ISRAEL JOURNAL OF MATHEMATICS **158** (2007), 225–247
DOI: 10.1007/s11856-007-0012-z

MEASURE THEORETICAL ENTROPY OF COVERS

BY

Uri Shapira

Institute of Mathematics, The Hebrew University of Jerusalem
Givat Ram, Jerusalem 91904, Israel
e-mail: ushapira@math.huji.ac.il

ABSTRACT

In this paper we introduce three notions of measure theoretical entropy of a measurable cover $\mathcal{U}$ in a measure theoretical dynamical system. Two of them were already introduced in [R] and the new one is defined only in the ergodic case. We then prove that these three notions coincide, thus answering a question posed in [R], and recover a variational inequality (proved in [GW]) and a proof of the classical variational principle based on a comparison between the entropies of covers and partitions.

1. Introduction

In this paper a measure theoretical dynamical system (m.t.d.s) is a four tuple $(X, \mathcal{B}, \mu, T)$, where $(X, \mathcal{B})$ is a standard space (i.e., isomorphic to $[0,1]$ with the Borel σ-algebra, μ is a probability measure on $(X, \mathcal{B})$ and T is an invertible measure preserving map from X to itself.

A topological dynamical system (t.d.s) is a pair (X, T), where X is a compact metric space and T is a homeomorphism from X to itself.

In [R] the author introduced two notions of measure theoretical entropy of a cover, both generalizing the definition of measure theoretical entropy of a partition and influenced by [BGH], namely,

1. $h_\mu^+(\mathcal{U}) = \inf_{\alpha \succeq \mathcal{U}} h_\mu(\alpha)$,
2. $h_\mu^-(\mathcal{U}) = \lim \frac{1}{n} \inf_{\alpha \succeq \mathcal{U}_0^{n-1}} H_\mu(\alpha)$.

It was shown there among other things that $h_\mu^-(\mathcal{U}) \leq h_\mu^+(\mathcal{U})$ and that in the topological case (i.e., a t.d.s and an open cover), one can always find an invariant measure μ such that $h_\mu^-(\mathcal{U}) = h_{top}(\mathcal{U})$. This generalizes the result from [BGH]

Received August 31, 2005

asserting that in the topological case one can always find an invariant measure μ such that $h^+_\mu(\mathcal{U}) \geq h_{top}(\mathcal{U})$.

The question whether $h^-_\mu(\mathcal{U}) = h^+_\mu(\mathcal{U})$ arose. In [HMRY] the authors continued the research on these concepts and proved, among other results, with the aid of the Jewett–Krieger theorem, that if there exists a t.d.s, an invariant measure μ and an open cover $\mathcal{U}$ such that $h^-_\mu(\mathcal{U}) < h^+_\mu(\mathcal{U})$, then one can find such a situation in a uniquely ergodic t.d.s.

Recently, B. Weiss and E. Glasner [GW] showed that if (X, T) is a t.d.s and $\mathcal{U}$ is any cover, then for any invariant measure μ, $h^+_\mu(\mathcal{U}) \leq h_{top}(\mathcal{U})$, and so combining these results one concludes that for a t.d.s and an open cover we have that $h^-_\mu(\mathcal{U}) = h^+_\mu(\mathcal{U})$.

The measure theoretical entropy of a partition α in an ergodic m.t.d.s can be defined as $\lim \frac{1}{n} \log \mathcal{N}(\alpha_0^{n-1}, \epsilon)$, where $0 < \epsilon < 1$ and $\mathcal{N}(\alpha_0^{n-1}, \epsilon)$ is the minimum number of atoms of α_0^{n-1} needed to cover X up to a set of measure less than ϵ. (See [Ru].)

In this paper we follow this line and in Section 4 define a notion of measure theoretical entropy for a cover $\mathcal{U}$ of an ergodic m.t.d.s as $h^e_\mu(\mathcal{U}) = \lim \frac{1}{n} \log \mathcal{N}(\mathcal{U}_0^{n-1}, \epsilon)$ (where $0 < \epsilon < 1$). We prove (Theorem 4.2) the existence of the limit and its independence of ϵ, in a different way from [Ru] using Strong Rohlin Towers. This can serve as an alternative proof of the fact that the above definition of measure theoretical entropy of a partition in an ergodic m.t.d.s is well defined.

We show in a direct way that in the ergodic case the three notions: $h^-_\mu(\mathcal{U})$, $h^+_\mu(\mathcal{U}), h^e_\mu(\mathcal{U})$ coincide (Theorems 4.4, 4.5), and from the ergodic decomposition for $h^-_\mu(\mathcal{U}), h^+_\mu(\mathcal{U})$, proved in [HMRY], we deduce that $h^-_\mu(\mathcal{U}) = h^+_\mu(\mathcal{U})$ in the general case (Corollary 5.2), and so we can denote this number by $h_\mu(\mathcal{U}, T)$ or $h_\mu(\mathcal{U})$.

We also get an immediate proof of a slight generalization of the inequality $h_\mu(\mathcal{U}) \leq h_{top}(\mathcal{U})$, mentioned earlier, from [GW], to the non-topological case (Theorem 6.1).

Acknowledgements: This paper was written as an M.Sc. thesis at The Hebrew University of Jerusalem under the supervision of Prof. Benjamin Weiss. I would like to thank Prof. Weiss for introducing me to the subject, and sharing with me his and Prof. Eli Glasner's valuable ideas.

2. Preliminaries

Recall that in the following a measure theoretical dynamical system (m.t.d.s) is a four tuple $(X, \mathcal{B}, \mu, T)$, where $(X, \mathcal{B})$ is a standard space, μ is a probability measure on $(X, \mathcal{B})$ and T is an invertible measure preserving transformation of X.

2.1 Definition:

- A cover of X is a finite collection of measurable sets that cover X.
- The collection of covers of X will be denoted by $\mathcal{C}_X$.
- A partition of X is a cover of X whose elements are mutually disjoint.
- The collection of partitions of X will be denoted by $\mathcal{P}_X$.
  Usually we denote covers by $\mathcal{U}, \mathcal{V}$ and partitions by α, β, γ etc.
- We say that a cover $\mathcal{U}$ is finer than $\mathcal{V}$ ($\mathcal{U} \succeq \mathcal{V}$) if any element of $\mathcal{U}$ is contained in an element of $\mathcal{V}$.
- For any $\mathcal{U} \in \mathcal{C}_X$ and $k \in \mathbb{Z}$ we denote by $T^k(\mathcal{U})$ the cover whose elements are the sets of the form $T^k(U)$ where $U \in \mathcal{U}$.
- We define the join, $\mathcal{U} \vee \mathcal{V}$, of two covers $\mathcal{U}, \mathcal{V}$ to be the cover whose elements are sets of the form $U \cap V$ where $U \in \mathcal{U}$ and $V \in \mathcal{V}$.
- When the transformation T is understood we denote, for $l > k$, the cover $T^{-k}(\mathcal{U}) \vee T^{-(k+1)}(\mathcal{U}) \vee \cdots \vee T^{-l}(\mathcal{U})$ by $\mathcal{U}_k^l$.

2.2 Definition: For $0 < \delta < 1$ define $H(\delta) = -\delta \log \delta - (1-\delta)\log(1-\delta)$.

Note that $\lim_{\delta \to 0} H(\delta) = 0$.

In the sequel, we will prove some combinatorial lemmas and often we will encounter the expression $\sum_{j \leq \delta K} \binom{K}{j}$. We shall make use of the next elementary lemma:

1.1 Lemma (Lemma 1.5.4 in [Sh1]): *If $\delta < \frac{1}{2}$ then $\sum_{j \leq \delta K} \binom{K}{j} \leq 2^{H(\delta)}$.*

2.4 Definition: A m.t.d.s $(X, \mathcal{B}, \mu, T)$ is said to be *aperiodic* if, for every $n \in \mathbb{N}$, $\mu(\{x | T^n x = x\}) = 0$.

An ergodic system which is not aperiodic is easily seen to be a cyclic permutation on a finite number of atoms.

One of our main tools in practice will be the Strong Rohlin Lemma ([Sh2] p. 15):

2.5 Lemma: *Let $(X, \mathcal{B}, \mu, T)$ be an ergodic, aperiodic system and let $\alpha \in \mathcal{P}_X$. Then for any $\delta > 0$ and $n \in \mathbb{N}$, one can find a set $B \in \mathcal{B}$ such that*

$B, TB, \ldots, T^{n-1}B$ are mutually disjoint, $\mu(\bigcup_0^{n-1} T^i B) > 1 - \delta$ and the distribution of α is the same as the distribution of the partition $\alpha|_B$ that α induces on B.

The data (n, δ, B, α) will be called a strong Rohlin tower of height n and error δ with respect to α and with B as a base.

3. Measure theoretical entropy for covers

Let $(X, \mathcal{B}, \mu, T)$ be a m.t.d.s. The definitions and proofs in this section were introduced in [R].

3.1 Definition: For $\mathcal{U} \in \mathcal{C}_X$ we define the entropy of $\mathcal{U}$ as

$$H_\mu(\mathcal{U}) = \inf_{\alpha \succeq \mathcal{U}} H_\mu(\alpha).$$

3.2 Proposition:

1. *If $\mathcal{U}, \mathcal{V} \in \mathcal{C}_X$, then $H_\mu(\mathcal{U} \vee \mathcal{V}) \leq H_\mu(\mathcal{U}) + H_\mu(\mathcal{V})$.*
2. *For every $\mathcal{U} \in \mathcal{C}_X$, $H_\mu(T^{-1}\mathcal{U}) = H_\mu(\mathcal{U})$.*

3.3 Corollary: *If $\mathcal{U} \in \mathcal{C}_X$, then the sequence $H_\mu(\mathcal{U}_0^{n-1})$ is sub-additive.*

3.4 Corollary: *If $\mathcal{U} \in \mathcal{C}_X$, then the sequence $\frac{1}{n} H_\mu(\mathcal{U}_0^{n-1})$ converges to $\inf_n \frac{1}{n} H_\mu(\mathcal{U}_0^{n-1})$.*

Two ways of generalizing the definition of measure theoretical entropy of a partition to a cover are:

3.5 Definition: If $\mathcal{U} \in \mathcal{C}_X$, define

1. $h_\mu^-(\mathcal{U}, T) = \lim \frac{1}{n} H_\mu(\mathcal{U}_0^{n-1})$;
2. $h_\mu^+(\mathcal{U}, T) = \inf_{\alpha \succeq \mathcal{U}} h_\mu(\alpha, T)$.

