Furstenberg x2x3 Conjecture and Rudolph's Theorem

by

Kobe Marshall-Stevens

MA4K9 Dissertation

Submitted to The University of Warwick

Mathematics Institute

April, 2020

Contents

1 Motivating the Conjecture and Rudolph's Theorem **1**

2 The Symbolic Representation and Correspondence with $\mathbb{T}$ **4**

3 Lifted Measures and Inverse Limits **9**

4 Disintegration of Measures, Symmetric Points and Weak Convergence **16**

5 Reducing to the Case of Zero Entropy **29**

6 Extension and Developments of Rudolph's Theorem **37**

1 Motivating the Conjecture and Rudolph's Theorem

The starting point of the conjecture of Furstenberg is in a theorem about dynamics of the circle under the action of non-lacunary semi-groups of the integers. Briefly, a non-lacunary semi-group of the integers is one not generated by powers of a single integer. The example we shall focus on, and indeed the one that the conjecture addresses, is that of the the non-lacunary semi-group generated by coprime natural numbers p and q.

Let us make the convention that $\mathbb{N} = \{0, 1, 2, ...\}$ and define the circle, $\mathbb{T}$, to be $[0, 1]$ with the endpoints identified, we will use additive notation. Throughout, we will consider coprime natural numbers $p, q > 1$ and maps $T, S : \mathbb{T} \to \mathbb{T}$ defined by $Tx = px \, (mod\, 1)$ and $Sx = qx \, (mod\, 1)$ respectively. We shall work primarily in the setting of what we shall hereafter call *Borel probability spaces*, $(X, \mathcal{B}, \mu)$, where X is a compact metric space, $\mathcal{B}$ is the σ-algebra of Borel sets (generated by the open sets in X) and μ is a probability measure defined on $\mathcal{B}$ (so $\mu(X) = 1$).

Remark 1.1. *For most results throughout our work the coprime assumption will not be necessary, though it does play a crucial role in establishing the final results for Rudolph's theorem.*

Originally proven in greater generality (see [F]) by Furstenberg we have the following theorem for dynamics of the circle. It is worth noting that the proof of this theorem is highly non-trivial, though a simpler proof may be found in [Bo], and so we shall only state the result here in order to motivate our work.

Theorem 1.2 (Furstenberg 67')**.** *Let A be an infinite, closed subset of $\mathbb{T}$ that is invariant under both T and S, in the sense that for any $x \in A$, $\{p^m q^n x \mid m, n \in \mathbb{N}\} \subseteq A$. Then $A = \mathbb{T}$.*

Note that there exists numerous finite invariant sets; for example, consider the orbits under T and S of any rational point in $\mathbb{T}$. In order to motivate Furstenberg's conjecture from the above theorem, we first recall the definition of the support of a Borel measure:

Definition 1.3. *For a Borel probability space, $(X, \mathcal{B}, \mu)$, the support of μ, $\operatorname{supp} \mu$, is defined to be the set of points in X such that every open neighbourhood of the point has positive measure.*

This set; $\operatorname{supp} \mu$ is closed by definition, as it's compliment is the union of all null open sets of X. If μ is non-atomic, $\operatorname{supp} \mu$ is necessarily infinite; to see this, suppose that μ is supported on finitely many points then, as it's compliment is a null set, at least one such point in the support has positive measure and hence μ is atomic.

We now establish a corollary of Furstenberg's dynamical result that serves as motivation for the conjecture we shall focus on.

Proposition 1.4. *Furstenberg's theorem (theorem 1.2) implies that every non-atomic, invariant, Borel probability measure on the circle has full support, i.e.* $\operatorname{supp}\mu = \mathbb{T}$.

Proof. It suffices to show that $\operatorname{supp}\mu$ is invariant under both T and S. For $x \in \operatorname{supp}\mu$ consider an open neighbourhood, V, of Tx. T is continuous so the set $T^{-1}(V)$ is open and contains x, hence it has positive measure. As μ is invariant, V also has positive measure, and so Tx is an element of $\operatorname{supp}\mu$. Similarly $\operatorname{supp}\mu$ is S-invariant. Thus by Furstenberg's theorem we have that $\operatorname{supp}\mu = \mathbb{T}$. □

It is interesting to note that the converse to this result is non-trivial, and likely as difficult to prove as the original theorem. Russell Lyons states in [Ly] that this should follow easily, however his intended argument relied on being able to find an irrational number in any such infinite set; though there exist closed infinite sets that contain no irrationals (e.g. $\{2^{-k} \mid k \in \mathbb{N}\} \cup \{0\}$).

We now extend the definition of ergodicity for measure preserving systems under the action of one transformation to that of two transformations, in order to formulate the conjecture.

Definition 1.5. *An invariant Borel probability measure for a dynamical system* $(X, \mathcal{B}, \mu, (T, S))$ *is said to be ergodic with respect to both* T *and* S *(or jointly ergodic) if, for any* $E \in \mathcal{B}$*, with* $T^{-1}(E) = E$ *and* $S^{-1}(E) = E$ *then either* $\mu(E) = 0$ *or* $\mu(E) = 1$.

It is worth noting that joint ergodicity is a weaker condition to assume than ergodicity with respect to a single transformation. Such an assumption would greatly simplify our following work; for further discussion we refer the reader to section 6.

This corollary of the dynamical theorem led Furstenberg to conjecture the following measure rigidity result.

Conjecture 1.6. *Any invariant, ergodic, Borel probability measure on* $\mathbb{T}$ *with respect to both transformations* T *and* S *is either atomic or the Lebesgue measure.*

The reason behind describing the conjecture as a measure rigidity result is that, if true, then the set of ergodic measures for the circle under the joint action of T and S is relatively small; consisting of the Lebesgue measure and linear combinations of Dirac point measures. However, if we consider each transformation T or S individually there are uncountably many fully supported ergodic measures on the circle; for example, by considering Bernoulli measures, on the one-sided full shift on p symbols, and the factor map given by base p expansion, we get an uncountable family of fully supported ergodic measures on the circle.

This conjecture of Furstenberg has received much attention and though it is still open, the work of Rudolph in [R], which generalised prior results of Lyons [Ly], provides the strongest known partial result, which we shall hereafter refer to as *Rudolph's theorem.*

Theorem 1.7. *(Rudolph 89') Let μ be an invariant, Borel probability measure on $\mathbb{T}$ that is ergodic with respect to both T and S. Then, if μ is not Lebesgue measure both T and S have zero entropy.*

This theorem reduces conjecture 1.6 to the case of measures for which T and S have zero measure theoretic entropy; however this does not solve the conjecture as there exist dynamical systems that admit non-atomic, invariant, ergodic, Borel probability measures with zero entropy. Two examples of such are provided by minimal systems of symbolic shifts of zero entropy, and so called Rank-1 transformations; for literature about these examples we refer the reader to the chapter on the Jewett-Kreiger theorem in [DGS] and the section "Rank-1 has zero entropy" in [K] respectively.

The goal of this report is to present a self contained, streamlined exposition of the original arguments given by Rudolph, expanding on details and clarity where possible. In order to do this we will need to introduce and utilise the techniques of inverse limits, disintegration of measures and entropy with respect to σ-algebras; all of which are likely unseen in a first course in ergodic theory. Though there exist various proofs of the above theorem, all other works utilise different constructions and techniques to the original work (e.g. Radon-Nikodym derivatives in [PY], invertible extensions and Pinsker algebras in [ELW]).

Throughout, we assume knowledge of basic notions and results of ergodic theory, particularly those of measure preserving transformations, the ergodic theorems and entropy for partitions. We shall also use results from measure theory and functional analysis, giving references where applicable. For a good introduction to these subjects we refer the reader to [EW], [ELW] and [C].

We now give an outline of the report; in section 2 we re-frame the dynamics of the circle, under the action of T and S. This is done in the setting of symbolic dynamics, in order to later construct an invertible dynamical system uniquely lifted from the circle, for a given measure μ. In order to do this we will need to introduce the notion of an inverse limit of dynamical systems, which is the focus of section 3. We use this new system in section 4 and establish the disintegration theorem, a central result we utilise frequently throughout sections 4 and 5. The latter part of section 4 is devoted to the notions of the so called δ probability distributions and symmetric points. These notions will allow us to conclude that in the case that almost every point is symmetric, our measure must be Lebesgue. Section 5 introduces and applies the theory of entropy with respect to σ-algebras; this is done in order to prove that under the assumption of positive entropy, almost every point is symmetric, which establishes Rudolph's theorem. Finally, section 6 concludes the report by collecting several extensions, generalisations and relaxations of Rudolph's theorem.

2 The Symbolic Representation and Correspondence with $\mathbb{T}$

The first step in the proof of Rudolph's theorem is to represent the dynamics of the circle, under the actions of T and S, in terms of arrays under the action of a left and down shift respectively. We will construct an almost (in a precise sense) 1-1 correspondence between points of the circle and such arrays, allowing us to view measures on the circle as measures on the set of arrays. From this symbolic description we are then able to obtain the results needed to prove the theorem.

We begin by defining a partition of the circle into pq intervals according to the pre-image of the circle under the actions of both T and S,

$$I_j := \left[\frac{j}{pq}, \frac{j+1}{pq}\right] \text{ for } j \in \{0, \ldots, pq-1\}.$$

For each of the transformations we now associate transition matrices, whose entries are determined by which intervals, I_j, of the partition are contained in the images of intervals, I_i, under the maps T and S. Precisely, we define a pq by pq transition matrix, $M_T = (a_{ij})$, for T where $a_{ij} = 1$ if $I_j \subset T(I_i)$, and zero otherwise. Similarly for S we define $M_S = (b_{ij})$, with $b_{ij} = 1$ if $I_j \subset S(I_i)$, and zero otherwise.

Thus in the simplest case, where we take $p = 2$ and $q = 3$, we get the following transition matrices:

$$M_T = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \quad M_S = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}$$

We now let $\Sigma = \{0, \ldots, pq-1\}$ and consider the one-sided shift space, $\Sigma^+_{M_T}$; namely the set of infinite one-sided sequences of elements of Σ where a symbol j may follow a symbol i if and only if $a_{ij} = 1$.

We now introduce arrays and show that there is an (almost) one-to-one correspondence between such arrays and the circle.

Define $Y \subset \Sigma^{\mathbb{N}^2}$ to be all arrays whose rows are sequences in $\Sigma^+_{M_T}$ and columns are sequences in $\Sigma^+_{M_S}$; we will denote by $y(i,j)$ the symbol determined by the pair $(i,j) \in \mathbb{N}^2$. With this notation, for any $y \in Y$ and fixed $i \in \mathbb{N}$, $(y(k,i))_{k\in\mathbb{N}} \in \Sigma^+_{M_T}$ and $(y(i,k))_{k\in\mathbb{N}} \in \Sigma^+_{M_S}$.

Given any point, $x \in \mathbb{T}$, we may associate to it an array, $y_x \in Y$, defined by $y_x(i,j) = k$

if $T^iS^j(x) \in I_k$. To see that such an array lies in Y, we note that if $y_x(i,j) = k$ and $y_x(i+1,j) = l$ for some $i,j \in \mathbb{N}$ then by definition we must have $T^iS^j(x) \in I_k$ and $T^{i+1}S^j(x) \in I_l$. Hence we conclude that $T(I_k) \cap I_l \neq \varnothing$, which by construction of the partition implies that $I_l \subset T(I_k)$; thus we see that the rows of y_x lie in $\Sigma^+_{M_T}$. A symmetric argument for S shows that the columns of y_x lie in $\Sigma^+_{M_S}$.

Such an array will arise uniquely from a point $x \in \mathbb{T}$, providing none of the forward images of x under T and S are contained in the intersection of two intervals of the partition given by the I_j. To see this let

$$V := \left\{ x \in \mathbb{T} \,\middle|\, x = \frac{t}{p^n q^m}, t, n, m \in \mathbb{N} \right\} \subsetneq \mathbb{T},$$

and note that for any $x = \frac{t}{p^n q^m} \in V$ we have that T^aS^bx is contained in two of the intervals in the partition given by the I_j when $a \geq n-1, b \geq m-1$. If we were to make a choice for the symbol $y_x(a,b)$ for some such a and b above, we would then determine all symbols for y_x. The reason for this is that the choice of symbol $y_x(a,b)$ corresponds to making a choice of one of two intervals of the partition given by the I_j, either to the left or to the right of the point $T^aS^b(x) \in [0,1)$. Such a left or right interval choice then forces a respective left or right interval choice for points $T^nS^m(x) \in [0,1)$ for each $n,m \in \mathbb{N}$. Thus, in order to ensure $y_x \in Y$, our choice of symbol $y_x(n,m)$ is forced by our initial left or right choice above; in order for the rows and columns of y_x to be in $\Sigma^+_{M_T}$ and $\Sigma^+_{M_S}$ respectively.

We've shown that given any point $x \in \mathbb{T} \setminus V$ we can construct a unique array $y_x \in Y$ as defined above, and for $x \in V$ we have two such arrays that may arise in the above construction. We now perform this construction in reverse.

Consider a finite sequence, $(x_0, \ldots, x_n)$, of elements in Σ such that $a_{x_i x_{i+1}} = 1$ for all $i \in \{0, \ldots, n-1\}$. As $T^{-(n-1)}(I_{x_n})$ is a disjoint union of intervals of length $p^{-n}q$, we have a corresponding interval (of length $p^{-n}q$) of the form,

$$\bigcap_{j=0}^{n-1} T^{-j}(I_{x_j}) = \left[\frac{t}{p^n q}, \frac{t+1}{p^n q}\right] \text{ for some } t \in \mathbb{N}.$$

Thus for any sequence, $(x_i)_{i=0}^{\infty}$, in $\Sigma^+_{M_T}$ we may (uniquely, provided the point constructed does not lie in V) associate a point on the circle, given by

$$x = \bigcap_{j=0}^{\infty} T^{-j}(I_{x_j}) \in \mathbb{T}.$$

Similarly, we perform the above construction in the same manner for S, except now we require $b_{x_i x_{i+1}} = 1$ for all $i \in \{0, \ldots, n-1\}$; so we see that each sequence in $\Sigma^+_{M_T}$ or $\Sigma^+_{M_S}$

gives rise to a corresponding point on the circle.

We will now show, similarly to the above, that from an array in Y we recover a corresponding point in $\mathbb{T}$ which is determined by the joint action TS. To do this we will need the following lemma:

Lemma 2.1. *For $k, l \in \Sigma$ and $i \in \mathbb{N}$, all $y \in Y$ such that $y(i,i) = k$ and $y(i+1,i+1) = l$ agree on the symbols $y(i,i+1)$ and $y(i+1,i)$. In other words every $y \in Y$ is determined by its diagonal symbols, $(y(i,i))_{i\in\mathbb{N}}$.*

Proof. To prove this lemma we consider directed graphs whose vertices are labelled by symbols in Σ; we show that for adjacency matrices corresponding to applications of the maps T and S there are unique intermediate vertices determining the symbols $y(i,i+1)$ and $y(i+1,i)$.

As the map $STx = TSx = pqx(mod 1)$ takes any interval I_i to the whole of $\mathbb{T}$, we have that the matrices $M_T M_S$ and $M_S M_T$ have positive entries. Now summing a row or column of $M_T M_S$ gives exactly pq and thus every entry must be equal to 1. A symmetric argument shows that $M_S M_T$ has every entry equal to 1.

We now consider two directed graphs, both with vertices labelled by symbols in Σ and each one with an adjacency matrix determined by M_T and M_S respectively. Suppose we are at a vertex labelled by the symbol k and travel to a vertex labelled by the symbol l, first moving via a single directed edge determined by M_S and then by a single directed edge determined by M_T. The number of possible ways of taking such a path between k and l is determined by $M_T M_S$, which by the above means we have a unique choice of intermediate vertex; the label of this intermediate vertex corresponds to the unique symbol for $y(i+1,i)$. Similarly, if we consider moving first by a single directed edge determined by M_T and then by a single directed edge determined by M_S, we have a unique symbol for $y(i,i+1)$. Thus we see that the diagonal symbols entirely determine the array. □

By the above lemma, any $y \in Y$ is determined entirely by the diagonal symbols, $(y(i,i))_{i\in\mathbb{N}}$, and so given any $y \in Y$ we may define

$$x = \bigcap_{i=0}^{\infty} S^{-i}T^{-i}(I_{y(i,i)}) \in \mathbb{T}.$$

Following our above setup we have a well defined map between our space, Y, of arrays and the circle, $\mathbb{T}$.

Definition 2.2. *Define the map $\varphi : Y \to \mathbb{T}$ by setting for any $y \in Y$*

$$\varphi(y) = \bigcap_{i=0}^{\infty} S^{-i}T^{-i}(I_{y(i,i)}) \in \mathbb{T}.$$

With this definition and our above lemma we have that φ is one-to-one onto $\mathbb{T} \setminus V$ and two-to-one onto V. We shall see later on that the set V is not an obstruction, as under sensible assumptions it will be measure theoretically negligible, allowing us to identify the measure preserving systems involving Y and $\mathbb{T}$.

We now prove a lemma that will aid us in section 4, ensuring our constructed sequence of δ measures converges weakly to the Lebesgue measure. First let us note that the set $\varphi^{-1}(V)$ is forward and backwards invariant under the action of the left and down shifts on the space Y, which we shall denote by $\widehat{T}$ and $\widehat{S}$ respectively.

