must have  $p_r+r+(p_r M) \leq n+1$ . Moreover, for every  $0 \leq m < p_r M$  we have that  $p_r+r+m \in B$ . In other words, the interval  $[r, r+p_r+p_r M)$  has at least  $p_r M$  elements in  $B$ , equivalently, at most  $p_r$  elements in  $E$ . Since this argument holds for every  $r \in E$  we conclude that  $M\#E \leq n+2$  and therefore

$$\#E \leq \frac{n+2}{M}. \tag{6.1}$$

From (6.1) it follows that every  $x \in \widehat{T}_n$  belongs to a cylinder of the form  $[x_0, \dots, x_{n+1}]$ , where  $x_0 \leq q$ ,  $x_{n+1} \leq q$  and

$$\#\{i \in \{0, 1, \dots, n+1\} : x_i \leq q\} \leq \frac{n+2}{M}.$$

This implies that  $\hat{z}_n(K) \leq z_n(M, q)$ , for every  $n \in \mathbb{N}$ . Therefore  $\hat{\delta}_\infty(K) \leq \delta_\infty(M, q)$ .  $\square$

Define  $\hat{\delta}_\infty(q) := \inf_{(a_i)_{i: a_0=q}} \hat{\delta}_\infty(K((a_i)_i))$ .

**Corollary 6.3.** *For every  $q \in \mathbb{N}$  we have  $\hat{\delta}_\infty(q) \leq \delta_\infty(q)$ .*

*Proof.* Combine Lemma 6.1 with Lemma 6.2.  $\square$

We now prove Theorem 1.1.

*Proof of Theorem 1.1.* Let  $(a_i)_i$  be a sequence of non-negative integers and  $K := K((a_i)_i)$  the corresponding compact set. We assume  $K$  large enough so that there exists a periodic measure  $\mu_p$  with  $\mu_p(K) > 0$ . We will prove that

$$\limsup_{n \rightarrow \infty} h_{\mu_n}(\sigma) \leq |\mu| h_{\mu/|\mu|}(\sigma) + (1 - \mu(K)) \hat{\delta}_\infty(K). \quad (6.2)$$

Let  $\mu'_n := (1 - \frac{1}{n})\mu_n + \frac{1}{n}\mu_p$ . Observe that for every  $n \in \mathbb{N}$  we have  $\mu'_n(K) > 0$ . It follows from Proposition 4.1 that there exists an ergodic measure  $\nu_n$  arbitrarily close in the weak\* topology to  $\mu'_n$  such that  $h_{\nu_n}(\sigma) > h_{\mu'_n}(\sigma) - \frac{1}{n}$ . In particular, we can assume that  $\nu_n(K(n, n)) > 0$  and that  $(\nu_n)_n$  converges on cylinders to  $\mu$ .

Let  $k, N \in \mathbb{N}$ . If  $n > \max\{k, N\}$  then  $K(n, n) \subset K(k, N)$ , therefore  $\nu_n(K(k, N)) > 0$ . It now follows from (5.7) that

$$\limsup_{n \rightarrow \infty} h_{\nu_n}(\sigma) \leq |\mu| h_{\mu/|\mu|}(\sigma) + (\hat{\delta}_\infty(K) + \epsilon)(1 - \mu(G_{k, N})) + \frac{1}{k}(\hat{\delta}_\infty(K) + \epsilon).$$

Letting  $k$  tend to infinity and  $\epsilon$  to zero we obtain

$$\limsup_{n \rightarrow \infty} h_{\nu_n}(\sigma) \leq |\mu| h_{\mu/|\mu|}(\sigma) + (1 - \mu(K)) \hat{\delta}_\infty(K).$$

Since  $h_{\nu_n}(\sigma) > h_{\mu'_n}(\sigma) - \frac{1}{n} = (1 - \frac{1}{n})h_{\mu_n} - \frac{1}{n}$ , then

$$\limsup_{n \rightarrow \infty} h_{\mu_n}(\sigma) \leq \limsup_{n \rightarrow \infty} h_{\nu_n}(\sigma),$$

from which (6.2) follows.

The argument above also holds for every set  $K' = K((b_i)_i)$ , where  $a_0 = b_0$  and  $a_i \leq b_i$ . Observe that  $\sup_{(b_i)_{i: b_0=a_0}} \mu(K((b_i)_i)) = \mu(K_{a_0})$ . Thus, it is a consequence of Corollary 6.3 that

$$\begin{aligned} \limsup_{n \rightarrow \infty} h_{\mu_n}(\sigma) &\leq |\mu| h_{\mu/|\mu|}(\sigma) + (1 - \mu(K_{a_0})) \hat{\delta}_\infty(a_0) \\ &\leq |\mu| h_{\mu/|\mu|}(\sigma) + (1 - \mu(K_{a_0})) \delta_\infty(a_0). \end{aligned}$$

Letting  $a_0$  tend to infinity concludes the proof of Theorem 1.1.  $\square$

## 7. VARIATIONAL PRINCIPLE FOR THE ENTROPIES AT INFINITY

In this section we prove Theorem 1.4. That is, we prove a variational principle at infinity: the measure theoretic entropy at infinity coincides with its topological counterpart.

For each pair  $(i, j) \in \mathbb{N}^2$  choose a non-empty cylinder  $w(i, j)$  of length  $\ell(i, j) + 1$  such that

$$w(i, j) := [i, \dots, j] = [(w(i, j)_0, \dots, w(i, j)_{\ell(i, j)})].$$

Let  $\varphi : \Sigma \rightarrow \mathbb{R}$  be a potential and define

$$Z_n(\varphi, a, b) := \sum_{x: \sigma^{n+\ell(b, a)}(x)=x} \exp(S_{n+\ell(b, a)}\varphi(x)) 1_{[a] \cap \sigma^{-n}w(b, a)}(x).$$

In the following lemma we show that the Gurevich pressure can be computed by means of the partition function  $Z_n(\varphi, a, b)$ ; this will be used in Lemma 7.3.

**Lemma 7.1.** *Let  $(\Sigma, \sigma)$  be a transitive CMS and  $\varphi : \Sigma \rightarrow \mathbb{R}$  a bounded potential with summable variations. Then for every pair  $(a, b) \in \mathbb{N}^2$  we have that*

$$P_G(\varphi) = \limsup_{n \rightarrow \infty} \frac{1}{n} \log Z_n(\varphi, a, b).$$

*Proof.* Let  $C = \|\varphi\|_0$  and  $D = \sum_{k=2}^{\infty} \text{var}_k(\varphi)$ . It follows from the definition of  $Z_n(\varphi, a, b)$  that

$$Z_{n+\ell(b, a)}(\varphi, a) = \sum_{x: \sigma^{n+\ell(b, a)}(x)=x} \exp(S_{n+\ell(b, a)}\varphi(x)) 1_{[a]}(x) \geq Z_n(\varphi, a, b).$$

In particular we obtain that

$$P_G(\varphi) = \limsup_{n \rightarrow \infty} \frac{1}{n} \log Z_n(\varphi, a) \geq \limsup_{n \rightarrow \infty} \frac{1}{n} \log Z_n(\varphi, a, b).$$

Let  $\mathbb{P}_n := w(a, b) \cap \sigma^{-n}w(b, a)$ . Note that

$$\begin{aligned} Z_n(\varphi, a, b) &\geq \sum_{x: \sigma^{n+\ell(b, a)}(x)=x} \exp(S_{n+\ell(b, a)}\varphi(x)) 1_{\mathbb{P}_n}(x) \\ &\geq e^{-(\ell(a, b) + \ell(b, a))C} \sum_{x: \sigma^{n+\ell(b, a)}(x)=x} \exp(S_{n-\ell(b, a)}\varphi(\sigma^{\ell(a, b)}x)) 1_{\mathbb{P}_n}(x). \end{aligned}$$