When T is understood, we usually omit it and write $h_\mu^-(\mathcal{U})$, $h_\mu^+(\mathcal{U})$.

We shall see later that in fact $h_\mu^-(\mathcal{U}) = h_\mu^+(\mathcal{U})$.

3.6 Proposition:

1. $h_\mu^-(\mathcal{U}) \leq h_\mu^+(\mathcal{U})$.
2. *For any $m \in \mathbb{N}$, $h_\mu^-(\mathcal{U}, T) = \frac{1}{m} h_\mu^-(\mathcal{U}_0^{m-1}, T^m)$.*
3. $h_\mu^-(\mathcal{U}, T) = \lim_n \frac{1}{n} h_\mu^+(\mathcal{U}_0^{n-1}, T^n)$.

4. The ergodic case

Throughout this section, $(X, \mathcal{B}, \mu, T)$ is an ergodic m.t.d.s.

For $\mathcal{U} \in \mathcal{C}_X$, we denote by $\mathcal{N}(\mathcal{U}, \epsilon, \mu)$ the minimum number of elements of $\mathcal{U}$ needed to cover all of X, up to a set of measure less than ϵ. When μ is understood we write $\mathcal{N}(\mathcal{U}, \epsilon)$.

By a straightforward calculation one deduces from [Sh1] p. 51, the following:

4.1 Theorem: *If $(X, \mathcal{B}, \mu, T)$ is an ergodic m.t.d.s and $\alpha \in \mathcal{P}_X$, then for any $0 < \epsilon < 1$, $h_\mu(\alpha, T) = \lim \frac{1}{n} \log \mathcal{N}(\alpha_0^{n-1}, \epsilon)$.*

In view of this result, a natural way to generalize the definition of measure theoretical entropy of a partition to covers will be the following:

$$h_\mu(\mathcal{U}, T) = \lim \frac{1}{n} \log \mathcal{N}(\mathcal{U}_0^{n-1}, \epsilon),$$

where $0 < \epsilon < 1$. In order to do so we have to show that the above limit exists and is independent of ϵ.

4.2 Theorem: *For any $0 < \epsilon < 1$, the sequence $\frac{1}{n} \log \mathcal{N}(\mathcal{U}_0^{n-1}, \epsilon)$ converges and the limit is independent of ϵ.*

In order to prove this theorem we shall need a combinatorial lemma. Let us first introduce some terminology (in first reading the reader may skip the following discussion and turn to the discussion held after the proof of Lemma 4.3):

- We say that two intervals in $\mathbb{N}$, I and J, are separated if there is $n \in \mathbb{N}$ such that for any $i \in I$, $j \in J$ we have $i < n < j$ or $j < n < i$.
- We say that a collection $\{I_i\}_{i \in A}$ of intervals in $\mathbb{N}$ is a separated collection if any two of its elements are separated.
- We say that a collection $\{I_i\}_{i \in A}$ of subintervals of an interval $[1, K]$ is a (λ, ϵ) separated cover of $[1, K]$ (for $0 < \lambda < 1$, $0 < \epsilon$), if it is separated and
$$\left| \frac{|\bigcup I_i|}{K} - \lambda \right| < \epsilon.$$
- Given a vector $\vec{\lambda} = (\lambda_1, \ldots, \lambda_l)$, we denote
$$\nu_r(\vec{\lambda}) = \prod_{j=r}^{l} (1 - \lambda_j)$$
or just ν_r when $\vec{\lambda}$ is understood. For $r > l$ we set $\nu_r = 1$. Note that for $j \leq l$ we have:
$$\sum_{r=j}^{l} \lambda_r \nu_{r+1} = 1 - \nu_j.$$

In the following combinatorial lemma, we will be given l separated collections $\{I_i^j\}_{i\in A_j}$, $j=1,\dots,l$ of subintervals of a very long interval $[1,K]$. Our knowledge about these collections is that the members of the jth collection all have the same length, N_j, $N_1 << N_2 << \cdots << N_l$ and every collection is very "equally distributed" in $[1,K]$ in some sense.

We would like to extract from these collections a separated collection that will cover as much as we can, from $[1,K]$.

Let us denote by λ_j the percentage of $[1,K]$ that is covered by the jth collection and by $\vec{\lambda}$, the corresponding vector. Then, $\lambda_l = 1-\nu_l$ percent of $[1,K]$ is covered by $\{I_i^l\}$. The complement is of size $K\nu_l$ and we could cover λ_{l-1} percent of it with the $\{I_i^{l-1}\}$'s. By now we covered $K(1-\nu_{l-1})$ and we could cover λ_{l-2} percent of the complement by the $\{I_i^{l-2}\}$'s. So by now we covered $K(1-\nu_{l-2})$ of $[1,K]$. We go on this way and extract a separated collection that covers $1-\nu_1$ percent of $[1,K]$. Let us now make these ideas precise.

4.3 Lemma: *For any $l>0$, there exists a positive function*

$$\varphi = \varphi(N_1,\dots,N_l,\eta_1,\dots,\eta_l,\epsilon)$$

(where $N_1 < N_2 < \cdots < N_l \in \mathbb{N}$, $\eta_i,\epsilon > 0$) such that

$$(*) \qquad \limsup_{\epsilon\to 0}\limsup_{N_1\to\infty}\limsup_{\eta_1\to 0}\cdots\limsup_{N_l\to\infty}\limsup_{\eta_l\to 0}\varphi(N_i,\eta_i,\epsilon)=0,$$

and such that if $0<\lambda_j<1$, $j=1,\dots,l$ and $\{I_i^j\}_{i\in A_j}$ are separated collections of subintervals of $[1,K]$ that satisfy:

(a) *for every $1\le j\le l$, $|I_i^j| = N_j$,*
(b) *for every $1\le j\le l$, $\{I_i^j\}$ is a (λ_j,ϵ)-separated cover of $[1,K]$,*
(c) *for every $0\le j<r\le l$, the number of subintervals, J, of $[1,K]$, of length N_r, which are not (λ_j,ϵ)-separately covered by $\{I_i^j\subset J\}$ is less than $\eta_r K$,*

then there are sets $\tilde{A}_j\subset A_j$, $j=1,\dots,l$, such that $\{\{I_i^j\}_{i\in\tilde{A}_j}\}_{j=1}^l$ is a separated collection and $[1,K]$ is $((1-\nu_1(\vec{\lambda})),\varphi(N_i,\eta_i,\epsilon))$-separately covered by $\{\{I_i^j\}_{i\in\tilde{A}_j}\}_{j=1}^l$.

Proof: We will build the $\tilde{A}_j$'s by recursion, starting with $j=l$. Define $\tilde{A}_l = A_l$. Then from (b) we have that

$$\left|\frac{N_l|\tilde{A}_l|}{K}-\lambda_l\right|<\epsilon.$$

So if we will define $f_l(N_i,\eta_i,\epsilon)=\epsilon$, then f_l satisfies $(*)$ and $[1,K]$ is $(\lambda_l\nu_{l+1}, f_l(N_i,\eta_i,\epsilon))$-separately covered by $\{I_i^l\}_{i\in\tilde{A}_l}$. Now, suppose we have defined $\tilde{A}_l,\dots,\tilde{A}_{j+1}$ and positive functions $f_l,\dots,f_{j+1}$, that satisfy $(*)$, such that

$\{\{I_i^r\}_{i\in\tilde{A}_r}\}_{r=j+1}^l$ is a separated collection and for every $j+1\le r\le l$, $[1,K]$ is $(\lambda_r\nu_{r+1}, f_r(N_i,\eta_i,\epsilon))$-separately covered by $\{I_i^r\}_{i\in\tilde{A}_r}$. Define now

$$\tilde{A}_j=\big\{i\in A_j|I_i^j \text{ is separated from } \{I_s^r\}_{s\in\tilde{A}_r}, r=j+1,\dots,l\big\}.$$

We want to estimate the size of $\tilde{A}_j$.

Estimation from below: Choose $j+1\le r\le l$ and divide the members of $\{I_i^r\}_{i\in\tilde{A}_r}$ into good ones and bad ones according to (c), i.e., I_s^r is good if it is (λ_j,ϵ)-separately covered by $\{I_i^j\subset I_s^r\}$. We have at most $\eta_r K$, I_i^r's, which are bad, and at most $|\tilde{A}_r|$, I_i^r's, which are good. Every bad I_i^r rules out at most $\frac{N_r}{N_j}+2$ i's in A_j from being in $\tilde{A}_j$. Every good I_i^r rules out at most $\frac{N_r}{N_j}(\lambda_j+\epsilon)+2$ i's in A_j from being in $\tilde{A}_j$. In total, the maximum number of i's in A_j that are not in $\tilde{A}_j$ is at most

$$\sum_{r=j+1}^{l}|\tilde{A}_r|\Big(\frac{N_r}{N_j}(\lambda_j+\epsilon)+2\Big)+\eta_r K\Big(\frac{N_r}{N_j}+2\Big)=(**).$$