Lemma 2.3. *Any Σ_{M_T} horizontal ray in $\Sigma^{\mathbb{N}^2}$, namely a sequence of symbols $(i_{(n,m)}, i_{(n+1,m)}, ...)$ for some $n, m \in \mathbb{N}$, determines the symbols $y(i,j)$ for any $i \geq n$ and $j \geq m$ of any $y \in Y$ such that $y(n+k,m) = i_{(n+k,m)}$ for all $k \in \mathbb{N}$, provided that $\varphi(y) \notin V$. A symmetric result holds for any Σ_{M_S} vertical ray in $\Sigma^{\mathbb{N}^2}$ with symbols to the right being determined.*

Proof. As we assume that $\varphi(y) \notin V$ and have that $\varphi^{-1}(V)$ is forward invariant under the actions of $\widehat{T}$ and $\widehat{S}$ we must have that $\varphi(\widehat{T}^n\widehat{S}^m(y)) \notin V$ also. Hence, knowledge of all symbols $y(n+t,m)$ for any $t \geq 0$ determines a unique point (by our above work)

$$x = \varphi(\widehat{T}^n\widehat{S}^m(y)) = \bigcap_{t=0}^{\infty} T^{-t}(I_{(n+t,m)}).$$

Then by construction $x \notin V$ and so x has a unique pre-image under φ, such that

$$y(i,j) = \varphi^{-1}(x)(i-n, j-m)$$

for $i \geq n$ and $j \geq m$. □

With this correspondence established, we introduce the main space we shall focus on for much of our work; before concluding this section with a discussion of the topologies we will endow our spaces with.

Define $\widehat{Y} \subset \Sigma^{\mathbb{Z}^2}$ to be the set of $\mathbb{Z}^2$ arrays with rows being two sided sequences in Σ_{M_T} and columns being two sided sequences in Σ_{M_S}; we will denote by $y(i,j)$ the symbol determined by the pair $(i,j) \in \mathbb{Z}^2$.

Define the map $\psi : \widehat{Y} \to Y$ to be the restriction of an array in $\Sigma^{\mathbb{Z}^2}$ to it's top right quadrant (i.e. restriction to $\Sigma^{\mathbb{N}^2}$), by setting $\psi(\hat{y})(i,j) = \hat{y}(i,j)$ for every $\hat{y} \in \widehat{Y}$ and each pair $(i,j) \in \mathbb{N}^2$. With this notation we define the map, $\widehat{\varphi} := \varphi \circ \psi : \widehat{Y} \to \mathbb{T}$ (i.e. the map that first restricts the $\mathbb{Z}^2$ array to an $\mathbb{N}^2$ array and then gives the corresponding point in $\mathbb{T}$).

Let us also denote by $\widehat{T}$ and $\widehat{S}$ the left and down shifts, respectively, on the space $\widehat{Y}$ and note that both transformations are invertible on this space. The main focus of the next section will be on measures on the space $\widehat{Y}$ that are invariant under $\widehat{T}$ and $\widehat{S}$.

Remark 2.4. *We note for future reference that $T \circ \hat{\varphi} = \hat{\varphi} \circ \widehat{T}$ and $S \circ \hat{\varphi} = \hat{\varphi} \circ \widehat{S}$.*

To conclude this section we shall discuss the topologies of the spaces that we shall work with, and their corresponding Borel σ-algebras. For the circle, $\mathbb{T}$, we shall endow it with the topology generated by the open intervals and denote its Borel σ-algebra by $\mathcal{B}_{\mathbb{T}}$.

For the spaces of arrays, Y and $\widehat{Y}$, we shall endow them with the topology generated by the cylinder sets, similarly to the case of one and two sided shift spaces. Each cylinder set is defined to be the set of arrays, in Y or $\widehat{Y}$, such that the symbols agree with a prescribed finite size array in Y or $\widehat{Y}$ respectively. With this topology generated by such cylinder sets, we shall denote the Borel σ-algebras of Y and $\widehat{Y}$ by $\mathcal{B}_Y$ and $\mathcal{B}_{\widehat{Y}}$ respectively.

Having defined topologies, we then note that each of our three spaces, $\mathbb{T}, Y$ and $\widehat{Y}$, are compact metric spaces (though we shall never refer explicitly to the metrics they are endowed with) with countably generated σ-algebras. With these defined σ-algebras, φ and $\hat{\varphi}$ are then also measurable maps between their respective spaces.

It now becomes useful to introduce the notion of an atom of a σ-algebra.

Definition 2.5. *For a Borel probability space $(X, \mathcal{B}, \mu)$ and a countably generated σ-algebra, $\mathcal{A} \subseteq \mathcal{B}$, we define the atom of $\mathcal{A}$ containing x to be*

$$[x]_{\mathcal{A}} = \cap_{x \in A \in \mathcal{A}} A = \bigcap_{x \in A_i} A_i \cap \bigcap_{x \notin A_i} X \backslash A_i$$

(i.e. the smallest element of $\mathcal{A}$ containing x).

Finally, by definition of the map ψ and the cylinder sets, we note for later use that by viewing $\mathcal{B}_Y$ as a sub-σ-algebra of $\mathcal{B}_{\widehat{Y}}$, we have that $[\hat{y}]_{\mathcal{B}_Y} = \psi^{-1}\big(\psi(\hat{y})\big)$ for each $\hat{y} \in \widehat{Y}$. This follows as if any two arrays give the same $\Sigma^{\mathbb{N}^2}$ restriction they must also belong to all of the same cylinder sets in Y, and hence the same atom of $\mathcal{B}_Y$.

3 Lifted Measures and Inverse Limits

Let us denote by $\mathcal{M}$ the space of all T and S invariant Borel probability measures on $\mathbb{T}$. We now show that the set V without the point 0 is indeed a null set.

Lemma 3.1. *For any $\mu \in \mathcal{M}$ we have that $\mu(V \setminus \{0\}) = 0$.*

Proof. Let us assume for contradiction that $\mu(V \setminus \{0\}) \neq 0$; then, as V is countable, there must exist $x \in V$ such that $\mu(x) > 0$. For such a point x there exist positive integers $m, n \in \mathbb{N}$ such that $S^m T^n(x) = 0$. We then conclude that $\mu(0) = \mu(S^{-m}T^{-n}(0)) \geq \mu(0) + \mu(x)$ (as both 0 and x are contained within $S^{-m}T^{-n}(0)$), however this implies that $\mu(x) \leq 0$, a contradiction. □

Under the assumption that $\mu(\{0\}) = 0$ we may now identify the two dynamical systems $(\mathbb{T}, \mathcal{B}_{\mathbb{T}}, \mu, (T, S))$ and $(Y, \mathcal{B}_Y, \hat{\mu}, (\widehat{T}, \widehat{S}))$; where $\hat{\mu}$ is the unique measure that lifts to the space Y via the map φ defined in section 2. Precisely, as the map φ is invertible when restricted to the compliment of the pre-image of V, $Y \setminus \varphi^{-1}(V)$, we may uniquely define a Borel measure $\hat{\mu}$ by setting $\hat{\mu}(E) := \mu(\varphi(E))$ each $E \in \mathcal{B}_Y$. As we have uniquely lifted μ to the measure $\hat{\mu}$ on Y, and in our following work shall only focus on the spaces Y and $\widehat{Y}$, it becomes useful to identify $\hat{\mu}$ with μ.

With this identification it is now notationally convenient to drop the hats from the transformations on both Y and $\widehat{Y}$, it will be clear from context whether the transformations T and S are being used on the spaces of arrays or on the circle.

For the rest of our work we will restrict our measures to the set, $\widehat{\mathcal{M}_0}$, which we define to be the set of T and S invariant, ergodic Borel probability measures on the space $\widehat{Y}$, excluding the measure, $\hat{\delta}_0$ (lifted from the Dirac point mass at 0, δ_0). This restriction is made so that we are in the setting of the above lemma, and as δ_0 has zero entropy with respect to both transformations, this is done without any loss of generality.

We now turn our attention to the two dynamical systems $\mathbf{Y} = (Y, \mathcal{B}_Y, \mu, (\widehat{T}, \widehat{S}))$ and $\widehat{\mathbf{Y}} = (\widehat{Y}, \mathcal{B}_{\widehat{Y}}, \hat{\mu}, (T, S))$; where $\hat{\mu}$ is the measure on $\mathcal{B}_{\widehat{Y}}$ defined by $\hat{\mu}(E) = \mu\big(\hat{\varphi}(E)\big)$ for each $E \in \mathcal{B}_{\widehat{Y}}$. In this manner, as we have that $\mathcal{B}_Y \subset \mathcal{B}_{\widehat{Y}}$ (where we are now considering cylinders only in the top right quadrant), we can view μ as the restriction of $\hat{\mu}$ to $\mathcal{B}_Y$. This perspective also allows us to identify the dynamical system $\mathbf{Y}$ with $(\widehat{Y}, \mathcal{B}_Y, \mu, (\widehat{T}, \widehat{S}))$, which will be helpful later on.

To proceed with the proof it becomes necessary to introduce the notion of an inverse limit of dynamical systems. It will turn out that our system $\widehat{\mathbf{Y}}$ is precisely this notion for the underlying system $\mathbf{Y}$, and the theory we will develop for inverse limit systems will mean that we may turn our attention to $\widehat{\mathbf{Y}}$ in order to complete our result.

We will provide the necessary definitions and set up before establishing the results we require for our purposes. As such, our introduction to inverse limits will omit a large amount of detail, more of which may found in Brown's paper, [Br], from which the treatment here is adapted (more results and application of inverse limits to dynamical systems may be found in [I]). In Brown's paper the concept of an inverse limit is introduced for systems under the action of just one transformation; however, the definitions extend naturally to our setting, where we are considering two (or more generally a semi-group of) transformations, and this is the generality we shall give here.

The notion of an inverse limit in a category is a construction that we shall exploit to great effect in our proof of Rudolph's theorem. In order to define the concept, we first need to establish the category of dynamical systems that we are to work with by describing its objects and morphisms. The objects we consider will be so called conjugacy classes of dynamical systems of the form $\Phi = (X, \mathcal{B}, \mu, (\phi, \psi))$; where X is some non-empty set, $\mathcal{B}$ is a σ-algebra of subsets of X, μ is a probability measure on $\mathcal{B}$ and (ϕ, ψ) is a pair of measure preserving transformations on the set X.

To motivate the definition of conjugacy for dynamical systems of the above form, we may look at our examples given by T and S in the theorem. For instance, in the case that $p = 2$ and $q = 3$; the dynamics on $\mathbb{T}$ are no different from those of the maps that square and cube complex numbers on the unit circle in the complex plane. Formally, we give the following definition:

Definition 3.2. *We shall say that two dynamical systems,* $\Phi = (X, \mathcal{B}, \mu, (\phi, \psi))$ *and* $\widetilde{\Phi} = (\widetilde{X}, \widetilde{\mathcal{B}}, \widetilde{\mu}, (\widetilde{\phi}, \widetilde{\psi}))$*, are conjugate if there exists a bijective map* $\Theta : \widetilde{\mathcal{B}} \to \mathcal{B}$ *that satisfies the two following conditions:*

1. $\mu(\Theta(\widetilde{B})) = \widetilde{\mu}(\widetilde{B})$ *for any* $\widetilde{B} \in \widetilde{\mathcal{B}}$
2. $\mu(\phi^{-1}(\Theta(\widetilde{B})) \,\Delta\, \Theta(\widetilde{\phi}^{-1}(\widetilde{B}))) = 0$
3. $\mu(\psi^{-1}(\Theta(\widetilde{B})) \,\Delta\, \Theta(\widetilde{\psi}^{-1}(\widetilde{B}))) = 0$

(where Δ *denotes the symmetric difference between sets).*

Remark 3.3. *If we have an invertible measure preserving map* $\Psi : X \to \widetilde{X}$ *(up to sets of zero measure) whose inverse is also measurable and is such that* $\Psi \circ \phi = \widetilde{\phi} \circ \Psi$ *(mod zero), then its inverse,* Ψ^{-1}*, effects such a conjugacy (in the case of the doubling/tripling and squaring/cubing maps above we may take* $\Psi : \mathbb{T} \to S^1 : x \mapsto exp(2\pi i x)$ *to see that the dynamical systems are conjugate).*

We now introduce the morphisms, which we shall call factor maps, between the conjugacy classes of dynamical systems.

Definition 3.4. *We say that a dynamical system* $\widetilde{\Phi} = (\widetilde{X}, \widetilde{\mathcal{B}}, \widetilde{\mu}, (\widetilde{\phi}, \widetilde{\psi}))$ *is a factor of the dynamical system* $\Phi = (X, \mathcal{B}, \mu, (\phi, \psi))$ *if there exists a measure preserving transformaton* $\Psi : X \to \widetilde{X}$ *such that* $\Psi \circ \phi = \widetilde{\phi} \circ \Psi$ *(mod zero) and* $\Psi \circ \psi = \widetilde{\psi} \circ \Psi$ *(mod zero). We then write that* $\Phi \xrightarrow{\Psi} \widetilde{\Phi}$ *or* $\widetilde{\Phi}|\Phi$.

With this definition, we shall no longer distinguish between dynamical systems that are conjugate to one another, and in the proceeding treatment we may work with a representative of each conjugacy class.

Remark 3.5. *If we have that* $\widetilde{\Phi}$ *is a factor of* Φ*, then we may assume (by suitable choice of conjugacy class representative) that* $\widetilde{X} \subseteq X$*,* $\widetilde{\mathcal{B}} \subseteq \mathcal{B}$ *and hence that* $\widetilde{\mu}$ *and the pair* $(\widetilde{\phi}, \widetilde{\psi})$ *may be viewed as the respective restrictions of* μ *and the pair* (ϕ, ψ) *to* $\widetilde{\mathcal{B}}$ *and* $\widetilde{X}$*. By taking this perspective, we then have that* $\widetilde{\mathcal{B}}$ *is an invariant sub-*σ*-algebra of* $\mathcal{B}$*, in the sense that* $\phi^{-1}(\widetilde{\mathcal{B}}) \subseteq \widetilde{\mathcal{B}}$ *(mod zero) and* $\psi^{-1}(\widetilde{\mathcal{B}}) \subseteq \widetilde{\mathcal{B}}$ *(mod zero). Furthermore, the system* $\widetilde{\Phi}$ *has a conjugacy class representative as an invertible system if and only if we have that* $\phi^{-1}(\widetilde{\mathcal{B}}) = \widetilde{\mathcal{B}}$ *(mod zero) and* $\psi^{-1}(\widetilde{\mathcal{B}}) = \widetilde{\mathcal{B}}$ *(mod zero). It is also important to note that even if* $\widetilde{\Phi}|\Phi$ *and* $\Phi|\widetilde{\Phi}$ *we do not necessarily have that* Φ *and* $\widetilde{\Phi}$ *are measure theoretically isomorphic (e.g. see [Le],[P]). In this case however one says that* Φ *and* $\widetilde{\Phi}$ *are weakly measure theoretically isomorphic.*

With the equivalence between dynamical systems and their factor maps defined, we now have a concrete category of measure preserving dynamical systems that we can work with. We now make a few preliminary definitions before constructing the notion of an inverse limit.

Definition 3.6. *An inverse system of dynamical systems is a triple,* $(\mathcal{J}, \Phi_\alpha, \Psi_{\alpha\widetilde{\alpha}})$*, such that:*

1. $\mathcal{J}$ *is a directed set (with pre-ordering denoted by* $<$*),*
2. Φ_α *is a dynamical system for all* $\alpha \in \mathcal{J}$*,*
3. *For any* $\alpha, \widetilde{\alpha} \in \mathcal{J}$ *with* $\alpha < \widetilde{\alpha}$ *we have that* $\Phi_{\widetilde{\alpha}} \xrightarrow{\Psi_{\alpha\widetilde{\alpha}}} \Phi_\alpha$ *(i.e.* $\Phi_\alpha|\Phi_{\widetilde{\alpha}}$*).*

Remark 3.7. *A directed set is a non-empty set with a reflexive and transitive binary operation; it is a more general concept than a partial ordering as we don't require that the binary operation is antisymmetric.*

Definition 3.8. *An upper bound for an inverse system* $(\mathcal{J}, \Phi_\alpha, \Psi_{\alpha\widetilde{\alpha}})$ *is a dynamical system* Φ *such that* $\Phi_\alpha|\Phi$ *(denoting the factor maps by* $\Phi \xrightarrow{\rho_\alpha} \Phi_\alpha$*) for any* $\alpha \in \mathcal{J}$ *and in addition whenever* $\alpha, \widetilde{\alpha} \in \mathcal{J}$ *are such that* $\alpha < \widetilde{\alpha}$*, the following diagram*

$$\begin{array}{ccc} \Phi_\alpha & \xleftarrow{\Psi_{\alpha\widetilde{\alpha}}} & \Phi_{\widetilde{\alpha}} \\ & \nwarrow^{\rho_\alpha} & \uparrow \rho_{\widetilde{\alpha}} \\ & & \Phi \end{array}$$

commutes. Equivalently Φ *is an upper bound of the inverse system if for each* $\alpha \in \mathcal{J}$ *we have that* Φ_α *is a factor of* Φ*.*

The inverse limit of an inverse system is in a sense analogous with the supremum of a set, and is a good comparison (as the least upper bound) to keep in mind for the following definition:

Definition 3.9. *An inverse limit of an inverse system* $(\mathcal{J}, \Phi_\alpha, \Psi_{\alpha\widetilde{\alpha}})$ *is an upper bound,* $\widehat{\Phi}$ *(with factor maps given by* $\widehat{\Phi} \xrightarrow{\widehat{\rho}_\alpha} \Phi_\alpha$*), such that* $\widehat{\Phi}$ *is a factor of every other upper bound for the inverse system. We then write that*

$$\widehat{\Phi} = \varprojlim_{\alpha \in \mathcal{J}} (\Phi_\alpha).$$

This is equivalent to saying that if $\bar{\Phi}$ *is another upper bound, with maps given by* $\bar{\Phi} \xrightarrow{\bar{\rho}_\alpha} \Phi_\alpha$*, for* $(\mathcal{J}, \Phi_\alpha, \Psi_{\alpha\widetilde{\alpha}})$ *we have a factor map* $\bar{\Phi} \xrightarrow{\sigma} \widehat{\Phi}$ *such that the following diagram*

$$\begin{array}{ccc} \Phi_\alpha & \xleftarrow{\widehat{\rho}_\alpha} & \widehat{\Phi} \\ & \nwarrow^{\bar{\rho}_\alpha} & \uparrow \sigma \\ & & \bar{\Phi} \end{array}$$

commutes.