Observe that if  $x = (x_0, x_1, \dots) \in \mathbb{P}_n$ , then  $x_{\ell(b, a)} = x_n = b$ . Define the periodic point  $y(x) := \overline{x_{\ell(b, a)} \dots x_{n-1}}$ . The function  $y$  establishes a one-to-one correspondence between points in  $x \in \mathbb{P}_n$  such that  $\sigma^{n+\ell(b, a)}(x) = x$ , and periodic points of length  $n - \ell(b, a)$  in  $[b]$ . Moreover, note that if  $x \in \mathbb{P}_n$ , then

$$\left| S_{n-\ell(b, a)}(\varphi(\sigma^{\ell(a, b)}x)) - S_{n-\ell(b, a)}(\varphi(y(x))) \right| \leq D.$$

We conclude that

$$\begin{aligned} &\sum_{x: \sigma^{n+\ell(b, a)}(x)=x} \exp(S_{n-\ell(b, a)}\varphi(\sigma^{\ell(a, b)}x)) 1_{\mathbb{P}_n}(x) \geq \\ &e^{-D} \sum_{x: \sigma^{n-\ell(b, a)}(y)=y} \exp(S_{n-\ell(b, a)}\varphi(y)) 1_{[b]}(x). \end{aligned}$$

That is  $Z_n(\varphi, a, b) \geq e^{-(\ell(a,b)+\ell(b,a))C-D} Z_{n-\ell(b,a)}(\varphi, b)$  and therefore

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log Z_n(\varphi, a, b) \geq P_G(\varphi).$$

□

*Remark 7.2.* Note that in Lemma 7.1 the assumption  $\|\varphi\|_0 < \infty$  is too strong for what is required: it suffices to assume that for every  $n \in \mathbb{N}$  we have  $\sup_{x \in [n]} |\varphi(x)| < \infty$ .

We say that a point  $x \in \Sigma$  belongs to the set  $Per(q, M, n)$  if the following properties hold:

- (a)  $\sigma^n(x) = x$ .
- (b) If  $x \in [x_0, \dots, x_{n-1}]$ , then  $x_0 \leq q$ , and  $\#\{k \in \{0, \dots, n-1\} : x_k \leq q\} \leq \frac{n}{M}$ .

The following lemma is important in our proof of Theorem 1.4 as it will allow us to find a sequence of invariant probability measures which converges to the zero measure and entropies approach the topological entropy at infinity.

**Lemma 7.3.** *Let  $\varphi : \Sigma \rightarrow \mathbb{R}$  be a bounded potential of summable variations such that*

$$\lim_{n \rightarrow \infty} \sup_{x \in [n]} |\varphi(x)| = 0.$$

Then  $P_G(\varphi) \geq \delta_\infty$ .

*Proof.* For every  $\epsilon > 0$  there exists  $N_0 = N_0(\epsilon) \in \mathbb{N}$  such that  $\sup_{x \in [n]} |\varphi(x)| \leq \epsilon$ , for every  $n \geq N_0$ . By Lemma 7.1, for sufficiently large values of  $n \in \mathbb{N}$ , since  $Z_n(a, n)(\varphi) \leq \exp(nP_G(\varphi) + \epsilon)$  there exists  $N' = N'(N_0) \in \mathbb{N}$  such that

$$N' \exp(nP_G(\varphi) + \epsilon) \geq \sum_{(a,b) \in \{1, \dots, N_0\}^2} Z_n(\varphi, a, b).$$

That is,

$$P_G(\varphi) \geq \limsup_{n \rightarrow \infty} \frac{1}{n} \log \sum_{(a,b) \in \{1, \dots, N_0\}^2} Z_n(\varphi, a, b).$$

Define

$$\mathbb{T}_n(a, b) := \sum_{x \in Per(N_0, M, n + \ell(b, a))} \exp(S_{n + \ell(b, a)} \varphi(x)) 1_{[a] \cap \sigma^{-n} w(b, a)}(x),$$

and observe that  $Z_n(\varphi, a, b) \geq \mathbb{T}_n(a, b)$ . Recall that  $x \in Per(N_0, M, n + \ell(b, a))$  implies that

$$\#\{k \in \{0, \dots, n + \ell(b, a) - 1\} : x_k \leq N_0\} \leq \frac{n + \ell(b, a)}{M}.$$

It follows from the choice of  $N_0$  that

$$S_{n + \ell(b, a)} \varphi(x) \geq -(n + \ell(b, a))\epsilon - \frac{\|\varphi\|_0}{M}(n + \ell(b, a)).$$

In particular

$$\mathbb{T}_n(a, b) \geq \#\{Per(N_0, M, n + \ell(b, a)) \cap [a] \cap \sigma^{-n} w(b, a)\} e^{-(n + \ell(b, a))(\epsilon + \frac{\|\varphi\|_0}{M})}.$$

Denote by  $\mathcal{W}_n(a, b, N_0, M)$  the collection of cylinders of the form  $[x_0, \dots, x_n]$ , where  $x_0 = a$ ,  $x_n = b$ , and  $\#\{k \in \{0, \dots, n\} : x_k \leq N_0\} \leq \frac{n+1}{2M}$ . In order to estimate the number of these using periodic points, to each cylinder  $[x_0, \dots, x_n] \in \mathcal{W}_n(a, b, N_0, M)$  we associate the cylinder

$$D = [y_0, \dots, y_{n+\ell(b,a)}] = [x_0, \dots, x_n, (w(b, a))_1, \dots, w(b, a)_{\ell(b,a)}].$$

Observe that  $y_0 = a$ ,  $y_{n+\ell(b,a)} = a$ , and

$$\#\{k \in \{0, \dots, n + \ell(b, a)\} : y_k \leq N_0\} \leq \frac{n+1}{2M} + \ell(b, a).$$

For  $n \in \mathbb{N}$  sufficiently large we can assume that  $\frac{n+1}{2M} + \ell(b, a) \leq \frac{n+\ell(b,a)}{M}$ . In particular the periodic point associated to  $D$  belongs to  $Per(N_0, M, n + \ell(b, a)) \cap [a] \cap \sigma^{-n}w(b, a)$ . We conclude that

$$\#\mathcal{W}_n(a, b, N_0, M) \leq \#\{Per(N_0, M, n + \ell(b, a)) \cap [a] \cap \sigma^{-n}w(b, a)\}.$$

Observe that  $\sum_{(a,b) \in \{1, \dots, N_0\}^2} \#\mathcal{W}_n(a, b, N_0, M) = z_{n-1}(2M, N_0)$ , which implies

$$z_{n-1}(2M, N_0) \leq \sum_{(a,b) \in \{1, \dots, N_0\}^2} \#\{Per(N_0, M, n + \ell(b, a)) \cap [a] \cap \sigma^{-n}w(b, a)\}.$$

Hence, writing  $\ell_{N_0} := \max_{(a,b) \in \{1, \dots, N_0\}^2} \ell(b, a)$ , we obtain that

$$\sum_{(a,b) \in \{1, \dots, N_0\}^2} Z_n(\varphi, a, b) \geq \sum_{(a,b) \in \{1, \dots, N_0\}^2} \mathbb{T}_n(a, b) \geq z_{n-1}(2M, N_0) e^{-(n+\ell_{N_0})(\epsilon + \frac{\|\varphi\|_0}{M})},$$

and therefore

$$P_G(\varphi) \geq \limsup_{n \rightarrow \infty} \frac{1}{n} \log \sum_{(a,b) \in \{1, \dots, N_0\}^2} Z_n(\varphi, a, b) \geq \delta_\infty(2M, N_0) - \epsilon - \frac{\|\varphi\|_0}{M}.$$

Letting  $M \rightarrow \infty$  we obtain that  $P(\varphi) \geq \delta_\infty(N_0) - \epsilon$ . Choosing  $N_0$  sufficiently large we can make  $\epsilon$  arbitrarily small, to conclude that  $P_G(\varphi) \geq \delta_\infty$ .  $\square$

Recall that the measure theoretic entropy at infinity of a transitive CMS of finite entropy  $(\Sigma, \sigma)$  is defined by

$$h_\infty := \sup_{(\mu_n)_{n \rightarrow 0}} \limsup_{n \rightarrow \infty} h_{\mu_n}(\sigma),$$

where the supremum is taken over all sequences of invariant probability measures converging on cylinders to the zero measure. An immediate consequence of Theorem 1.1 is the following upper bound for the measure theoretic entropy at infinity of  $(\Sigma, \sigma)$ :

$$h_\infty \leq \delta_\infty \tag{7.1}$$

We will now prove that in fact equality holds. This is equivalent to the sharpness of the inequality in Theorem 1.1.