Note that because $[1,K]$ is $(\lambda_r\nu_{r+1},f_r)$-separately covered by $\{I_i^r\}_{i\in\tilde{A}_r}$, we must have

$$|\tilde{A}_r|\le\frac{K}{N_r}(\lambda_r\nu_{r+1}+f_r).$$

Using this we get

$$\begin{aligned}(**)\le&\sum_{r=j+1}^{l}\frac{K}{N_r}(\lambda_r\nu_{r+1}+f_r)\Big(\frac{N_r}{N_j}(\lambda_j+\epsilon)+2\Big)+\eta_r K\Big(\frac{N_r}{N_j}+2\Big)\\
=&\sum_{r=j+1}^{l}\frac{K}{N_j}\lambda_r\nu_{r+1}(\lambda_j+\epsilon)+\frac{K}{N_j}(\lambda_j+\epsilon)f_r+\frac{2K}{N_r}(\lambda_r\nu_{r+1}+f_r)\\
&+\frac{K}{N_j}\eta_r N_r+2\eta_r K\\
=&\frac{K}{N_j}\lambda_j\Big(\sum_{r=j+1}^{l}\lambda_r\nu_{r+1}\Big)\\
&+\frac{K}{N_j}\sum_{r=j+1}^{l}\Big\{\epsilon\lambda_r\nu_{r+1}+(\lambda_j+\epsilon)f_r+2\frac{N_j}{N_r}\big(\lambda_r\nu_{r+1}+f_r\big)+\eta_r(N_r+2N_j)\Big\}\\
=&(\aleph),\end{aligned}$$

as mentioned earlier $\sum_{j+1}^{l} \lambda_r \nu_{r+1} = 1 - \nu_{j+1}$, so we have that

$$|\tilde{A}_j| \geq |A_j| - (\aleph) \geq \frac{K}{N_j}(\lambda_j - \epsilon) - (\aleph)$$
$$= \frac{K}{N_j}\Big\{\lambda_j \nu_{j+1} - \Big\{\epsilon + \sum_{r=j+1}^{l} \Big\{\epsilon\lambda_r\nu_{r+1} + (\lambda_j + \epsilon) f_r + 2\frac{N_j}{N_r}(\lambda_r \nu_{r+1} + f_r) + \eta_r(N_r + 2N_j)\Big\}\Big\}\Big\}.$$

Note that

$$\Big|\epsilon + \sum_{r=j+1}^{l} \Big\{\epsilon\lambda_r\nu_{r+1} + (\lambda_j + \epsilon) f_r + 2\frac{N_j}{N_r}(\lambda_r \nu_{r+1} + f_r) + \eta_r(N_r + 2N_j)\Big\}\Big|$$
$$\leq \epsilon + \sum_{r=j+1}^{l} \Big\{\epsilon + (1+\epsilon) f_r + 2\frac{N_j}{N_r}(1 + f_r) + \eta_r(N_r + 2N_j)\Big\},$$

so if we denote the last expression by $\tilde{f}_j(N_i, \eta_i, \epsilon)$, then we see that $\tilde{f}_j$ satisfies $(*)$ and $|\tilde{A}_j| \geq \frac{K}{N_j}(\lambda_j \nu_{j+1} - \tilde{f}_j)$.

Estimation from above: For every $j + 1 \leq r \leq l$, we have that $|\tilde{A}_r| \geq \frac{K}{N_r}(\lambda_r \nu_{j+1} - f_r)$ and the number of bad I_i^r's is at most $\eta_r K$, so we must have at least $\frac{K}{N_r}(\lambda_r \nu_{j+1} - f_r) - \eta_r K$ good I_i^r's. Every good I_i^r rules out at least $\frac{N_r}{N_j}(\lambda_j - \epsilon)$ i's in A_j from being in $\tilde{A}_j$. So the number of i's in A_j that are not in $\tilde{A}_j$ is at least

$$\sum_{r=j+1}^{l} \frac{N_r}{N_j}(\lambda_j - \epsilon)\Big\{\frac{K}{N_r}(\lambda_r \nu_{r+1} - f_r) - \eta_r K\Big\}$$

and so

$$|\tilde{A}_j| \leq |A_j| - \sum_{r=j+1}^{l} \frac{N_r}{N_j}(\lambda_j - \epsilon)\Big\{\frac{K}{N_r}(\lambda_r \nu_{r+1} - f_r) - \eta_r K\Big\}$$
$$\leq \frac{K}{N_j}(\lambda_j + \epsilon) - \sum_{r=j+1}^{l} \Big\{\frac{K}{N_j}\big(\lambda_j(\lambda_r \nu_{r+1} - f_r) - \epsilon(\lambda_r \nu_{r+1} - f_r)\big) - \frac{K}{N_j}\eta_r N_r(\lambda_j - \epsilon)\Big\}$$
$$= \frac{K}{N_j}\Big\{\lambda_j\Big(1 - \sum_{r=j+1}^{l} \lambda_r \nu_{r+1}\Big) + \epsilon + \sum_{r=j+1}^{l} (\lambda_j f_r + \epsilon(\lambda_r \nu_{r+1} - f_r) + \eta_r N_r(\lambda_j - \epsilon))\Big\}$$

$$\leq \frac{K}{N_j}\Big\{\lambda_j\nu_{j+1}+\epsilon+\sum_{r=j+1}^{l}(f_r+\epsilon(1+f_r)+\eta_r N_r(1+\epsilon))\Big\},$$

so if we denote

$$\hat{f}_j(N_i,\eta_i,\epsilon)=\epsilon+\sum_{r=j+1}^{l}(f_r+\epsilon(1+f_r)+\eta_r N_r(1+\epsilon)),$$

then $\hat{f}_j$ satisfies $(*)$ and $|\tilde{A}_j|\leq\frac{K}{N_j}(\lambda_j\nu_{j+1}+\hat{f}_j)$. Define $f_j=\max(\tilde{f}_j,\hat{f}_j)$; then we have that f_j satisfies $(*)$ and

$$\left|\frac{|\tilde{A}_j|N_j}{K}-\lambda_j\nu_{j+1}\right|\leq f_j.$$

We have defined $\tilde{A}_j\subset A_j$ and a positive function f_j that satisfies $(*)$, such that $\{\{I_i^r\}_{i\in\tilde{A}_r}\}_{r=j}^{l}$ is a separated collection and $[1,K]$ is $(\lambda_j\nu_{j+1},f_j)$-separately covered by $\{I_i^j\}_{i\in\tilde{A}_j}$.

We continue this way and define sets $\tilde{A}_j\subset A_j$ and positive functions f_j, $j=1,\dots,l$, such that $\{\{I_i^j\}_{i\in\tilde{A}_j}\}_{j=1}^{l}$ is a separated collection and $[1,K]$ is $(\lambda_j\nu_{j+1},f_j)$-separately covered by $\{I_i^j\}_{i\in\tilde{A}_j}$.

Note that this means

$$K\Big(\sum_{j=1}^{l}\lambda_j\nu_{j+1}-\sum_{j=1}^{l}f_r\Big)\leq\Big|\bigcup_{j=1}^{l}\bigcup_{i\in\tilde{A}_j}I_i^j\Big|\leq K\Big(\sum_{j=1}^{l}\lambda_j\nu_{j+1}+\sum_{j=1}^{l}f_r\Big)$$

and so, if we define $\varphi=\sum f_j$, then φ satisfies $(*)$ and $\{\{I_i^j\}_{i\in\tilde{A}_j}\}_{j=1}^{l}$ is a $(1-\nu_1,\varphi)$-separated cover of $[1,K]$. ∎

Before turning to the proof of Theorem 4.2, let us present some terminology. In the following $\mathcal{U}=\{U_1,\dots,U_M\}$ is a cover of X. For any $\rho>0$, we can find a partition $\beta\succeq\mathcal{U}$, such that $\mathcal{N}(\mathcal{U},\rho)=\mathcal{N}(\beta,\rho)$. Namely, we choose a subset of $\mathcal{U}$, of $N=\mathcal{N}(\mathcal{U},\rho)$ elements, that covers X up to a set of measure $<\rho$, $\{U_{i1},\dots,U_{iN}\}$, and define $C_1=U_{i1}$, $C_j=U_{ij}\setminus\bigcup_{m=1}^{j-1}U_{im}$, $j=2,\dots,N$. The C_j's are disjoint, $C_j\subset U_{ij}$ and $\bigcup_1^N C_j=\bigcup_{j=1}^N U_{ij}$. Extend the collection $\{C_j\}_{j=1}^N$ to a partition, β, refining $\mathcal{U}$ in some way. Then, because $\beta\succeq\mathcal{U}$, we have $\mathcal{N}(\beta,\rho)\geq N$ and, from our construction, it follows that $\mathcal{N}(\beta,\rho)\leq N$.

- We call such a partition a ρ-good partition for $\mathcal{U}$.

If $(X,\mathcal{B},\mu,T)$ is aperiodic and $N\in\mathbb{N},\rho,\delta>0$ are given; then for a ρ-good partition β, for $\mathcal{U}_0^{N-1}$, we can construct a strong Rohlin tower with height $N+1$ and error $<\delta$. Let $\tilde{B}$ denote the base of the tower and let $B\subset\tilde{B}$ be a union of $\mathcal{N}(\beta,\rho)$ atoms of $\beta|_{\tilde{B}}$ that covers $\tilde{B}$ up to a set of measure, less than $\rho\mu(\tilde{B})$.

- We call $(\beta, \tilde{B}, B)$ a good base for $(\mathcal{U}, N, \rho, \delta)$.
- For a set $J \subset \mathbb{N}$, a $(\mathcal{U}, J)$-name is a function $f : J \to \{1, \dots, M\}$.
- f is a name of $x \in X$, if $x \in \bigcap_{j\in J} T^{-j}U_{f(j)}$.
- We denote the set of elements of X with f as a name by S_f.
- A set of $(\mathcal{U}, J)$-names, $\{f_i\}$, covers a set $C \in \mathcal{B}$, if $C \subset \bigcup_i S_{f_i}$.

In the sequel, we will want to estimate the number of elements of $\mathcal{U}_0^{N-1}$ needed to cover a set $C \in \mathcal{B}$, i.e., we will want to estimate the number of $(\mathcal{U}, [0, N-1])$-names needed to cover C. The usual way to do so is to find a collection of disjoint sets $J_i \subset [0, N-1]$, $i = 1, \dots, m$, that covers most of $[0, N-1]$, such that we can bound the number of $(\mathcal{U}, J_i)$-names needed to cover C. If we can cover C by R_i $(\mathcal{U}, J_i)$-names, $\{f^i_m\}_{m=1}^{R_i}$, then the set $\Gamma = \{f\colon [0, N-1] \to \{1, \dots, M\}|\ f|_{J_i} \in \{f^i_m\}_{m=1}^{R_i}\}$ of $(\mathcal{U}, [0, N-1])$-names covers C and contains $\prod R_i \cdot M^{N-\sum |J_i|}$ elements.