By considering remark 3.5, if we have a dynamical system $\Phi = (X, \mathcal{B}, \mu, (\phi, \psi))$ that is an upper bound for some inverse system, $(\mathcal{J}, \Phi_\alpha, \Psi_{\alpha\widetilde{\alpha}})$, of dynamical systems, then for each $\alpha \in \mathcal{J}$ we have a conjugacy class representative such that $\Phi_\alpha = (X, \mathcal{B}_\alpha, \mu, (\phi, \psi))$. Furthermore we have that the $\mathcal{B}_\alpha$ are increasing, with respect to the indexing by the directed set $\mathcal{J}$, invariant sub-σ-algebras of $\mathcal{B}$ (hence for any $\alpha, \widetilde{\alpha} \in \mathcal{J}$ with $\alpha < \widetilde{\alpha}$ the maps $\Psi_{\alpha\widetilde{\alpha}}$ and ρ_α are the identity on X). Therefore, if the inverse limit, say $\widehat{\Phi}$, exists for the inverse system, then it has a conjugacy class representation as $\widehat{\Phi} = (X, \widehat{\mathcal{B}}, \mu, (\phi, \psi))$ where $\widehat{\mathcal{B}} \supseteq \bigcup_{\alpha \in \mathcal{J}} \mathcal{B}_\alpha$.

Now that we have established the formal definition of an inverse limit, we will collect the results that we will require for the proof of Rudolph's theorem, before showing that $\widehat{Y}$ is indeed the appropriate space to be working with. We shall not deal with the general existence of inverse limits, as we shall explicitly show that $\widehat{Y}$ is the underlying set for the inverse limit for an appropriate inverse system. Instead, we first simultaneously tackle a characterisation, as well as uniqueness, of the inverse limit. We first recall a definition from measure theory and make analogous definition for dynamical systems.

Definition 3.10. *For a σ-algebra $\mathcal{B}$, we define the join of a collection of sub-σ-algebras, $\{\mathcal{B}_\alpha\}_{\alpha\in\mathcal{A}}$, to be the smallest σ-algebra, denoted by $\bigvee_{\alpha\in\mathcal{A}}\mathcal{B}_\alpha$, containing their union, $\bigcup_{\alpha\in\mathcal{A}}\mathcal{B}_\alpha$.*

Now suppose that $\{\Phi_\alpha\}_{\alpha\in\mathcal{A}}$ is a collection of factors (with σ-algebras $\mathcal{B}_\alpha$ and factor maps ρ_α), of a dynamical system $\Phi = (X, \mathcal{B}, \mu, (\phi, \psi))$. We then define their join, denoted by $\bigvee_{\alpha\in\mathcal{A}}\Phi_\alpha$, to be the dynamical system given by

$$\left(X, \bigvee_{\alpha\in\mathcal{A}} \rho_\alpha^{-1}(\mathcal{B}_\alpha), \mu, (\phi, \psi)\right).$$

We now show that if we take our indexing set, $\mathcal{A}$, to be $\mathcal{J}$ for an inverse system, $(\mathcal{J}, \Phi_\alpha, \Psi_{\alpha\tilde{\alpha}})$, of dynamical systems we have that the join and inverse limit coincide, namely:

Theorem 3.11. *For an inverse system $(\mathcal{J}, \Phi_\alpha, \Psi_{\alpha\tilde{\alpha}})$ that has an upper bound Φ we have that*

$$\varprojlim_{\alpha\in\mathcal{J}} (\Phi_\alpha) = \bigvee_{\alpha\in\mathcal{A}} \Phi_\alpha.$$

Furthermore, if the inverse limit exists it is uniquely determined by the join construction.

Proof. By choice of a conjugacy class representative we view the upper bound as $\Phi = (X, \mathcal{B}, \mu, (\phi, \psi))$, with factor maps denoted by $\Phi \xrightarrow{\rho_\alpha} \Phi_\alpha$; remark 3.5 then means that for all $\alpha \in \mathcal{J}$ we have representations $\Phi_\alpha = (X, \mathcal{B}_\alpha, \mu, (\phi, \psi))$ with $\mathcal{B}_\alpha \subseteq \mathcal{B}$, and hence that $\Psi_{\alpha\tilde{\alpha}}$ and ρ_α are the identity on X. If we then define $\widehat{\Phi} = (X, \bigvee_{\alpha\in\mathcal{A}}\mathcal{B}_\alpha, \mu, (\phi, \psi))$ it is evident that $\widehat{\Phi}$ is an upper bound for the inverse system; thus it remains to show that $\widehat{\Phi}$ is a factor of every other upper bound in order to conclude that it is indeed the inverse limit.

Suppose now that $\bar{\Phi} = (\bar{X}, \bar{\mathcal{B}}, \bar{\mu}, (\bar{\phi}, \bar{\psi}))$ is another upper bound for the inverse system, with factor maps $\bar{\Phi} \xrightarrow{\bar{\rho}_\alpha} \Phi_\alpha$. As the maps $\Psi_{\alpha\tilde{\alpha}}$ are the identity on the space X we see that by considering the following commutative diagram,

$$\begin{array}{ccc} \Phi_\alpha & \xleftarrow{\Psi_{\alpha\widetilde{\alpha}}} & \Phi_{\widetilde{\alpha}} \\ & \nwarrow_{\bar{\rho}_\alpha} & \uparrow{\scriptstyle \bar{\rho}_{\widetilde{\alpha}}} \\ & & \bar{\Phi} \end{array}$$

that for any $\alpha, \widetilde{\alpha} \in \mathcal{J}$ we have $\bar{\rho}_\alpha = \bar{\rho}_{\widetilde{\alpha}}$ as maps from $\bar{X}$ to X; let us then denote these maps simply by $\bar{\rho}$. Now we observe that as $\bar{\rho}$ is measure preserving and $\bar{\rho}^{-1}(\bigcup_{\alpha\in\mathcal{A}} \mathcal{B}_\alpha) = \bigcup_{\alpha\in\mathcal{A}} \bar{\rho}^{-1}(\mathcal{B}_\alpha) \subseteq \bar{\mathcal{B}}$ we have that $\bar{\rho}^{-1}(\bigvee_{\alpha\in\mathcal{A}} \mathcal{B}_\alpha) \subseteq \bar{\mathcal{B}}$. Thus we may conclude that $\bar{\Phi} \xrightarrow{\bar{\rho}} \widehat{\Phi}$ and hence $\widehat{\Phi}$ is an inverse limit.

As remark 3.5 above shows, we have not established uniqueness as yet; all we have shown is that any two inverse limits are weakly measure theoretically isomorphic. To prove uniqueness, we may suppose without loss of generality that our original upper bound Φ is another inverse limit for the inverse system. Thus by definition of the inverse limit there must exist a factor map $\widehat{\Phi} \xrightarrow{\sigma} \Phi$ such that the following diagram

$$\begin{array}{ccc} \Phi_\alpha & \xleftarrow{\rho_\alpha} & \Phi \\ & \nwarrow_{\widehat{\rho}_\alpha} & \uparrow{\scriptstyle \sigma} \\ & & \widehat{\Phi} \end{array}$$

commutes. Again note that for any $\alpha \in \mathcal{J}$ we have that $\rho_\alpha, \widehat{\rho}_\alpha$ are the identity on X and hence σ must be the identity also. As σ is measure preserving we also have that $\mathcal{B} = \sigma^{-1}(\mathcal{B}) \subseteq \bigvee_{\alpha\in\mathcal{A}} \mathcal{B}_\alpha \subseteq \mathcal{B}$ and hence $\mathcal{B} = \bigvee_{\alpha\in\mathcal{A}} \mathcal{B}_\alpha$. Therefore we conclude that the dynamical systems $\widehat{\Phi}$ and Φ are measure theoretically isomorphic, and so the inverse limit is unique. □

We now are ready to return to the setting of our dynamical systems $\mathbf{Y}$ and $\widehat{\mathbf{Y}}$. We wish to establish a suitable inverse system, containing $\mathbf{Y}$, such that the dynamical system $\widehat{\mathbf{Y}}$ is the inverse limit. Recall that we may view μ as the restriction of $\hat{\mu}$ to the σ-algebra $\mathcal{B}_Y$.

Let us define a pre-ordering, $<$, on the set $\mathbb{N}^2$ by $(i,j) < (k,l)$ if and only if $i \leq k$ or $j \leq l$. Then by defining for each $(i,j) \in \mathbb{N}^2$ the dynamical system $\mathbf{Y}_{(i,j)} = (\widehat{Y}, T^iS^j(\mathcal{B}_Y), \hat{\mu}, (T,S))$ (i.e. $\mathbf{Y}_{(0,0)} \cong \mathbf{Y}$) we may construct an inverse system by defining the maps $\Psi_{(i,j),(k,l)}(\widehat{y}) = T^{i-k}S^{j-l}(\widehat{y})$ for each $\widehat{y} \in \widehat{Y}$ whenever $(i,j) < (k,l)$. Here we are using that both transformations T and S are invertible on $\widehat{\mathbf{Y}}$, as well as viewing $\hat{\mu}$ restricted to $T^iS^j(\mathcal{B}_Y)$. It is then clear that such maps are indeed factor maps as the σ-algebras are nested; in fact, for each $(i,j) \in \mathbb{N}^2$ the dynamical systems $\mathbf{Y}_{(i,j)}$ are measure theoretically isomorphic to the original system $\mathbf{Y}$ via the map $\Psi_{(0,0),(i,j)}$. We now have an inverse system given by $(\mathbb{N}^2, \mathbf{Y}_{(i,j)}, \Psi_{(i,j),(k,l)})$ and proceed to show that $\widehat{\mathbf{Y}}$ is the inverse limit of this system, via the join construction above.

Let us denote by $\mathcal{B}_{(i,j)}$ the σ-algebra $T^iS^j(\mathcal{B}_Y)$ and recall that the σ-algebra for $\widehat{\mathbf{Y}}$ is generated by cylinders in $\mathbb{Z}^2$ that define vertical and horizontal two sided sequences in Σ_{M_T} and Σ_{M_S} respectively. Note that for each $(i,j) \in \mathbb{N}^2$ the σ-algebra $\mathcal{B}_{(i,j)}$, for the system $\mathbf{Y}_{(i,j)}$, is generated by cylinders in the restriction to $(a,b) \in \mathbb{Z}^2$ where $a \geq i$ and $b \geq j$. Therefore $\mathcal{B}_{\widehat{Y}} = \bigvee_{(i,j)\in\mathbb{N}^2} \mathcal{B}_{(i,j)}$, and hence we may apply theorem 3.11 to conclude that $\widehat{\mathbf{Y}}$ is indeed the inverse limit of $(\mathbb{N}^2, \mathbf{Y}_{(i,j)}, \Psi_{(i,j),(k,l)})$.

We now prove two results that carry over to our inverse limit system, $\widehat{\mathbf{Y}}$, to utilise in our proof of Rudolph's theorem; namely that the system $\widehat{\mathbf{Y}}$ is jointly ergodic with respect to T and S and has the same entropy as its factors, i.e. the same entropy as $\mathbf{Y}$. The following two lemmas are in fact much more general and hold for arbitrary inverse limits of dynamical systems.

For use in the proof of the following lemma, we remark that a variant of Von Neumann's mean ergodic theorem that holds for our definition of joint ergodicity. In particular we have the following characterisation:

Theorem 3.12. *An invariant Borel probability measure for a dynamical system* $(X, \mathcal{B}, \mu, (T,S))$ *is jointly ergodic if and only if for all* $A, B \in \mathcal{B}$ *we have that*

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{i,j=0}^{n-1} \mu(A \cap T^{-i}S^{-j}(B)) \to \mu(A)\mu(B).$$

Lemma 3.13. *The dynamical system* $\widehat{\mathbf{Y}}$ *is jointly ergodic with respect to the transformations* T *and* S.

Proof. By the above characterisation of joint ergodicity we have that $\widehat{\mathbf{Y}}$ is ergodic if and only if we have that for any $A, B \in \mathcal{B}_{\widehat{Y}}$

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{i,j=0}^{n-1} \widehat{\mu}(A \cap T^{-i}S^{-j}(B)) \to \widehat{\mu}(A)\widehat{\mu}(B).$$

Via the join characterisation of the inverse limit $\widehat{\mathbf{Y}}$ and that fact that $\bigcup_{(i,j)\in\mathbb{N}^2} \mathcal{B}_{(i,j)}$ is dense in $\mathcal{B}_{\widehat{Y}}$ (in the sense that for any $A \in \mathcal{B}_{\widehat{Y}}$ and $\epsilon > 0$ there is some $\widetilde{A} \in \bigcup_{(i,j)\in\mathbb{N}^2} \mathcal{B}_{(i,j)}$ such that $\widehat{\mu}(A \,\Delta\, \widetilde{A}) < \epsilon$), it is sufficient to show that the above limit holds for any $\widetilde{A}, \widetilde{B} \in \bigcup_{(i,j)\in\mathbb{N}^2} \mathcal{B}_{(i,j)}$. To see this, let $\epsilon > 0$ and $\widetilde{A}, \widetilde{B} \in \bigcup_{(i,j)\in\mathbb{N}^2} \mathcal{B}_{(i,j)}$ approximate $A, B \in \mathcal{B}_{\widehat{Y}}$ (i.e. $\widehat{\mu}(A \,\Delta\, \widetilde{A}) < \frac{\epsilon}{2}$ and $\widehat{\mu}(B \,\Delta\, \widetilde{B}) < \frac{\epsilon}{2}$), then by the following calculation;

$$\begin{aligned}|\widehat{\mu}(A \cap T^{-i}S^{-j}(B)) - \widehat{\mu}(\widetilde{A} \cap T^{-i}S^{-j}(\widetilde{B}))| &\leq \widehat{\mu}([A \cap T^{-i}S^{-j}(B)] \,\Delta\, [\widetilde{A} \cap T^{-i}S^{-j}(\widetilde{B})]) \\ &\leq \widehat{\mu}([A\Delta\widetilde{A}] \cup [T^{-i}S^{-j}(B)\Delta T^{-i}S^{-j}(\widetilde{B})]) < \epsilon.\end{aligned}$$

As we know that each factor $\mathbf{Y}_{(i,j)}$ is isomorphic to the ergodic system $\mathbf{Y}$, the above limit indeed holds for each $\widetilde{A}, \widetilde{B} \in \bigcup_{(i,j)\in\mathbb{N}^2} \mathcal{B}_{(i,j)}$. Thus, by the calculation above, the limit holds for the system $\widehat{\mathbf{Y}}$, and hence it is ergodic. □

Lemma 3.14. *The measure theoretic entropies of the inverse limit system $\widehat{\mathbf{Y}}$ and the system $\mathbf{Y}$ coincide.*

Proof. As we are viewing the factors $\mathbf{Y}_{(i,j)}$ as being the same system as $\widehat{\mathbf{Y}}$ equipped with different (sub-)σ-algebras there is no ambiguity about the measure involved. Thus, let us for this proof denote the entropy of $\widehat{\mathbf{Y}}$ as $\widehat{h}((T,S))$ and that of $\mathbf{Y}_{(i,j)}$ as $h_{(i,j)}((T,S))$. Furthermore let us set $h_{(0,0)}((T,S)) = h((T,S))$ to be the entropy of $\mathbf{Y}$. Let us note that as the $\mathcal{B}_{(i,j)}$ are sub-σ-algebras of $\mathcal{B}_{\widehat{Y}}$ we have that $\widehat{h}((T,S)) \geq h_{(i,j)}((T,S))$ for any $(i,j) \in \mathbb{N}^2$.