*Proof of Theorem 1.4.* As observed in (7.1), it suffices to prove the inequality  $\delta_\infty \leq h_\infty$ . Let  $\varphi : \Sigma \rightarrow \mathbb{R}$  be a bounded, strictly negative locally constant potential depending only on the first coordinate such that

$$\lim_{n \rightarrow \infty} \sup_{x \in [n]} |\varphi(x)| = 0.$$

By Lemma 7.3, for every  $t \in \mathbb{R}$  we have  $P(t\varphi) \geq \delta_\infty$ . Now consider a sequence of measures  $(\mu_n)_n$  such that

$$h_{\mu_n}(\sigma) + n \int \varphi d\mu_n > P(n\varphi) - \frac{1}{n}.$$

The existence of such a sequence of invariant probability measures is guaranteed by the variational principle. Then

$$h_{\mu_n}(\sigma) + n \int \varphi d\mu_n > \delta_\infty - \frac{1}{n}.$$

Since the potential  $\varphi$  is strictly negativity and bounded we conclude that the sequence  $(\mu_n)_n$  converges on cylinders to the zero measure. Since  $h_{\mu_n}(\sigma) \geq \delta_\infty - \frac{1}{n}$ ,

$$\limsup_{n \rightarrow \infty} h_{\mu_n}(\sigma) \geq \delta_\infty.$$

In particular,  $\delta_\infty \leq h_\infty$ . □

## 8. APPLICATIONS

In this section we discuss several consequences of Theorem 1.1. Among the consequences we obtain the upper semi-continuity of the entropy map, the entropy density of the space of ergodic measures, the stability of the measure of maximal entropy in the SPR case, existence of equilibrium states for potentials in  $C_0(\Sigma)$ , a relationship between the entropy at infinity and the dimension of the set of recurrent points that escape on average and a bound on the amount of mass that can escape for measures with large entropy.

**8.1. Upper semi-continuity of the entropy map.** Starting in the early 1970s with the work of Bowen [Bo1] many results describing the continuity properties of the entropy map have been obtained. More precisely, given a dynamical system  $T : X \rightarrow X$ , the map  $\mu \mapsto h_\mu(T)$  defined on the space  $\mathcal{M}(X, T)$  endowed with the weak\* topology is called *entropy map*. In general it is not continuous [Wa, p.184]. However, it was soon realised that that if  $X$  is compact and  $T$  expansive then the entropy map is upper-semi continuous [Wa, Theorem 8.2]. This result has been extended to a wide range of dynamical systems exhibiting weak forms of expansion or hyperbolicity, but always assuming the compactness of  $X$ . Indeed, there exist examples of expansive maps  $T$  defined on non-compact spaces for which the entropy map is not upper semi-continuous. We discuss some of them in this section (see Remark 8.2). We recently proved in [ITV, Corollary 1.2] that if  $(\Sigma, \sigma)$  is a finite entropy transitive CMS then the entropy map is upper semi-continuous when restricted to ergodic measures. The method of proof used in [ITV] does not seem to generalise to handle the non-ergodic case. However, the general case can be obtained directly as a corollary of Theorem 1.1.

**Theorem 8.1.** *Let  $(\Sigma, \sigma)$  be a transitive CMS of finite topological entropy and  $(\mu_n)_n$  a sequence of  $\sigma$ -invariant probability measures converging weak\* to  $\mu$ . Then*

$$\limsup_{n \rightarrow \infty} h_{\mu_n}(\sigma) \leq h_\mu(\sigma).$$

*That is, the entropy map is upper semicontinuous.*

The proof follows immediately from Theorem 1.1, the fact that  $|\mu| = 1$  and Lemma 2.5.

*Remark 8.2.* We now describe the situation in the infinite entropy case.

- (a) Without the finite entropy assumption, Theorem 8.1 is false, as we demonstrate here. If  $(\Sigma, \sigma)$  is a topologically transitive infinite entropy CMS then there exists a sequence  $(\nu_n)_n$  and  $\mu$  in  $\mathcal{M}(\Sigma, \sigma)$  such that  $h_\mu(\sigma) < \infty$ , and  $\lim_{n \rightarrow \infty} h_{\nu_n}(\sigma) = \infty$ . Let  $(\mu_n)_n$  be the sequence of invariant probability measures defined by

$$\mu_n := \left(1 - \frac{1}{\sqrt{h_{\nu_n}(\sigma)}}\right) \mu + \frac{1}{\sqrt{h_{\nu_n}(\sigma)}} \nu_n.$$

Notice  $\mu_n$  is well defined for large enough  $n$ . Then  $(\mu_n)_n$  converges weak\* to  $\mu$  and

$$h_\mu(\sigma) < \lim_{n \rightarrow \infty} h_{\mu_n}(\sigma) = \infty.$$

Therefore, the entropy map is not upper semi-continuous at any finite entropy measure.

- (b) Examples of sequences of ergodic measures with finite entropy uniformly bounded above converging weak\* to an ergodic measure (with finite entropy) in the full-shift on a countable alphabet, for which the entropy map is not upper-semi continuous can be found in [JMU, p.774] and [ITV, Remark 3.11].
- (c) The entropy map is trivially upper semi-continuous at any measure of infinite entropy.

We conclude this subsection with a consequence of Theorem 1.1 and Remark 8.2.

**Proposition 8.3.** *Let  $(\Sigma, \sigma)$  be a transitive CMS. Then  $h_{top}(\sigma)$  is finite if and only if  $\delta_\infty$  is finite.*

*Proof.* We only need to prove that if  $\delta_\infty$  is finite, then  $h_{top}(\sigma)$  is finite; the other direction follows directly from the inequality  $\delta_\infty \leq h_{top}(\sigma)$ .

First assume that  $(\Sigma, \sigma)$  does not satisfy the  $\mathcal{F}$ -property. It follows directly from the definition of  $\delta_\infty$  that in this situation we have  $\delta_\infty = \infty$ . As mentioned above there is nothing to prove in this case.