This situation occurs in our proofs in the following way: Let $(\beta, \tilde{B}, B)$ be a good base for $(\mathcal{U}, N, \rho, \delta)$ and $K >> N$. Set C to be the set of elements of X that visits B at times $i_1 < \cdots < i_m$ between 0 to $K - N$ (under the action of T). Then, we can cover C by no more than $\mathcal{N}(\beta, \rho)$, $(\mathcal{U}, [i_j, i_j + N - 1])$-names. We can now turn to the proof of Theorem 4.2.

Proof of Theorem 4.2: If $(X, \mathcal{B}, \mu, T)$ is periodic, it follows from the ergodicity, that the system is a cyclic permutation on a finite set of atoms and for every $0 < \epsilon < 1$ we have $\lim \frac{1}{n} \log \mathcal{N}(\mathcal{U}_0^{n-1}, \epsilon) = 0$. We assume, then, that the system is aperiodic and thus we are able to use the Strong Rohlin Lemma. Given $0 < \rho_2 < \rho_1 < 1$, we need to show that the limits $\lim \frac{1}{n} \log \mathcal{N}(\mathcal{U}_0^{n-1}, \rho_i)$, $i = 1, 2$, exist and are equal. Note that for every n, we have that $\mathcal{N}(\mathcal{U}_0^{n-1}, \rho_1) \leq \mathcal{N}(\mathcal{U}_0^{n-1}, \rho_2)$ and thus $\limsup \frac{1}{n} \log \mathcal{N}(\mathcal{U}_0^{n-1}, \rho_1) \leq \liminf \frac{1}{n} \log \mathcal{N}(\mathcal{U}_0^{n-1}, \rho_2)$, so it is enough to prove that $\limsup \frac{1}{n} \log \mathcal{N}(\mathcal{U}_0^{n-1}, \rho_2) \leq \liminf \frac{1}{n} \log \mathcal{N}(\mathcal{U}_0^{n-1}, \rho_1)$. Let $0 < \epsilon_0 < \frac{1}{2}$ be given and denote

$$h_0 = \liminf \frac{1}{n} \log \mathcal{N}(\mathcal{U}_0^{n-1}, \rho_1), \quad L = \{n \in \mathbb{N} | |h_0 - \frac{1}{n} \log \mathcal{N}(\mathcal{U}_0^{n-1}, \rho_1)| < \epsilon_0\},$$

so L contains arbitrarily large numbers. Choose $\ell \in \mathbb{N}$ large enough so that

$$(*) \qquad \Big(\frac{1}{2}(1+\rho_1)\Big)^{\ell} \log M < \epsilon_0, \quad \Big(\frac{1}{2}(1+\rho_1)\Big)^{\ell} + \epsilon_0 < \frac{1}{2}.$$

THE TOWERS CONSTRUCTION: Remember the function φ from the combinatorial lemma (Lemma 4.3). It satisfies

$$\limsup_{\epsilon\to 0}\limsup_{N_1\to\infty}\limsup_{\eta_1\to 0}\cdots\limsup_{N_\ell\to\infty}\limsup_{\eta_\ell\to 0}\varphi(N_i,\eta_i,\epsilon)=0$$

so we can choose $\epsilon>0$ small enough such that

$$\limsup_{N_1\to\infty}\limsup_{\eta_1\to 0}\cdots\limsup_{N_\ell\to\infty}\limsup_{\eta_\ell\to 0}\varphi(N_i,\eta_i,\epsilon)<\epsilon_0.$$

Choose a small enough $\delta>0$ (in a manner specified later). Choose $N_1\in L$ large enough such that

$$\limsup_{\eta_1\to 0}\cdots\limsup_{N_\ell\to\infty}\limsup_{\eta_\ell\to 0}\varphi(N_i,\eta_i,\epsilon)<\epsilon_0.$$

Find a good base $(\beta_1,\tilde{B}_1,B_1)$ for $(\mathcal{U},N_1,\rho_1,\delta)$. Choose $\eta_1>0$ small enough such that

$$\limsup_{N_2\to\infty}\limsup_{\eta_2\to 0}\cdots\limsup_{N_\ell\to\infty}\limsup_{\eta_\ell\to 0}\varphi(N_i,\eta_i,\epsilon)<\epsilon_0.$$

From the ergodicity, we can choose $N_2\in L$ large enough such that

- $\limsup_{\eta_2\to 0}\cdots\limsup_{N_\ell\to\infty}\limsup_{\eta_\ell\to 0}\varphi(N_i,\eta_i,\epsilon)<\epsilon_0$;
- $\mu\{x||\frac{1}{N_2}\sum_{r=0}^{N_2-N_1}\chi_{B_1}(T^rx)-\mu(B_1)|<\frac{\epsilon}{N_1}\}>1-\eta_1$.

Find a good base $(\beta_2,\tilde{B}_2,B_2)$ for $(\mathcal{U},N_2,\rho_1,\delta)$. Choose $\eta_2>0$ small enough such that

$$\limsup_{N_3\to\infty}\limsup_{\eta_3\to 0}\cdots\limsup_{N_\ell\to\infty}\limsup_{\eta_\ell\to 0}\varphi(N_i,\eta_i,\epsilon)<\epsilon_0.$$

Again, from the ergodicity, we can choose $N_3\in L$ such that

- $\limsup_{\eta_3\to 0}\cdots\limsup_{N_\ell\to\infty}\limsup_{\eta_\ell\to 0}\varphi(N_i,\eta_i,\epsilon)<\epsilon_0$;
- $\mu\{x||\frac{1}{N_3}\sum_{r=0}^{N_3-N_j}\chi_{B_j}(T^rx)-\mu(B_j)|<\frac{\epsilon}{N_j}j=1,2\}>1-\eta_2$.

In this way we construct, inductively, $N_1<N_2<\cdots<N_\ell$ (all from L), $\eta_1,\ldots,\eta_\ell$ and good bases $(\beta_j,\tilde{B}_j,B_j)$ for $(\mathcal{U},N_j,\rho_1,\delta)$ such that $\varphi(N_i,\eta_i,\epsilon)<\epsilon_0$; and if we denote

$$F_j=\left\{x\left|\left|\frac{1}{N_j}\sum_{r=0}^{N_j-N_i}\chi_{B_i}(T^rx)-\mu(B_i)\right|<\frac{\epsilon}{N_i},i=1,\ldots,j-1\right.\right\}$$

then $\mu(F_j)>1-\eta_j$.

Define

$$E_K=\left\{x\left|\frac{1}{K}\sum_{r=0}^{K-N_j}\chi_{F_j}(T^rx)>1-\eta_j,\left|\frac{1}{K}\sum_{r=0}^{K-N_j}\chi_{B_j}(T^rx)-\mu(B_j)\right|<\frac{\epsilon}{N_j},\right.\right.$$
$$\left. j=1,\ldots,\ell\right\}.$$

From the ergodicity, we know that there is a K_0 such that, for any $K > K_0$, we have $\mu(E_K) > \rho_2$. Fix $K > K_0$; we shall show that we can cover E_K by "few" $(\mathcal{U}, [0, K-1])$-names. For a fixed $x \in E_K$ denote

$$A_j = \{0 \leq m \leq K - N_j | T^m x \in B_j\}$$

and, for every $i \in A_j$, let $I_i^j = [i, i + N_j - 1]$. We claim that the collections $\{I_i^j\}_{i \in A_j}$, $j = 1, \dots, \ell$, satisfy conditions (a),(b),(c) from the combinatorial lemma (Lemma 4.3), with $\lambda_j = N_j \mu(B_j)$. To see this, note first that because the height of the jth tower was $N_j + 1$, we have that each collection $\{I_i^j\}_{i \in A_j}$ is separated.

(a) By definition $|I_i^j| = N_j$.

(b) Because $x \in E_K$, we know that $|\frac{1}{K} \sum_{r=0}^{K-N_j} \chi_{B_j}(T^r x) - \mu(B_j)| < \frac{\epsilon}{N_j}$ and thus $|\frac{N_j |A_j|}{K} - \lambda_j| < \epsilon$. So the $\{I_i^j\}_{i \in A_j}$ forms a (λ_j, ϵ)-separated cover of $[0, K-1]$.

(c) For $1 < r \leq \ell$ we know, from the fact that $x \in E_K$, that

$$\frac{1}{K} \sum_{s=0}^{K-N_r} \chi_{F_r}(T^s x) > 1 - \eta_r$$

and thus we have $\frac{1}{K} \sum_{s=0}^{K-N_r} \chi_{F_r^c}(T^s x) < \eta_r$. If we use the definition of F_r, this becomes

$$\frac{1}{K} \# \Bigg\{ 0 \leq s \leq K - N_r | \exists 1 \leq j \leq r - 1 \\ \left| \frac{1}{N_r} \sum_{i=0}^{N_r - N_j} \chi_{B_j}(T^{i+s} x) - \mu(B_j) \right| \geq \frac{\epsilon}{N_j} \Bigg\} < \eta_r$$

or equivalently

$$\# \Big\{ 0 \leq s \leq K - N_r | \exists 1 \leq j \leq r - 1 \left| \frac{N_j}{N_r} \# \{ i | i + s \in A_j \} - \lambda_j \right| \geq \epsilon \Big\} < \eta_r K,$$

so if we choose $1 \leq j < r \leq \ell$, we must have

$$\# \Big\{ J \subset [0, K-1] | |J| = N_r, \left| \frac{N_j}{N_r} \# \{ i | I_i^j \subset J \} - \lambda_j \right| \geq \epsilon \Big\} < \eta_r K.$$

In words, the number of subintervals of $[0, K-1]$ of length N_r, J, which are not (λ_j, ϵ)-separately covered by those I_i^j which are contained in J is less than $\eta_r K$, as we wanted.