By definition of entropy, for any $\epsilon > 0$ there exists some finite partition α in $\mathcal{B}_{\widehat{Y}}$ such that $\widehat{h}((T,S),\alpha) \geq \widehat{h}((T,S)) - \frac{\epsilon}{2}$. Now as the $\mathcal{B}_{(i,j)}$ are a nested and increasing (with respect to the pre-ordering $<$) sequence that generates $\mathcal{B}_{\widehat{Y}}$, we may find a finite partition β, belonging to some $\mathcal{B}_{(i,j)}$, such that $h_{(i,j)}((T,S),\beta) \geq \widehat{h}((T,S),\alpha) - \frac{\epsilon}{2}$. Thus we may conclude that $h_{(i,j)}((T,S)) \geq \widehat{h}((T,S)) - \epsilon$, but as the choice of ϵ was arbitrary, $h_{(i,j)}((T,S)) \geq \widehat{h}((T,S))$. As all factors $\mathbf{Y}_{(i,j)}$ are isomorphic to $\mathbf{Y}$, in particular they share the same entropy and hence $h((T,S)) = \widehat{h}((T,S))$. □

To conclude this section we make the following observation. As the inverse limit system $\widehat{\mathbf{Y}}$ arises uniquely from the space $\mathbf{Y}$, we may now focus on these two spaces of arrays in order to obtain results about $\mathbb{T}$; via our correspondence that was established in section 2.

4 Disintegration of Measures, Symmetric Points and Weak Convergence

In this section, we will introduce the notions of disintegration of measures and that of symmetric points for the arrays in $\widehat{Y}$. This is done in order to show that in the case that an array is symmetric, an appropriately defined sequence of measures on the circle converge in the weak sense to the Lebesgue measure. We then conclude this section by showing that if almost every point is symmetric, then our measure on the circle is Lebesgue. In section 5 we will establish that in the case of positive entropy, almost every point is symmetric, completing our result.

The role that the symmetric points play in the proof of Rudolph's theorem is crucial, and though a property about arrays in $\widehat{Y}$, they act as the key property distinguishing the Lebesgue measure on the circle. In order to construct our desired sequence of measures on the circle, we will need to first introduce the notion of disintegration of measures. To aid us in the construction and characterisation of these so called disintegrated measures let us recall the notion of conditional expectation from probability theory.

For a probability space $(X, \mathcal{B}, \mu)$, we shall denote by $L^1(X, \mathcal{B}, \mu)$ the equivalence classes of integrable functions on X, with equivalence defined by equality almost everywhere. We also let $\mathcal{L}^1(X, \mathcal{B}, \mu)$ denote the set of integrable functions on X and $C(X)$ denote the continuous functions on X.

Definition 4.1. *Let $(X, \mathcal{B}, \mu)$ be a probability space, $\mathcal{A} \subseteq \mathcal{B}$ a σ-algebra and $f \in L^1(X, \mathcal{B}, \mu)$. The conditional expectation of the function f with respect to the σ-algebra $\mathcal{A}$ is the function $\mathbb{E}(f \mid \mathcal{A}) \in L^1(X, \mathcal{A}, \mu)$, such that for any $A \in \mathcal{A}$ we have that*

$$\int_A f d\mu = \int_A \mathbb{E}(f \mid \mathcal{A}) d\mu.$$

(We shall, when necessary, write $\mathbb{E}_\mu(f \mid \mathcal{A})$ in order to specify which measure we are working with.)

We now collect some properties of the above defined conditional expectation that will aid us in the proof of the disintegration theorem; we will not provide proofs for these results as they are all well established and may be found in numerous textbooks on the subject, for example in [EW].

Proposition 4.2. *Let $(X, \mathcal{B}, \mu)$ be a probability space, $\mathcal{A} \subseteq \mathcal{B}$ a σ-algebra and $f \in L^1(X, \mathcal{B}, \mu)$. The conditional expectation of f with respect to $\mathcal{A}$ satisfies the following properties:*

1. *The function $\mathbb{E}(f \mid \mathcal{A}) \in L^1(X, \mathcal{A}, \mu)$ exists and is unique.*
2. *The mapping $\mathbb{E}(\_\_ \mid \mathcal{A}) : L^1(X, \mathcal{B}, \mu) \to L^1(X, \mathcal{A}, \mu)$ defines a positive, continuous linear operator with operator norm equal to 1.*
3. *$\inf_{x \in X} f(x) \leq \mathbb{E}(f \mid \mathcal{A})(x) \leq \sup_{x \in X} f(x)$ for almost every $x \in X$.*

The following theorem is pervasive throughout probability theory and may be established in more generality than we will do so here; details of which may be found for example in [EW] and [Hoc]. The formulation that we give here is adapted from chapter 5 of [EW] and section 7 of [Mo].

Theorem 4.3. *(The disintegration of measures) Let $(X, \mathcal{B}, \mu)$ be a Borel probability space and $\mathcal{A} \subseteq \mathcal{B}$ be a countably generated σ-algebra. Then there exists a set $D \subseteq X$ of full measure ($\mu(D) = 1$) such that for every $x \in D$ there is a probability measure, denoted by μ_x (though we shall denote it by $\mu_x^{\mathcal{A}}$ in order to specify the σ-algebra), called the disintegration* *of μ with respect to $\mathcal{A}$, and the following properties hold:*

1.
$$\mathbb{E}(f \mid \mathcal{A})(x) = \int_X f(y)d\mu_x(y)$$
for $x \in D$ and for any $f \in \mathcal{L}^1(X, \mathcal{B}, \mu)$. Furthermore, we have that the mapping $\left[x \mapsto \int_X f(y)d\mu_x(y)\right]$ is $\mathcal{A}$-measurable and for any $A \in \mathcal{A}$ we have
$$\int_A \int_X f(y)d\mu_x(y)d\mu(x) = \int_A f d\mu.$$
2. *For $x \in D$ we have that $\mu_x([x]_{\mathcal{A}}) = 1$. Furthermore, for any $x, y \in Y$ such that $[x]_{\mathcal{A}} = [y]_{\mathcal{A}}$ we have that $\mu_x = \mu_y$.*
3. *The first property, restricted to a dense subset of functions in $C(X)$, uniquely determines μ_x for μ almost every $x \in X$.*

(The reason property 3. is not stated for the set D is as follows: If we have two systems of measures, μ_x and $\widetilde{\mu}_x$ defined on full measure sets D and $\widetilde{D}$ respectively, then 3. states there is some full measure set $\widehat{D} \subseteq D \cap \widetilde{D}$ such that for $x \in \widehat{D}$ we have $\mu_x = \widetilde{\mu}_x$.)

Proof. Firstly, let us suppose that $\{f_n\}_{n\in\mathbb{N}}$, with $f_0 \equiv 1$, is a dense set of functions in $C(X)$ that forms a vector space over the rational numbers. Now for every $i \geq 0$ we have that $\mathbb{E}(f_i|\mathcal{A})$ is a well defined function for every point in a full measure set, $D \subseteq X$. Applying the above recorded properties of conditional expectation, by defining $\Lambda_x(f_i) = \mathbb{E}(f_i|\mathcal{A})(x)$ for every $x \in D$ we see that we may extend Λ_x to a positive, continuous linear functional on $C(X)$ with $||\Lambda_x||_{op} \leq 1$. Thus, for each $x \in D$, by applying the Riesz-Markov-Kakutani representation theorem (a short proof of which may be found in [S]) to the functional Λ_x, we have that there exists a Borel measure, μ_x, on X such that for each $f \in C(X)$,
$$\Lambda_x(f) = \int_X f d\mu_x.$$
Now noting that for each $x \in D$ we have $\Lambda_x(1) = \mathbb{E}(1|\mathcal{A}) \equiv 1$, we see that the μ_x are indeed probability measures on X. Thus we have established that property 1. holds for the functions $\{f_n\}_{n\in\mathbb{N}}$.

For each $f \in C(X)$ there is some subsequence, (f_{n_k}), of the functions $\{f_n\}_{n\in\mathbb{N}}$ that converges uniformly to f. Now we may use the dominated convergence theorem to establish that property 1. holds for all functions in $C(X)$; this follows by property 1. of the conditional expectation in proposition 5.2.

If property 1. holds for indicator functions for Borel sets, then by considering simple functions and applying the monotone convergence theorem, property 1. holds for any positive $\mathcal{L}^1(X, \mathcal{B}, \mu)$ function. For each $f \in \mathcal{L}^1(X, \mathcal{B}, \mu)$ we may write $f = f^+ - f^-$, where $f^+, f^- \geq 0$ are in $\mathcal{L}^1(X, \mathcal{B}, \mu)$, and hence it is sufficient to prove property 1. for indicator functions of Borel sets. To do this we will need to use the monotone class theorem, which we briefly introduce now; a detailed proof of the theorem may be found in chapter 6 of [Li].

A monotone class is a class of sets that is closed under taking countable monotone unions and intersections; the monotone class theorem then states that for an algebra, $\mathcal{A}$, the smallest monotone class that contains $\mathcal{A}$ is $\sigma(\mathcal{A})$ (the smallest σ-algebra generated by $\mathcal{A}$). Note now that for any open or closed set, $A \subseteq X$, we may find an increasing subsequence of the $\{f_n\}_{n\in\mathbb{N}}$ that converge to the indicator function, χ_A; so property 1. holds for χ_A. For any open set, O, and closed set, C, observe that $O \cap C$ is the countable intersection of open sets (a G_δ set). Hence by the linearity of property 1., we have that for any set of the form $R = \sqcup_{i=1}^n O_i \cap C_i$ (for open sets, O_i, and closed sets, C_i), the indicator function, χ_R satisfies property 1.. It is not hard to show that sets, R, of the above form form an algebra (see page 140 of [EW] for details). We shall denote this algebra by $\mathcal{R}$, and then note that $\mathcal{B} = \sigma(\mathcal{R})$. By defining the set

$$\mathcal{M} = \{B \in \mathcal{B} \mid \chi_B \text{ satisfies property 1.}\}$$

we then have, by the monotone convergence theorem, that $\mathcal{M}$ is a monotone class. Applying the monotone class theorem we see that $\mathcal{B} \subseteq \mathcal{M}$; establishing property 1. holds for general $f \in \mathcal{L}^1(X, \mathcal{B}, \mu)$.

(Note that we have implicitly used property 2. of the conditional expectation from proposition 5.2 in order to guarantee that the mapping $\left[x \mapsto \int_X f(y) d\mu_x(y)\right]$ is $\mathcal{A}$-measurable.)

We now turn our attention to the second property. As $\mathcal{A}$ is countably generated, we have that $\mathcal{A} = \sigma(\{A_i\}_{i\in\mathbb{N}})$, for some $A_i \in \mathcal{A}$, and by property 1. we see that for every $x \in D$ and $i \in \mathbb{N}$

$$\mu_x(A_i) = \mathbb{E}(\chi_{A_i} \mid \mathcal{A})(x) = \chi_{A_i}(x).$$

Hence by writing the atom, $[x]_{\mathcal{A}} = \bigcap_{x\in A_i} A_i \cap \bigcap_{x\notin A_i} X \backslash A_i$, we see that indeed $\mu_x([x]_{\mathcal{A}}) = 1$. Now if $[x]_{\mathcal{A}} = [y]_{\mathcal{A}}$ for some $x, y \in D$, then by $\mathcal{A}$-measurability of the mapping $\left[x \mapsto \int_X f(y) d\mu_x(y)\right]$, we must have that (as the sub-σ-algebra $\mathcal{A}$ doesn't distinguish between such x and y) for any $f \in C(X)$

$$\int_X f d\mu_x = \int_X f d\mu_y.$$

By again appealing to the Riesz-Markov-Kakutani representation theorem, we have that

$\mu_x = \mu_y$ as desired.

Finally we establish the third property. Suppose that $\{f_n\}_{n\in\mathbb{N}}$ is dense in $C(X)$ and that we have two systems of measures, $\{\mu_x \mid x \in D\}$ and $\{\nu_x \mid x \in \widetilde{D}\}$, satisfying property 1. for two full measure sets $D, \widetilde{D} \subseteq X$ respectively. There then exists some full measure set $\widehat{D} \subseteq D \cap \widetilde{D}$, such that for every $x \in \widehat{D}$ and $n \in \mathbb{N}$ we have

$$\int_X f_n d\mu_x = \mathbb{E}(f_n \mid \mathcal{A})(x) = \int_X f_n d\nu_x.$$

We may then use density of the $\{f_n\}_{n\in\mathbb{N}}$, along with the dominated convergence theorem, to see that for every $x \in \widehat{D}$ and $f \in C(X)$ we have

$$\int_X f d\mu_x = \mathbb{E}(f \mid \mathcal{A})(x) = \int_X f d\nu_x.$$

So arguing as before for property 2. we conclude that $\mu_x = \nu_x$ for every $x \in \widehat{D}$. □

Let us also establish a corollary that will be helpful in allowing us to establish key properties for our later defined sequence of measures.

Corollary 4.3.1. *(Corollary 5.24 in [EW]) If $\phi : (X, \mathcal{B}_X, \mu) \to (Z, \mathcal{B}_Z, \nu)$ is a measure preserving map between two Borel probability spaces, and $\mathcal{C} \subseteq \mathcal{B}_Z$ is a countably generated σ-algebra, then for almost every $x \in X$ we have that*

$$\phi_* \mu_x^{\phi^{-1}(\mathcal{C})} = \nu_{\phi(x)}^{\mathcal{C}}.$$

(where ϕ_* denotes push-forward of the measure by ϕ, i.e. $\phi_*\mu(E) := \mu(\phi^{-1}(E))$).

Proof. First note that by the defining property of conditional expectation (property 1. of proposition 4.2.) and the fact that, for each $f \in L^1(Z, \mathcal{B}_Z, \nu)$, $\mathbb{E}(f \mid \mathcal{C}) \circ \phi$ is $\phi^{-1}(\mathcal{C})$-measurable we have

$$\mathbb{E}_\nu(f \mid \mathcal{C}) \circ \phi = \mathbb{E}_\mu(f \circ \phi \mid \phi^{-1}\mathcal{A}).$$

Thus, by applying the disintegration theorem, we have that for μ almost every $x \in X$ and every f in a countable dense subset of $C(Y)$:

$$\int_X f \, d\nu_{\phi(x)}^{\mathcal{A}} = \mathbb{E}_\mu(f \mid \mathcal{C})(\phi(x)) = \int_X f \circ \phi \, d\mu_x^{\phi^{-1}\mathcal{A}} = \int_X f \, d(\phi_* \mu_x^{\phi^{-1}\mathcal{A}}).$$

And so, by arguing as in the proof of theorem 4.3., we establish the desired equality. □

By viewing $\mathcal{B}_Y$ as a sub-σ-algebra of $\mathcal{B}_{\widehat{Y}}$ we may apply the disintegration theorem to produce measures $\hat{\mu}_{\hat{y}}$, supported entirely on the atom $[\hat{y}]_{\mathcal{B}_Y}$, for $\hat{\mu}$ almost every $\hat{y} \in \widehat{Y}$. Recall that by considering the restriction mapping $\psi : \widehat{Y} \to Y$ we have that $[\hat{y}]_{\mathcal{B}_Y} = \psi^{-1}(\psi(\hat{y}))$; equivalently $[\hat{y}]_{\mathcal{B}_Y}$ is the set of all arrays in $\widehat{Y}$ that give the same $\Sigma^{\mathbb{N}^2}$ restriction.

Remark 4.4. *We have that $\psi^{-1}(\psi(\hat{y})) = \widehat{\varphi}^{-1}(\widehat{\varphi}(\hat{y}))$ for $\hat{\mu}$ almost every $\hat{y} \in \widehat{Y}$. This follows from the definition of $\widehat{\varphi}$, as we know that φ is a one-to-one correspondence between Y and $\mathbb{T}$ outside of the set V, which has zero measure.*

We now consider our invertible dynamical system $\widehat{\mathbf{Y}} = (\widehat{Y}, \mathcal{B}_{\widehat{Y}}, \hat{\mu}, (T, S))$. For a given $\hat{y} \in \widehat{Y}$ we now ask the following; for $n \in \mathbb{N}$, what possible arrays in $\widehat{Y}$ after an application of T^{-n} give the same $\Sigma^{\mathbb{N}^2}$ restriction as our original array $\hat{y}$ after an application of T^{-n}?

This question above is equivalent to asking which arrays in $\widehat{Y}$ give the same pre-image of $\hat{\varphi}(\hat{y})$ under the action of T^{-n} on the circle. Each of these pre-images may be written in the form $\hat{\varphi}(T^{-n}(\hat{y})) + \frac{t}{p^n}$ $(mod\, 1)$ for integer values $0 \leq t \leq (p^n - 1)$ (where we are implicitly using remark 2.4.).

