JOURNAL OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 8, Number 3, July 1995

QUASI-FACTORS OF ZERO ENTROPY SYSTEMS

ELI GLASNER AND BENJAMIN WEISS

0. Introduction

The basic numerical invariant associated to dynamical systems is the entropy which divides them into two classes—the *chaotic* systems where the entropy is positive and the *stochastically deterministic* systems where the entropy is zero. It follows immediately from the definitions that factors of zero entropy systems have zero entropy. This holds in both the topological category for topological entropy as well as in the category of measure preserving transformations for the Kolmogorov-Sinai entropy.

There is a natural generalization of factor systems called *quasi-factors* and for these it turns out that there is a surprisingly sharp difference between the two categories.

For a topological system (X, T), X a compact metric space and T a homeomorphism, a *factor* is a continuous surjection $\pi : X \to Y$, which intertwines T with a homeomorphism S of Y, $S\pi = \pi T$. When π is an open map this factor can also be viewed as a subsystem of 2^X, the compact space of closed subsets of X with the Hausdorff topology and the natural action that T induces on it. Indeed as y ranges over Y, $\pi^{-1}(y)$ ranges over points of 2^X and T acts on these points just like S does on Y (when π is not open the subset $\{\pi^{-1}(y) : y \in Y\}$ is not closed in 2^X and its closure is only an almost 1-1 approximation of Y). In general a *quasi-factor* of (X, T) is any subsystem of $(2^X, T)$.

For a measure preserving transformation $(X, \mathscr{B}, \mu, T)$ a factor system $(Y, \mathscr{C}, \nu, S)$ is one for which there is a measurable map $\pi : X \to Y$ satisfying $\pi\mu = \nu$, $S\pi = \pi T$. To see Y as part of X, disintegrate the measure μ along the fibers of $\pi^{-1}(\mathscr{C})$,

$$(*) \qquad \mu = \int_Y \mu_y \, d\nu(y)$$

and observe that the T-invariance of μ implies that $T\mu_y = \mu_{Sy}$. Thus the action of S on Y with measure ν is exactly mirrored by the action of T on the space of probability measures on X, $M(X)$, with ν viewed now as a measure on $M(X)$. The connection with μ is given by $(*)$ which says that μ is the *barycenter* of ν. A general *quasi-factor* of $(X, \mathscr{B}, \mu, T)$ is any T-invariant measure on $M(X)$ whose barycenter is μ.

Received by the editors June 3, 1993 and, in revised form, April 1, 1994.
1991 *Mathematics Subject Classification.* Primary 28D05, 54H20.

Topological quasi-factors of zero entropy systems may have positive entropy. This can be seen already in the very simple example (X_0, T_0) where $X_0 = \mathbb{Z} \cup \{\infty\}$ is the one-point compactification of the integers and T_0 is translation by one on $\mathbb{Z}$ and $T_0\infty = \infty$. This example might suggest that the problem is with the lack of minimality on the part of (X_0, T_0). Recall that (X, T) is *minimal* if every T-orbit is dense. A more restricted version of the question would be: if (X, T) is minimal and $Y \subset 2^X$ is a minimal quasi-factor does $h_{top}(X, T) = 0$ imply $h_{top}(Y, T) = 0$? One of the main results of this paper is that the answer to this question is no! After discovering this fact we were sure that the measure theory analogue would have a similar answer, i.e. we expected that the vanishing of the entropy of (X, T) would not imply the same for $(M(X), T)$. It turned out that we were wrong and that indeed we could prove:

Theorem A. *If $h_{top}(X, T) = 0$ and $M(X)$ denotes the probability measures on X with the weak* topology then $h_{top}(M(X), T) = 0$.*

We found two quite different proofs of this fact. The first (given in section 1), goes via a measure theoretical result concerning quasi-factors of zero entropy systems $(X, \mathscr{B}, \mu, T)$ and uses ideas from ergodic theory such as K-automorphisms and disjointness. This method of proof leads to some further results concerning distal systems that we will describe in detail later on in that section.

The second proof (in section 2) uses some combinatorial tools as well as the precise information available concerning the dimension of almost Hilbertian sections of the unit balls in l_p-spaces. This reveals some new unexpected connections between dynamics and the local theory of Banach spaces.

This proof, being of a constructive nature, enables one to give a quantitative aspect to the qualitative statement of Theorem A. In addition it can be used to prove an analogous result for actions of amenable groups. The disjointness theory used in the ergodic theoretic proof of section 1 has not yet been established in that context and hence cannot be easily extended to such groups. We shall not pursue these directions here. However, in section 3, we do give another application of some of these combinatorial tools, to provide a topological characterization of K-automorphisms which are usually defined by probabilistic concepts.

Returning to the topological category, the examples that we construct (in section 4), of a minimal positive entropy quasi-factor of a minimal zero entropy system, enable us to obtain some new results related to the recent developments [B,1], [B,2], [B-L], [G-W,2], [B,3] in topological dynamics. Briefly put F. Blanchard singled out a nice class of systems, u.p.e., with uniformly positive entropy (see definitions below) and showed that they were a good topological analogue of K-automorphisms. In particular a u.p.e. system has no non-trivial factors of zero entropy (such systems are said to be of completely positive entropy; c.p.e.), and a u.p.e. system is disjoint from every minimal zero entropy system. The notions of c.p.e. and u.p.e. systems have natural generalizations to c.p.e. and u.p.e. extensions. The way in which we show (in section 5) that in our basic construction the minimal quasi-factor has positive topological entropy is by showing it to be a u.p.e. extension of its Kronecker factor. In so doing we also provide an example of a u.p.e. extension which is not relatively disjoint from a minimal zero entropy system. Further related results concerning these relative notions will appear elsewhere [G-W,4].

An outline of the two proofs of Theorem A appears in [G-W,3]. We express our thanks to Y. Lindenstrauss for his invaluable assistance in the geometric proof of Theorem A, and to M. Boyle for his contribution to the ergodic theoretical proof of that theorem.

1. THE ERGODIC THEORETIC PROOF OF THEOREM A

Given a system $(X, \mathscr{B}, \mu, T)$ and a factor $(X, \mathscr{B}, \mu, T) \xrightarrow{\pi} (Y, \mathscr{A}, \nu, S)$, one has the standard disintegration of μ over ν, which represents μ as

$$(*) \qquad \mu = \int_Y \mu_y d\nu(y)$$

where μ_y is a probability measure concentrated on $\pi^{-1}(y)$. The essential uniqueness of this representation and the invariance of μ under T lead to the relationship $T\mu_y = \mu_{Sy}$. Now the action of S on Y is represented by the action of T on the space $M(X)$ of probability measures on X.

Equipped with the weak* topology, $M(X)$ is a compact metric space and the map $y \mapsto \mu_y$ is measurable with respect to the corresponding Borel structure on $M(X)$. Via this map the measure ν is transferred onto a measure, also denoted ν, which is concentrated on the μ_y's. The relation $(*)$ expresses the fact that the barycenter of ν, as a measure on $M(X)$, is μ. The notion of a *quasi-factor*, [G], is a generalization of this way of looking at factors and is simply any probability measure ν on $M(X)$, invariant under T, and with barycenter equal to μ:

$$\int_{M(X)} \theta d\nu(\theta) = \mu.$$

As is well known, if the entropy of (X, μ, T) is zero so is the entropy of any factor. Our goal in this section is to prove the following:

Theorem 1.1. *Any quasi-factor of a zero entropy transformation has zero entropy.*

Assuming Theorem 1.1 a proof of Theorem A is obtained as follows. Let ν be a T-invariant probability measure on $M(X)$ and let $\mu = \int \theta d\nu(\theta)$ be the barycenter of ν so that μ is a T-invariant probability measure on X. We assume that (X, T) has zero topological entropy and it follows that the measure entropy of μ is zero. By Theorem 1.1 the measure entropy of ν is zero. Now use the variational principle [D-G-S] to deduce that $h_{top}(M(X), T) = 0$. □

The ergodic decomposition of μ and the fact that entropy is an affine function of the measure show that we may restrict in the sequel to ergodic μ's and ergodic quasi-factors. Our proof of this theorem relies on the concept of disjointness introduced by H. Furstenberg [F,1]. A *joining* of two systems $\mathscr{X}_1 = (X_1, \mathscr{B}_1, \mu_1, T_1), \mathscr{X}_2 = (X_2, \mathscr{B}_2, \mu_2, T_2)$ is a $T_1 \times T_2$ invariant measure on $X_1 \times X_2$ whose projection onto X_i equals μ_i $(i = 1, 2)$. The systems $\mathscr{X}_1, \mathscr{X}_2$ are *disjoint* if their only joining is $\mu_1 \times \mu_2$. Relativizing this notion, if the X_i's have a common factor, say $\mathscr{Z} = (Z, \mathscr{C}, R, \eta)$ with maps $\pi_i : X_i \to Z$, then we say that the $\mathscr{X}_i$'s are *disjoint relative to* $\mathscr{Z}$ if the only joining of $\mathscr{X}_1$ with $\mathscr{X}_2$ that identifies this common factor $\mathscr{Z}$ is the relatively independent one λ. This is obtained by disintegrating μ_i over η:

$$\mu_i = \int \mu_{i,z} d\eta(z), \qquad i = 1, 2,$$

and then forming λ as follows:

$$\lambda = \int_Z \mu_{1,z} \times \mu_{2,z} d\eta(z).$$

It is well known that K-automorphisms are disjoint from zero entropy systems. At the heart of our proof of Theorem 1.1 is a relativized version of this fact.