Now assume that  $(\Sigma, \sigma)$  satisfies the  $\mathcal{F}$ -property. In the proof of Theorem 1.1 we did not use the fact that the topological entropy of  $(\Sigma, \sigma)$  is finite, we only used that our CMS has the  $\mathcal{F}$ -property and that  $\delta_\infty$  is finite—those follow trivially under the finite entropy assumption. The  $\mathcal{F}$ -property is crucially used in Proposition 5.10 and Lemma 6.2. If  $\delta_\infty$  is finite, then Theorem 1.1 implies that the entropy map is upper semi-continuous, which would contradict Remark 8.2 if  $h_{top}(\sigma)$  is infinite. We conclude that the topological entropy of  $(\Sigma, \sigma)$  is finite.  $\square$

**8.2. Suspension flows.** Let  $(\Sigma, \sigma)$  be a transitive, finite entropy CMS and  $\tau : \Sigma \rightarrow \mathbb{R}^+$  a potential bounded away from zero. Let

$$Y := \{(x, t) \in \Sigma \times \mathbb{R} : 0 \leq t \leq \tau(x)\},$$

with the points  $(x, \tau(x))$  and  $(\sigma(x), 0)$  identified for each  $x \in \Sigma$ . The *suspension flow* over  $\Sigma$  with *roof function*  $\tau$  is the semi-flow  $\Phi = (\varphi_t)_{t \in \mathbb{R}_{\geq 0}}$  on  $Y$  defined by  $\varphi_t(x, s) = (x, s + t)$  whenever  $s + t \in [0, \tau(x)]$ . Denote by  $\mathcal{M}(Y, \Phi)$  the space of flow invariant probability measures. In this section we prove that in this continuous time, non-compact setting again the entropy map is upper semi-continuous. This

generalises [ITV, Proposition 5.2] in which upper semi-continuity of the entropy map was proven for ergodic measures. Let

$$\mathcal{M}_\sigma(\tau) := \left\{ \mu \in \mathcal{M}_\sigma : \int \tau d\mu < \infty \right\}. \quad (8.1)$$

A result by Ambrose and Kakutani [AK] implies that the map  $M: \mathcal{M}_\sigma \rightarrow \mathcal{M}_\Phi$ , defined by

$$M(\mu) = \frac{(\mu \times \text{Leb})|_Y}{(\mu \times \text{Leb})(Y)},$$

where  $\text{Leb}$  is the one-dimensional Lebesgue measure, is a bijection. The following result proved in [ITV, Lemma 5.1] describes the relation between weak\* convergence in  $\mathcal{M}(Y, \Phi)$  with that in  $\mathcal{M}(\Sigma, \sigma)$ .

**Lemma 8.4.** *Let  $(\nu_n), \nu \in \mathcal{M}(Y, \Phi)$  be flow invariant probability measures such that*

$$\nu_n = \frac{\mu_n \times \text{Leb}}{\int \tau d\mu_n} \quad \text{and} \quad \nu = \frac{\mu \times \text{Leb}}{\int \tau d\mu}$$

where  $(\mu_n)_n, \mu \in \mathcal{M}(\Sigma, \sigma)$ . If the sequence  $(\nu_n)_n$  converges weak\* to  $\nu$  then  $(\mu_n)_n$  converges weak\* to  $\mu$  and  $\lim_{n \rightarrow \infty} \int \tau d\mu_n = \int \tau d\mu$ .

**Proposition 8.5.** *Let  $(\Sigma, \sigma)$  be a transitive CMS of finite topological entropy. Let  $\tau$  be a potential bounded away from zero and  $(Y, \Phi)$  the suspension flow of  $(\Sigma, \sigma)$  with roof function  $\tau$ . Then the entropy map of  $(Y, \Phi)$  is upper semi-continuous.*

The proof directly follows from Abramov's formula [Ab], Lemma 8.4 and Theorem 1.1. Because of the similarities between the geodesic flow and the suspension flow over a Markov shift it is reasonable to expect that, under suitable assumptions on the roof function  $\tau$ , the suspension flow also satisfies an entropy inequality like Theorem 1.1. This is in fact the case and will be discussed in [Vel2]. The space of invariant measures for the suspension flow was already investigated and described in [IV, Section 6].

**8.3. Entropy density of ergodic measures.** The structure of the space of invariant measures for finite entropy (non-compact) CMS was studied in [IV]. In this non-compact setting it is well known that the space of ergodic measures is still dense in  $\mathcal{M}(\Sigma, \sigma)$  (see [CSc, Section 6]). A natural question is whether the approximation by ergodic measures can be arranged so that the corresponding entropies also converge. If this is the case we say that the set of ergodic measures is *entropy dense*. More precisely,

**Definition 8.6.** A subset  $\mathcal{L} \subset \mathcal{M}(\Sigma, \sigma)$  is *entropy dense* if for every measure  $\mu \in \mathcal{M}(\Sigma, \sigma)$  there exists a sequence  $(\mu_n)_n$  in  $\mathcal{L}$  such that

- (a)  $(\mu_n)_n$  converges to  $\mu$  in the weak\* topology.
- (b)  $\lim_{n \rightarrow \infty} h_{\mu_n}(\sigma) = h_\mu(\sigma)$ .

Results proving that certain classes of measures are entropy dense have been obtained for different dynamical systems defined on compact spaces by Katok [Ka], Orey [Or], Föllmer and Orey [FO], Eizenberg, Kifer and Weiss [EKW] and by Gorodetski and Pesin [GP] among others. In this section we prove, for the non-compact setting of finite entropy CMS, that the set of ergodic measures  $\mathcal{E}(\Sigma, \sigma)$  is entropy dense.

**Theorem 8.7.** *Let  $(\Sigma, \sigma)$  be a finite entropy, transitive CMS and  $\mu \in \mathcal{M}(\Sigma, \sigma)$ . Then there exists a sequence  $(\mu_n)_n$  of ergodic measures such that  $(\mu_n)_n$  converges to  $\mu$  in the weak\* topology and  $\lim_{n \rightarrow \infty} h_{\mu_n}(\sigma) = h_\mu(\sigma)$ , i.e.,  $\mathcal{E}(\Sigma, \sigma)$  is entropy dense. Moreover, it is possible to choose the sequence so that each  $\mu_n$  has compact support.*

The proof of this result directly follows combining Theorem 8.1, where the upper semi-continuity of the entropy map is proved, and Proposition 4.1, where we proved a weak form of entropy density of the set of ergodic measures. Note that the entropy density property of ergodic measures is an important tool in proving large deviations principles via the orbit-gluing technique (see, for example, [EKW] and [FO]).

**8.4. Points that escape on average.** In this section we relate the Hausdorff dimension of the set of recurrent points that escape on average with the entropy at infinity of  $(\Sigma, \sigma)$ . Recall we have fixed an identification of the alphabet of  $(\Sigma, \sigma)$  with  $\mathbb{N}$ .

**Definition 8.8.** Let  $(\Sigma, \sigma)$  be a CMS, the set of points that *escape on average* is defined by

$$E := \left\{ x \in \Sigma : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} 1_{[a]}(\sigma^i x) = 0, \text{ for every } a \in \mathbb{N} \right\}.$$

We say that  $x \in \Sigma$  is a *recurrent point* if there exists an increasing sequence  $(n_k)_k$  such that  $\lim_{k \rightarrow \infty} \sigma^{n_k}(x) = x$ . The set of recurrent points is denoted by  $\mathcal{R}$ .

A version of the set  $E$  has been considered in the context of homogeneous dynamics. Interest in that set stems from work of Dani [D] in the mid 1980s who proved that singular matrices are in one-to-one correspondence with certain divergent orbits of one parameter diagonal groups on the space of lattices. For example, Einsiedler and Kadyrov [EK, Corollary 1.7] computed the Hausdorff dimension of that set in the setting of  $SL_3(\mathbb{Z}) \backslash SL_3(\mathbb{R})$ . In the context of unimodular  $(n+m)$ -lattices an upper bound for the Hausdorff dimension of the set of points that escape on average has been obtained in [KKLM, Theorem 1.1]. More recently, for the Teichmüller geodesic flow, in [AAEKMU, Theorem 1.8] the authors prove an upper bound for the Hausdorff dimension of directions in which Teichmüller geodesics escape on average in a stratum. In all the above mentioned work, either explicitly or not, the bounds are related to the entropy at infinity of the system. Our next result establishes an analogous result for CMS. In this case the upper bound is the entropy at infinity divided by  $\log 2$ . This latter constant comes from the metric we consider in the space (see (2.1)) and can be thought of as the Lyapunov exponent of the system.