We then note that for any such t above, $T^n(\hat{\varphi}^{-1}(\hat{\varphi}(T^{-n}(\hat{y})) + \frac{t}{p^n}))$ is the set of arrays in $\widehat{Y}$ that after an application of T^{-n} give the pre image, $\hat{\varphi}(T^{-n}(\hat{y})) + \frac{t}{p^n}$, when viewed as points on the circle. Furthermore, let us note that any of the above arrays will belong to the same atom of $\mathcal{B}_Y$ as our original array $\hat{y}$. Therefore, for distinct values of t, the above sets are disjoint and partition $\psi^{-1}(\psi(\hat{y})) = [\hat{y}]_{\mathcal{B}}$, and so we have that

$$\bigsqcup_{t=0}^{p^n-1} \left[T^n \left(\hat{\varphi}^{-1} \left(\hat{\varphi}\big(T^{-n}(\hat{y})\big) + \frac{t}{p^n} \right) \right) \right] = \psi^{-1}(\psi(\hat{y})) = [\hat{y}]_{\mathcal{B}}.$$

As mentioned above, we apply the disintegration theorem to the Borel probability space $(\widehat{Y}, \mathcal{B}_{\widehat{Y}}, \hat{\mu})$ with the sub-σ-algebra $\mathcal{B}_Y$, which is countably generated by the cylinder sets in $\Sigma^{\mathbb{N}^2}$. In doing so we find a full measure set, D, of arrays in $\widehat{Y}$, such that for each $\hat{y} \in D$

$$\hat{\mu}_{\hat{y}} \left(\bigsqcup_{t=0}^{p^n-1} \left[T^n \left(\hat{\varphi}^{-1} \left(\hat{\varphi}\big(T^{-n}(\hat{y})\big) + \frac{t}{p^n} \right) \right) \right] \right) = \hat{\mu}_{\hat{y}}([\hat{y}]_{\mathcal{B}}) = 1.$$

We are finally ready to introduce our sequence of measures that will play a crucial role in distinguishing the Lebesgue measure. For each $0 \leq t \leq (p^n - 1)$ we consider the value of $\hat{\mu}_{\hat{y}}(T^n(\hat{\varphi}^{-1}(\hat{\varphi}(T^{-n}(\hat{y})) + \frac{t}{p^n})))$ in $[0, 1]$ in order to define the following probability distribution, $\delta(\hat{y}, n)$, on the points $\{0, \frac{1}{p^n}, ..., \frac{p^n-1}{p^n}\}$ by setting:

$$\delta(\hat{y}, n)\left(\frac{t}{p^n}\right) := \hat{\mu}_{\hat{y}} \left(T^n \left(\hat{\varphi}^{-1} \left(\hat{\varphi}\big(T^{-n}(\hat{y})\big) + \frac{t}{p^n} \right) \right) \right).$$

By property 1. of the disintegration of measure theorem we may also interpret the δ distributions as conditional expectations, in the following sense:

$$\delta(\hat{y},n)\left(\frac{t}{p^n}\right) = \mathbb{E}\left(\chi_{T^n(\hat{\varphi}^{-1}(\hat{\varphi}(T^{-n}(\hat{y}))+\frac{t}{p^n}))} \,\middle|\, \mathcal{B}_Y\right)(\hat{y}).$$

With this perspective, for fixed $\hat{y} \in D$ we may view the δ distributions as encoding the probability that given a specific point on the circle, $\hat{\varphi}(\hat{y})$, it came from a the pre-image $\hat{\varphi}(T^{-n}(\hat{y})) + \frac{t}{p^n}$, under the action of T^n on the circle. We shall therefore call the value $\delta(\hat{y},n)\big(\frac{t}{p^n}\big)$ the probability of seeing the pre-image $\hat{\varphi}(T^{-n}(\hat{y})) + \frac{t}{p^n}$ (given the point $\hat{\varphi}(\hat{y})$).

We now prove several properties of these newly defined δ distributions, before introducing the notion of symmetric points that will be essential for us to establish our desired results.

Lemma 4.5. *The δ probability distributions satisfy the following properties for any $\hat{y} \in D$ and $n \in \mathbb{N}$:*

1. *If $\hat{y}_1, \hat{y}_2 \in D$ are such that $[\hat{y}_1]_{\mathcal{B}_Y} = [\hat{y}_2]_{\mathcal{B}_Y}$ (i.e. the two arrays give the same point in $\mathbb{T}$ under $\hat{\varphi}$) then $\delta(\hat{y}_2,n)$ differs from $\delta(\hat{y}_1,n)$ by a translation in the argument modulo 1, given by*
$$\hat{\varphi}(T^{-n}(\hat{y}_1)) - \hat{\varphi}(T^{-n}(\hat{y}_2))$$
(i.e. we have that $\delta(\hat{y}_1,n)\big(\frac{t}{p^n}\big) = \delta(\hat{y}_2,n)\big(\frac{t}{p^n} + \hat{\varphi}(T^{-n}(\hat{y}_1)) - \hat{\varphi}(T^{-n}(\hat{y}_2))\big)$.
2. *$\delta(\hat{y},n)$ determines $\delta(\hat{y},k)$ for all $k \leq n$ via the formula*
$$\delta(\hat{y},n-1)\left(\frac{t}{p^{n-1}}\right) = \sum_{s=t\ (mod\, p^{n-1})} \delta(\hat{y},n)\left(\frac{s}{p^n}\right).$$
3. $$\delta(S(\hat{y}),n)\left(\frac{qt}{p^n}\right) = \delta(\hat{y},n)\left(\frac{t}{p^n}\right)$$
(where $\frac{qt}{p^n}$ is taken $(mod\,1)$).
4. *The distribution $\delta(T^n(\hat{y}),2n)$ determines $\delta(T^i(\hat{y}),n+i)$ for $0 \leq i \leq n$.*

(We note that only property 3. will require the coprime assumption on p and q).

Proof. If $\hat{y}_1$ and $\hat{y}_2$ are such that $[\hat{y}_1]_{\mathcal{B}} = [\hat{y}_2]_{\mathcal{B}}$, then by property 2. of theorem 4.3 on the disintegration of measures, we have that $\mu_{\hat{y}_1} = \mu_{\hat{y}_2}$. Thus by construction of the δ distributions we see that

$$\begin{aligned}\delta(\hat{y}_1, n)\left(\frac{t}{p^n}\right) &= \hat{\mu}_{\hat{y}_1}\left(T^n\left(\hat{\varphi}^{-1}\left(\hat{\varphi}\big(T^{-n}(\hat{y}_1)\big) + \frac{t}{p^n}\right)\right)\right) \\ &= \hat{\mu}_{\hat{y}_2}\left(T^n\left(\hat{\varphi}^{-1}\left(\hat{\varphi}\big(T^{-n}(\hat{y}_1)\big) + \frac{t}{p^n}\right)\right)\right) \\ &= \hat{\mu}_{\hat{y}_2}\left(T^n\left[\hat{\varphi}^{-1}\left(\hat{\varphi}(T^{-n}(\hat{y}_2)) + \left[\frac{t}{p^n} + \hat{\varphi}(T^{-n}(\hat{y}_1)) - \hat{\varphi}(T^{-n}(\hat{y}_2))\right]\right)\right]\right) \\ &= \delta(\hat{y}_2, n)\left(\frac{t}{p^n} + \hat{\varphi}\big(T^{-n}(\hat{y}_1)\big) - \hat{\varphi}\big(T^{-n}(\hat{y}_2)\big)\right),\end{aligned}$$

establishing that property 1. holds.

To prove the other three properties, we will need to make use of the conditional expectation perspective of the δ distributions as discussed above. We first fix $\hat{y} \in D$ and $n \geq 0$.

For each fixed $0 \leq t \leq p^{n-1}$, consider all pre-images, $\hat{\varphi}(T^{-n}(\hat{y})) + \frac{s}{p^n}$, that map forward under the action of T to the pre-image $\hat{\varphi}(T^{-(n-1)}(\hat{y})) + \frac{t}{p^{n-1}}$ in question. Let us denote by $T^{-(n-1)}(\hat{y})_t$ one possible array with $\hat{\varphi}(T^{-(n-1)}(\hat{y})_t) = \hat{\varphi}(T^{-(n-1)}(\hat{y})) + \frac{t}{p^{n-1}}$ (i.e. an array in $\widehat{Y}$ whose point on $\mathbb{T}$ corresponds to the point $\hat{\varphi}(T^{-(n-1)}(\hat{y})) + \frac{t}{p^{n-1}}$ via $\hat{\varphi}$). The values of s considered above are precisely the p such values of $s = t\,(mod\,p^{n-1})$; therefore, we note that

$$[T^{-(n-1)}(\hat{y})_t]_{\mathcal{B}} = \bigsqcup_{s=t\,(mod\,p^{n-1})}\left[T^n\left(\hat{\varphi}^{-1}\left(\hat{\varphi}\big(T^{-n}(\hat{y})\big) + \frac{s}{p^n}\right)\right)\right].$$

Now for each s in question, $\delta(\hat{y}, n)\left(\frac{s}{p^n}\right)$ is the probability that $\hat{\varphi}(\hat{y})$ arises from a specific pre-image $\hat{\varphi}(T^{-n}(\hat{y})) + \frac{s}{p^n}$. With our probabilistic interpretation, this is then equal to the probability of seeing the pre-image $\hat{\varphi}(T^{-(n-1)}(\hat{y})) + \frac{t}{p^{n-1}}$ multiplied by the probability of seeing $\hat{\varphi}(T^{-n}(\hat{y})) + \frac{s}{p^n}$ given that we have seen $\hat{\varphi}(T^{-(n-1)}(\hat{y})) + \frac{t}{p^{n-1}}$. Interpreting this in terms of the δ distributions we then have that

$$\delta(\hat{y}, n)\left(\frac{s}{p^n}\right) = \delta(\hat{y}, n-1)\left(\frac{t}{p^{n-1}}\right) \cdot \mu_{T^{-(n-1)}(\hat{y})_t}\left(T^n\left(\hat{\varphi}^{-1}\left(\hat{\varphi}\big(T^{-n}(\hat{y})\big) + \frac{s}{p^n}\right)\right)\right)$$

We then establish that property 2. holds, for a fixed value of t, by summing over all values of $s = t\,(mod\,p^{n-1})$ and noting that

$$\mu_{T^{-(n-1)}(\hat{y})_t}\left(\bigsqcup_{s=t\,(mod\,p^{n-1})}\left[T^n\left(\hat{\varphi}^{-1}\left(\hat{\varphi}\big(T^{-n}(\hat{y})\big) + \frac{s}{p^n}\right)\right)\right]\right) = 1.$$

For the third property, we first note that the map $\hat{\varphi} \circ T^{-n} : \widehat{Y} \to \mathbb{T}$ is a measure preserving transformation between two Borel probability spaces and so we may apply corollary 4.3.1. to the countably generated σ-algebra $p^{-n}\mathcal{B}_{\mathbb{T}} \subseteq \mathcal{B}_{\mathbb{T}}$ to see that

$$(\hat{\varphi} \circ T^{-n})_* \hat{\mu}_{\hat{y}}^{\mathcal{B}_{\widehat{Y}}} = \mu_{\hat{\varphi}(T^{-n}(\hat{y}))}^{p^{-n}\mathcal{B}_{\mathbb{T}}}.$$

In the interests of clarity we make the following notational changes; let us omit the superscript σ-algebras in the measures and write $\frac{y}{p^n} = p^{-n}y$ for $\hat{\varphi}(T^{-n}(\hat{y}))$. With these changes, we then note that $\mu_{p^{-n}y}$ is supported on the points $\left\{\frac{y}{p^n} + \frac{t}{p^n}\right\}_{t=0}^{p^n-1}$, and so by the above equality we have that

$$\mu_{p^{-n}y}\left(\frac{y}{p^n} + \frac{t}{p^n}\right) = \hat{\mu}_{\hat{y}}\left(T^n\left(\hat{\varphi}^{-1}\left(\frac{y}{p^n} + \frac{t}{p^n}\right)\right)\right) = \delta(\hat{y}, n)\left(\frac{t}{p^n}\right)$$

Now by the co-primality assumption on p and q we have that $q^{-1}p^{-n}\mathcal{B}_{\mathbb{T}} = p^{-n}\mathcal{B}_{\mathbb{T}}$; so applying corollary 4.3.1 to the measure preserving transformation $S : \mathbb{T} \to \mathbb{T}$ (multiplication by q (mod 1)) we see that $\mu_{qp^{-n}y} = S_*\mu_{p^{-n}y}$.

Then, as application of S in $\widehat{Y}$ corresponds to multiplication by q on $\mathbb{T}$ (via $\hat{\varphi} \circ T^{-n}$), we also have that $\delta(S(\hat{y}), n)\left(\frac{qt}{p^n}\right) = \mu_{qy}\left(\frac{qy}{p^n} + \frac{qt}{p^n}\right)$ and so we conclude that

$$\delta(S(\hat{y}), n)\left(\frac{qt}{p^n}\right) = \mu_{qy}\left(\frac{qy}{p^n} + \frac{qt}{p^n}\right) = S_*\mu_y\left(\frac{qy}{p^n} + \frac{qt}{p^n}\right) = \mu_y\left(\frac{y}{p^n} + \frac{t}{p^n}\right) = \delta(\hat{y}, n)\left(\frac{t}{p^n}\right)$$

as desired.

We again take the probabilistic perspective on the δ distributions. Writing $\delta(T^n(\hat{y}), 2n)\left(\frac{t}{p^{2n}}\right)$ as the probability of seeing the pre-image $\hat{\varphi}\big(T^{-n}(\hat{y})\big) + \frac{t}{p^{2n}}$, given that we've seen the point $\hat{\varphi}(T^i(\hat{y}))$, multiplied by the probability of seeing that point, $\hat{\varphi}(T^i(\hat{y}))$, as a pre-image, given that we've seen $\hat{\varphi}(T^n(\hat{y}))$. Writing this in terms of the δ distributions we have that

$$\delta(T^n(\hat{y}), 2n)\left(\frac{t}{p^{2n}}\right) = \delta(T^i(\hat{y}), n+i)\left(\frac{k}{p^{n+i}}\right) \cdot \delta(T^n(\hat{y}), n-i)(0).$$

Where we are considering values of $t = kp^{n-i}$ for each $0 \le k \le p^{n+i} - 1$.

If we know $\delta(T^n(\hat{y}), 2n)$, then by property 2. proven above we also have knowledge of $\delta(T^n(\hat{y}), n-i)$, and so therefore by the above relation we may determine $\delta(T^i(\hat{y}), n+i)$. □

We are now ready to define the symmetric points for our space $\widehat{Y}$ before proving the following properties; the set of symmetric points is invariant under the action of T and S (hence of measure 0 or 1 by our ergodicity assumption), that if $\hat{y}$ is symmetric then $\delta(\hat{y}, n)$

converges weakly to the Lebesgue measure on $\mathbb{T}$, and finally, that if almost every array in $\widehat{Y}$ is symmetric, then our measure is Lebesgue.

Definition 4.6. *We say that an array $\hat{y} \in \widehat{Y}$ is symmetric if there exist distinct arrays $\hat{y}_1 \neq \hat{y}_2$ such that the following two properties hold:*

1. $\hat{\varphi}(\hat{y}) = \hat{\varphi}(\hat{y}_1) = \hat{\varphi}(\hat{y}_2)$,
2. *For any $n, m \in \mathbb{N}$ we have that $\delta(T^m(\hat{y}_1), n) = \delta(T^m(\hat{y}_2), n)$.*

The reasoning behind calling such an array $\hat{y}$ symmetric, is due to the δ distribution's behaviour, with respect to the pre-images under the action of T, for the two arrays $\hat{y}_1$ and $\hat{y}_2$ when viewed as points on the circle. As the arrays are distinct, there must exist some $n \in \mathbb{N}$ such that $\hat{\varphi}(T^{-n}(\hat{y}_1)) \neq \hat{\varphi}(T^{-n}(\hat{y}_2))$. However, property 2. in the definition of a symmetric point guarantees that the δ distributions assign the same weighting to these two pre-images. With this in mind, we may, loosely speaking, think of symmetric arrays as corresponding to pairs of distinct arrays that are indistinguishable as far as the δ distributions are concerned.

Lemma 4.7. *The set of symmetric points in $\widehat{Y}$ is invariant under the action of T and S and so, by ergodicity of $\widehat{\mathbf{Y}}$, is of measure 0 or 1.*

Proof. To see that the set of symmetric points is invariant under T, note that for a given symmetric point $\hat{y}$ we may take the points $T(\hat{y}_1)$ and $T(\hat{y}_2)$ as the points in the definition to ensure that $T(\hat{y})$ is symmetric.

To establish S invariance we need to exploit property 3. of the δ probability distributions from lemma 4.5.; for a given symmetric point $\hat{y}$ we take the points $S(\hat{y}_1)$ and $S(\hat{y}_2)$ and note that for any $0 \leq t \leq (p^n - 1)$,

$$\delta(T^m S(\hat{y}_1), n)\left(\frac{qt}{p^n}\right) = \delta(T^m(\hat{y}_1), n)\left(\frac{t}{p^n}\right) = \delta(T^m(\hat{y}_2), n)\left(\frac{t}{p^n}\right) = \delta(T^m S(\hat{y}_2), n)\left(\frac{qt}{p^n}\right).$$

Thus, as p and q are coprime, we conclude that $\delta(T^m S(\hat{y}_1), n) = \delta(T^m S(\hat{y}_2), n)$ and hence $S(\hat{y})$ is symmetric. Note here that we have implicitly used the fact that the transformations T and S commute.