Theorem 1.2. *If $\mathscr{X}_i = (X_i, \mathscr{B}_i, \mu_i, T_i)$, $i = 1, 2$, both have $\mathscr{Z} = (Z, \mathscr{C}, \eta, R)$ as a factor, and if $\mathscr{X}_1$ has relative zero entropy, i.e. $h(T_1, \mu_1) = h(R, \eta)$, while $\mathscr{X}_2$ has relative completely positive entropy, i.e. any factor of $\mathscr{X}_2$ that properly includes $(Z, \mathscr{C}, \eta)$ has entropy strictly greater than $h(R, \eta)$, then $\mathscr{X}_1$ is disjoint from $\mathscr{X}_2$ relative to $\mathscr{Z}$.*

We postpone the proof of Theorem 1.2 to the end of this section. We shall next use it to prove a proposition describing how any joining between a zero entropy system and a positive entropy one arises; this proposition in turn will be used to prove Theorem 1.1. Note that every system $(Y, \mathscr{A}, \nu, R)$ can be trivially represented as a quasi-factor of itself by: $\nu = \int_{M(Y)} \delta_y d\nu(y)$, where as usual δ_y is the point mass at y.

Proposition 1.3. *Let $\mathscr{X} = (X, \mathscr{B}, \mu, T)$ be a zero entropy system, $\mathscr{Y} = (Y, \mathscr{A}, \nu, S)$ a system of positive entropy. Let λ be a joining of these two systems. We let $\mathscr{Z} = (Z, \mathscr{C}, \eta, R)$ denote the maximal zero entropy factor of $\mathscr{Y}$, so that $\mathscr{C}$ lifts up to the Pinsker algebra of $\mathscr{Y}$. Denote by ρ the projection of λ on $Z \times X$. Then λ is the trivial lift of ρ. More precisely, if $\nu = \int \nu_z d\eta(z)$ is the disintegration of ν over Z, and $\rho = \int \delta_z \times \rho_z d\eta(z)$ is the disintegration of ρ over Z, where ρ_z is a measure on X, then*

$$\lambda = \int_Z \nu_z \times \rho_z d\eta(z).$$

Proof. Let $\pi : Y \to Z$ be the canonical projection. We let

$$\lambda = \int \delta_y \times \lambda_y d\nu(y),$$

where the λ_y's are measures on X, be the disintegration of λ over Y. Observe that we now have the following representation for ρ—which is a joining of $\mathscr{Z}$

and $\mathscr{X}$:

$$\begin{aligned}\rho &= (\pi \times id)(\lambda) \\ &= \int \delta_{\pi(y)} \times \lambda_y d\nu(y) \\ &= \int \delta_z \times \left(\int \lambda_y d\nu_z(y)\right) d\eta(z) \\ &= \int \delta_z \times \rho_z d\eta(z),\end{aligned}$$

and thus we see that $\rho_z = \int \lambda_y d\nu_z(y)$,

Clearly the system $\tilde{\mathscr{X}} = (Z \times X, \rho, \tilde{T})$ is an extension of $\mathscr{X}$ and has zero entropy. We can now form the joining $\tilde{\lambda}$ of $\mathscr{Y}$ and $\tilde{\mathscr{X}}$, by letting

$$\tilde{\lambda} = \int \delta_y \times \delta_{\pi(y)} \times \lambda_y d\nu(y).$$

We claim that $\tilde{\lambda}$ is the relatively independent product $\int_Z \nu_z \times \delta_z \times \rho_z d\eta(z)$ of ν and ρ over Z. In fact, since $\tilde{X}$ has zero entropy and since $\mathscr{Z}$ corresponds to the Pinsker algebra of $\mathscr{Y}$, our claim follows directly from Theorem 1.2. Now we have

$$\begin{aligned}\tilde{\lambda} &= \int \delta_y \times \delta_{\pi(y)} \times \lambda_y d\nu(y) \\ &= \int_Z \int \delta_y \times \delta_z \times \lambda_y d\nu_z(y) d\eta(z) \\ &= \int_Z \nu_z \times \delta_z \times \rho_z d\eta(z).\end{aligned}$$

The uniqueness of disintegration (of $\tilde{\lambda}$ over η) implies now that for η-almost every z,

$$\int \delta_y \times \lambda_y d\nu_z(y) = \nu_z \times \rho_z = \int \delta_y \times \rho_z d\nu_z(y).$$

Again, by the uniqueness of disintegration, we get for ν_z-almost every y that $\lambda_y = \rho_z$. This latter equality yields

$$\lambda = \int_Z \nu_z \times \rho_z d\eta(z),$$

and our proof is complete. □

A proof of Theorem 1.1. Consider an ergodic system $(X, \mathscr{B}, \mu, T)$ of zero entropy and an ergodic quasi-factor ν on $Y = M(X)$, so that $\int_{M(X)} \theta d\nu(\theta) = \mu$. Assuming that the system $(Y, \mathscr{A}, \nu, T)$ has positive entropy (where $\mathscr{A}$ is the

Borel σ-algebra on $M(X)$) we shall get a contradiction. Let λ be the measure on $M(X) \times X$ defined by

$$\lambda = \int_{M(X)} \delta_\theta \times \theta d\nu(\theta).$$

Then clearly, λ is a joining of ν and μ. The notations are as in Proposition 1.3 and by that proposition we get:

$$\begin{aligned}\lambda &= \int \nu_z \times \rho_z d\eta(z) \\ &= \int\int \delta_\theta \times \rho_z d\nu_z(\theta) d\eta(z) \\ &= \int \delta_\theta \times \rho_{\pi(\theta)} d\nu(\theta) \\ &= \int \delta_\theta \times \theta d\nu(\theta).\end{aligned}$$

Uniqueness of disintegration implies that for ν almost every θ,

$$\theta = \rho_{\pi(\theta)} = \int \theta' d\nu_{\pi(\theta)}(\theta').$$

This clearly implies that π is an isomorphism so that $(M(X), \mathscr{A}, \nu, T)$ is its own Pinsker factor. Since we assumed that $(M(X), \mathscr{A}, \nu, T)$ has positive entropy, this is the desired contradiction and the proof of Theorem 1.1 is complete. □

A proof of Theorem 1.2. From the formulation of the theorem it follows that we may assume that $\mathscr{Z}$ has finite entropy. By Krieger's theorem, [D-G-S], we can assume, therefore, that $\mathscr{Z}$ is given by some finite valued stationary stochastic process $\{z_n\}_{-\infty}^{+\infty}$. For such processes we shall use the notation z_j^i to denote the σ-algebra generated by the variables z_n for $j \le n \le i$, and for brevity $z^- = z_{-\infty}^{-1}$ represents the past while $z^\infty = z_{-\infty}^{+\infty}$ represents the full process.

We shall need some standard results concerning the conditional entropy of processes; these may be found, for example, in [P]. All processes are finite valued and ergodic, $\mathscr{X}$ denotes now the stochastic process $\{x_n\}$, and all processes occurring together are assumed to be defined on a common probability space.

$$h(\mathscr{X} \vee \mathscr{Y}) = H(x_0 | x^- \vee y^\infty) + H(y_0|y^-) \tag{1}$$

$$= H(x_0|x^- \vee y^\infty) + \lim_{n \to +\infty} H(y_0|y^- \vee x_{-\infty}^{-n}), \tag{2}$$

whence

$$H(y_0|y^-) = \lim_{n\to+\infty} H(y_0|y^- \vee x_{-\infty}^{-n}). \tag{3}$$

Likewise for the relative entropy $h(\mathscr{X}|\mathscr{Z}) = H(x_0|x^- \vee z^\infty)$, we have:

$$h(\mathscr{X} \vee \mathscr{Y}|\mathscr{Z}) = H(x_0|x^- \vee y^\infty \vee z^\infty) + H(y_0|y^- \vee z^\infty) \tag{4}$$

$$= H(x_0|x^- \vee y^\infty \vee z^\infty) + \lim_{n\to+\infty} H(y_0|y^- \vee x_{-\infty}^{-n} \vee z^\infty), \tag{5}$$

whence

$$H(y_0|y^- \vee z^\infty) = \lim_{n\to+\infty} H(y_0|y^- \vee x_{-\infty}^{-n} \vee z^\infty). \tag{6}$$

Lemma 1.4. *Suppose that any process that is not already measurable with respect to z^∞ adds entropy to $\mathscr{Z}$ when it is added to $\mathscr{Z}$. Then for any process $\mathscr{X} = \{x_n\}$:*

$$\lim_{n\to+\infty} H(x_0|x_\infty^{-n} \vee z^\infty) = H(x_0|z^\infty).$$

Proof. By the martingale convergence theorem it suffices to show that

$$\bigcap_n (x_{-\infty}^{-n} \vee z^\infty) = z^\infty.$$

Notice that this intersection is an invariant σ-algebra, thus it suffices to show that for any process $\{y_n\}$ measurable with respect to this intersection, $\{y_n\}$ is actually measurable with respect to z^∞.