**Theorem 8.9.** *Let  $(\Sigma, \sigma)$  be a finite entropy transitive CMS. Then*

$$\dim_H(E \cap \mathcal{R}) \leq \frac{\delta_\infty}{\log 2}$$

where  $\dim_H$  denotes the Hausdorff dimension with respect to the metric (2.1).

Before initiating the proof of Theorem 8.9 let us set up some notation. Given natural numbers  $a, b, q, m$  and  $N$  we define  $S_{a,b}^q(N, m)$  as the collection of cylinders of the form  $[x_0, \dots, x_{L-1}]$ , where  $L \geq Nm$ ,  $x_0 = a, x_{L-1} = b$ , and the number of

indices  $i \in \{0, \dots, L-1\}$  such that  $x_i \leq q$  is exactly  $N$ . It will be convenient to define

$$H_{a,b}^q(n, m) := \bigcup_{N \geq n} S_{a,b}^q(N, m).$$

Finally define

$$\mathcal{L}_b := \{x \in \Sigma : \exists (n_k)_k \text{ strictly increasing such that } \sigma^{n_k}(x) \in [b], \forall k \in \mathbb{N}\},$$

and  $\mathcal{L} = \bigcup_{b \in \mathbb{N}} \mathcal{L}_b$ .

*Remark 8.10.* Let  $a, b, q$  and  $m$  be natural numbers. Assume that  $q \geq b$ . Note that if  $x \in (E \cap \mathcal{L}_b \cap [a])$ , then there exists  $s_0 \in \mathbb{N}$  such that

$$\#\{i \in \{0, \dots, s-1\} : x_i \leq q\} \leq \frac{s}{m},$$

for every  $s \geq s_0$ . Moreover, there exists an increasing sequence  $(n_k)_k$  such that  $x_{n_k} = b$ . Define  $\mathcal{T}_k(x) = \#\{i \in \{0, \dots, n_k-1\} : x_i \leq q\}$ . Since  $q \geq b$  we get that  $\mathcal{T}_k(x) \geq k$ . Observe that if  $n_k \geq s_0$ , then

$$m\mathcal{T}_k(x) = m\#\{i \in \{0, \dots, n_k-1\} : x_i \leq q\} \leq n_k.$$

We conclude that

$$[x_0, \dots, x_{n_k-1}] \in S_{a,b}^q(\mathcal{T}_k(x), m) \subset \bigcup_{p \geq k} S_{a,b}^q(p, m) = H_{a,b}^q(k, m).$$

This gives us the inclusion

$$(E \cap \mathcal{L}_b \cap [a]) \subset \bigcup_{C \in H_{a,b}^q(k, m)} C, \quad (8.2)$$

for every  $k \in \mathbb{N}$ .

*Proof of Theorem 8.9.* First observe that  $(E \cap \mathcal{R}) \subset \bigcup_{b \in \mathbb{N}} (E \cap \mathcal{L}_b \cap [b])$ . In particular it suffices to prove that  $\dim_H(E \cap \mathcal{L}_b \cap [a]) \leq \delta_\infty / \log 2$ , for every pair of natural numbers  $a$  and  $b$ . Fix  $t > \delta_\infty / \log 2$ . Recall that  $\delta_\infty = \inf_{m,q} \delta_\infty(m, q)$  (see equation (2.4)). Choose  $m$  and  $q$  large enough so that  $t > \delta_\infty(q, m) / \log 2$ , and that  $q \geq \max\{a, b\}$ . Observe that we are now in the same setup as in Remark 8.10.

In order to estimate the Hausdorff dimension of  $E \cap \mathcal{L}_b \cap [a]$  we will use the covering given by (8.2). Thus, it is enough to bound  $\sum_{C \in H_{a,b}^q(k, m)} \text{diam}(C)^t$ . First observe that since  $q \geq \max\{a, b\}$ , a cylinder  $C \in H_{a,b}^q(k, m)$  has length  $\ell(C) \geq k$ . Recall that  $\text{diam}(C) \leq 2^{-\ell(C)} = e^{-(\log 2)\ell(C)}$ . Therefore, as  $k \in \mathbb{N}$  increases the diameter of the covering given by (8.2) converges to zero. Now observe that

$$\begin{aligned} \sum_{C \in H_{a,b}^q(k, m)} \text{diam}(C)^t &\leq \sum_{C \in H_{a,b}^q(k, m)} e^{-t(\log 2)\ell(C)} \\ &= \sum_{l \geq k} e^{-t(\log 2)l} \#\{C : C \in H_{a,b}^q(k, m) \text{ and } \ell(C) = l\} \\ &\leq \sum_{l \geq k} e^{-t(\log 2)l} z_{l-2}(m, q). \end{aligned}$$

In the last inequality we used that

$$\#\{C \in H_{a,b}^q(k, m) \text{ and } \ell(C) = l\} \leq z_{l-2}(m, q).$$

Indeed, if  $C \in H_{a,b}^q(k, m)$  and  $\ell(C) = \ell$ , then  $C$  is a cylinder of the form  $[x_0, \dots, x_{\ell-1}]$  where  $x_0 = a$ ,  $x_{\ell-1} = b$ , and

$$\#\{i \in \{0, \dots, \ell-1\} : x_i \leq q\} = k \leq \frac{\ell}{m}.$$

Since  $\max\{a, b\} \leq q$  we conclude that  $C$  is one of the cylinders counted in the definition of  $z_{\ell-2}(m, q)$  (see Definition 1.2).

By the definition of  $\delta_\infty(m, q)$  the series  $Z(s) := \sum_{\ell=2}^{\infty} e^{-s\ell} z_{\ell-2}(m, q)$  is convergent for  $s > \delta_\infty(m, q)$ . In particular since  $t \log 2 > \delta_\infty(m, q)$  we have that  $Z(t \log 2)$  is finite. This implies that the tail of  $Z(t \log 2)$  converges to zero. We conclude that  $\sum_{C \in H_{a,b}^q(k, m)} \text{diam}(C)^t$  goes to zero as  $k \rightarrow \infty$ . This implies that  $\dim_H(E \cap \mathcal{L}_b \cap [a]) \leq t$ , but  $t$  was an arbitrary number larger than  $\delta_\infty / \log 2$ .  $\square$

*Remark 8.11.* It is proved in [I, Theorem 3.1] that if  $(\Sigma, \sigma)$  is a transitive CMS with finite topological entropy, then  $\dim_H(\mathcal{R}) = h_{\text{top}}(\sigma) / \log 2$ . In particular if  $(\Sigma, \sigma)$  is SPR, then  $\dim_H(E \cap \mathcal{R}) < \dim_H(\mathcal{R})$ .

**8.5. Measures of maximal entropy.** An invariant measure  $\mu \in \mathcal{M}(\Sigma, \sigma)$  is called a *measure of maximal entropy* if  $h_\mu(\sigma) = h_{\text{top}}(\sigma)$ . It follows from work by Gurevich [Gu1, Gu2] that if  $h_{\text{top}}(\sigma) < \infty$  then there exists at most one measure of maximal entropy. Note that a direct consequence of the variational principle (see [Gu2] or Theorem 2.3) is that there exists a sequence of invariant probability measures  $(\mu_n)_n$  such that  $\lim_{n \rightarrow \infty} h_{\mu_n}(\sigma) = h_{\text{top}}(\sigma)$ . Moreover, if the sequence has a weak\* accumulation point  $\mu$  then it follows from the upper semi-continuity of the entropy map, see Theorem 8.1, that  $h_\mu(\sigma) = h_{\text{top}}(\sigma)$ . Since the space  $\mathcal{M}(\Sigma, \sigma)$  is not compact there are cases in which the sequence  $(\mu_n)_n$  does not have an accumulation point. In fact, there exist transitive finite entropy CMS that do not have measures of maximal entropy (see [R2] for a wealth of explicit examples). Our next result follows directly from Theorem 1.1 and Theorem 2.7. Recall that  $(\Sigma, \sigma)$  is SPR if and only if  $\delta_\infty < h_{\text{top}}(\sigma)$  (see Proposition 2.20).