Finally, as both transformations T and S are invertible on the space $\widehat{Y}$ and the set of symmetric points is invariant under both transformations, the ergodicity assumption on the measure $\hat{\mu}$ ensures the set is of measure 0 or 1. $\square$

We now establish a small number theoretic lemma to aid us in showing the weak convergence of the δ measures to the Lebesgue measure for symmetric arrays.

Lemma 4.8. *If $a = \sum_{i=1}^{n} \frac{a_i}{p^i}$ (mod 1) for some $n \geq 1$ with $a_i \in \mathbb{Z}$ for all $i \in \{1, ..., n\}$, $a_n \neq 0$ and $-p < a_i < p$, then if $a = \frac{u}{v}$ in least terms, $v \geq 2^n$.*

Proof. We will prove this result by induction on n; note that in the case $n = 1$ we have that $a = \frac{a_1}{p}$ and by assumption $p \geq 2$, so the result holds. Now assuming the result holds for some natural number $n - 1$ we see that

$$ap = p \cdot \sum_{i=1}^{n} \frac{a_i}{p^i} = \sum_{i=1}^{n-1} \frac{a_{i+1}}{p^i} \ (mod\, 1).$$

If in least terms $ap = \frac{\tilde{u}}{\tilde{v}}$ then, by the inductive assumption, we have that $\tilde{v} \geq 2^{n-1}$. Note that if $a = \frac{u}{v}$ in least terms, any prime divisor of v must divide p by the definition of a. We then see that $\tilde{v}$ has at least one fewer prime in its prime decomposition than v does. Therefore $v \geq 2\tilde{v} \geq 2^n$ and so by induction the result holds for all $n \in \mathbb{N}$. □

With all our prior work we are now able to prove a crucial lemma, to do this we will need to utilise the Riemann integral of continuous functions. We note that we may view any continuous function, f, on $\mathbb{T}$ as a continuous function defined on the interval $[0, 1]$, subject to the condition that $f(0) = f(1)$.

Lemma 4.9. *If $\hat{y} \in \widehat{Y}$ is a symmetric point then the measures $\delta(\hat{y}, n)$ weakly converge to the Lebesgue measure, λ, on $\mathbb{T}$ as $n \to \infty$.*

Proof. Let $\hat{y}_1$ and $\hat{y}_2$ be two points satisfying the conditions for $\hat{y}$ to be a symmetric point. As $\hat{y}_1 \neq \hat{y}_2$ there exists some smallest natural number, $i \geq 0$, independent of n, such that $\hat{y}_1(-i, 0) \neq \hat{y}_2(-i, 0)$, or we would contradict lemma 2.3. about horizontal rays determining above symbols. By property 1. of the δ distributions we know that $\delta(\hat{y}, n)$ differs from the distribution $\delta(\hat{y}_1, n)$ by an argument translation (mod 1) of the form, $\hat{\varphi}(T^{-n}(\hat{y}_1)) - \hat{\varphi}(T^{-n}(\hat{y}))$. We also know by the symmetric condition on $\hat{y}$ that in particular $\delta(\hat{y}_1, n) = \delta(\hat{y}_2, n)$ for all $n \geq 0$. Thus as $\delta(\hat{y}, n)$ also differs from $\delta(\hat{y}_2, n)$ by a translation (mod 1) of the form $\hat{\varphi}(T^{-n}(\hat{y}_2)) - \hat{\varphi}(T^{-n}(\hat{y}))$ this implies that the distribution $\delta(\hat{y}, n)$ is invariant under translation by $\hat{\varphi}(T^{-n}(\hat{y}_2)) - \hat{\varphi}(T^{-n}(\hat{y}_1))$ (mod 1). Explicitly, we have that

$$\delta(\hat{y}, n)\left(\frac{t}{p^n} + \hat{\varphi}(T^{-n}(\hat{y}_1)) - \hat{\varphi}(T^{-n}(\hat{y}))\right) = \delta(\hat{y}, n)\left(\frac{t}{p^n} + \hat{\varphi}(T^{-n}(\hat{y}_2)) - \hat{\varphi}(T^{-n}(\hat{y}))\right).$$

Now $\hat{\varphi}(T^{-n}(\hat{y}_2)) - \hat{\varphi}(T^{-n}(\hat{y}_1))$ is of the form $a_n = \sum_{i=1}^{n} \frac{a_i}{p^i}$ $(mod\, 1)$ (as both $\hat{\varphi}(T^{-n}(\hat{y}_1))$

and $\hat{\varphi}(T^{-n}(\hat{y}_2))$ correspond to pre-images in the set $T^{-n}(\hat{\varphi}(\hat{y}))$ and hence differ (mod 1) by some integer multiple of $\frac{1}{p^n}$). We then apply lemma 4.6. and see that the group of transformations (mod 1) that act invariantly on $\delta(\hat{y}, n)$ is of order at least 2^{n-i-1}. Furthermore, this group of transformations has a least element, $m_n \leq \frac{1}{2^{n-i-1}}$. Let us denote by I_n the finite additive group generated by m_n, and note that translation in the argument by any element of I_n leaves $\delta(\hat{y}, n)$ invariant.

If we denote by R_α the rotation by $\alpha \in I_n$ (mod 1) on $\mathbb{T}$, we then have that for any continuous function, f, on the circle

$$\int_{\mathbb{T}} f(x)d\delta(\hat{y}, n)(x) = \int_{\mathbb{T}} R_\alpha f(x)d\delta(\hat{y}, n)(x) = \frac{1}{|I_n|} \sum_{\alpha \in I_n} \int_{\mathbb{T}} R_\alpha f(x)d\delta(\hat{y}, n)(x).$$

By the above discussion we have that the order, $|I_n|$, of the group of I_n is bounded below by 2^{n-i-1}. By viewing f as a continuous function on $[0, 1]$ (with $f(0) = f(1)$) with a partition of the interval given by $\frac{1}{|I_n|}$ (i.e. into $|I_n|$ equal intervals of length $|I_n|$) we see that

$$\lim_{n\to\infty} \frac{1}{|I_n|} \sum_{\alpha \in I_n} R_\alpha f(x) = \int_{[0,1]} f d\lambda = \int_{\mathbb{T}} f d\lambda$$

by definition of the Riemann integral. Hence we have that for any continuous function f on $\mathbb{T}$

$$\lim_{n\to\infty} \int_{\mathbb{T}} f(x)d\delta(\hat{y}, n)(x) = \int_{\mathbb{T}} f d\lambda,$$

which is precisely the definition of $\delta(\hat{y}, n)$ weakly converging to λ (swapping the integral and limit here is justified by the dominated convergence theorem, as f is bounded on $\mathbb{T}$). □

We now present the final result of this section; first let us recall that for any Borel probability space $(X, \mathcal{B}, \mu)$ we have the following chain of equalities

$$\mu(E) = \int_E 1d\mu(x) = \int_X \chi_E(x)d\mu(x) = \int_X \delta_x(E)d\mu(x).$$

Where χ_E denotes the indicator function of some measurable set E, and δ_x denotes the Dirac point measures for $x \in X$.

Theorem 4.10. *If $\hat{\mu}$ almost every $\hat{y} \in \widehat{Y}$ is a symmetric point, then the corresponding measure, μ, on the circle is the Lebesgue measure, λ.*

Proof. Let us again denote by R_α the rotation by $\alpha \in \mathbb{R}$ (mod 1) on $\mathbb{T}$. For almost every $\hat{y} \in \widehat{Y}$ and $n \geq 0$, the probability measure $\delta(\hat{y}, n)$ is supported on $\left\{0, \frac{1}{p^n}, ..., \frac{p^n-1}{p^n}\right\}$, and so the measure $\big(R_{\hat{\varphi}(T^{-n}(\hat{y}))}\big)_* \delta(\hat{y}, n)$ is supported on $\Big\{\hat{\varphi}(T^{-n}(\hat{y})), \hat{\varphi}(T^{-n}(\hat{y})) + \frac{1}{p^n}, ..., \hat{\varphi}(T^{-n}(\hat{y})) + \frac{p^n-1}{p^n}\Big\}$. By lemma 4.9., and the translation invariance of the Lebesgue measure, we conclude that, for almost every $\hat{y} \in \widehat{Y}$, $\big(R_{\hat{\varphi}(T^{-n}(\hat{y}))}\big)_* \delta(\hat{y}, n)$ converges weakly to the Lebesgue measure, λ, as $n \to \infty$.

Recall that in the proof of property 4. in lemma 4.5. we established that $(\hat{\varphi} \circ T^{-n})_* \hat{\mu}_{\hat{y}}^{\mathcal{B}_{\widehat{Y}}} = \mu_{\hat{\varphi}(T^{-n}(\hat{y}))}^{p^{-n}\mathcal{B}_{\mathbb{T}}}$, which in particular establishes that for $E \in \mathcal{B}_{\mathbb{T}}$ we have

$$\mu_{\hat{\varphi}(T^{-n}(\hat{y}))}^{p^{-n}\mathcal{B}_{\mathbb{T}}}(E) = \big(R_{\hat{\varphi}(T^{-n}(\hat{y}))}\big)_* \delta(\hat{y}, n)(E)$$

(this is merely a rephrasing of the calculation in the middle of page 24; where $\mu_{\hat{\varphi}(T^{-n}(\hat{y}))}^{p^{-n}\mathcal{B}_{\mathbb{T}}}(E)$ is the sum over the values $\mu_{\hat{\varphi}(T^{-n}(\hat{y}))}^{p^{-n}\mathcal{B}_{\mathbb{T}}}\Big(\hat{\varphi}(T^{-n}(\hat{y})) + \frac{t}{p^n}\Big)$ for which $\hat{\varphi}\big(T^{-n}(\hat{y})\big) + \frac{t}{p^n} \in E$).

By property 1. of theorem 4.3. we may write, for each $n \in \mathbb{N}$ and $E \in \mathcal{B}_{\mathbb{T}}$, that

$$\mu(E) = \int_{\mathbb{T}} \int_{\mathbb{T}} \chi_E(z) d\mu_x^{p^{-n}\mathcal{B}_{\mathbb{T}}}(z) d\mu(x) = \int_{\mathbb{T}} \mu_x^{p^{-n}\mathcal{B}_{\mathbb{T}}}(E) d\mu(x).$$

Now as the map $\hat{\varphi} : \widehat{Y} \to \mathbb{T}$ takes the measure $\hat{\mu}$ on $\widehat{Y}$ to the measure μ on $\mathbb{T}$ we have

$$\mu(E) = \int_{\mathbb{T}} \mu_x^{p^{-n}\mathcal{B}_{\mathbb{T}}}(E) d\mu(x) = \int_{\widehat{Y}} \mu_{\hat{\varphi}(\hat{y})}^{p^{-n}\mathcal{B}_{\mathbb{T}}}(E) d\hat{\mu}(\hat{y});$$

furthermore, as we have that the map $T^{-n} : \widehat{Y} \to \widehat{Y}$ is a bijection and $\hat{\mu}$ is T invariant (so T^{-n} preserves $\hat{\mu}$),

$$\mu(E) = \int_{\widehat{Y}} \mu_{\hat{\varphi}(\hat{y})}^{p^{-n}\mathcal{B}_{\mathbb{T}}}(E) d\mu(\hat{y}) = \int_{\widehat{Y}} \mu_{\hat{\varphi}(T^{-n}(\hat{y}))}^{p^{-n}\mathcal{B}_{\mathbb{T}}}(E) d\hat{\mu}(\hat{y}).$$

Therefore, for each $n \in \mathbb{N}$ we have

$$\mu(E) = \int_{\widehat{Y}} \big(R_{\hat{\varphi}(T^{-n}(\hat{y}))}\big)_* \delta(\hat{y}, n)(E) d\hat{\mu}(\hat{y}) \xrightarrow{n\to\infty} \int_{\widehat{Y}} \lambda(E) d\hat{\mu}(\hat{y}) = \lambda(E)$$

(where we are using that $\lim_{n\to\infty} \int_{\mathbb{T}} \chi_E(x) d\big(R_{\hat{\varphi}(T^{-n}(\hat{y}))}\big)_* \delta(\hat{y}, n)(x) = \int_{\mathbb{T}} \chi_E(x) d\lambda(x) = \lambda(E)$ as well as the dominated convergence theorem in order to swap the integral and the limit).

So, for any $E \in \mathcal{B}_{\mathbb{T}}$ we have $\mu(E) = \lambda(E)$, i.e. $\mu = \lambda$. □

5 Reducing to the Case of Zero Entropy

We would now like to show that under the assumption of positive entropy of one of the transformations, $\hat{\mu}$ almost every point $\hat{y} \in \widehat{Y}$ is symmetric. Applying the theorem from the previous section, we conclude that under this assumption our measure is Lebesgue, completing our work. In order to show our desired results, we will need to extend the theory of entropy from partitions to that of σ-algebras; for which we shall exploit our theorem on the disintegration of measures.

The entropy theory for σ-algebras extends from that of entropy for partitions in a straightforward manner, via use of the disintegration theorem. Therefore, for the sake of brevity, and also not to detract from the flow of the proof, we quote results regarding entropy theory with respect to σ-algebras from the freely available and detailed book on the subject, [ELW].

Recall that we denote the disintegration, given by theorem 5.3, of a measure μ onto a σ-algebra, $\mathcal{A}$, by measures $\mu_x^{\mathcal{A}}$. For another σ-algebra $\mathcal{C}$ we shall still denote by $[x]_{\mathcal{C}}$, as in definition 2.5., the atom of $\mathcal{C}$ containing x. In the following work, justified by example 2.9 in [ELW], we do not distinguish between a partition ξ and the σ-algebra that it generates, $\sigma(\xi)$.

Definition 5.1. *(2.8. and 2.18. in [ELW]) For a Borel probability space, $(X, \mathcal{B}, \mu)$, and two countably generated sub-σ-algebras $\mathcal{A}, \mathcal{C} \subseteq \mathcal{B}$ we define the entropy of $\mathcal{C}$ given $\mathcal{A}$ to be*

$$H_\mu(\mathcal{C} \mid \mathcal{A}) = -\int_X \log \left[\mu_x^{\mathcal{A}}([x]_{\mathcal{C}})\right] d\mu(x).$$

Furthermore, if $\mathcal{A}$ is strictly invariant, i.e. such that $T^{-1}(\mathcal{A}) = \mathcal{A}$, then by first defining

$$h_\mu(T, \xi \mid \mathcal{A}) = \lim_{n\to\infty} \frac{1}{n} H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}(\xi) \,\middle|\, \mathcal{A} \right)$$

we may define the conditional entropy of T given $\mathcal{A}$ to be

$$h_\mu(T \mid \mathcal{A}) = \sup_{\{\xi \mid H_\mu(\xi) < \infty\}} h_\mu(T, \xi \mid \mathcal{A}).$$

We now will state the properties of entropy with respect to σ-algebras that we need, most of which are in analogy with results about entropy with respect to partitions.