Using the combinatorial lemma, we can choose for every $x \in E_K$ a separated collection $\{\{I_i^j(x)\}_{i\in\tilde{A}_j}\}_{j=1}^{\ell}$ that covers at least $K(1-\nu_1(\vec{\lambda})-\epsilon_0)$ elements of $[0, K-1]$. Because these collections are separated, there is a $1-1$ correspondence between them and their complements. Hence, the number of such covers is less than

$$(**) \qquad \psi(K,\lambda_j,\epsilon_0) = \sum_{j\le(\nu_1+\epsilon_0)K} \binom{K}{j}.$$

Fix such a collection $\{\{I_i^j\}_{i\in\tilde{A}_j}\}_{j=1}^{\ell}$ and set

$$C = \{x \in E_K | \{I_i^j(x)\} = \{I_i^j\}\}.$$

From the construction we see that, for every $1 \le j \le \ell$ we can cover B_j by no more than $2^{N_j(h_0+\epsilon_0)}$ $(\mathcal{U}, [0, N_j-1])$-names, thus we can cover C by no more than $2^{N_j(h_0+\epsilon_0)}$ $(\mathcal{U}, I_i^j)$-names. So the number of $(\mathcal{U}, [0, K-1])$-names needed to cover C is at most

$$\begin{aligned}\prod_{j=1}^{\ell}(2^{N_j(h_0+\epsilon_0)})^{|\tilde{A}_j|}\cdot M^{K(\nu_1+\epsilon_0)} &= 2^{(\sum_j N_j|\tilde{A}_j|)(h_0+\epsilon_0)}\cdot M^{K(\nu_1+\epsilon_0)}\\ &\le 2^{K(h_0+\epsilon_0)}\cdot M^{K(\nu_1+\epsilon_0)}.\end{aligned}$$

Finally, we get from this and $(**)$ that

$$\mathcal{N}(\mathcal{U}_0^{K-1},\rho_2) \le \psi(K,\lambda_j,\epsilon_0)\cdot 2^{K(h_0+\epsilon_0)}\cdot M^{K(\nu_1+\epsilon_0)}$$

and so

$$\frac{1}{K}\log\mathcal{N}(\mathcal{U}_0^{K-1},\rho_2) \le \frac{1}{K}\log\psi(K,\lambda_j,\epsilon_0) + h_0 + \epsilon_0 + \nu_1\log M + \epsilon_0\log M.$$

If, in the construction of the towers, we choose δ small enough and N_1 large enough, we can ensure that $\lambda_j = N_j\mu(B_j) > (1-\rho_1)/2$ and thus $1-\lambda_j < (1+\rho_1)/2 \Rightarrow \nu_1 < (\frac{1+\rho_1}{2})^{\ell}$ and so, from $(*)$, we have that

$$\nu_1\log M < \epsilon_0, \quad \nu_1+\epsilon_0 \le \frac{1}{2};$$

hence, from Lemma 2.3,

$$\psi(K,\lambda_j,\epsilon_0) \le 2^{K\cdot H((\frac{1+\rho_1}{2})^{\ell}+\epsilon_0)}$$

so

$$\begin{gathered}\frac{1}{K}\log\mathcal{N}(\mathcal{U}_0^{K-1},\rho_2) \le h_0 + \epsilon_0(2+\log M) + H\Big(\Big(\frac{1+\rho_1}{2}\Big)^{\ell}+\epsilon_0\Big) \Rightarrow\\ \limsup_K \frac{1}{K}\log\mathcal{N}(\mathcal{U}_0^{K-1},\rho_2) \le h_0 + \epsilon_0(2+\log M) + H\Big(\Big(\frac{1+\rho_1}{2}\Big)^{\ell}+\epsilon_0\Big).\end{gathered}$$

Letting $\ell \to \infty$ and $\epsilon_0 \to 0$ we get

$$\limsup_K \frac{1}{K} \log \mathcal{N}(\mathcal{U}_0^{K-1}, \rho_2) \leq h_0$$

as desired. ∎

After proving Theorem 4.2, we can define, for an ergodic m.t.d.s, $(X, \mathcal{B}, \mu, T)$ and a cover $\mathcal{U} = \{U_1, \ldots, U_M\}$ of X, a notion of measure theoretical entropy in the following way:

$$h_\mu^e(\mathcal{U}, T) = \lim \frac{1}{n} \log \mathcal{N}(\mathcal{U}_0^{n-1}, \epsilon) \quad \text{where } 0 < \epsilon < 1.$$

Often we omit T and write $h_\mu^e(\mathcal{U})$.

4.4 Theorem: $h_\mu^e(\mathcal{U}) = h_\mu^+(\mathcal{U})$.

Proof: As before, if the system is periodic then $h_\mu^e(\mathcal{U}) = h_\mu^+(\mathcal{U}) = 0$. We assume then that the system is aperiodic. For every partition $\alpha \succeq \mathcal{U}, n \in \mathbb{N}$ and $0 < \epsilon < 1$, we have that $\mathcal{N}(\mathcal{U}_0^{n-1}, \epsilon) \leq \mathcal{N}(\alpha_0^{n-1}, \epsilon)$ and therefore

$$h_\mu^e(\mathcal{U}) = \lim \frac{1}{n} \log \mathcal{N}(\mathcal{U}_0^{n-1}, \epsilon) \leq \lim \frac{1}{n} \log \mathcal{N}(\alpha_0^{n-1}, \epsilon) = h_\mu(\alpha)$$
$$\Rightarrow h_\mu^e(\mathcal{U}) \leq h_\mu^+(\mathcal{U}).$$

To prove the other inequality, we shall show that for a given $0 < \epsilon < \frac{1}{4}$ and $n \in \mathbb{N}$ we have

$$(*) \qquad h_\mu^+(\mathcal{U}) \leq \frac{1}{n} \log \mathcal{N}(\mathcal{U}_0^{n-1}, \epsilon) + \sqrt{\epsilon} \cdot \log M + H(\sqrt{\epsilon}).$$

Once we prove $(*)$, we are done, for letting $n \to \infty$ we get $h_\mu^+(\mathcal{U}) \leq h_\mu^e(\mathcal{U}) + \sqrt{\epsilon} \cdot \log M + H(\sqrt{\epsilon})$ and now, letting $\epsilon \to 0$, we get $h_\mu^+(\mathcal{U}) \leq h_\mu^e(\mathcal{U})$ as desired.

Proof of $(*)$: Choose $\delta > 0$ such that $\epsilon + \delta < \frac{1}{4}$ and find a good base $(\beta, \tilde{B}, B)$ for $(\mathcal{U}, n, \epsilon, \delta)$. (Now we take $\tilde{B}$ to be a base for a strong Rohlin tower of height N and error $< \delta$ and not of height $N+1$ as before.) Set $N = \mathcal{N}(\mathcal{U}_0^{n-1}, \epsilon)$, so B is the union of N elements of $\beta|_{\tilde{B}}$. We index these elements by sequences $i_0, \ldots, i_{n-1}$ such that if $B_{i_0,\ldots,i_{n-1}}$ is one, then $T^j(B_{i_0,\ldots,i_{n-1}}) \subset U_{i_j}$ for every $0 \leq j \leq n-1$. We have that $\mu(X \setminus \bigcup_{i=0}^{n-1} T^i(B)) \leq \epsilon + \delta$. Let $\hat{\alpha} = \{\hat{A}_1, \ldots, \hat{A}_M\}$ be the partition of

$$E = \bigcup_{i=0}^{n-1} T^i(B)$$

defined by

$$\hat{A}_m = \bigcup \{T^j(B_{i_0,\dots,i_{n-1}}) | j \in [0, n-1], i_j = m\}.$$

Note that $\hat{A}_m \subset U_m$ for every $1 \le m \le M$. Extend $\hat{\alpha}$ to a partition α of X, refining $\mathcal{U}$ in some way. Set $\eta^2 = \epsilon + \delta$ and define, for every $k > n$, $f_k(x) = \frac{1}{k}\sum_{i=0}^{k-1} \chi_E(T^j x)$. We have that $0 \le f_k \le 1$ and $\int f_k > 1 - \eta^2$, so if we denote

$$G_k = \{x | f_k(x) > 1 - \eta\}$$

then

$$\eta \cdot \mu(G_k^c) \le \int_{G_k^c} 1 - f_k \le \int 1 - f_k \le \eta^2$$
$$\Rightarrow \mu(G_k) \ge 1 - \eta.$$

We shall show that we can cover G_k by "few" $(\alpha, [0, k-1])$-names. Partition G_k according to the values of $0 \le i \le k - n$ such that $T^i x \in B$. Note that if $x \in G_k$ and $0 \le i_1 < \cdots < i_m \le k-n$ are the times at which x visits B, then the collection $\{[i_j, i_j + n - 1]\}_{j=1}^m$ covers all but at most $\eta k + 2n$ elements of $[0, k-1]$. Because each element of this partition defines a collection of subintervals of $[0, k-1]$, of length n, that covers all but at most $\eta k + 2n$ elements of $[0, k-1]$ in a $1-1$ manner, we have that the number of elements in the partition of G_k is at most

$$\psi(k, n, \eta) = \sum_{j < (\eta + \frac{2n}{k})k} \binom{k}{j}.$$

We fix an element C of this partition of G_k and want to estimate the number of $(\alpha, [0, k-1])$-names needed to cover it. If $0 \le i_1 < \cdots < i_m \le k - n$ are the times elements of C visit B, then we need at most N $(\alpha, [i_j, i_j + n - 1])$-names to cover C. Because the size of $[0, k-1] \setminus \bigcup_j [i_j, i_j + n - 1]$ is at most $\eta k + 2n$, we need at most $N^{k/n} \cdot M^{\eta k + 2n}$ $(\alpha, [0, k-1])$-names to cover C. Finally, we have that we can cover G_k by no more than

$$\psi(k, n, \eta) \cdot N^{k/n} \cdot M^{\eta k + 2n}$$

$(\alpha, [0, k-1])$-names. Because $\mu(G_k) > 1 - \eta$, this means that

$$\frac{1}{k} \log \mathcal{N}(\alpha_0^{k-1}, \eta) \le \frac{1}{k} \log \psi(k, n, \eta) + \frac{1}{n} \log N + \left(\eta + \frac{2n}{k}\right) \log M.$$