**Theorem 8.12.** *Let  $(\Sigma, \sigma)$  be a SPR CMS and  $(\mu_n)_n$  a sequence of  $\sigma$ -invariant probability measures such that*

$$\lim_{n \rightarrow \infty} h_{\mu_n}(\sigma) = h_{\text{top}}(\sigma).$$

*Then the sequence  $(\mu_n)_n$  converges in the weak\* topology to the unique measure of maximal entropy.*

*Proof.* Note that the inequality  $\delta_\infty < h_{\text{top}}(\sigma)$  immediately implies that  $(\Sigma, \sigma)$  has finite topological entropy (see Proposition 8.3). Since  $\mathcal{M}_{\leq 1}(\Sigma, \sigma)$  is compact (see Theorem 2.7) there exists a subsequence  $(\mu_{n_k})_k$  which converges on cylinders to  $\mu \in \mathcal{M}_{\leq 1}(\Sigma, \sigma)$ . It follows directly from Theorem 1.1 that

$$h_{\text{top}}(\sigma) = \limsup_{k \rightarrow \infty} h_{\mu_{n_k}}(\sigma) \leq |\mu| h_{\mu/|\mu|}(\sigma) + (1 - |\mu|) \delta_\infty.$$

Recall that  $\delta_\infty < h_{\text{top}}(\sigma)$ . If  $|\mu| < 1$  then the right hand side of the equation is a convex combination of numbers, one of each is strictly smaller than  $h_{\text{top}}(\sigma)$ . Since this is not possible we have that  $|\mu| = 1$ . In particular

$$h_{\text{top}}(\sigma) \leq h_\mu(\sigma).$$

That is,  $\mu$  is a measure of maximal entropy. We conclude that  $(\Sigma, \sigma)$  has a measure of maximal entropy. The same argument holds for every subsequence of  $(\mu_n)_n$ ,

this implies that the entire sequence  $(\mu_n)_n$  converges in the weak\* topology to the unique measure of maximal entropy.  $\square$

In fact Theorem 1.1 also gives a complete description of non strongly positive recurrence, as follows. Some of these results were originally proved in [GS, Theorem 6.3] by different methods.

**Theorem 8.13.** *Let  $(\Sigma, \sigma)$  be a transitive CMS of finite entropy.*

- (a) *Suppose  $(\Sigma, \sigma)$  does not admit a measure of maximal entropy. Let  $(\mu_n)_n$  be a sequence of  $\sigma$ -invariant probability measures such that  $\lim_{n \rightarrow \infty} h_{\mu_n}(\sigma) = h_{top}(\sigma)$ . Then  $(\mu_n)_n$  converges on cylinders to the zero measure and  $\delta_\infty = h_{top}(\sigma)$ .*
- (b) *Suppose that  $(\Sigma, \sigma)$  is positive recurrent, but  $h_{top}(\sigma) = \delta_\infty$ . Let  $(\mu_n)_n$  be a sequence of  $\sigma$ -invariant probability measures such that  $\lim_{n \rightarrow \infty} h_{\mu_n}(\sigma) = h_{top}(\sigma)$ . Then the accumulation points of  $(\mu_n)_n$  lie in the set  $\{\lambda\mu_{max} : \lambda \in [0, 1]\}$ , where  $\mu_{max}$  is the measure of maximal entropy. Moreover, every measure in  $\{\lambda\mu_{max} : \lambda \in [0, 1]\}$  can be realised as such limit.*

*Proof.* Note that part (a) directly follows from Theorem 1.1. Indeed, if a sequence  $(\mu_n)_n$  with  $\lim_{n \rightarrow \infty} h_{\mu_n}(\sigma) = h_{top}(\sigma)$  converges in cylinder to a measure  $\mu \in \mathcal{M}_{\leq 1}(\Sigma, \sigma)$  different from the zero measure then  $\mu/|\mu|$  would be a measure of maximal entropy. This argument also gives us the first part of (b), that is, the accumulation points of  $(\mu_n)_n$  lie in  $\{\lambda\mu_{max} : \lambda \in [0, 1]\}$ . As for the second part of (b), by Theorem 1.4 there exists a sequence  $(\mu_n)_n$  in  $\mathcal{M}(\Sigma, \sigma)$  with  $\lim_{n \rightarrow \infty} h_{\mu_n}(\sigma) = h_{top}(\sigma)$  such that  $(\mu_n)_n$  converges on cylinders to the zero measure. Since there exist a measure of maximal entropy  $\nu$  we have that for every  $\lambda \in [0, 1]$  the sequence  $\rho_n := \lambda\nu + (1 - \lambda)\mu_n$  converges on cylinders to  $\lambda\nu$  and  $\lim_{n \rightarrow \infty} h_{\rho_n}(\sigma) = h_{top}(\sigma)$ .  $\square$

**8.6. Existence of equilibrium states.** In this section we will always assume that  $(\Sigma, \sigma)$  is a transitive CMS with finite entropy. In Section 2.4 we described the thermodynamic formalism developed by Sarig in the setting of CMS and functions (potentials) of summable variations. It turns out that the same methods can be extended and thermodynamic formalism can be developed for functions with weaker regularity assumptions (for example functions satisfying the Walters condition [S4]). However, these methods can not be extended much further. In this section we propose an alternative definition of pressure that generalises the Gurevich pressure to the space of functions  $C_0(\Sigma)$  (see Definition 2.6). We stress that these functions are just uniformly continuous. Making use of Theorem 1.1 we can ensure the existence of equilibrium states.

The following result is a direct consequence of Theorem 1.1 and the continuity of the map  $\mu \mapsto \int F d\mu$ , when  $F \in C_0(\Sigma)$  and  $\mu$  ranges in  $\mathcal{M}_{\leq 1}(\Sigma, \sigma)$  endowed with the cylinder topology.

**Theorem 8.14.** *Let  $(\Sigma, \sigma)$  be a transitive CMS with finite entropy and  $F \in C_0(\Sigma)$ . Let  $(\mu_n)_n$  be a sequence in  $\mathcal{M}(\Sigma, \sigma)$  converging on cylinders to  $\lambda\mu$ , where  $\lambda \in [0, 1]$  and  $\mu \in \mathcal{M}(\Sigma, \sigma)$ . Then*

$$\limsup_{n \rightarrow \infty} \left( h_{\mu_n}(\sigma) + \int F d\mu_n \right) \leq \lambda \left( h_\mu(\sigma) + \int F d\mu \right) + (1 - \lambda)\delta_\infty.$$

For a continuous, bounded potential  $F$  define the (*variational*) *pressure* of  $F$  by

$$P_{var}(F) := \sup_{\mu \in \mathcal{M}(\Sigma, \sigma)} \left( h_{\mu}(\sigma) + \int F d\mu \right).$$

A measure  $\mu$  is an equilibrium state for  $F$  if  $P_{var}(F) = h_{\mu}(\sigma) + \int F d\mu$ . Recall that since  $F$  needs not to be of summable variations then the classifications of potentials (see Definition 2.10) and the uniqueness of equilibrium states (Theorem 2.12) do not necessarily hold.