Proposition 5.2. *(2.13., 2.17., 2.19., 2.20. and 2.21. in [ELW]) For a dynamical system, $(X, \mathcal{B}, \mu, T)$, on a Borel probability space and countably generated sub-σ-algebras $\mathcal{A}, \mathcal{C}$ and*

$\widetilde{\mathcal{C}}$ *of* $\mathcal{B}$*, the following properties of conditional entropy with respect to* σ*-algebras hold:*

1. *(Additivity)* $H_\mu(\mathcal{C} \vee \widetilde{\mathcal{C}} \mid \mathcal{A}) = H_\mu(\mathcal{C} \mid \mathcal{A}) + H_\mu(\widetilde{\mathcal{C}} \mid \mathcal{A} \vee \mathcal{C})$*.*
2. *(Monotonicity)* $H_\mu(\widetilde{\mathcal{C}} \mid \mathcal{A} \vee \mathcal{C}) \leq H_\mu(\widetilde{\mathcal{C}} \mid \mathcal{A})$*.*
3. *(Invariance)* $H_\mu(\mathcal{C} \mid \mathcal{A}) = H_\mu(T^{-1}(\mathcal{C}) \mid T^{-1}(\mathcal{A}))$*.*
4. *(Future formula) If* $\mathcal{A}$ *is strictly invariant and* ξ *is a countable partition of finite entropy then*

$$h_\mu(T, \xi \mid \mathcal{A}) = H_\mu\left(\xi \,\middle|\, \bigvee_{i=0}^{\infty} T^{-i}(\xi) \vee \mathcal{A}\right).$$

5. *(Kolmogorv-Sinai Theorem) If* $(\xi_k)_{k\in\mathbb{N}}$ *is a sequence of finite entropy partitions such that* $\xi_k \subseteq \sigma(\xi_{k+1})$ *for all* $k \in \mathbb{N}$*, and up to sets of zero measure either* $\mathcal{B} = \bigvee_{k=1}^{\infty} \bigvee_{i=0}^{\infty} T^{-i}(\xi_k)$ *or* $\mathcal{B} = \bigvee_{k=1}^{\infty} \bigvee_{i=-\infty}^{\infty} T^{-i}(\xi_k)$*, then we have that*

$$h_\mu(T) = \lim_{k\to\infty} h_\mu(T, \xi_k).$$

Furthermore, under the same assumptions, if $\mathcal{A}$ *is strictly invariant then*

$$h_\mu(T \mid \mathcal{A}) = \lim_{k\to\infty} h_\mu(T, \xi_k \mid \mathcal{A}).$$

6. *(Abramov-Rokhlin formula) If* $(X, \mathcal{B}, \mu, T) \xrightarrow{\Psi} (Z, \mathcal{B}_Z, \nu, S)$ *is a factor map (defined analogously to definition 3.4.) then we have that*

$$h_\mu(T) = h_\nu(S) + h_\mu(T \mid \mathcal{A}),$$

where we identify $(Z, \mathcal{B}_Z, \nu, S)$ *with the strictly invariant* σ*-algebra* $\mathcal{A} = \Psi^{-1}(\mathcal{B}_Z)$*.*

With the necessary definitions and properties established, we return to the setting of our proof. Let us denote by $\mathcal{P}$ the partition of the set $\widehat{Y}$ determined by the symbol $\hat{y}(0,0)$, and make the convention that every element of this partition has strictly positive measure with respect to $\hat{\mu}$. Note that this partition directly corresponds to the partition given by the I_j (from section 2) of $\mathbb{T}$ (i.e. the partition $\mathcal{P}$ tells us which I_j the point $\hat{\varphi}(\hat{y})$ is contained in). We shall first show that the partition $\mathcal{P}$, while not generating for $\mathbf{Y}$, gives the entropy for the system $\widehat{\mathbf{Y}}$ in the following sense:

Lemma 5.3. *(9.10. from [ELW] phrased in our terminology.) The partition* $\mathcal{P}$*, while not necessarily a generator for the dynamical system* $\widehat{\mathbf{Y}} = (\widehat{Y}, \mathcal{B}_{\widehat{Y}}, \hat{\mu}, (T, S))$ *is such that*

$$h_{\hat{\mu}}(T) = h_{\hat{\mu}}(T, \mathcal{P}),\ h_{\hat{\mu}}(S) = h_{\hat{\mu}}(S, \mathcal{P}).$$

Furthermore, by the Kolmogorov-Sinai theorem we have that for any strictly invariant σ-algebra $\mathcal{A}$,

$$h_{\hat{\mu}}(T \mid \mathcal{A}) = h_{\hat{\mu}}(T, \mathcal{P} \mid \mathcal{A}),\ h_{\hat{\mu}}(S \mid \mathcal{A}) = h_{\hat{\mu}}(S, \mathcal{P} \mid \mathcal{A}).$$

Proof. We first note that the σ-algebra $\mathcal{C} = \bigvee_{i=-\infty}^{\infty} T^{-i}\mathcal{P}$ generates the Borel σ-algebra, $\mathcal{B}_{\widehat{Y}}$, under the action of S. Hence we may apply the Kolmogorov-Sinai theorem (property 5. of propisition 5.2.) to see that $h_{\hat{\mu}}(T) = \lim_{n\to\infty} h_{\hat{\mu}}(T, S^n\mathcal{P})$.

As the transformation S is invertible on $\widehat{\mathbf{Y}}$, we have that $\hat{\mu}(E) = \hat{\mu}(S^{-1}S(E)) = \hat{\mu}(S(E))$ by S invariance of $\hat{\mu}$; therefore, for each $n \in \mathbb{N}$ we have $h_{\hat{\mu}}(T, S^n\mathcal{P}) = h_{\hat{\mu}}(T, \mathcal{P})$. Thus we conclude that $h_{\hat{\mu}}(T) = h_{\hat{\mu}}(T, \mathcal{P})$; the argument for the entropy of S is symmetric to that of T.

The final statement of the lemma follows by an identical argument; first conditioning on the σ-algebra $\mathcal{A}$ and then exploiting the second statement of the Kolmogorov-Sinai theorem. □

Our next objective for the this section is to relate the entropy of the two transformations on the space $\widehat{Y}$ with respect to the measure $\hat{\mu}$. Then, if we can show that we have zero entropy with respect to one of the transformations, we must have zero entropy for the other. Consider the dynamical system $(\widehat{Y}, \mathcal{A}, \hat{\mu}, T)$ for some strictly invariant sub-σ-algebra $\mathcal{A} \subseteq \mathcal{B}$; let us denote the entropy of this system by $h_{\hat{\mu}}^{\mathcal{A}}(T)$. Similarly we write $h_{\hat{\mu}}^{\mathcal{A}}(S)$ for the entropy of the dynamical system $(\widehat{Y}, \mathcal{A}, \hat{\mu}, S)$ (we still denote by $h_{\hat{\mu}}(T)$ the entropy of the system $(\widehat{Y}, \mathcal{B}_{\widehat{Y}}, \hat{\mu}, T)$, and similarly for $h_{\hat{\mu}}(S)$).

Lemma 5.4. *For any T and S strictly invariant sub-σ-algebra, $\mathcal{A}$, of the system $\widehat{\mathbf{Y}}$ we have the following relation,*

$$h_{\hat{\mu}}^{\mathcal{A}}(T) = \frac{\log(p)}{\log(q)} h_{\hat{\mu}}^{\mathcal{A}}(S).$$

Proof. (proof of claim is 9.11 from [ELW]) Firstly, we claim that it is sufficient to prove that $h_{\hat{\mu}}(T, \mathcal{P} \mid \mathcal{A}) = \frac{\log(p)}{\log(q)} h_{\hat{\mu}}(S, \mathcal{P} \mid \mathcal{A})$. We now show that the claim implies the desired result. We use the Abramov-Rokhlin formula, property 6. of proposition 5.2, in order to write

$$h_{\hat{\mu}}(T, \mathcal{P}) = h_{\hat{\mu}}^{\mathcal{A}}(T) + h_{\hat{\mu}}(T, \mathcal{P} \mid \mathcal{A}),\ h_{\hat{\mu}}(S, \mathcal{P}) = h_{\hat{\mu}}^{\mathcal{A}}(S) + h_{\hat{\mu}}(S, \mathcal{P} \mid \mathcal{A})$$

(where we've implicitly used remark 3.5. from section 3).

Assuming that $h_{\hat{\mu}}(T,\mathcal{P}|\mathcal{A}) = \frac{\log(p)}{\log(q)} h_{\hat{\mu}}(S,\mathcal{P}|\mathcal{A})$ and rearranging, we then have the following:

$$h_{\hat{\mu}}^{\mathcal{A}}(T) = \frac{\log(p)}{\log(q)} h_{\hat{\mu}}^{\mathcal{A}}(S) + \left(h_{\hat{\mu}}(T,\mathcal{P}) - \frac{\log(p)}{\log(q)} h_{\hat{\mu}}(S,\mathcal{P}) \right).$$

Then, by considering the trivial σ-algebra, $\mathcal{N} = \{\widehat{Y}, \emptyset\}$, we have that (see page 54 in [ELW] for details)

$$h_{\hat{\mu}}(T,\mathcal{P} \mid \mathcal{N}) = h_{\hat{\mu}}(T,\mathcal{P}),\ h_{\hat{\mu}}(S,\mathcal{P} \mid \mathcal{N}) = h_{\hat{\mu}}(S,\mathcal{P}).$$

Therefore we establish that $h_{\hat{\mu}}^{\mathcal{A}}(T) = \frac{\log(p)}{\log(q)} h_{\hat{\mu}}^{\mathcal{A}}(S)$, proving the result.

In order to prove our claim we first bound the entropy of a finite partition; suppose that $\alpha = \{A_1, ..., A_n\}$ is a finite partition of some measure space $(X, \mathcal{B}, \mu)$ into strictly positive measure sets. We then note that the entropy of the partition satisfies

$$H_\mu(\alpha) = -\sum_{i=1}^{n} \mu(A_i) \log(\mu(A_i)) \leq \log(n).$$

This follows by applying the weighted arithmetic-geometric mean inequality (a simple consequence of Jensen's inequality), as we have that

$$\exp(H_\mu(\alpha)) = \prod_{i=1}^{n} \left(\frac{1}{\mu(A_i)}\right)^{\mu(A_i)} \leq \sum_{i=1}^{n} \mu(A_i) \left(\frac{1}{\mu(A_i)}\right) = n.$$

We now consider the partition $\mathcal{P}$ as defined above. Note that for any $n, m \geq 1$ the partitions $\bigvee_{i=0}^{n-1} S^{-i}\mathcal{P}$ and $\bigvee_{i=0}^{m-1} T^{-i}\mathcal{P}$ are comprised of intervals of length $\mathcal{L}_q(n) := \frac{1}{pq^n}$ and $\mathcal{L}_p(m) := \frac{1}{qp^m}$ respectively.

Now as p and q are coprime, $\frac{\log(q)}{\log(p)}$ is irrational, and so for each $\epsilon > 0$ we may find an $n \in \mathbb{N}$ such that $\left\{\frac{\log(q)}{\log(p)} n\right\} < \epsilon$ (where $\{x\}$ denotes the fractional part of x). We then define $m = \left\lfloor \frac{\log(q)}{\log(p)} n \right\rfloor$ (where $\lfloor x \rfloor$ is the largest integer less than or equal to x); allowing us to conclude that

$$p^m \leq q^n \leq p^{m+\epsilon}$$

(where we have used that $\frac{\log(q)}{\log(p)} = \log_p(q)$ and that $\log_p(q)n = m + \{\log_p(q)n\}$).

For $\epsilon > 0$ such that $p^\epsilon \leq \frac{q}{p}$, by defining n and m as above we have that, as $p < q$,

$$\frac{\mathcal{L}_p(m)}{\mathcal{L}_q(n)} = \frac{pq^n}{qp^m} \leq \frac{q^n}{p^m} \leq p^\epsilon \leq \frac{q}{p}$$

and

$$\frac{\mathcal{L}_q(n)}{\mathcal{L}_p(m)} = \frac{qp^m}{pq^n} \leq \frac{q}{p}.$$

Thus, setting $k := \lceil \frac{q}{p} \rceil$, we conclude that each interval in $\bigvee_{i=0}^{n-1} S^{-i}\mathcal{P}$ is contained in at most k intervals, and hence belongs to a partition of size at most k, of $\bigvee_{i=0}^{m-1} T^{-i}\mathcal{P}$; and visa versa. Hence by our above bound on entropy for finite partitions we have that

$$H_{\hat{\mu}}\left(\bigvee_{i=0}^{n-1} S^{-i}\mathcal{P} \,\middle|\, \bigvee_{i=0}^{m-1} T^{-i}\mathcal{P} \right) \leq \log(k)$$

and

$$H_{\hat{\mu}}\left(\bigvee_{i=0}^{m-1} T^{-i}\mathcal{P} \,\middle|\, \bigvee_{i=0}^{n-1} S^{-i}\mathcal{P} \right) \leq \log(k).$$

We now exploit the additivity and monotonicity of conditional entropy, given by properties 1. and 2. in proposition 5.2., to see that

$$\begin{aligned}
H_{\hat{\mu}}\left(\bigvee_{i=0}^{n-1} S^{-i}\mathcal{P} \,\middle|\, \mathcal{A} \right) &\leq H_{\hat{\mu}}\left(\bigvee_{i=0}^{n-1} S^{-i}\mathcal{P} \vee \bigvee_{i=0}^{m-1} T^{-i}\mathcal{P} \,\middle|\, \mathcal{A} \right) \\
&= H_{\hat{\mu}}\left(\bigvee_{i=0}^{m-1} T^{-i}\mathcal{P} \,\middle|\, \mathcal{A} \right) + H_{\hat{\mu}}\left(\bigvee_{i=0}^{n-1} S^{-i}\mathcal{P} \,\middle|\, \bigvee_{i=0}^{m-1} T^{-i}\mathcal{P} \vee \mathcal{A} \right) \\
&\leq H_{\hat{\mu}}\left(\bigvee_{i=0}^{m-1} T^{-i}\mathcal{P} \,\middle|\, \mathcal{A} \right) + H_{\hat{\mu}}\left(\bigvee_{i=0}^{n-1} S^{-i}\mathcal{P} \,\middle|\, \bigvee_{i=0}^{m-1} T^{-i}\mathcal{P} \right) \\
&\leq H_{\hat{\mu}}\left(\bigvee_{i=0}^{m-1} T^{-i}\mathcal{P} \,\middle|\, \mathcal{A} \right) + \log(k)
\end{aligned}$$

(using non-negativity of entropy alongside additivity for the first inequality, additivity for the equality, monotonicity for the second inequality and finally our bound for the third inequality).

By dividing by n and taking the limit in the above calculation, we conclude

$$h_{\hat{\mu}}(S, \mathcal{P} \mid \mathcal{A}) \leq \lim_{n\to\infty} \frac{m}{n} h_{\hat{\mu}}(T, \mathcal{P} \mid \mathcal{A}) = \frac{\log(q)}{\log(p)} h_{\hat{\mu}}(T, \mathcal{P} \mid \mathcal{A}).$$

Switching the roles of S and T in the above calculations see that

$$h_{\hat{\mu}}(T, \mathcal{P} \mid \mathcal{A}) \leq \lim_{m \to \infty} \frac{n}{m} h_{\hat{\mu}}(S, \mathcal{P} \mid \mathcal{A}) = \frac{\log(p)}{\log(q)} h_{\hat{\mu}}(S, \mathcal{P} \mid \mathcal{A}).$$

Therefore we have that

$$h_{\hat{\mu}}(S, \mathcal{P} \mid \mathcal{A}) = \frac{\log(q)}{\log(p)} h_{\hat{\mu}}(T, \mathcal{P} \mid \mathcal{A}),$$

proving the claim. □

We now define $\mathcal{H}$ to be the smallest T and S strictly invariant σ-algebra for which, for any $n \geq 0$ and almost every $\hat{y} \in \widehat{Y}$, all of the $\delta(\hat{y}, n)$ probability distributions are measurable when viewed as functions.

For each $n \in \mathbb{N}$ we define $\mathcal{H}_n$ to be the smallest σ-algebra such that, for any $\hat{y} \in \widehat{Y}$, the $\delta(T^n(\hat{y}), 2n)$ are measurable when viewed as functions. Property 4. of the δ distributions in lemma 4.5 implies that $\delta(T^n(\hat{y}), 2n)$ determines $\delta(T^i(\hat{y}), n+i)$ for any $0 \leq i \leq n$ (in particular $\delta(\hat{y}, n)$). Therefore, the σ-algebras $\mathcal{H}_n$ are nested and refine in n to the above defined σ-algebra, $\mathcal{H}$.

Lemma 5.5. *For any $n \geq 0$, the transformation S acts periodically on the σ-algebra $\mathcal{H}_n$; in the sense that there exists a natural number i_n such that for any $E \in \mathcal{H}_n$ we have that $S^{i_n}(E) = E$.*

Proof. As the transformation S acts by multiplication by q, for every $n \in \mathbb{N}$ the pigeonhole principle guarantees that there exists two natural numbers $k > l$ (dependent on n) such that $q^k = q^l \, (mod\, p^{2n})$. By setting $i_n := k - l > 0$ we have that S^{i_n} acts by the identity, i.e. for any $0 \leq t \leq p^{2n} - 1$ we have that $q^{i_n} t = t \, (mod\, p^{2n})$. Exploiting property 3. of the δ distributions inductively, for $\hat{\mu}$ almost every $\hat{y} \in \widehat{Y}$ we have that

$$\delta(S^{i_n}(T^n(\hat{y})), 2n)\left(\frac{q^{i_n} t}{p^{2n}}\right) = \delta(T^n(\hat{y}), 2n)\left(\frac{t}{p^{2n}}\right).$$

But by the above we see that $\frac{q^{i_n} t}{p^{2n}} = \frac{t}{p^{2n}} \, (mod\, 1)$ and hence that

$$\delta(S^{i_n}(T^n(\hat{y})), 2n) = \delta(T^n(\hat{y}), 2n).$$

If application of the transformation S^{i_n} to an array in $\widehat{Y}$ leaves the δ distributions unchanged as functions, this means that the action of S^{i_n} leaves the σ-algebra $\mathcal{H}_n$ unchanged (by their construction as the smallest σ-algebra such that these δ distributions are measurable). Hence, for any $E \in \mathcal{H}_n$ we have that $S^{i_n}(E) = E$. □

As in lemma 5.4, letting $h_{\hat{\mu}}^{\mathcal{H}}(S)$ denote the entropy of the dynamical system $(\widehat{Y}, \mathcal{H}, \hat{\mu}, S)$, we establish the following corollary.

Corollary 5.5.1. *For the dynamical system $(\widehat{Y}, \mathcal{H}, \hat{\mu}, S)$ we have that $h_{\hat{\mu}}^{\mathcal{H}}(S) = 0$.*

Proof. We first note that the identity map on any measure space gives zero entropy; which follows from the property that for $k \in \mathbb{N}$ we have that $h_\mu(T^k) = kh_\mu(T)$ for any dynamical system $(X, \mathcal{B}, \mu, T)$. By lemma 5.4., as S^{i_n} acts by the identity, Id, on all $\mathcal{H}_n$ measurable sets and as $i_n \geq 1$;

$$i_n \cdot h_{\hat{\mu}}^{\mathcal{H}_n}(S) = h_{\hat{\mu}}^{\mathcal{H}_n}(S^{i_n}) = h_{\hat{\mu}}^{\mathcal{H}_n}(Id) = 0.$$

We then have that $h_{\hat{\mu}}^{\mathcal{H}_n}(S) = 0$ and now show that this allows us to conclude $h_{\hat{\mu}}^{\mathcal{H}}(S) = 0$.