Using formula (6) we see that $H(y_0|y^- \vee z^\infty) = 0$ and then formula (1), replacing $\mathscr{X}, \mathscr{Y}$ by $\mathscr{Y}, \mathscr{Z}$, gives $h(\mathscr{Y} \vee \mathscr{Z}) = h(\mathscr{Z})$ which by hypothesis implies that $\mathscr{Y}$ is $\mathscr{Z}$ measurable. □

Suppose now that $\mathscr{Y}$ adds no entropy to $\mathscr{Z}$, while any process measurable with respect to $\mathscr{X}$ that is not measurable with respect to $\mathscr{Z}$ does add entropy to $\mathscr{Z}$. We claim that $\mathscr{X}$ and $\mathscr{Y}$ are relatively independent over $\mathscr{Z}$. We shall use the formulas (1)-(6) replacing the time shift T by T^n.

If z_0 is replaced by $(z_0, z_1, \dots, z_{n-1})$ then z^∞ remains the same. Now $y_{-\infty}^{-n}$ should be interpreted as using only the variables $y_{-n}, y_{-2n}, \dots$. Nonetheless it is easy to see that we still have $H(y_0|y_{-\infty}^{-n} \vee z^\infty) = 0$. Using (4) we get

$$h(\mathscr{X}|\mathscr{Z}) \le h(\mathscr{X} \vee \mathscr{Y}|\mathscr{Z}) = H(x_0|x_{-\infty}^{-n} \vee y^\infty \vee z^\infty) + H(y_0|y_{-\infty}^{-n} \vee z^\infty),$$

which gives $H(x_0|x_{-\infty}^{-n} \vee z^\infty) \le H(x_0|y^\infty \vee z^\infty)$.

By Lemma 1.4 the left-hand side converges to $H(x_0|z^\infty)$, and so we deduce

$$H(x_0|z^\infty) = H(x_0|y^\infty \vee z^\infty).$$

The same conclusion would hold for all finite blocks $(x_{-N}, \dots, x_N)$ and so we conclude that indeed $\mathscr{X}$ is independent of $\mathscr{Y}$ relative to $\mathscr{Z}$. Applying the above reasoning to finite valued stationary processes defined on the systems $\mathscr{X}$ and $\mathscr{Y}$ we conclude the proof of Theorem 1.2. □

Remarks. (a) It is possible to prove Theorem 1.1 from Theorem A. For this one constructs for a zero entropy measure preserving transformation a topological model with zero topological entropy. [D-G-S] for example, after a preliminary reduction to the ergodic case.

(b) The converse of the claim in Theorem 1.2 is also true and can be deduced from [T]. Thus we have:

Theorem 1.2′. *If $\mathscr{X}_i = (X_i, \mathscr{B}_i, \mu_i, T_i)$, $i = 1, 2$, where both have $\mathscr{Z} = (Z, \mathscr{C}, \eta, R)$ as a factor, and $\mathscr{X}_2$ has relative completely positive entropy over*

$\mathscr{Z}$, then $\mathscr{X}_1$ is disjoint from $\mathscr{X}_2$ relative to $\mathscr{Z}$ if and only if $\mathscr{X}_1$ has relative zero entropy over $\mathscr{Z}$.

The above proof of Theorem 1.1 can be applied whenever one encounters a situation in which a relative disjointness claim like Theorem 1.2 can be proved. One such case is when we replace, in Theorem 1.2, zero measure entropy by measure distal (generalized discrete spectrum) and relative K-extension by relative weakly mixing extension. The corresponding theorem we obtain asserts that every ergodic quasi-factor of a distal system is distal. The topological analog of this result for minimal distal systems is well known, see e.g. [K]. Recall that an ergodic system $(X, \mathscr{B}, \mu, T)$ is *measure distal* or has *generalized discrete spectrum* if it belongs to the smallest class of ergodic systems which contains all the Kronecker systems (compact group rotations) and is closed under isometric extensions and (countable) inverse limits, [Z,1], [Z,2], [F,2], [F,3]. Such a system has a canonical description as a tower of successive isometric extensions and, when necessary, inverse limits.

An extension $\pi : X \to Z$ where $\mathscr{X} = (X, \mathscr{B}, \mu, T)$ and $\mathscr{Z} = (Z, \mathscr{C}, \eta, R)$ are ergodic systems is said to be *relatively distal* if X can be obtained from Z by a series of isometric extensions and inverse limits. We shall need the following analogue of Theorem 1.2 which is essentially due to H. Furstenberg.

Theorem 1.5. *If $\mathscr{X}_1$ is relatively distal over $\mathscr{Z}$ and $\mathscr{X}_2$ is relatively weakly mixing over $\mathscr{Z}$, then $\mathscr{X}_1$ and $\mathscr{X}_2$ are relatively disjoint over $\mathscr{Z}$.*

Proof. The proof is accomplished by going step by step up the tower (of $\mathscr{X}_1$ over $\mathscr{Z}$). For the main step assume that $\mathscr{X}_1$ is an extension of $\mathscr{Z}$ by a compact group G. By Theorem 7.5 of [F,2] the relatively independent joining $\mathscr{X}_1 \underset{Z}{\times} \mathscr{X}_2$ is ergodic. Then if θ is any joining of $\mathscr{X}_1$ to $\mathscr{X}_2$ over Z so is $g\theta$ for all $g \in G$. Averaging $g\theta$ with respect to Haar measure clearly gives the relatively independent joining which being ergodic forces θ itself to be relatively independent. □

Any ergodic system $\mathscr{Y} = (Y, \mathscr{A}, \nu, S)$ has a maximal distal factor $\mathscr{Z}$ and $\mathscr{Y}$ is relatively weak mixing over $\mathscr{Z}$. Using separating sieves Zimmer showed that a factor of a distal system is distal. An alternate proof is as follows. If $\mathscr{X}$ is distal and $\mathscr{Y}$ is a factor let $\mathscr{Z}$ be the maximal distal factor of $\mathscr{Y}$. It is easy to see that $\mathscr{X}$ is a distal extension of $\mathscr{Z}$ while $\mathscr{Y}$ is relatively weakly mixing and therefore by Theorem 1.5 they are relatively disjoint—which contradicts the tower structure $\mathscr{X} \to \mathscr{Y} \to \mathscr{Z}$.

Finally it follows immediately from the tower description that an ergodic joining of two distal systems is itself distal.

Theorem 1.6. *Every ergodic quasi-factor of a distal system is distal.*

Proof. The proof is almost verbatim that of Theorem 1.1 via Proposition 1.3 with the obvious necessary changes. In that proposition we now assume that $\mathscr{X} = (X, \mathscr{B}, \mu, T)$ is distal and that $\mathscr{Y} = (Y, \mathscr{A}, \nu, S)$ is a non-distal ergodic system with $\mathscr{Z} = (Z, \mathscr{C}, \eta, R)$ as its largest distal factor. The claim of the proposition is the same and so is the proof. The only place where some care has to be exercised is when we use Theorem 1.5 (instead of Theorem 1.2

in the original proof). In fact we want to deduce the relative disjointness of the weakly mixing extension $\pi : \mathscr{Y} \to \mathscr{Z}$ from the system $\widetilde{\mathscr{X}} = (Z \times X, \rho, \widetilde{T})$ over their common factor $\mathscr{Z}$; i.e. the relative independence of $\widetilde{\lambda}$. However since we are not assuming the ergodicity of the joining λ, neither ρ nor the system $\widetilde{\mathscr{X}}$ need be ergodic. To overcome this difficulty we observe that each ergodic component of $\widetilde{\mathscr{X}}$ as a joining of the distal systems $\mathscr{Z}$ and $\mathscr{X}$ is itself a distal system, the relative independence of this ergodic component from $\mathscr{Y}$ over $\mathscr{Z}$ follows by Theorem 1.5 and the relative independence of $\widetilde{\lambda}$ is readily deduced as well. □

2. The geometric proof of Theorem A

The discussion simplifies when X is zero dimensional and so let us begin with that extra assumption. If $\mathscr{P}$ is a finite partition of X into closed and open sets and $\epsilon > 0$ is fixed let

$$\mathscr{U}(\mathscr{P}, \epsilon) = \{(\mu, \nu) \in M(X) \times M(X) : \max_{P \in \mathscr{P}} |\mu(P) - \nu(P)| < \epsilon\}.$$

For a sequence of partitions into sets whose diameter tends to zero and epsilons tending to zero the sets $\mathscr{U}(\mathscr{P}, \epsilon)$ define a metric on $M(X)$ which can be used to compute the topological entropy of $(M(X), T)$. Indeed if we denote for fixed $\mathscr{P}$,

$$||\mu - \nu|| = \max_{P \in \mathscr{P}} |\mu(P) - \nu(P)|$$

and declare that a set of measures $\{\mu_i : 1 \le i \le k\}$ is (n-$\mathscr{U}$)-separated if for any $1 \le a < b \le k$ there is some $0 \le i < n$ so that

$$||T^i \mu_a - T^i \mu_b|| \ge \epsilon, \tag{1}$$

then the topological entropy of $(M(X), T)$ is given by:

$$\lim_{\mathscr{U}(\mathscr{P}, \epsilon)} \limsup\, (1/n) \log\, c(n, \mathscr{U}),$$

where $c(n, \mathscr{U})$ is the cardinality of a maximal (n-$\mathscr{U}$)-separated set and the limit over $\mathscr{U}(\mathscr{P}, \epsilon)$ is as we described above, with $\operatorname{diam}(\mathscr{P})$ and ϵ tending to zero.