Recall that once $(\eta + 2n/k) < \frac{1}{2}$, we have $\psi(k, n, \eta) \le 2^{k \cdot H(\eta + 2n/k)}$ and so

$$h_\mu(\alpha) = \lim \frac{1}{k} \log \mathcal{N}(\alpha_0^{k-1}, \eta) \le \frac{1}{n} \log \mathcal{N}(\mathcal{U}_0^{n-1}, \epsilon) + \eta \cdot \log M + H(\eta),$$

hence

$$h_\mu^+(\mathcal{U}) \le \frac{1}{n}\log\mathcal{N}(\mathcal{U}_0^{n-1},\epsilon) + \sqrt{\epsilon+\delta}\cdot\log M + H(\sqrt{\epsilon+\delta}).$$

Letting $\delta\to 0$ we get

$$h_\mu^+(\mathcal{U}) \le \frac{1}{n}\log\mathcal{N}(\mathcal{U}_0^{n-1},\epsilon) + \sqrt{\epsilon}\cdot\log M + H(\sqrt{\epsilon}),$$

as desired. ∎

4.5 Theorem: $h_\mu^+(\mathcal{U}) = h_\mu^-(\mathcal{U})$.

We already know that $h_\mu^+(\mathcal{U}) \ge h_\mu^-(\mathcal{U})$ (Proposition 3.6), so we only need to prove the other inequality. Before we turn to the proof, let us present some terminology and prove a combinatorial lemma.

Let Λ be a finite alphabet of M letters, $k,n\in\mathbb{N}$, $k >> n$, $0<\delta<1$ and $\omega=\omega_0^{k-1}$, a word of length k on Λ. (The symbol a_r^s stands for $a_r,\ldots,a_s$.) Denote $\Gamma=\Lambda^n$.

- An (n,k,δ)-packing is a pair $\mathcal{C}=(i_0^{m-1},\gamma_0^{m-1})$ where $0\le i_j\le k-n$, $\gamma_j\in\Gamma$, $j=0,\ldots,m-1, i_j+n-1<i_{j+1}$ and $\frac{m\cdot n}{k}>1-\delta$. (We think of an (n,k,δ)-packing as instructions to "almost" write a word of length k; we just fill it with the γ_j's, where γ_j starts at the i_j letter and there will be no more than δk letters to add.)
- An (n,k,δ)-packing for ω is an (n,k,δ)-packing, $\mathcal{C}=(i_0^{m-1},\gamma_0^{m-1})$, such that $\omega_{i_j}^{i_j+n-1}=\gamma_j$.
- If μ_1,μ_2 are probability distributions on Γ then

$$||\mu_1-\mu_2|| = \max_\gamma|\mu_1(\gamma)-\mu_2(\gamma)|.$$

- An (n,k,δ)-packing, $\mathcal{C}=(i_0^{m-1},\gamma_0^{m-1})$, induces a probability distribution on Γ, denoted by $P_{\mathcal{C}}$, by the formula $P_{\mathcal{C}}(\gamma)=\frac{1}{m}\#\{0\le j\le m-1|\gamma=\gamma_j\}$.
- If μ is a probability distribution on Γ and $\mathcal{C}$ is an (n,k,δ)-packing, then we say that $\mathcal{C}$ is (n,k,δ,μ), if $||\mu-P_{\mathcal{C}}||<\delta$. We say that ω is (n,k,δ,μ), if there is an (n,k,δ)-packing for ω, which is (n,k,δ,μ).

4.6 Lemma: *If μ is a probability distribution on Γ, with "average entropy"*

$$h_0 = -\frac{1}{n}\sum_{\gamma\in\Gamma}\mu(\gamma)\log\mu(\gamma),$$

then there exists a positive function $\varphi(\delta)$ such that $\varphi(\delta)\to 0$ as $\delta\to 0$ and such that if $0<\delta<\frac{1}{2}$ then, for any $k>n$, the number of words $\omega\in\Lambda^k$, which are (n,k,δ,μ), is at most $2^{k(h_0+\varphi(\delta))}$.

Proof: Fix $k > n$. We want to estimate the number of words $\omega = \omega_0^{k-1} \in \Lambda^k$, that are (n,k,δ,μ). For every such word ω we can choose an (n,k,δ)-packing $\mathcal{C} = (i_0^{m-1}, \gamma_0^{m-1})$ which is (n,k,δ,μ). In this way we define a map

$$\pi : \{\omega \in \Lambda^k | \omega \text{ is } (n,k,\delta,\mu)\} \to \{\mathcal{C} | \mathcal{C} \text{ is an } (n,k,\delta,\mu)\text{-packing}\}.$$

If $\mathcal{C} = (i_0^{m-1}, \gamma_0^{m-1})$ is an (n,k,δ)-packing, then $\frac{n \cdot m}{k} > 1 - \delta$. This means that $|\pi^{-1}(\mathcal{C})| \le |\Lambda|^{\delta k} = M^{\delta k}$. So we have that

$$\#\{\omega \in \Lambda^k | \omega \text{ is } (n,k,\delta,\mu)\} \le M^{\delta k} \#\{\mathcal{C} | \mathcal{C} \text{ is an } (n,k,\delta,\mu)\text{-packing}\}.$$

Let us now estimate the number of (n,k,δ,μ)-packings, $\mathcal{C} = (i_0^{m-1}, \gamma_0^{m-1})$: The number of sequences, i_0^{m-1}, such that $0 \le i_j \le k-n$, $i_j + n - 1 < i_{j+1}$ and $\frac{m \cdot n}{k} > 1 - \delta$ is at most $\sum_{j < \delta k} \binom{k}{j}$. From Lemma 2.3 we know that for $\delta < \frac{1}{2}$, this sums to something $\le 2^{H(\delta)k}$.

Fix such a sequence i_0^{m-1}. Let us now estimate the number of sequences, γ_0^{m-1}, such that the (n,k,δ)-packing $\mathcal{C} = (i_0^{m-1}, \gamma_0^{m-1})$ is (n,k,δ,μ).

Denote $\nu = \bigotimes_1^m \mu$, the product measure on Γ^m. If $\gamma_0^{m-1} \in \Gamma^m$, then

$$\begin{aligned}\nu(\gamma_0^{m-1}) &= \prod_{\gamma \in \Gamma} \mu(\gamma)^{\#\{0 \le j \le m-1 | \gamma = \gamma_j\}} = 2^{\sum_{\{\gamma | \mu(\gamma) \neq 0\}} \#\{0 \le j \le m-1 | \gamma = \gamma_j\} \cdot \log \mu(\gamma)} \\ &= 2^{m \sum_{\{\gamma | \mu(\gamma) \neq 0\}} \frac{1}{m} \#\{0 \le j \le m-1 | \gamma = \gamma_j\} \cdot \log \mu(\gamma)}.\end{aligned}$$

Now the function f: $\{(x_\gamma)_{\gamma \in \Gamma} \in \mathbb{R}^\Gamma \,|\, \sum x_\gamma = 1\} \to \mathbb{R}$, defined by

$$f(\vec{x}_\gamma) = \sum_{\{\gamma | \mu(\gamma) \neq 0\}} x_\gamma \cdot \log \mu(\gamma),$$

is continuous and so there is a positive function $\psi(\delta)$ such that $\psi(\delta) \to 0$ as $\delta \to 0$ and, if $\max_\gamma |x_\gamma - \mu(\gamma)| < \delta$, then $|f(\vec{x}_\gamma) - f(\vec{\mu(\gamma)})| < \psi(\delta)$ (note that ψ depends only on n, μ). So if $\gamma_0^{m-1} \in \Gamma^m$ is such that $\mathcal{C} = (i_0^{m-1}, \gamma_0^{m-1})$ is a (n,k,δ,μ)-packing, it follows that

$$\begin{aligned}\nu(\gamma_0^{m-1}) &= 2^{m \sum_{\{\gamma | \mu(\gamma) \neq 0\}} \frac{1}{m} \#\{0 \le j \le m-1 | \gamma = \gamma_j\} \cdot \log \mu(\gamma)} \\ &\ge 2^{m(\sum_{\{\gamma | \mu(\gamma) \neq 0\}} \mu(\gamma) \log \mu(\gamma) - \psi(\delta))} \ge 2^{k(-h_0 - \frac{\psi(\delta)}{n})},\end{aligned}$$

where the last inequality follows from the fact that $m < k/n$ and the definition of h_0. We conclude that an upper bound for the number of such sequences γ_0^{m-1} is $2^{k(h_0 + \psi(\delta)/n)}$. If we collect these estimations, we reach the conclusion that for $0 < \delta < \frac{1}{2}$

$$\begin{aligned}\#\{\omega \in \Lambda^k | \omega \text{ is } (n,k,\delta,\mu)\} &\le M^{\delta k} \cdot 2^{H(\delta)k} \cdot 2^{k(h_0 + \psi(\delta)/n)} \\ &\le 2^{k(h_0 + \psi(\delta)/n + H(\delta) + \delta \cdot \log M)},\end{aligned}$$

so $\varphi(\delta) = \psi(\delta)/n + H(\delta) + \delta \cdot \log M$ is our desired function. ∎

Proof of Theorem 4.5: We want to show that for an ergodic system $(X, \mathcal{B}, \mu, T)$ and a cover $\mathcal{U} = \{U_1, \ldots, U_M\}$ of X, we have $h_\mu^+(\mathcal{U}) \le h_\mu^-(\mathcal{U})$. As before, if the system is periodic, then, from the ergodicity, it must be a cyclic permutation on a finite set of atoms. Therefore $h_\mu^+(\mathcal{U}) = h_\mu^-(\mathcal{U}) = 0$. In the aperiodic case we can use the Strong Rohlin Lemma.