To do this we will exploit the fact that entropy of a transformation can be computed by considering only finite partitions (lemma 1.19 in [ELW]). Thus, it is sufficient to prove that for an arbitrary finite partition, ξ, in $\mathcal{H}$ we have $h_{\hat{\mu}}^{\mathcal{H}}(S, \xi) = 0$. We first fix $\epsilon > 0$ and note that the σ-algebra generated by the union of the $\mathcal{H}_n$ generates $\mathcal{H}$. As ξ is a finite partition in $\mathcal{H}$, we may use theorem 4.16. from [W] (equivalently exercises 1.1.5. and 1.1.6. in [ELW]) to find a finite partition, ξ_n, in the union of the $\mathcal{H}_n$, and hence in $\mathcal{H}_n$ for some $n \in \mathbb{N}$, such that

$$H_{\hat{\mu}}^{\mathcal{H}}(\xi \mid \xi_n) \leq H_{\hat{\mu}}^{\mathcal{H}}(\xi \mid \xi_n) + H_{\hat{\mu}}^{\mathcal{H}}(\xi_n \mid \xi) < \epsilon.$$

Now, as $\xi_n \subset \mathcal{H}_n$ we also have that $h_{\hat{\mu}}^{\mathcal{H}}(S, \xi_n) \leq h_{\hat{\mu}}^{\mathcal{H}_n}(S) = 0$. By proposition 1.17. in [ELW] (the continuity bound) we have that

$$h_{\hat{\mu}}^{\mathcal{H}}(S, \xi) \leq h_{\hat{\mu}}^{\mathcal{H}}(S, \xi_n) + H_{\hat{\mu}}^{\mathcal{H}}(\xi \mid \xi_n) < \epsilon.$$

But as ϵ above was arbitrary, we have that $h_{\hat{\mu}}^{\mathcal{H}}(S, \xi) = 0$ for any finite partition, ξ, in $\mathcal{H}$; hence we conclude that $h_{\hat{\mu}}^{\mathcal{H}}(S) = 0$. □

Lemma 5.6. *If $\hat{\mu}$ almost every $\hat{y} \in \widehat{Y}$ is not symmetric then*

$$T(\mathcal{P}) \subseteq \mathcal{H} \vee \bigvee_{i=0}^{\infty} T^{-i}(\mathcal{P}).$$

Proof. As we've assumed that $\mathcal{P}$ consists only of $\hat{\mu}$ positive measure sets, $T(\mathcal{P})$ must also. If we assume that the inclusion doesn't hold, then there must exist some set of strictly positive measure, $S \in T(\mathcal{P})$, such that for each point $\hat{y} \in S$, knowing $\delta(\hat{y}, n)$ for all $n \geq 0$ and $\hat{\varphi}(\hat{y})$ doesn't determine the element of $T(\mathcal{P})$ that contains $\hat{y}$.

For each such $\hat{y} \in S$ and fixed $n \in \mathbb{N}$ there must exist two distinct arrays, say $\hat{y}_1 \neq \hat{y}_2$, such that $\hat{\varphi}(\hat{y}) = \hat{\varphi}(\hat{y}_1) = \hat{\varphi}(\hat{y}_2)$, $\delta(\hat{y}_1, n) = \delta(\hat{y}_2, n)$ and that $\hat{\varphi}(T^{-1}(\hat{y}_1)) \neq \hat{\varphi}(T^{-1}(\hat{y}_2))$. Note that by construction of the δ distributions, knowing $\delta(\hat{y}, n)$ for $n \geq 0$ also determines $\delta(T^m(\hat{y}), n)$ for any $m \geq 0$; as $\delta(\hat{y}, n)$ encodes information about the pre-images of $\hat{y}$ under T. For the above arrays $\hat{y}_1$ and $\hat{y}_2$, we then also have that $\delta(T^m(\hat{y}_1), n) = \delta(T^m(\hat{y}_2), n)$ for any $m \geq 0$.

We now consider, for each $n \in \mathbb{N}$, the equivalence class of arrays, with $\hat{y}_1$ and $\hat{y}_2$ deemed equivalent if they satisfy $\delta(T^m(\hat{y}_1), n) = \delta(T^m(\hat{y}_2), n)$ for all $m \in \mathbb{N}$. Such equivalence classes of arrays are non-empty, closed (as intersections of cylinder sets) so compact, and decreasingly nested in n (by property 2. in lemma 4.5. of the δ distributions). Now as there are only p choices for the value of $\hat{\varphi}\big(T^{-1}(\hat{y})\big)$, we may choose a subsequence of these equivalence classes (indexing them by $n \in \mathbb{N}$), and apply Cantor's intersection theorem to extract two arrays $\hat{y}_1 \neq \hat{y}_2$ with the property that $\hat{\varphi}(\hat{y}) = \hat{\varphi}(\hat{y}_1) = \hat{\varphi}(\hat{y}_2)$, and for each $n > 0$, $m \geq 0$ we have that $\delta(T^m(\hat{y}_1), n) = \delta(T^m(\hat{y}_2), n)$ (i.e. $\hat{y}$ is symmetric). We have shown that every point $\hat{y} \in S$ is symmetric; but S has strictly positive measure and so, by lemma 4.7., almost every $\hat{y} \in \widehat{Y}$ is symmetric, contradicting our assumption. $\square$

We have now collected all of the results needed to prove Rudolph's theorem, which will complete our work. Note that the following theorem implies theorem 1.7. holds.

Theorem 5.7. *For any $\hat{\mu} \in \widehat{\mathcal{M}_0}$ such that $\hat{\mu} \neq \hat{\lambda}$ (i.e. μ does not arise from the Lebesgue measure, λ, on $\mathbb{T}$) we have that $h_{\hat{\mu}}(T) = h_{\hat{\mu}}(S) = 0$.*

Proof. Firstly we use the Abramov-Rokhlin Formula, property 6. of proposition 5.2., along with lemma 5.3 to write

$$h_{\hat{\mu}}(T, \mathcal{P}) = h_{\hat{\mu}}(T) = h_{\hat{\mu}}^{\mathcal{H}}(T) + h_{\hat{\mu}}(T \mid \mathcal{H}) = h_{\hat{\mu}}^{\mathcal{H}}(T) + h_{\hat{\mu}}(T, P \mid \mathcal{H}).$$

As our measure does not correspond to the Lebesgue measure on the circle, $\hat{\mu}$ almost every $\hat{y} \in \widehat{Y}$ is not symmetric. Therefore, we may then apply lemma 5.6 and the future formula, property 4. of proposition 5.2., to see that

$$h_{\hat{\mu}}(T, \mathcal{P} \mid \mathcal{H}) = H_{\hat{\mu}}\left(\mathcal{P} \,\middle|\, \bigvee_{i=1}^{\infty} T^{-i}(P) \vee \mathcal{H}\right) = H_{\hat{\mu}}\left(T(\mathcal{P}) \,\middle|\, \bigvee_{i=0}^{\infty} T^{-i}(\mathcal{P}) \vee \mathcal{H}\right) = 0$$

Note that the second equality used the invariance property of conditional entropy,property 3. of proposition 5.2., combined with the fact that T is invertible and $\mathcal{H}$ is T invariant. Thus we have that $h_{\hat{\mu}}(T, \mathcal{P}) = h_{\hat{\mu}}^{\mathcal{H}}(T)$.

By applying lemma 5.4. and corollary 5.5.1. we conclude that

$$0 = h^{\mathcal{H}}_{\hat{\mu}}(S) = \frac{\log(q)}{\log(p)} h^{\mathcal{H}}_{\hat{\mu}}(T) = \frac{\log(q)}{\log(p)} h_{\hat{\mu}}(T, \mathcal{P}) = h_{\hat{\mu}}(S, \mathcal{P})$$

(where as in the proof of lemma 5.4. for the final equality we use that $h_{\hat{\mu}}(T, \mathcal{P} \mid \mathcal{N}) = h_{\hat{\mu}}(T, \mathcal{P})$, $h_{\hat{\mu}}(S, \mathcal{P} \mid \mathcal{N}) = h_{\hat{\mu}}(S, \mathcal{P})$).

From this we see that $h_{\hat{\mu}}(T, \mathcal{P}) = h_{\hat{\mu}}(S, \mathcal{P}) = 0$, which by lemma 5.3 allows us to conclude that $h_{\hat{\mu}}(T) = h_{\hat{\mu}}(S) = 0$. □

Thus we have established the theorem of Rudolph, concluding that under the assumption of positive entropy of one of our transformations, T or S, our measure must be Lebesgue. To briefly recapitulate our argument; we first encapsulated the dynamics of the circle under the maps T and S, in a space of two dimensional one sided symbolic dynamics. Via the technique of inverse limits of dynamical systems, we constructed an invertible dynamical system $\widehat{\mathbf{Y}}$ that lifted uniquely from $\mathbf{Y}$ and inherited properties of measures, erogdicity and entropy directly from the circle. We then focused on $\widehat{\mathbf{Y}}$, constructing the sequence of δ measures, via the disintegration theorem, that were crucial in distinguishing the Lebesgue measure through the notion of symmetric points. After establishing lemmas about symmetric points, we showed that our measure must be Lebesgue in the case that the set of symmetric points is of full measure. Finally, we showed that under the assumption of positive entropy, almost every point is symmetric and so our original measure was Lebesgue.

6 Extension and Developments of Rudolph's Theorem

To conclude the report, in this final section we discuss some improvements and a relaxation of Rudolph's theorem, before mentioning its influence on an open problem in number theory.

The first immediate extensions of Rudolph's theorem were to powers of products of co-prime numbers in his original paper (corollary 4.11 in [R]), though this was generalised shortly after by Johnson, in [J], to the following stronger result:

Theorem 6.1. *(Theorem B in [J]) Let Σ be a non-lacunary semi-group of the integers and let μ be an invariant Borel probability measure on $\mathbb{T}$ that is ergodic for the transformations induced by multiplication of elements of Σ. Then either μ is the Lebesgue measure, or each element of Σ gives zero entropy.*

This extension of Rudolph's theorem is in analogy with the more general version of theorem 1.2. from section 1, which was proven in [F].

Two more recent papers; [HS1] from 2012 and [HS2] from 2015, of Hochman and Shmerkin establish two results of different flavours, both of which are shown directly to give short proofs of Rudolph's theorem. The first of these results (theorem 1.3. in [HS1]) is about the Hausdorff dimension of measures on $\mathbb{T}$ invariant under the transformations T and S, and the second of which (theorem 10.4. in [HS2]) is about measures on $\mathbb{T}$ that are T invariant and supported on points that are normal base q. The latter of these results generalises theorem 1. from the earlier paper of Host, [Hos].

A weakened version of Rudolph's theorem has also been established, utilising different methods to those in the original paper and this report. Under the stronger assumption that we have ergodicity with respect to a single transformation, the blog post [Ma] by Matheus presents a proof of the Rudolph's theorem using techniques from Fourier analysis; in a similar manner to the original result of Lyons (in [Ly]), which Rudolph's theorem strengthens. For the interested reader, a discussion and extension, to higher dimensional toral endomorphisms, of the results in the above mentioned blog post may be found in Zickert's masters thesis, [Z].

Finally, we briefly discuss the implications of Rudolph's theorem, as mentioned in [Ma]. In recent years Rudolph's theorem has served, in a qualitative sense, to motivate approaches towards an open problem in number theory, namely that of the Littlewood conjecture; an accessible introduction and survey of which may be found in [V]. This open conjecture concerns the simultaneous approximation of two real numbers by rationals, and is stated as follows:

Conjecture 6.2. *(Littlewood 30') For every $x, y \in \mathbb{R}$,*

$$\liminf_{n\to\infty} n\{nx\}\{ny\} = 0$$

(where $\{x\}$ denotes the fractional part of x, i.e. the distance to the nearest integer).

Though this conjecture is still open, the strongest result towards it is obtained in [EKL]. Precisely, the result obtained is the following:

Theorem 6.3. *(Theorem 1.5 in [EKL]) Let*

$$\Xi := \left\{ (x,y) \in \mathbb{R}^2 \;\middle|\; \liminf_{n\to\infty} n\{nx\}\{ny\} > 0 \right\}$$

Then Ξ is of Hausdorff dimension zero.

Without us explicitly defining the notion of Hausdorf dimension, we note that this theorem establishes that the set of exceptions to the Littlewood conjecture, Ξ, is relatively small; in particular of zero Lebesgue measure. This theorem is established as a corollary to the central results of [EKL]; which are qualitatively similar to Rudolph's theorem, in that they reduce another conjecture (1.1 in [EKL]) to looking at measures of zero entropy.

In conclusion, while conjectures 1.6 and 6.2 are still both open; the entropy theoretic approaches that Rudolph's theorem has afforded and motivated are certainly significant.

Bibliography

[F] Harry Furstenberg, *Disjointness in Ergodic Theory, Minimal Sets, and a Problem in Diophantine Approximation*, Mathematical Systems Theory (1967), 1, 1-49.

[Bo] Michael D. Boshernitzan, *Elementary Proof of Furstenberg's Diophantine Result*, Proceedings of the American Mathematical Society (1994), Volume 122, Number 1, 67-70.

[Ly] Russell Lyons, *On Measures Simultaneously 2- and 3-Invariant*, Israel Journal of Mathematics (1988), Volume 61, Number 2, 219-224.

[R] Daniel J. Rudolph, *x2x3 Invariant Measures and Entropy*, Ergodic Theory and Dynamical Systems (1990), 10, 396-406.

[DGS] Manfred Denker, Christian Grillenberger and Karl Sigmund, *Ergodic Theory on Compact Spaces*, Springer-Verlag Lecture Notes in Mathematics (1976), Volume 527.

[K] Jonathan L. F. King, *Entropy in Ergodic Theory*, Encyclopedia of Complexity and Systems Science (2009).

[PY] M. Pollicott and M. Yuri, *Dynamical Systems and Ergodic Theory*, Cambridge University Press (1998).

[ELW] Manfred Einsiedler, Elon Lindenstrauss, and Thomas Ward, *Entropy in ergodic theory and homogeneous dynamics*, `https://tbward0.wixsite.com/books/entropy`.

[EW] Manfred Einsiedler and Thomas Ward, *Ergodic Theory with a view towards number theory*, Springer-Verlag (2011).

[C] Donald L. Cohn, *Measure Theory*, Second Edition, Birkhäuser (2013).

[Br] James R. Brown, *Inverse Limits, Entropy and Weak Isomorphism for Discrete Dynamical Systems*, Transactions of the American Mathematical Society (1972), Volume 164, 55-66.

[I] W.T. Ingram, *Inverse Limits and Dynamical Systems*, Open Problems in Topology II (2007), 289-301.

[Le] M. Lemańczyk, *Weakly Isomorphic Transformations that are not Isomorphic*, Probability Theory and Related Fields (1988), Volume 78, 491–507.

[P] S. Polit, *Weakly Isomorphic maps need not be Isomorphic*, Ph. D. dissertation, Stanford (1974).

[Hoc] Michael Hochman, *Notes on Ergodic Theory*, `http://math.huji.ac.il/~mhochman/courses/ergodic-theory-2012/notes.final.pdf`.

[Mo] Joel Moreira, *Ergodic Theory - Notes*, `https://sites.math.northwestern.edu/~jmoreira/ErgodicNotes`.

[S] V.S. Sunder, *The Riesz Representation Theorem*, https://www.imsc.res.in/~sunder/rrt1.pdf.

[Li] Tom Lindstrom, *Mathematical Analysis*, https://www.uio.no/studier/emner/matnat/math/MAT2400/v14/mathanalbook.pdf.

[W] Peter Walters, *An Introduction to Ergodic Theory*, Springer-Verlag (1982).

[J] Aimee S. A. Johnson, *Measures on the circle invariant under multiplication by a non-lacunary subsemigroup of the integers*, Israel Journal of Mathematics (1992), 77, 211-240.

[HS1] Michael Hochman and Pablo Shmerkin, *Local entropy averages and projections of fractal measures*, Annals of Mathematics (2012), Volume 175, 1001-1059.

[HS2] Michael Hochman and Pablo Shmerkin, *Equidistribution from fractal measures*, Inventiones mathematicae (2015), Volume 202, 27–479.

[Hos] Bernard Host, *Nombres normaux, entropie, translations*, Israel Journal of Mathematics (1995), Volume 91, 419–428.

[Ma] Carlos Matheus, *Furstenberg's 2x, 3x (mod 1) problem*, https://matheuscmss.wordpress.com/2009/02/19/furstenbergs-2x-3x-mod-1-problem/ (2009).

[Z] Gustav Zickert, *Furstenberg's conjecture and measure rigidity for some classes of non-abelian affine actions on tori*, https://www.diva-portal.org/smash/get/diva2:845399/FULLTEXT01.pdf.

[V] Akshay Venkatesh, *The work of Einsiedler, Katok and Lindenstrauss on the Littlewood conjecture*, Bulletin of the American Mathematical Society (2008), Volume 45, 117-134.

[EKL] Manfred Einsiedler, Anatole Katok, and Elon Lindenstrauss, *Invariant measures and the set of exceptions to Littlewood's conjecture*, Annals of Mathematics (2006), Volume 164, 513–560.