In the calculation of (1) all that one needs to know is the measures of atoms in $\bigvee_0^{n-1} T^{-i}\mathscr{P} = \mathscr{P}_0^{n-1}$. Since the sets of $\mathscr{P}$ are closed and open there is some positive distance $\delta > 0$ between them and if L_n denotes the number of non-empty atoms of $\mathscr{P}_0^{n-1}$, then $h(X, T) = 0$ implies

$$\lim_{n \to 0} (1/n) \log L_n = 0. \tag{2}$$

To see the essence of the problem let us suppose that $\mathscr{P} = \{P(0), P(1)\}$ consists of two sets. Then for fixed n, the relevant data about the measures μ on X is how much mass they assign to each atom A_l of $\mathscr{P}_0^{n-1}$; denote this by

$\mu(l)$, $1 \le l \le L_n$. Define an $L_n \times n$, 0-1 matrix ϕ by the formula

$$A_l = \bigcap_{0}^{n-1} T^{-i} P(\phi(l, i))$$

where A_l is the l-th atom of $\mathscr{P}_0^{n-1}$. Then

$$\sum_{l=1}^{L-n} \mu(l)\phi(l, i) = (T^i \mu)(P(1))$$

and ϕ defines in this way a mapping from $l_1^{L_n}$ to l_∞^n. Furthermore, the measures $\{\mu_k\}$ on X are n-separated if and only if the images $\{\mu_k \phi\}$ are $(n\text{-}\mathscr{U})$-separated in the l_∞-norm. What we shall need is

Proposition 2.1. *For given constants $\epsilon > 0$ and $b > 0$, there is an n_0 and a constant $c > 0$ so that for all $n \ge n_0$, if ϕ is a linear mapping from $l_1^{L_n}$ to l_∞^n of norm $\|\phi\| \le 1$, and if $\phi(B_1(l_1^{L_n}))$ contains more than 2^{bn} points that are ϵ-separated, then $L_n \ge 2^{cn}$.*

For the proof of this proposition we need a simple combinatorial result due to N. Sauer [Sa] and Perles and Shelah [Sh], and a geometric result about ϵ-separated sets in l_∞^n.

Lemma 2.2 ([Sa],[Sh]). *Given $b > 0$, there is some constant $c > 0$ and n_0 so that for all $n \ge n_0$ if $\mathscr{A} \subset \{0, 1\}^I$ satisfies $|\mathscr{A}| \ge 2^{bn}$ then there is some $I \subset \{1, 2, \dots, n\}$ satisfying:*

(1) $|I| \ge cn$,

(2) $\mathscr{A}|I = \{0, 1\}^I$.

In fact the precise relationship between b and c is known but since it is not needed for our purposes we formulated a cruder result.

Lemma 2.3. *For constants $\epsilon > 0$ and $b > 0$ there are constants $d > 0$ and $\delta > 0$ such that for all sufficiently large n, if $A \subset B_1(l_\infty^n)$, the unit ball of l_∞^n, is ϵ-separated and $|A| \ge 2^{bn}$, then there is some value y_0 and a set $I \subset \{1, 2, \dots, n\}$ such that:*

(1) $|I| \ge dn$,

(2) *for every element $f \in \{0, 1\}^I$, there is some $a \in A$ such that for all $i \in I$*

$$a(i) \ge y_0 + (\delta/2) \quad \textit{if} \quad f(i) = 1 \quad \textit{and}$$
$$a(i) \le y_0 - (\delta/2) \quad \textit{if} \quad f(i) = 0.$$

Proof of Lemma 2.3. (1) For a $\delta > 0$ to be determined later (much smaller than ϵ), consider the ϵ/δ different grids

$$G_k = \bigcup_{m=-(1/\epsilon)}^{1/\epsilon} [k\delta + m\epsilon, k\delta + m\epsilon + \delta], \qquad 0 \le k < K = [\epsilon/\delta].$$

For each $a \in A$ there is some $0 \leq k < K$ such that

$$|\{1 \leq i \leq n : a(i) \in G_k\}| \leq (1/K)n \tag{1}$$

since the G_k's are disjoint. If A_k denotes those a's in A that satisfy (1) then for some $0 \leq k < K$,

$$|A_k| \geq (1/K)|A|. \tag{2}$$

Fix some k_0 that satisfies (2), and to ease notation denote A_k by A. Note that since K is fixed the exponential size of A_k is the same as before (for large n).

(2) For each interval $[k\delta + m\epsilon, k\delta + m\epsilon + \delta]$ that constitutes G_k, we define a map from A to $\{0, 1, *\}^n$ by sending $a(i)$ to 0 if $a(i) \leq k\delta + m\epsilon$, to 1 if $a(i) \geq k\delta + m\epsilon + \delta$ and to $*$ if neither holds, i.e. if $a(i)$ is in the interval. Let $\mathscr{U}_m$ denote the image of A under this mapping. Each element $u \in \mathscr{U}_m$ has the value $*$ on some set of indices $\beta \subset \{1, 2, \dots, n\}$ whose size is at most n/k. Denote by $\mathscr{U}_m^\beta$ those u's in $\mathscr{U}_m$ whose $*$ values occur precisely at β. The number of different β's where $\mathscr{U}_m^\beta$ is non-empty is less then $\sum_{0 \leq x < 1/k} \binom{n}{x} = f(n, 1/k)$ which is exponentially small with δ. Let

$$D = \max_{\beta, m} |\mathscr{U}_m^\beta|.$$

(3) We claim that $D \geq |A|^{1/M} f(n, 1/K)^{-1}$ where $M = [2/\epsilon] + 1$. Indeed if two elements $a, a' \in A$ are mapped to the same $\mathscr{U}_m^{\beta_m}$ for all $-(1/\epsilon) < m < 1/\epsilon$ then they clearly cannot be ϵ-separated. Therefore

$$\sum_{\beta_m\text{'s}} \prod_m |\mathscr{U}_m^{\beta_m}| \leq D^M f(n, 1/k)^M$$

and extracting the M-th root of both sides gives us our claim. We fix now some m and β with $|\mathscr{U}_m^\beta| \geq |A|^{1/M} f(n, 1/K)^{-1}$. If δ is sufficiently small then this is some fixed exponential size and Lemma 2.2 may be applied to give the desired conclusion. □

We turn now to the proof of Proposition 2.1.

Proof of Proposition 2.1. Apply Lemma 2.3 to the image $\phi(B_1(l_1^L))$, and noting that this image is also convex and symmetric we conclude that there is some fixed d and δ, which depend only on the constants ϵ and b, and $I \subset \{1, 2, \dots, n\}$ such that if $\pi : l_\infty^L \to l_\infty^I$ denotes the canonical projection then

$$\pi(\phi(B_1(l_1^L))) \supset B_{\delta/2}(l_\infty^I).$$

The dual $(\pi \circ \phi)^*$ is an injection of $(l_\infty^I)^* = l_1^I$ into $(l_1^L)^* = l_\infty^L$ and the norm of this injection is bounded, as is the norm of its inverse, by fixed bounds.

Now it is known [F-L-M] that l_1^n has almost Euclidean sections of dimension a constant times n. On the other hand l_∞^m has almost Euclidean sections of dimension that are at most a constant times $\log m$. It follows that $\log L$ must be at least some constant times n which is exactly the assertion of the proposition. □

In case $\mathscr{P}$ has more than two sets, we need to map the measures into l_∞^{np} where p is the number of atoms of $\mathscr{P}$, in order to capture the appropriate notion of separation. For fixed p this does not affect the argument at all and thus we see that for zero dimensional spaces our proof is complete.

The easiest way to deal with the general case is to make use of the following proposition whose proof was indicated to us by M. Boyle:

Proposition 2.4. *If T is a homeomorphism of a compact metric space X, there is a zero dimensional space $\widetilde{X}$ with a homeomorphism $\widetilde{T}$ and a continuous map $\pi : \widetilde{X} \to X$ satisfying:*

(1) $T\pi = \pi\widetilde{T}$,
(2) $h_{top}(\widetilde{T}) = h_{top}(T)$.

Proof. The space $\widetilde{X}$ will be built as an inverse limit of symbolic dynamical systems. Begin with an open cover $\mathscr{U}_1$, with sets whose closure has diameter at most $1/2$. Find N_1 such that $\bigvee_{i=1}^{N_1} \mathscr{U}_1$ has a subcover $\mathscr{V}_1$ with at most $e^{N_1/2}$ sets. Let A_1 denote the index set of the subcover $\mathscr{V}_1$, and let $Y_1^1 \subset (A_1 \cup \{*\})^{\mathbb{Z}}$ consist of those sequences of the form $\cdots \alpha *^{N_1-1} \beta *^{N_1-1} \gamma *^{N_1-1} \cdots$ that represent points in X, i.e. such that there exists some $x \in X$ with $x \in \overline{V}_\alpha$, $T^{N_1}x \in \overline{V}_\beta$, $T^{2N_1}x \in \overline{V}_\gamma$, etc. Clearly Y_1^1 is closed and shift invariant, and we have a mapping from Y_1^1 to 2^X that commutes with the shift, and has for its image sets of diameter at most 1.