Let $\epsilon > 0$. We shall show that $h_\mu^+(\mathcal{U}) \le h_\mu^-(\mathcal{U}) + 2\epsilon$. From the definition of $h_\mu^-(\mathcal{U})$, we can find $n \in \mathbb{N}$ and a partition $\beta \succeq \mathcal{U}_0^{n-1}$ such that $\frac{1}{n} H_\mu(\beta) \le h_\mu^-(\mathcal{U}) + \epsilon$. As $\beta \succeq \mathcal{U}_0^{n-1}$, we can index the elements of β by sequences $i_0^{n-1} = i_0, \ldots, i_{n-1}$, such that if $\tilde{B}_{i_0^{n-1}}$ is one, then $T^j \tilde{B}_{i_0^{n-1}} \subset U_{i_j}$, $j = 0, \ldots, n-1$. We can assume that each sequence i_0^{n-1} corresponds to at most one element of β, for otherwise, we could unite these elements and get a coarser partition β', still refining $\mathcal{U}_0^{n-1}$, such that $\frac{1}{n} H_\mu(\beta') \le \frac{1}{n} H_\mu(\beta) \le h_\mu^-(\mathcal{U}) + \epsilon$. Set $\Gamma = \{1, \ldots, M\}^n$. So the elements of β are indexed by Γ. (If $\gamma \in \Gamma$ does not correspond to an element of β, in the above way, we set $\tilde{B}_\gamma = \emptyset$.) In this way, the partition β defines a probability distribution ν on Γ, defined by $\nu(\gamma) = \mu(\tilde{B}_\gamma)$, and we have that $h_0 = \frac{1}{n} H_\mu(\beta)$ is the "average entropy" (see Lemma 4.6) of ν.

Choose $\delta > 0$ (in a manner specified later) and let F be a base for a strong Rohlin tower (with respect to β) of height n and error$\le \delta^2$. Denote the atoms of $\beta|_F$ by B_γ, $\gamma \in \Gamma$ (where $B_\gamma = \tilde{B}_\gamma \cap F$) and define a partition $\tilde{\alpha} = \{\tilde{A}_1, \ldots, \tilde{A}_M\}$ of $E = \bigcup_{j=0}^{n-1} T^j F$ by $\tilde{A}_m = \bigcup \{T^j B_{i_0^{n-1}} | j \in \{0, \ldots, n-1\}, i_j = m\}$. Note that $\tilde{A}_m \subset U_m$. Extend $\tilde{\alpha}$ to a partition α of X refining $\mathcal{U}$, in some way. The set of indices of elements of α, Λ (the alphabet in which α-names are written), contains $\{1, \ldots, M\}$ and we can always build α such that $|\Lambda| \le 2M$. We slightly abuse our notation and denote $\Gamma = \Lambda^n$. In this way, ν is still a probability distribution on Γ.

Claim: *If δ is small enough, then $h_\mu(\alpha) \le h_0 + \epsilon$.*

Once we prove this claim, we are done, because then

$$h_\mu^+(\mathcal{U}) \le h_\mu(\alpha) \le h_0 + \epsilon \le h_\mu^-(\mathcal{U}) + 2\epsilon.$$

Proof of Claim: For $k >> n$, we look at the function $f_k(x) = \frac{1}{k} \sum_0^{k-1} \chi_E(T^j x)$. We have that $0 \le f_k \le 1$ and $\int f_k > 1 - \delta^2$. Therefore

$$\delta \cdot \mu(\{x | 1 - f_k(x) > 1 - \delta\}) \le \int_{\{x|1-f_k(x)>1-\delta\}} 1 - f_k \le \int 1 - f_k \le \delta^2$$
$$\Rightarrow \mu(\{x | f_k(x) \ge 1 - \delta\}) \ge 1 - \delta.$$

Denote $G_1^k = \{x|f_k(x) \geq 1-\delta\}$. For $x \in G_1^k$, there are at most δk times $0 \leq i \leq k-1$ such that $T^i x \notin E$. Define

$$G_2^k = \left\{x \middle| \left|\frac{1}{k}\sum_0^{k-n} \chi_A(T^i x) - \mu(A)\right| < \delta, A \in \beta|_F \cup \{F\}\right\}.$$

Let us see what we can say about the $(\alpha, [0, k-1])$-name of an element x of $G_1^k \cap G_2^k$. Fix such an x and denote by $i_0 < \cdots < i_{m-1}$ the times between 0 to $k-n$ at which x visits F. We have that $0 \leq i_j \leq k-n$, $i_j + n - 1 < i_{j+1}$ (that is because the height of the tower is n). Except for at most $2n$ times (n at the beginning and n at the end), x visits E exactly at the times $i_j, \ldots, i_j + n - 1$, $j = 1, \ldots, m-1$. Therefore, we must have

$$n \cdot m \geq (1-\delta)k - 2n \Rightarrow \frac{n \cdot m}{k} \geq 1 - \left(\delta + \frac{2n}{k}\right).$$

Denote the $(\alpha, [0, k-1])$-name of x by $\omega = \omega_0^{k-1}$ ($\omega_i \in \Lambda$), and $\gamma_j = \omega_{i_j} \cdots \omega_{i_j+n-1} \in \Gamma$, $j = 0, \ldots, m-1$. We have that $\mathcal{C} = (i_0^{m-1}, \gamma_0^{m-1})$ is an $(n, k, \delta + 2n/k)$-packing for ω. Let us now see what we can say about the distribution $P_{\mathcal{C}}$ this packing induces on Γ.

For $0 \leq r \leq k-n$, we have that $T^r x \in B_\gamma$ if and only if there is a $0 \leq j \leq m-1$ such that $r = i_j$ and $\gamma = \gamma_j$. Therefore, because $x \in G_2^k$,

- $\forall \gamma \in \Gamma \quad |\frac{1}{k}\#\{0 \leq j \leq m-1 | \gamma = \gamma_j\} - \mu(B_\gamma)| < \delta$;
- $|\frac{m}{k} - \mu(F)| < \delta$.

Note that $\mu(F) > (1-\delta)/n$, so if δ is sufficiently small, we can guarantee that $|k/m - 1/\mu(F)|$ would be arbitrarily small and in turn we can guarantee that for every $\gamma \in \Gamma$,

$$\left|\frac{k}{m} \cdot \frac{1}{k}\#\{0 \leq j \leq m-1 | \gamma = \gamma_j\} - \frac{\mu(B_\gamma)}{\mu(F)}\right| = |P_{\mathcal{C}}(\gamma) - \nu(\gamma)|$$

would be arbitrarily small. This is to say that $||P_{\mathcal{C}} - \nu||$ is arbitrarily small. We see that there is a positive function $\psi(\delta)$, independent of k, such that $\psi(\delta) \to 0$ as $\delta \to 0$ and such that, if $x \in G_1^k \cap G_2^k$ and ω is its $(\alpha, [0, k-1])$-name, then ω is $(n, k, \psi(\delta) + 2n/k, \nu)$.

Recall the function φ from Lemma 4.6. There is an $\eta_0 > 0$ such that for every $0 < \eta < \eta_0$, $\varphi(\eta) < \epsilon$. Choose k to be large enough so that $2n/k < \eta_0/2$ and the error δ of the tower to be so small such that $\psi(\delta) < \eta_0/2$, and conclude, from Lemma 4.6, that the number of $(\alpha, [0, k-1])$-names of elements of $G_1^k \cap G_2^k$ is at most $2^{k(h_0+\epsilon)}$. From the ergodicity, we know that for large enough k,

$\mu(G_1^k \cap G_2^k) > 1 - 2\delta$, so we have

$$h_\mu(\alpha) = \lim \frac{1}{k} \log \mathcal{N}(\alpha_0^{k-1}, 2\delta) \leq h_0 + \epsilon,$$

as desired. ∎

Remarks:

- If (X, T) is totally ergodic, i.e., (X, T^n) is ergodic for every $n \in \mathbb{N}$, then we can look at expressions like $h_\mu^e(\mathcal{U}_0^{n-1}, T^n)$. It follows from the definition that $h_\mu^e(\mathcal{U}, T) = \frac{1}{n} h_\mu^e(\mathcal{U}_0^{n-1}, T^n)$. This enables us to prove the last theorem without any hard work. We know from Theorem 4.4 that $h_\mu^e(\mathcal{U}, T) = h_\mu^+(\mathcal{U}, T)$ and therefore $h_\mu^+(\mathcal{U}, T) = \frac{1}{n} h_\mu^+(\mathcal{U}_0^{n-1}, T^n)$. But then, Proposition 3.6, (which is elementary), gives $h_\mu^-(\mathcal{U}, T) = \lim \frac{1}{n} h_\mu^+(\mathcal{U}_0^{n-1}, T^n) = h_\mu^+(\mathcal{U}, T)$ and this gives the desired result.
- The definitions of $h_\mu^+(\mathcal{U}), h_\mu^-(\mathcal{U})$ were introduced in [R] and discussed also in [Ye], [HMRY]. There, a proof of their equality was given only in the case where (X, T) is a t.d.s and $\mathcal{U}$ is an open cover. The proof was based on a reduction to a uniquely ergodic case and then a use of a variational inequality, proved in [GW].
- The definition of $h_\mu^e(\mathcal{U})$ is new. This definition helps us to prove directly a slight generalization of the variational inequality, proved in [GW] and mentioned above, to the non-topological case (Theorem 6.1).
- The proofs of Theorems 4.2, 4.4, 4.5 and Lemma 4.6 are based on ideas of B. Weiss and E. Glasner

5. Ergodic decomposition for h_μ^+, h_μ^-

5.1 Theorem (Proposition 5 in [HMRY]): *Let $\mathcal{U} = \{U_1, \dots, U_M\}$ be a cover of X, and $\mu = \int \mu_x d\mu(x)$ the ergodic decomposition of μ with respect to T. Then*

$$h_\mu^+(\mathcal{U}, T) = \int h_{\mu_x}^+(\mathcal{U}, T) d\mu(x), \quad h_\mu^-(\mathcal{U}, T) = \int h_{\mu_x}^-(\mathcal{U}, T) d\mu(x).$$

5.2 Corollary: $h_\mu^+(\mathcal{U}) = h_\mu^-(\mathcal{U})$.

Proof: It follows immediately from the above and the ergodic case (Theorem 4.5). ∎

From now on we will denote the number $h_\mu^+(\mathcal{U}, T) = h_\mu^-(\mathcal{U}, T) (= h_\mu^e(\mathcal{U}, T)$ in the ergodic case) simply by $h_\mu(\mathcal{U}, T)$ or $h_\mu(\mathcal{U})$ or $h(\mathcal{U})$, when no ambiguity can occur.