Note that if  $F \in C_0(\Sigma)$ , then  $P_{var}(F) \geq \delta_{\infty}$ . Indeed, let  $(\mu_n)_n$  be a sequence of measures in  $\mathcal{M}(\Sigma, \sigma)$  converging on cylinders to the zero measure and such that  $\lim_{n \rightarrow \infty} h_{\mu_n}(\sigma) = \delta_{\infty}$ . Since  $F \in C_0(\Sigma)$ , then  $\lim_{n \rightarrow \infty} \int F d\mu_n = 0$ . We conclude that

$$P_{var}(F) \geq \limsup_{n \rightarrow \infty} \left( h_{\mu_n}(\sigma) + \int F d\mu_n \right) = \delta_{\infty}.$$

Our next result follows directly from Theorem 8.14 and Theorem 2.7, as Theorem 8.12 follows from Theorem 1.1 and Theorem 2.7.

**Theorem 8.15.** *Let  $(\Sigma, \sigma)$  be a transitive CMS with finite entropy and  $F \in C_0(\Sigma)$ . Assume that  $P_{var}(F) > \delta_{\infty}$ . Then there exists an equilibrium state for  $F$ . Moreover, if  $(\mu_n)_n$  is a sequence in  $\mathcal{M}(\Sigma, \sigma)$  such that*

$$\lim_{n \rightarrow \infty} \left( h_{\mu_n}(\sigma) + \int F d\mu_n \right) = P_{var}(F),$$

*then every limiting measure of  $(\mu_n)_n$  is an equilibrium state of  $F$ .*

In Theorem 8.15, if we further assume that  $F$  has summable variations, then the sequence  $(\mu_n)_n$  converges in the weak\* topology to the unique equilibrium state of  $F$ . For the description of the pressure map  $t \mapsto P_{var}(tF)$  we refer the reader to [RV, Theorem 5.7].

**8.7. Entropy and escape of mass.** In this subsection we show that for a SPR CMS  $(\Sigma, \sigma)$  it is possible to bound the escape of mass of sequences of measures with sufficiently large entropy. In the setting of homogenous flows an analogous result was proven in [EKP, Corollary of Theorem A].

**Theorem 8.16.** *Let  $(\Sigma, \sigma)$  be a SPR CMS. Let  $(\mu_n)_n$  be a sequence in  $\mathcal{M}(\Sigma, \sigma)$  such that  $h_{\mu_n}(\sigma) \geq c$ , for every  $n \in \mathbb{N}$ , and  $c \in (\delta_{\infty}, h_{top}(\sigma))$ . Then every limiting measure  $\mu$  of  $(\mu_n)_n$  with respect to the cylinder topology satisfies*

$$\mu(\Sigma) \geq \frac{c - \delta_{\infty}}{h_{top}(\sigma) - \delta_{\infty}}.$$

*Proof.* From Theorem 1.1 we have that

$$c \leq \limsup_{n \rightarrow \infty} h_{\mu_n}(\sigma) \leq \mu(\Sigma) h_{\mu/|\mu|}(\sigma) + (1 - \mu(\Sigma)) \delta_{\infty} \leq \mu(\Sigma) (h_{top}(\sigma) - \delta_{\infty}) + \delta_{\infty}.$$

The result then follows. □

## REFERENCES

- [Ab] L. Abramov, *On the entropy of a flow*. Dokl. Akad. Nauk SSSR 128 (1959), 873–875.  
 [AAEKMU] H. Al-Saqban, P. Apisa, A. Erchenko, O. Khalil, S. Mirzadeh, C. Uyanik, *Exceptional directions for the Teichmüller geodesic flow and Hausdorff dimension*. to appear in J. Eur. Math. Soc. (JEMS).

- [AK] W. Ambrose, S. Kakutani, *Structure and continuity of measurable flows*. Duke Math. J. 9 (1942), 25–42.
- [Bi] P. Billingsley, *Convergence of probability measures*. Second edition. Wiley Series in Probability and Statistics: Probability and Statistics. A Wiley-Interscience Publication. John Wiley and Sons, Inc., New York, 1999. x+277 pp.
- [B] V.I. Bogachev, *Measure theory. Vol. I, II*. Springer-Verlag, Berlin, 2007. Vol. I: xviii+500 pp., Vol. II: xiv+575 pp.
- [Bo1] R. Bowen, *Entropy-expansive maps*. Trans. Amer. Math. Soc. 164 (1972), 323–331.
- [Bo2] R. Bowen, *Topological entropy for noncompact sets*. Trans. Amer. Math. Soc. 184 (1973), 125–136.
- [Bo3] R. Bowen, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*. Lecture Notes in Mathematics, Vol. 470. Springer-Verlag, Berlin-New York, 1975. i+108 pp.
- [BBG] M. Boyle, J. Buzzi, R. Gómez, *Almost isomorphism for countable state Markov shifts*. J. Reine Angew. Math. 592 (2006), 23–47.
- [BS] J. Buzzi, O. Sarig, *Uniqueness of equilibrium measures for countable Markov shifts and multidimensional piecewise expanding maps*. Ergodic Theory Dynam. Systems 23 (2003), no. 5, 1383–1400.
- [B1] J. Buzzi, *Puzzles of quasi-finite type, zeta functions and symbolic dynamics for multi-dimensional maps*. Annales de l’Institut Fourier 60 (2010), 801–852.
- [B2] J. Buzzi, *Bowen factors, their degree, and codings of surface diffeomorphisms*. arXiv:1807.04017
- [BCS] J. Buzzi, S. Crovisier, O. Sarig, *Measures of maximal entropy for surface diffeomorphisms*. ArXiv:1811.02240
- [CS] V. Cyr, O. Sarig, *Spectral Gap and Transience for Ruelle Operators on Countable Markov Shifts*. Comm. Math. Phys. 292 (2009), no. 3, 637–666.
- [CSc] Y. Coudène, B. Schapira, *Generic measures for hyperbolic flows on non-compact spaces*. Israel J. Math. 179 (2010), 157–172.
- [D] S.G. Dani, *Divergent trajectories of flows on homogeneous spaces and Diophantine approximation*. J. Reine Angew. Math. 359 (1985), 55–89.
- [Di] E. I. Dinaburg, *A correlation between topological entropy and metric entropy*. Dokl. Akad. Nauk SSSR, 190 (1970), 19–22.
- [ELMV] M. Einsiedler, E. Lindenstrauss, P. Michel, A. Venkatesh, *The distribution of closed geodesics on the modular surface, and Duke’s theorem*. Enseign. Math. (2) 58 (2012), no. 3–4, 249–313.
- [EK] M. Einsiedler, S. Kadyrov, *Entropy and escape of mass for  $SL_3(\mathbb{Z}) \backslash SL_3(\mathbb{R})$* . Israel J. Math. 190 (2012), 253–288.
- [EKP] M. Einsiedler, S. Kadyrov, A. D. Pohl, *Escape of mass and entropy for diagonal flows in real rank one situations*, Israel J. Math. 210 (2015), no. 1, 245–295.
- [EKW] A. Eizenberg, Y. Kifer, B. Weiss, *Large deviations for  $Z^d$ -actions*. Comm. Math. Phys. 164 (1994), no. 3, 433–454.
- [FO] H. Föllmer, S. Orey, *Large deviations for the empirical field of a Gibbs measure*. Ann. Probab. 16 (1988), no. 3, 961–977.
- [GP] A. Gorodetski, Y. Pesin, *Path connectedness and entropy density of the space of hyperbolic ergodic measures*. Modern theory of dynamical systems, 111–121, Contemp. Math., 692, Amer. Math. Soc., Providence, RI, 2017.
- [Gu1] B. Gurevic, *Shift entropy and Markov measures in the space of paths of a countable graph*. Dokl. Akad. Nauk SSSR 192 (1970), 963–965.
- [Gu2] B. Gurevic, *Topological entropy of enumerable Markov chains*. Dokl. Akad. Nauk SSSR, 187 (1969), 715–718; English translation: Soviet Math. Dokl., 10(4) (1969), 911–15.
- [GZ] B. Gurevich, A. Zargaryan, *Conditions for the existence of a maximal measure for a countable symbolic Markov chain*. Vestnik Moskov. Univ. Ser. I Mat. Mekh. 1988, no. 5, 14–18, 103; translation in Moscow Univ. Math. Bull. 43 (1988), no. 5, 18–23.
- [GS] B. Gurevich, S. Savchenko, *Thermodynamic formalism for countable symbolic Markov chains* (Russian), Uspekhi Mat. Nauk, 53(2) (1998), 3–106; English translation: Russian Math. Surveys, 53(2) (1998), 245–344.
- [GK] B. Gurevich, S. Katok, *Arithmetic coding and entropy for the positive geodesic flow on the modular surface*. Mosc. Math. J. 1 (2001), no. 4, 569–582, 645.