Next let $\mathscr{U}_2$ refine $\mathscr{V}_1$ with sets whose closures have diameter at most $1/2^2$. Since also the entropy of T^{N_1} is zero, we can find an N_2 so that $\bigvee_{i=1}^{N_2} T^{iN_1}\mathscr{U}_2$ has a subcover $\mathscr{V}_2$ with at most $e^{N_2/2^2}$ elements. Note that if A_2 is the index set of this cover then in a natural way each element of A_2 maps onto a block of length N_2 of elements of A_1. Thus if $Y_2^2 \subset (A_2 \cup \{*\})^{\mathbb{Z}}$ is defined in an analogous way—elements of A_2 separated by $N_1N_2 - 1$ asterisks that represent points in X—there is a natural projection of Y_2^2 *into* Y_1^1 that commutes with the shift. Denote by Y_1^2 the image of Y_2^2 under this mapping. Note that the topological entropy of Y_1^2 is smaller than $(1/2^2)1/N_1$. Also Y_2^2 maps into 2^X, and its image consists of sets whose diameter is at most $1/2$.

Continuing this scheme we get $Y_1^1 \supset Y_1^2 \supset Y_1^3 \supset \cdots$ and we denote $\bigcap_m Y_1^m$ by Y_1, $Y_2^2 \supset Y_2^3 \supset Y_2^4 \supset \cdots$, $Y_2 = \bigcap_m Y_2^m$, etc. The space $\widetilde{X}$ is the inverse limit of the Y_n's. The image of $\widetilde{X}$ in 2^X consists of singletons hence can be considered as a map to X. Finally it is clear that in the limit the topological entropy of each Y_n is zero and hence the same is true for $\widetilde{X}$. □

Since π clearly also induces a factor mapping from $M(\widetilde{X})$ to $M(X)$ we see that the general case of Theorem A follows from the special case dealt with above. □

3. A TOPOLOGICAL-COMBINATORIAL CHARACTERIZATION OF K-AUTOMORPHISMS

Let Ω_a denote the space of sequences with values in the finite set $\{0, 1, \ldots, a-1\}$, and $Y \subset \Omega_a$. A subset $I \subset \mathbb{Z}$ is called an *interpolating set for* Y if $Y|I = \Omega_a|I$. More concretely, for each choice $\{b_i : i \in I, b_i \in \{0, 1, \ldots, a-1\}\}$ there is some $\omega \in Y$ such that $\omega_i = b_i$ for all $i \in I$. Now suppose that (X, T, μ) is a measure preserving transformation of a finite measure space, and that $\mathscr{P} = \{P_0, P_1, \ldots, P_{a-1}\}$ is a finite measurable partition of X. Construct a set $Y_{\mathscr{P}} \subset \Omega_a$ as follows:

$$Y_{\mathscr{P}} = \{\omega \in \Omega_a : \text{for all finite subsets } J \subset \mathbb{Z},\ \mu(\bigcap_{j\in J} T^{-j} P_{\omega_j}) > 0\}.$$

If $\mu_{\mathscr{P}}$ is the image of μ under the mapping $\theta : X \to \Omega_a$ defined by $(\theta x)_n =$ that index b such that $T^n x \in P_b$, i.e. the distribution of the stochastic process defined by $(T, \mu, \mathscr{P})$, then $Y_{\mathscr{P}}$ is simply the closed support of $\mu_{\mathscr{P}}$. Finally, recall that a set $I \subset \mathbb{Z}$ has *positive density* if

$$\lim_{n\to\infty} \frac{|I \cap \{-n, \ldots, n\}|}{2n+1} > 0.$$

The results in this section were obtained in response to some questions posed by H. Furstenberg. The context of those questions consists of results connecting the dynamical nature of $X \subset \Omega_a$ with properties of its interpolating sets (cf. also [G-W,2]). The main result of this section is the following theorem which characterizes K-automorphisms in terms of these concepts.

Theorem 3.1. *For every non-trivial partition $\mathscr{P}$, the set $Y_{\mathscr{P}}$ has interpolating sets of positive density if and only if (X, T, μ) is a K-automorphism, or equivalently has completely positive entropy.*

We can quickly dispose of one direction, namely if T has a zero entropy factor then one can construct non-trivial partitions $\mathscr{P}$ such that the topological entropy of $Y_{\mathscr{P}}$ is zero and this rules out the existence of interpolating sets of positive density. In the other direction one can prove a stronger result, namely:

Theorem 3.2. *If $\mathscr{P}$ has two elements and $h(T, \mathscr{P}) > 0$, or more generally, if $\mathscr{P}$ has a elements and $h(T, \mathscr{P}) > \log(a-1)$, then $Y_{\mathscr{P}}$ has interpolating sets of positive density.*

Here no assumption is made about the dynamic nature of T—but of course rather a strong assumption about the size of $h(T, \mathscr{P})$ is involved. If $\mathscr{P}$ has two elements, then we can also show that there is a direct relation between how close $h(T, \mathscr{P})$ is to $\log 2$ and how close one can get to density one for the interpolating sequences.

The case $a = 2$ relies on Lemma 2.2 while for $a > 2$ one needs the analogous lemma which can be found in [K-M].

Proof of Theorem 3.2. Standard reductions allow us to assume that T is ergodic. Now the Shannon-McMillan theorem, [E-F], gives that for any $h' < h(T, \mathscr{P})$, and N sufficiently large the set of elements

$$B_N = \{b \in \{0, 1, \ldots, a-1\}^N : \text{ for which } \mu(\bigcap_{j=1}^{N} T^{-j} P_{b(j)}) > 0\}$$

has at least $2^{h'N}$ elements. Choosing $h' > \log(a-1)$, and applying Stirling's formula to approximate the constants in Lemma 2.2 (or its analogue for $a > 2$), we see that for some constant d, and all N sufficiently large, B_N has interpolating sets with size dN.

Observe that by identifying subsets of $\mathbb{Z}$ with elements of $\{0, 1\}^{\mathbb{Z}} = \Omega_2$ via their indicator functions, the collection of interpolating sets for $Y_{\mathscr{P}}$ is a closed, shift invariant subset of Ω_2 which we denote by $\mathscr{I}$. The existence of arbitrarily large finite interpolating sets of density d yields the existence of a shift invariant measure, ρ, on $\mathscr{I}$ such that $\rho([1]) \geq d$ where $[1] = \{\omega : \omega(0) = 1\}$. Take, on $\mathscr{I}$, an ergodic component ρ_0 of ρ such that $\rho_0([1]) \geq d$, and then any generic point for ρ_0 gives an infinite interpolating set with density at least $d > 0$. This completes the proof of Theorem 3.2. □

Proof of Theorem 3.1. Let us begin with a special case, in which $\mu(P_i) = 1/a$ for $0 \leq i < a$. In this case, since T is a K-automorphism it follows that

$$\lim_{k\to\infty} h(T^k, \mathscr{P}) = H(\mathscr{P}) = \log a$$

and thus for sufficiently large k_0, $h(T^{k_0}, \mathscr{P}) > \log(a-1)$ and using Theorem 3.2 we get the desired conclusion.

For a general $\mathscr{P}$, we will refine $\mathscr{P}$ to a partition $\mathscr{R}$ whose atoms are nearly equal, and use the fact that if $\mathscr{R}$ refines $\mathscr{P}$ any interpolating set for $Y_{\mathscr{R}}$ is also an interpolating set for $Y_{\mathscr{P}}$. The only place where care must be exercised is in making the size of the atoms of $\mathscr{R}$ nearly equal enough so that $H(\mathscr{R}) > \log(n-1)$ where n is the number of atoms of $\mathscr{R}$. Let $p_i = \mu(P_i)$, $0 \leq i < a$, and by standard Diophantine approximation find integers u_i, v so that

(i) $$v \geq (10 \cdot a)^{2a}$$

and

(ii) $$|p_i - \frac{u_i}{v}| \leq \frac{1}{v^{1+(1/(a-1))}}, \qquad 1 \leq i < a.$$

Deduce from (ii) that

$$|p_0 - \frac{u_0}{v}| \leq \frac{a}{v^{1+(1/(a-1))}},$$

where $u_0 = v - (u_1 + u_2 + \cdots + u_{a-1})$.

For each i divide p_i into u_i atoms, giving $u_i - 1$ of them the mass $1/v$, and the last atom the mass $p_i - (u_{i-1}/v)$, thus defining $\mathscr{R}$. Now an easy calculation reveals that $H(\mathscr{R}) > \log(v-1)$, where clearly $\mathscr{R}$ has v atoms and hence once again Theorem 3.2 completes the proof. □

4. The construction of a minimal zero entropy flow with a minimal positive entropy quasi-factor

A general description. The minimal flow (X, T) will be a subset of $\{0, 1\}^{\mathbb{Z}}$ with the shift. The construction will be inductive. At stage n, we will have a set of allowable words $\mathscr{W}_n$, each a concatenation of words from $\mathscr{W}_{n-1}$. The length of the words in $\mathscr{W}_n$ is fixed and equals l_n. Thus if X_n denotes all infinite concatenations of words from $\mathscr{W}_n$, we have $X_0 \supset X_1 \supset X_2 \supset \cdots$ and $X = \bigcap_0^\infty X_n$.