6. Variational relations

As always, let $\mathcal{U} = \{U_1, \ldots, U_M\}$ be a cover of the m.t.d.s $(X, \mathcal{B}, \mu, T)$. We can define the "*combinatorial entropy*" of $\mathcal{U}$ as

$$h_c(\mathcal{U}, T) = \lim_n \frac{1}{n} \log \mathcal{N}(\mathcal{U}_0^{n-1}),$$

where $\mathcal{N}(\mathcal{V})$ is the minimum number of elements of $\mathcal{V}$ needed to cover the whole space. Note that the sequence $\log \mathcal{N}(\mathcal{U}_0^{n-1})$ is sub-additive, hence the limit exists. If (X, T) is a t.d.s and $\mathcal{U}$ is an open cover, then we denote $h_{top}(\mathcal{U}, T) = h_c(\mathcal{U}, T)$.

The next theorem was proved in [GW] for topological dynamical systems and measurable covers. We give here a simple proof for the non-topological case that uses the definition of $h_\mu^e(\mathcal{U})$.

6.1 Theorem: $h_\mu(\mathcal{U}) \leq h_c(\mathcal{U})$.

Proof: First, if the system is ergodic, then $h_\mu(\mathcal{U}) = \lim \frac{1}{n} \log \mathcal{N}(\mathcal{U}_0^{n-1}, \frac{1}{2})$ and, as $\mathcal{N}(\mathcal{U}_0^{n-1}, \frac{1}{2}) \leq \mathcal{N}(\mathcal{U}_0^{n-1})$, we have

$$h_\mu(\mathcal{U}) \leq \lim \frac{1}{n} \log \mathcal{N}(\mathcal{U}_0^{n-1}) = h_{top}(\mathcal{U}),$$

as desired. In the non-ergodic case, let $\mu = \int \mu_x d\mu(x)$ be the ergodic decomposition of μ. By Theorem 5.1, $h_\mu(\mathcal{U}) = \int h_{\mu_x}(\mathcal{U}) d\mu(x)$, so from the first part we see that $h_\mu(\mathcal{U}) \leq h_c(\mathcal{U})$. ∎

Remark: Another simple proof of the above uses the definition of $h_\mu^-(\mathcal{U})$:

$$H_\mu(\mathcal{U}_0^{n-1}) = \inf_{\alpha \succeq \mathcal{U}_0^{n-1}} H_\mu(\alpha) \leq \inf_{\alpha \succeq \mathcal{U}_0^{n-1}} \log |\alpha| \leq \log \mathcal{N}(\mathcal{U}_0^{n-1})$$

$$\Rightarrow h_\mu(\mathcal{U}) = \lim \frac{1}{n} H_\mu(\mathcal{U}_0^{n-1}) \leq \lim \frac{1}{n} \log \mathcal{N}(\mathcal{U}_0^{n-1}) = h_c(\mathcal{U}).$$

From this stage until the end of this paper we assume that (X, T) is a t.d.s. We denote by $\mathcal{M}_T(X)$ the set of T-invariant probability measures on X, and by $\mathcal{M}_T^e(X)$ the set of ergodic ones. Also, $\mathcal{C}_X^o$ will denote the set of finite open covers of X.

In [BGH], the following theorem was proved:

6.2 Theorem (Theorem 1 in [BGH]): *If $\mathcal{U} \in \mathcal{C}_X^o$, then there exists $\mu \in \mathcal{M}_T(X)$ such that $h_\mu(\mathcal{U}) \geq h_{top}(\mathcal{U})$.*

In light of Theorem 6.1 we have that, for every $\mathcal{U} \in \mathcal{C}_X^o$, one can find a measure $\mu \in \mathcal{M}_T(X)$ such that $h_\mu(\mathcal{U}) = h_{top}(\mathcal{U})$. In fact, theorem 7 in [HMRY] now becomes

6.3 Corollary: *For every $\mathcal{U} \in \mathcal{C}_X^o$, one can find a measure $\mu \in \mathcal{M}_T^e(X)$ such that $h_\mu(\mathcal{U}) = h_{top}(\mathcal{U})$.*

Proof: Choose $\mu \in \mathcal{M}_T(X)$ such that $h_\mu(\mathcal{U}) = h_{top}(\mathcal{U})$, and let $\mu = \int \mu_x d\mu(x)$ be its ergodic decomposition. We know that

$$h_{top}(\mathcal{U}) = h_\mu(\mathcal{U}) = \int h_{\mu_x}(\mathcal{U}) d\mu(x)$$

and that $h_{\mu_x}(\mathcal{U}) \leq h_{top}(\mathcal{U})$. So we must have $h_{\mu_x}(\mathcal{U}) = h_{top}(\mathcal{U})$ for $[\mu]$ a.e. x. ∎

We conclude from the above the classical variational principle: First we state a technical lemma, taken from [Ye].

6.4 Lemma: *For any $\epsilon > 0$, $\mu \in \mathcal{M}_T(X)$ and $\alpha = \{A_1, \ldots, A_M\} \in \mathcal{P}_X$, there exists an open cover $\mathcal{U} \in \mathcal{C}_X^o$ such that for every partition $\beta \succeq \mathcal{U}$ one has $H_\mu(\alpha|\beta) < \epsilon$.*

6.5 Theorem (The Variational Principle):

(a) *For every $\mu \in \mathcal{M}_T(X)$, $h_\mu(T) \leq h_{top}(T)$.*

(b) $\sup_{\mu \in \mathcal{M}_T^e(X)} h_\mu(T) = h_{top}(T)$.

Proof: To prove (a), we first show that for each $\mu \in \mathcal{M}_T(X)$, $h_\mu(T) = \sup_{\mathcal{U} \in \mathcal{C}_X^o} h_\mu(\mathcal{U}, T)$. If this is done, then from Theorem 6.1 we get

$$h_\mu(T) \leq \sup_{\mathcal{U} \in \mathcal{C}_X^o} h_{top}(\mathcal{U}, T) = h_{top}(T).$$

It follows from the definition that for any cover $\mathcal{U}$ of X, we have $h_\mu(\mathcal{U}, T) \leq h_\mu(T)$, so one inequality is clear. For the other inequality, fix a partition $\alpha = \{A_1, \ldots, A_M\}$ of X and $\epsilon > 0$. We need to find an open cover $\mathcal{U}$ of X such that $h_\mu(\alpha, T) \leq h_\mu(\mathcal{U}, T) + \epsilon$. By the preceding lemma and from the fact that for any $\beta \in \mathcal{P}_X$ one has $h_\mu(\alpha) \leq h_\mu(\beta) + H(\alpha|\beta)$, we have $\mathcal{U} \in \mathcal{C}_X^o$ such that

$$h_\mu(\mathcal{U}, T) = \inf_{\beta \succeq \mathcal{U}} h_\mu(\beta, T) \geq \inf_{\beta \succeq \mathcal{U}} (h_\mu(\alpha, T) - H_\mu(\alpha|\beta)) \geq h_\mu(\alpha, T) - \epsilon.$$

To prove (b), note that from (6.3) we know that for any $\mathcal{U} \in \mathcal{C}_X^o$ we can find $\mu \in \mathcal{M}_T^e(X)$ such that $h_\mu(\mathcal{U}, T) = h_{top}(\mathcal{U}, T)$. This gives us

$$\sup_{\mu \in \mathcal{M}_T^e(X)} h_\mu(T) \geq h_{top}(\mathcal{U}, T) \Rightarrow \sup_{\mu \in \mathcal{M}_T^e(X)} h_\mu(T) \geq \sup_{\mathcal{U} \in \mathcal{C}_X^o} h_{top}(\mathcal{U}, T) = h_{top}(T).$$

Together with (a), we get equality, which is (b). ∎

References

[BGH] F. Blanchard, E. Glasner and B. Host, *A variation on the variational principle and applications to entropy pairs,* Ergodic Theory and Dynamical Systems **17** (1997), 29–43.

[GW] E. Glasner and B. Weiss, *On the interplay between measurable and topological dynamics,* Handbook of dynamical systems Vol. 1B, Elsevier B. V., Amsterdam, 2006, pp. 597–648.

[HMRY] W. Huang, A. Maass, P. P. Romagnoly and X. Ye, *Entropy pairs and local Abramov formula for measure-theoretic entropy for a cover,* Ergodic Theory and Dynamical Systems **24** (2004), 1127–1153.

[R] P. P. Romagnoli, *A local variational principle for the topological entropy,* Ergodic Theory and Dynamical Systems **23** (2003), 1601–1610.

[Ru] D. J. Rudolph, *Fundamentals of Measurable Dynamics: Ergodic Theory on Lebesgue Spaces,* Oxford University Press, Clarendon Press, 1990.

[Sh1] P. C. Shields, *The Ergodic Theory of Discrete Sample Paths,* Graduate Studies in Mathematics Vol. 13, American Mathematical Society, Providence, RI, 1996.

[Sh2] P. C. Shields, *The Theory of Bernoulli Shifts,* Reprint of 1973 University of Chicago Press edition (available on the Web).

[Ye] W. Huang, X. Ye and G. Zhang, *A local variational principle for conditional entropy,* Ergodic Theory and Dynamical Systems **26** (2006), 219–245.