- [I] G. Iommi, *Multifractal analysis for countable Markov shifts*. Ergodic Theory Dynam. Systems 25 (2005), no. 6, 1881–1907.
- [IJT] G. Iommi, T. Jordan, M. Todd, *Recurrence and transience for suspension flows*. Israel J. Math. 209 (2015), no. 2, 547–592.
- [IRV] G. Iommi, F. Riquelme, A. Velozo, *Entropy in the cusp and phase transitions for geodesic flows*. Israel J. Math. 225 (2018), no. 2, 609–659.
- [ITV] G. Iommi, M. Todd, A. Velozo, *Upper semi-continuity of entropy in non-compact settings*. To appear in Math. Res. Lett.
- [IV] G. Iommi, A. Velozo, *The space of invariant measures for countable Markov shifts*. To appear in J. Anal. Math.
- [JMU] O. Jenkinson, R.D. Mauldin, M. Urbański, *Zero temperature limits of Gibbs equilibrium states for countable alphabet subshifts of finite type*. J. Stat. Phys. 119 (2005), no. 3–4, 765–776.
- [KKLM] S. Kadyrov, D. Kleinbock, E. Lindenstrauss, G. Margulis, *Singular systems of linear forms and non-escape of mass in the space of lattices*. J. Anal. Math. 133 (2017), 253–277.
- [KP] S. Kadyrov, A. Pohl, *Amount of failure of upper-semicontinuity of entropy in noncompact rank-one situations, and Hausdorff dimension*. Ergodic Theory Dynam. Systems 37 (2017), no. 2, 539–563.
- [Ka] A. Katok, *Lyapunov exponents, entropy and periodic orbits for diffeomorphisms*, Inst. Hautes Études Sci. Publ. Math. 51 (1980) 137–173.
- [Ki] B.P. Kitchens, *Symbolic dynamics. One-sided, two-sided and countable state Markov shifts*. Universitext. Springer-Verlag, Berlin, 1998. x+252 pp
- [Kr] U. Krengel, *Entropy of conservative transformations*. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 7 (1967) 161–181.
- [LM] Y. Lima, C. Matheus, *Symbolic dynamics for non-uniformly hyperbolic surface maps with discontinuities*. Ann. Sci. Ec. Norm. Super. 51, (2018) no.1, 1–38.
- [M] R. Mañé, *A proof of Pesin’s formula*. Ergodic Theory Dynam. Systems 1 (1981), no. 1, 95–102.
- [N] S. Newhouse, *Continuity properties of entropy*. Ann. of Math. (2) 129 (1989), 215–235.
- [Or] S. Orey, *Large deviations in ergodic theory*. Seminar on stochastic processes, 1984 (Evanston, Ill., 1984), 195–249, Progr. Probab. Statist., 9, Birkhauser Boston, Boston, MA, 1986.
- [Ov] S.B. Ovadia, *Symbolic dynamics for non uniformly hyperbolic diffeomorphisms of compact smooth manifolds*, J. Mod. Dyn. 13 (2018), 43–113.
- [P] K. R. Parthasarathy, *Probability measures on metric spaces*. Probability and Mathematical Statistics, No. 3 Academic Press, Inc., New York-London 1967 xi+276 pp
- [R1] S. Ruelle, *On the Vere-Jones classification and existence of maximal measures for countable topological Markov chains*. Pacific J. Math. 209 (2003), no. 2, 366–380.
- [R2] S. Ruelle, *Topological Markov chains of given entropy and period with or without measure of maximal entropy period*. To appear in Pacific J. Math.
- [Ri] F. Riquelme, *Ruelle’s inequality in negative curvature*. Discrete Contin. Dyn. Syst. 38 (2018), no. 6, 2809–2825.
- [RV] F. Riquelme, A. Velozo, *Escape of mass and entropy for geodesic flows*. Ergodic Theory Dynam. Systems 39 (2019), no. 2, 446–473.
- [Sal] I. Salama, *Topological entropy and recurrence of countable chains*. Pacific J. Math., 134(2) (1988), 325–341.
- [S1] O. Sarig, *Thermodynamic formalism for countable Markov shifts*. Ergodic Theory Dynam. Systems 19 (1999), 1565–1593.
- [S2] O. Sarig, *Phase transitions for countable Markov shifts*. Comm. Math. Phys., 217(3) (2001), 555–577.
- [S3] O. Sarig, *Symbolic dynamics for surface diffeomorphisms with positive entropy*. J. Amer. Math. Soc. 26 (2013), no. 2, 341–426.
- [S4] O. Sarig, *Thermodynamic formalism for countable Markov shifts*. Hyperbolic dynamics, fluctuations and large deviations, 81–117, Proc. Sympos. Pure Math. 89, Amer. Math. Soc., 2015.
- [ST] B. Schapira, S. Tapie, *Regularity of entropy, geodesic currents and entropy at infinity*. To appear in Annales scientifiques de l’ENS.

- [Vel1] A. Velozo, *Thermodynamic formalism and the entropy at infinity of the geodesic flow*. arXiv:1711.06796
- [Vel2] A. Velozo, *Pressure at infinity on countable Markov shifts*. In preparation.
- [Ve1] D. Vere-Jones, *Geometric ergodicity in denumerable Markov chains*. Quart. J. Math. Oxford Ser. (2) 13 (1962), 7–28.
- [Ve2] D. Vere-Jones, *Ergodic properties of nonnegative matrices*. I. Pacific J. Math. 22 (1967), 361–386.
- [Wa] P. Walters, *An introduction to ergodic theory*. Graduate Texts in Mathematics, 79. Springer-Verlag, New York-Berlin, 1982 ix+250 pp

FACULTAD DE MATEMÁTICAS, PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE (PUC), AVENIDA VICUÑA MACKENNA 4860, SANTIAGO, CHILE

*E-mail address:* [giommi@mat.puc.cl](mailto:giommi@mat.puc.cl)

*URL:* [www.mat.uc.cl/~giommi](http://www.mat.uc.cl/~giommi)

MATHEMATICAL INSTITUTE, UNIVERSITY OF ST ANDREWS, NORTH HAUGH, ST ANDREWS, KY16 9SS, SCOTLAND

*E-mail address:* [m.todd@st-andrews.ac.uk](mailto:m.todd@st-andrews.ac.uk)

*URL:* [www.mcs.st-and.ac.uk/~miket](http://www.mcs.st-and.ac.uk/~miket)

DEPARTMENT OF MATHEMATICS, YALE UNIVERSITY, NEW HAVEN, CT 06511, USA.

*E-mail address:* [anibal.velozo@yale.edu](mailto:anibal.velozo@yale.edu)

*URL:* <https://gauss.math.yale.edu/~av578/>