The minimality of (X, T) is guaranteed by incorporating in each word $w \in \mathscr{W}_n$ an initial segment which contains all pairs of words from $\mathscr{W}_{n-1}$. Thus in X, any block of length l_{n-1} that occurs at all will be seen in every block of length $2l_n$; this implies that (X, T) is minimal. The fact that (X, T) has zero entropy will follow from the fact that $\frac{1}{l_n} \log |\mathscr{W}_n| \to \infty$.

This procedure will establish a hierarchical block structure in elements of X which will, most probably, be determined uniquely by the initial segment in each of the basic blocks. If we want to ensure that this block structure is indeed uniquely determined by encoding this information into the content of the blocks, our construction will become more complicated and may hinder the reader from concentrating on the more essential points of the construction. It is, however, always possible to couple our system with a Kronecker system which consists of an inverse limit of cyclic groups of orders l_n, to mark the block structure of elements of X. For that reason we shall, from now on, ignore the problem of recognition of the block frameworks and assume that the block structure can be uniquely recognized in elements of X. Thus there is now a natural Kronecker factor $(X, T) \xrightarrow{\pi} (Z, T)$, where $Z \subset \prod_{n=1}^\infty \{0, 1, \dots, l_n - 1\}$ is defined by $Z = \{z : z(n+1) \equiv z(n)(mod\ l_n), n = 1, 2, \dots\}$.

So far we have described a general procedure for manufacturing zero entropy minimal sets. Now we have to describe the minimal set in 2^X. Our minimal set Ω in 2^X will be in 2^π, and that means that each $E \in \Omega \subset 2^X$ is contained in $\pi^{-1}(z)$ for some $z \in Z$. We will define a set $\Omega_n \subset 2^{X_n}$ by giving a collection $\mathscr{L}_n$ of lists of words from $\mathscr{W}_n$ of size $\le t_n$, i.e.

$$\mathscr{L}_n \subset \{A \subset \mathscr{W}_n : |A| \le t_n\}$$

(we think of a list as a matrix whose rows are formed by the words of the list). Given such a collection of lists, we declare that $E \in \Omega_n$ if:

(1) E lies in a single fiber over Z; i.e. all points in E have the same block framework (note that there are only l_n-possibilities at this stage).

(2) At each position of the framework the set of words in $\mathscr{W}_n$ that we see as we range over the points of E belongs to our collection of lists $\mathscr{L}_n$.

The lists $\mathscr{L}_n$ will be constructed in a fashion that is consistent with $\mathscr{L}_{n-1}$, which means that $\Omega_n \subset \Omega_{n-1}$ (this makes sense since $2^{X_n} \subset 2^{X_{n-1}}$).

To understand the topology on 2^X we recall that if X is a compact metric space and $\{U_j\}_{j=1}^\infty$ a basis for the topology on X, then a basis for the topology on 2^X is obtained as follows. For every finite subset $L = \{j_1, j_2, \dots, j_k\}$ of $\mathbb{N}$, let

$$\begin{aligned} V(L) &= \langle U_{j_1}, U_{j_2}, \dots, U_{j_k} \rangle \\ &= \{A \in 2^X : A \subset \bigcup_{i=1}^k U_{j_i} \text{ and } A \cap U_{j_i} \neq \varnothing, i = 1, \dots, k\}. \end{aligned}$$

The family $\{V(L) : L \subset \mathbb{N}, L \text{ finite }\}$ is a basis for the Hausdorff topology on 2^X.

Back to our example, let $\{U_j\}_{j=1}^{\infty}$ be the basis for the topology on X given by the cylinder sets defined by words in $\mathscr{W}_n$ at the l_n different positions they can occupy, $n = 1, 2, \ldots$. It is then clear that, letting L run through all lists of words in $\mathscr{W}_n$ (i.e. all subsets of $\mathscr{W}_n$), the resulting family $\{V(L)\}$ will yield a basis for the topology on 2^X. In particular two closed subsets in 2^π are near each other, when for a large n there exists a list L of words in $\mathscr{W}_n$ such that both sets are in $V(L)$.

In the same way that words in $\mathscr{W}_n$ are conveniently used to signify both a proper word as well as a neighborhood in X, so a list $L \in \mathscr{L}_n$ will mean both the list itself and the neighborhood it defines in 2^X.

To ensure the minimality of $\Omega = \bigcap_1^\infty \Omega_n$ we need to ensure that each list in $\mathscr{L}_n$ has some initial segment (i.e. the list obtained by considering a certain portion of each word in the list; this will be aligned with the initial segment of words in $\mathscr{W}_n$ mentioned above), wherein one sees *all* concatenations of pairs of lists in $\mathscr{L}_{n-1}$. Note that there are many ways of concatenating lists. If L, L' are lists in $\mathscr{L}_{n-1}$ then a collection of words J of length $2l_{n-1}$ is a *concatenation* of L and L' if the initial l_{n-1}-words of J, as a set, coincide with L while the final l_{n-1}-words of J, as a set, coincide with L'. We will call the concatenation $J = \{ww : w \in L\}$ of a list L with itself the *diagonal* concatenation, and, for two lists L, L', the *product* concatenation of L and L' will be the list $J = \{ww' : w \in L, w' \in L'\}$.

Finally the fact that (Ω, T) will have positive entropy will be deduced from an exponential growth rate for the size of $\mathscr{L}_n$ with respect to l_n. In fact it is not simply the size of $\mathscr{L}_n$ that matters but rather the size of $\mathscr{L}_n$ when we identify lists if their projections onto the "zero lists" are the same; i.e. if $l_n = k \cdot l_0$, L is a list in $\mathscr{L}_n$ and Λ^1 the list in $\mathscr{L}_0$ we see when looking at the first l_0-places, Λ^2 the list in $\mathscr{L}_n$ we see when looking at the next l_0-places, etc., then we count L's as different only if, for some j, the Λ^j differ. It is this fixed "mesh size" we look at with its exponential growth that gives positive entropy to the quasi-factor (Ω, T). We now go over to the explicit description of the construction.

The zero stage. The set $\mathscr{W}_0$ of admissible words at stage zero will consist of ten words of length $l_0 = 10$:

$$w_0 = 100\ldots0\ ;\ w_1 = 110\ldots0\ ;\ w_2 = 101\ldots0\ ;\ \ldots\ w_9 = 100\ldots1.$$

At this stage we simply let $\mathscr{L}_0$ be the collection of all 2^9 possible lists of words from $\mathscr{W}_0$ containing the word w_0. Note that the list $L_0 = \{w_0\}$ is a sublist of every list in $\mathscr{L}_0$.

The first stage. In describing the first stage of the construction we will have to deal simultaneously with words and lists; the words of $\mathscr{W}_1$ will be implicitly defined as all the words involved in $\mathscr{L}_1$.

The words of $\mathscr{W}_1$ will have two parts, the *initial segment* and the *main part.* The initial segment will have two *sections*—one for the minimality of words and the other for the minimality of lists. The same terminology—initial segment, main part, etc.—will apply to lists.

To construct the first section of the initial segment of words in $\mathscr{W}_1$ define

$$\alpha = (w_0 w_0)(w_0 w_1)\cdots(w_9 w_9),$$

i.e. α is formed by concatenating all pairs w_i, w_j, $0 \leq i, j \leq 9$. Let α_0 have the same length as α but be composed entirely of w_0. Now all words in $\mathscr{W}_1$ will begin with either $\alpha\alpha_0$ or $\alpha_0\alpha$. We need two kinds of words because all lists in $\mathscr{L}_0$ contain w_0; to make $\mathscr{L}_1$ consistent with $\mathscr{L}_0$ we have to see the word w_0 in each position of the list. Thus there is just one first-section-list, and this list contains just two words.

Now consider the second section of the initial segment. For lists L_1, L_2 in $\mathscr{L}_0$ let $L_1 \circ L_2$ denote some concatenation of these two lists. Write out a list of the form

$$(L_i \circ L_j)(L_i' \circ L_j')\cdots$$

in which all possible concatenations of lists from $\mathscr{L}_0$ occur. Thus for a fixed pair i, j there will be in this list all possible concatenations of L_i and L_j (how we concatenate the pairs is immaterial). The second section of the initial segment of each list in $\mathscr{L}_1$ will consist of this single and fixed list. The total length of the initial segment we denote by r_1.

Finally we go over to the main part (whose length will be s_1 so that $l_1 = r_1 + s_1$). The main part of words in $\mathscr{W}_1$ will consist exclusively of words of the type:

$$w_0^a w_i w_0^b \quad \text{with} \quad \begin{Bmatrix} l_0 \cdot (a+b+1) = s_1 \\ a, b \geq 0 \\ 0 \leq i \leq 9 \end{Bmatrix}.$$

In constructing the main part of the lists in $\mathscr{L}_1$, our main purpose is to produce exponentially many lists—with concatenation of words that is rather small in size—namely linear in s_1. First we choose k_1 so that $s_1 = l_0 \cdot k_1$ satisfies $1 - s_1/l_1 \leq 1/10$ (we will, eventually, make $s_n/l_n \to 1$). Then for every sequence $L_{i_1} L_{i_2} \ldots L_{i_{k_1}}$ in $(\mathscr{L}_0)^{k_1}$—at least one of whose entries is L_0—we form the list

$$\begin{pmatrix} L_{i_1} & L_0 & \cdots & L_0 \\ L_0 & L_{i_2} & \cdots & L_0 \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ L_0 & L_0 & \cdots & L_{i_{k_1}} \end{pmatrix}.$$

There are now one first-section-list, one second-section-list and exponentially many main-part-lists. We let the product of the first and second section-lists be the unique initial-segment-list, and for our lists in $\mathscr{L}_1$ we take some concatenation of this initial-segment-list and each of the main-part-lists. There are therefore as many lists in $\mathscr{L}_1$ as there are main-part-lists. The words of $\mathscr{W}_1$ are all the words appearing in these lists; they all have the same length l_1. The special list that results when we take the constant sequence $L_0 L_0 \ldots L_0$ in the main part, we call the *zero list* and denote it by $L_0^{(1)}$. Note that we made sure that every list in $\mathscr{L}_1$ contains this list as a sublist.

The second stage. The second stage of the construction will capture already all the features of the general case. When referring to words of $\mathscr{W}_1$ we will use the following notation. We will denote all words in $L_0^{(1)}$ as "zero" words of $\mathscr{W}_1$, $w_0^{(1)}$—even though their initial segments come in several types. Similarly, $w_1^{(1)}$ will denote words where in the main part we see $w_0^a w_i w_0^b$ $(w_i \in \mathscr{W}_0, i \neq 0)$, regardless of the initial segment. Thus there are several kinds of $w_1^{(1)}$, etc.

Now form the word $\alpha^{(2)}$, as before, by concatenating all possible pairs of words from $\mathscr{W}_1$. Form also the words of type $\alpha_0^{(2)}$, each of the same length as $\alpha^{(2)}$, by concatenating a single word of type $w_0^{(1)}$. There are now many types of $\alpha_0^{(2)}$'s. The first section of the initial segment of words in $\mathscr{W}_2$ will be any one of the words $\alpha^{(2)}\alpha_0^{(2)}$ or $\alpha_0^{(2)}\alpha^{(2)}$. The first-section-list in $\mathscr{L}_2$ will consist of all these words.

Next we describe the second-section-list; as in stage one there will be only one such list. This will be formed by concatenating all possible pairs $L_i^{(1)} \circ L_j^{(1)}$ of lists in $\mathscr{L}_1$, where for each pair i, j we take all possible concatenations of the two lists. Again the initial-segment-list will be the product of the first and second-section-lists.

Finally for the main part; first choose k_2 so that $s_2 = l_1 \cdot k_2$ satisfies $1 - s_2/l_2 \leq 1/(10)^2$, where $l_2 = r_2 + s_2$ and r_2 is the length of the initial segment defined above. The lists in the main part will be formed again by assigning to every sequence $L_{i_1}^{(1)} L_{i_2}^{(1)} \cdots L_{i_{k_2}}^{(1)}$ in $(\mathscr{L}_1)^{k_2}$—at least one of whose entries is $L_0^{(1)}$—the list

$$\begin{pmatrix} L_{i_1}^{(1)} & L_0^{(1)} & \dots & L_0^{(1)} \\ L_0^{(1)} & L_{i_2}^{(1)} & \dots & L_0^{(1)} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ L_0^{(1)} & L_0^{(1)} & \dots & L_{i_{k_2}}^{(1)} \end{pmatrix}.$$

However, unlike the first stage, the above list is not really well defined, because we now have more than one zero word in the list $L_0^{(1)}$, so that the concatenations with this list should be described explicitly. In order to keep the cardinality of $\mathscr{W}_2$ small we choose, in each row, the diagonal concatenation between any two copies of $L_0^{(1)}$, and some arbitrary concatenation for the triples $L_0^{(1)} L_{i_j}^{(1)} L_0^{(1)}$. The main part of words in $\mathscr{W}_2$ will now have the form

$$(w_0^{(1)})^a w_i^{(1)} (w_0^{(1)})^b$$

where $a + b + 1 = k_2$, and $(w_0^{(1)})^a$ means the a-fold concatenation of a single word of type $w_0^{(n-1)}$, and we use the same zero word for $(w_0^{(n-1)})^b$. The main part lists are now defined and we let the lists in $\mathscr{L}_2$ be formed by taking some concatenation of the initial-segment-list and each of the main-part-lists. As before, $\mathscr{W}_2$ is defined as the set of words appearing in these lists. The zero list $L_0^{(2)}$ is the one obtained by concatenating the initial-segment-list with the main-part-list which corresponds to the sequence $L_0^{(1)} L_0^{(1)} \cdots L_0^{(1)}$ (k_2 times).

General stage. The collection of lists $\mathscr{L}_n$ contains a distinguished list called the zero list and denoted $L_0^{(n)}$—the words there are denoted by $w_0^{(n)}$. All other lists $L_i^{(n)}$ contain $L_0^{(n)}$. All words of the form $w_0^{(n)}$ have in the main part only words of the type $w_0^{(n-1)}$. All words of the form $w_i^{(n)}$, $i \neq 0$, have in the main part exactly one word of the form $w_i^{(n-1)}$. The initial segment—i.e. what we see as $(n-1)$-lists in the initial segment of all lists in $\mathscr{L}_n$—is the same and hence equals the $L_0^{(n)}$ list there. In the main part, to keep the entropy down, whenever we write $(w_0^{(n-1)})^a$ we mean a concatenation of a single word of type $w_0^{(n-1)}$, and we use the same zero word for $(w_0^{(n-1)})^b$.

If $u_{n-1} = |L_0^{(n-1)}|$, $v_{n-1} = |\{w_i^{(n-1)} : i \geq 1\}|$, then $|\mathscr{W}_{n-1}| = u_{n-1} + v_{n-1}$. Let $s_n = a+b+1$, $a, b \geq 0$, $1 - s_n/l_n \leq 1/(10)^n$; then, there are s_n choices for a, b, then $u_{n-1} + v_{n-1}$ choices for the $w_i^{(n-1)}$ (including $i = 0$), then u_{n-1} choices for the padding by the various zero words. Thus there are $(s_n) \cdot |\mathscr{W}_{n-1}| \cdot u_{n-1}$ different main parts. Clearly for s_n large this makes $\frac{1}{l_n} \log |\mathscr{W}_n| \leq 1/(10)^{-n}$ and guarantees that eventually $h(X, T) = 0$. The calculations involved in showing that (Ω, T) has positive topological entropy will be given in the next section when we show the existence of entropy pairs in (Ω, T).

5. The extension σ

It is clear from our construction that a natural homomorphism $\sigma : (\Omega, T) \to (Z, T)$ is defined, simply by mapping $A \in \Omega$ onto the unique point $z \in Z$ for which $A \subset \pi^{-1}(z)$.

Proposition 5.1. *The homomorphism σ is almost one to one; i.e. there exists a dense G_δ subset, Z_0 of Z, such that $|\sigma^{-1}(z)| = 1$ for $z \in Z_0$.*

Proof. Let

$$Z_0 = \{z \in Z : \text{ for some } n_i \to \infty, \lim z(n_i) = \lim r_{n_i} - z(n_i) = \infty\},$$

where $z \in \prod_{n=1}^{\infty} \{0, 1, \ldots, l_n - 1\}$, and r_n is the length of the initial segment in words of $\mathscr{W}_n$.

Since there is a unique initial-segment-list at each stage, it is now clear that for each $z \in Z_0$ there is one and only one $A \in \Omega$ for which $\sigma(A) = z$. It is easily seen that, by minimality, the set of points $z \in Z$ for which $|\pi^{-1}(z)| = 1$—since it is not empty—is a dense G_δ subset of Z. Of course this subset cannot exhaust Z as (Ω, T) has positive entropy and hence is not isomorphic to (Z, T). $\square$

Next we briefly recall the definitions of entropy pairs and u.p.e. flows. For more details see [B,1], [B,2], [B-L], [G-W,1], [B,3] and [G-W,4].

Let (X, T) be a flow; an open cover $\mathscr{U} = \{U, V\}$ of X is called a *standard cover* if both U and V are none-dense in X. (X, T) has *uniform positive entropy* (*u.p.e.*) if for every standard cover $\mathscr{U}$ of X, the topological entropy $h(\mathscr{U}, T) > 0$. A pair $(x, x') \in X \times X$ is an *entropy pair* if for every standard cover $\mathscr{U}$ with $x \in int(U^c)$ and $x' \in int(V^c)$, $h(\mathscr{U}, T) > 0$. Thus (X, T) is u.p.e. iff every nondiagonal pair in $X \times X$ is an entropy pair. Of course the existence of entropy pairs implies positive entropy and the converse is also true.

In an arbitrary flow (Y, T), let us call a pair of points $(y_0, y_1) \in Y \times Y$ an *E-pair* if $y_0 \neq y_1$, and for every pair of disjoint neighborhoods U_0, U_1 of y_0, y_1 respectively, there exist $\delta > 0$ and k_0 such that for every $k \geq k_0$, there exists a sequence $0 \leq n_1 < n_2 < \cdots < n_k < k/\delta$ such that for every $s \in \{0, 1\}^k$, there exists $y \in Y$ with

$$T^{n_1}y \in U_{s(1)}, T^{n_2}y \in U_{s(2)}, \dots, T^{n_k}y \in U_{s(k)}.$$

It follows directly from the definition that for every pair U_0, U_1 as above the entropy $h(\mathscr{U}, T) \geq \delta$, where $\mathscr{U}$ is the open cover $\{(\overline{U_0})^c, (\overline{U_1})^c\}$ of Y. Hence every E-pair is an entropy-pair. We say that the extension $(Y, T) \xrightarrow{\sigma} (Z, T)$ is a *u.p.e. extension* (or an *entropy extension*), if every pair of distinct points $y, y' \in Y$ with $\sigma(y) = \sigma(y')$, is an entropy pair.

Proposition 5.2. *Every pair of points A_0, A_1 in Ω, such that $A_0 \neq A_1$ and $\sigma(A_0) = \sigma(A_1)$, is an E-pair, hence an entropy pair; in particular, the extension σ is a u.p.e. extension.*

Proof. Notice first that the conditions $1 - s_n/l_n < 1/(10)^n$ imposed in the construction imply that

$$\prod_{n=1}^{\infty} s_n/l_n = \gamma > 0.$$

Now since $A_0 \neq A_1$ and $\sigma(A_0) = \sigma(A_1)$, there exist an n and lists $L_0 \neq L_1$ in $\mathscr{L}_n$ such that $A_0 \in V(L_0)$ and $A_1 \in V(L_1)$ (note that $L_0 \neq L_1$ implies $V(L_0) \cap V(L_1) = \varnothing$). In the definition of an E-pair, take $k_0 = k_n$. Now given $k \geq k_0$ and $s \in \{0, 1\}^k$, let m be determined by $k_m \leq k < k_{m+1}$. Consider the sequence $L_{s(1)}L_{s(2)} \cdots L_{s(k)}$ in $\{L_0, L_1\}^k$. This is a sequence of lists in $\mathscr{L}_n$ and we can group them into sequences of length k_{n+1} to form the main parts of $[k/k_{n+1}] + 1$, $\mathscr{L}_{n+1}$ lists. Use this sequence of $\mathscr{L}_{n+1}$ lists to build the main parts of a certain number of $\mathscr{L}_{n+1}$ lists. These, in turn, we use to build the main parts of a sequence of lists in $\mathscr{L}_{n+2}$, etc., until we get in the last step, a portion of the main part of a single $\mathscr{L}_{m+1}$ list, which we complete in an arbitrary way. Let us denote the $\mathscr{L}_{m+1}$ list obtained in this way by Λ, and let $n_1 < n_2 < \cdots < n_k$ be the places in Λ where the $L_{s(j)}$, $j = 1, \dots, k$, begin. Then it is easy to see that the density of $\{n_j\}_{j=1}^k$ in the interval $[0, n_k]$ is bounded below by

$$(1/l_n) \cdot (s_{n+1}/l_{n+1}) \cdots (s_{m+1}/l_{m+1}) \geq \gamma \cdot 1/l_n.$$

Taking $\delta = \gamma \cdot 1/l_n$ completes the proof of the proposition. □

The following corollary shows that the relative version of Proposition 6 of [B,2] is false.

Corollary 5.3. *There exist a u.p.e. extension* $(\Omega, T) \xrightarrow{\sigma} (Z, T)$ *of minimal flows and an extension* $(X, T) \xrightarrow{\pi} (Z, T)$, *where* (X, T) *is a minimal flow of zero entropy, such that* (Ω, T) *and* (X, T) *are not disjoint over* (Z, T).

Proof. To say that (Ω, T) and (X, T) are disjoint over (Z, T) means that the subset

$$X \underset{Z}{\times} \Omega =: \{(x, A) : x \in X, A \in \Omega, \pi(x) = \sigma(A)\}$$

of $X \times \Omega$ is minimal. Now for our example: as we have seen σ is u.p.e., yet clearly the subset

$$\{(x, A) : x \in A \in \Omega\}$$

is a proper closed and T-invariant subset of $X \underset{Z}{\times} \Omega$. □

Remark. We observe that Corollary 5.3 also yields an example of a pair of minimal flows (X, T) and (Ω, T) which are not disjoint over their common factor (Z, T) but, at the same time, have no proper common factor over (Z, T).

REFERENCES

[B-S] W. Bauer and K. Sigmund, *Topological dynamics of transformations induced on the space of probability measures*, Monatsh. Math. **79** (1975), 81–92.

[B,1] F. Blanchard, *Fully positive topological entropy and topological mixing*, Symbolic Dynamics and Applications, Contemp. Math., vol. 135, Amer. Math. Soc., Providence, RI, 1992, pp. 95–105.

[B,2] ——, *A disjointness theorem involving topological entropy*, Bull. Soc. Math. France **121** (1993), 465–478.

[B-L] F. Blanchard and Y. Lacroix, *Zero entropy factors of topological flows*, Proc. Amer. Math. Soc. **119** (1993), 985–992.

[B,3] F. Blanchard, B. Host, A. Maass, and D. J. Rudolph, *Entropy pairs for a measure*, preprint.

[D-G-S] M. Denker, C. Grillenberger, and K. Sigmund, *Ergodic theory on compact spaces*, Lecture Notes in Math., vol. 527, Springer-Verlag, Berlin and New York, 1976.

[E-F] N. F. G. Martin and J. W. England, *Mathematical theory of entropy*, Addison-Wesley, Reading, MA, 1981.

[F-L-M] T. Figiel, J. Lindenstrauss, and V. D. Milman, *The dimension of almost spherical sections of convex bodies*, Acta Math. **139** (1977), 53–94.

[F,1] H. Furstenberg, *Disjointness in ergodic theory, minimal sets, and a problem in diophantine approximation*, Math. Systems Theory **1** (1967), 1–55.

[F,2] ——, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Analyse Math. **31** (1977), 204–256.

[F,3] ——, *Recurrence in ergodic theory and combinatorial number theory*, Princeton Univ. Press, Princeton, NJ, 1981.

[G] S. Glasner, *Quasi-factors in ergodic theory*, Israel J. Math. **45** (1983), 198–208.

[G-W,1] E. Glasner and B. Weiss, *Interpolation sets for subalgebras of* $l^\infty(\mathbb{Z})$, Israel J. Math. **44** (1983), 345–360.

[G-W,2] ——, *Strictly ergodic uniform positive entropy models*, Bull. Soc. Math. France (to appear).

[G-W,3] ——, *Dynamics and entropy of the space of measures*, C.R. Acad. Sci. Paris Sér. I **317** (1993), 239–243.

[G-W,4] ——, *Topological entropy of extensions*, Ergodic Theory and Harmonic Analysis (Petersen and Salama, eds.), Cambridge Univ. Press (to appear).

[K-M] M. G. Karpovsky and V. D. Milman, *Coordinate density of sets of vectors*, Discrete Math. **24** (1978), 177–184.

[K] A. W. Knapp, *Functions behaving like almost automorphic functions*, Proc. Sympos. Topological Dynamics, Benjamin, New York, 1968, pp. 299–317.

[P] W. Parry, *Topics in ergodic theory*, Cambridge Univ. Press, London and New York, 1980.

[Sa] N. Sauer, *On the density of families of sets*, J. Combin. Theory Ser. A **13** (1972), 145–147.

[Sh] S. Shelah, *A combinatorial problem; Stability and order for models and theories in infinitary languages*, Pacific J. Math. **41** (1972), 247–261.

[T] J. P. Thouvenot, *Quelque propriétés des systèmes dynamiques etc.*, Israel J. Math. **21** (1975), 177–207.

[Z,1] R. J. Zimmer, *Extensions of ergodic group actions and*, Illinois J. Math. **20** (1976), 373–409.

[Z,2] ——, *Ergodic actions with generalized discrete spectrum*, Illinois J. Math. **20** (1976), 555–588.

ABSTRACT. For minimal systems (X, T) of zero topological entropy we demonstrate the sharp difference between the behavior, regarding entropy, of the systems $(M(X), T)$ and $(2^X, T)$ induced by T on the spaces $M(X)$ of probability measures on X and 2^X of closed subsets of X. It is shown that the system $(M(X), T)$ has itself zero topological entropy. Two proofs of this theorem are given. The first uses ergodic theoretic ideas . The second relies on the different behavior of the Banach spaces l_1^n and l_∞^n with respect to the existence of almost Hilbertian central sections of the unit ball. In contrast to this theorem we construct a minimal system (X, T) of zero entropy with a minimal subsystem (Y, T) of $(2^X, T)$ whose entropy is positive.

SCHOOL OF MATHEMATICS, TEL AVIV UNIVERSITY, RAMAT AVIV, ISRAEL
E-mail address: glasner@math.tau.ac.il

INSTITUTE OF MATHEMATICS, HEBREW UNIVERSITY, JERUSALEM, ISRAEL
E-mail address: weiss@math.huji.ac.il