Complex Systems **2** (1988) 357–386

On the Entropy Geometry of Cellular Automata

John Milnor
Institute for Advanced Study, Princeton University,
Princeton, NJ 08540, USA

Abstract. We consider *configurations* which assign some element of a fixed finite alphabet to each point of an n-dimensional lattice. An n-dimensional *cellular automaton map* assigns a new configuration $\mathbf{a}' = f(\mathbf{a})$ to each such configuration $\mathbf{a}$, in a translation invariant manner, and in such a way that the values of $f(\mathbf{a})$ throughout any finite subset of the lattice depend only on the values of $\mathbf{a}$ throughout some larger finite subset. If we iterate such a map f over and over, then the complete history of the resulting configurations throughout time can be described as a new configuration over an $(n+1)$-dimensional "space-time" lattice. This note will describe the distribution and flow of information throughout this $(n+1)$-dimensional lattice by introducing an *n-dimensional entropy function* which measures the density of information in very large finite sets.

1. Introduction

Let K be some fixed finite alphabet with $k \geq 2$ elements, and let L be an n-dimensional *lattice*, that is, a free abelian group isomorphic to $\mathbf{Z}^n$. A function $\mathbf{a} : L \to K$ which assigns an alphabet element to each lattice point is called a *configuration* over L. For example when $n = 1$ such a configuration can be described as a doubly infinite sequence (or briefly "bisequence") of symbols from K, and when $n = 2$ it can be described as an infinite two-dimensional array of symbols from K. We will use the notation K^L for the space consisting of all such configurations. This space has a natural compact topology, in which two configurations are close to each other whenever they coincide throughout some large finite set. Note that the lattice L acts naturally as a group of continuous translations of K^L. This compact space K^L together with the group L of homeomorphisms is called the *n-dimensional full k-shift*.

Definition. A continuous map $f : K^L \to K^L$ which commutes with translation by lattice elements is called an *n-dimensional cellular automaton map*, or briefly *CA-map*. (In the terminology of Hedlund, f would be called an "endomorphism of the n-dimensional k-shift".) The triple $\{K, L, f\}$ consisting of alphabet K, lattice L, and CA-map f is called a *cellular automaton*.

More explicitly, such a map f assigns to each configuration $\mathbf{a} \in K^L$ a new configuration $\mathbf{a}' = f(\mathbf{a})$ which can be computed by some formula of the form

$$\mathbf{a}'(\ell) = \Phi(\mathbf{a}(\ell + v_1), \ldots, \mathbf{a}(\ell + v_r)). \tag{1.1}$$

Here $\Phi : K \times \cdots \times K \to K$ is to be some fixed function of r variables, and $v_1, \ldots, v_r$ are to be fixed vectors in the lattice. (See for example [19,6].) We will call Φ a *local map* or *block map*, and will call f the *associated CA-map*.

In the one-dimensional case, the *topological entropy* $h_{\mathrm{top}}(f)$, as defined by Adler, Konheim, and McAndrew, is an interesting and useful numerical invariant which measures "information per unit time". (Compare [4,9,10, 11], and see [18] as a general reference for topological and measure-theoretic entropy.) If we consider not only the entropy of f itself, but also the collection of all entropies of compositions of f with lattice translations, then we obtain a much richer structure ([13,17]). If we are also given a suitably invariant probability measure on K^L, then we can also consider corresponding measure-theoretic entropies (see [3,16,8]).

In the higher-dimensional case, $n > 1$, the topological and measure-theoretic entropies are usually infinite. In order to get a useful theory in the n-dimensional case, we must introduce a corresponding theory of "n-dimensional entropy." It is hoped that this concept of n-dimensional entropy will give a better picture of what is going on, even in the classical case $n = 1$.

The paper is organized as follows. Section 2 describes the topological or measure-theoretic information content $H(S)$ associated with a finite set S in the given lattice L or in the "space-time lattice" $\mathbf{Z} \times L$, for a given CA-map f. Section 3 constructs the d-dimensional entropy $\eta_d(X)$ associated with such an information function H, and also the d-dimensional directional entropy $h_d(\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_d)$. Here d is an integer determined by H, X is any compact subset of the ambient vector space $(\mathbf{Z} \times L) \otimes \mathbf{R}$, and $\mathbf{v}_1, \ldots, \mathbf{v}_d$ are linearly independent vectors in this space. Section 4 applies these constructions to cellular automaton limit sets. The last three sections, which depend heavily on oral communications to the author by Lind and Smillie, study the n-dimensional entropy of complete histories in the $(n+1)$-dimensional lattice $\mathbf{Z} \times L$. Section 5 studies "causal cones" associated to a cellular automaton, and Section 6 gives some one-dimensional examples. Section 7 discusses directional entropies, making use of normal vectors and dual cones, and makes a particular study of invertible or "quasi-invertible" CA-maps. The Appendix generalizes some of these constructions to commuting maps on an arbitrary compact metric space, and to commuting measure preserving transformations.

I am grateful to Lyman Hurd and Kyewon Park for helpful suggestions.

2. The information function $S \mapsto H(S)$

Definition. Let L be an n-dimensional lattice, and let S range over all finite subsets of L. By an *information function* H on L will be meant a function $S \mapsto H(S)$ which assigns a real number $H(S)$ to each such S, and which is monotone, subadditive, and translation-invariant; that is,

$$H(S) \le H(S') \qquad \text{whenever } S \subset S', \tag{2.1}$$

$$H(S \cup S') \le H(S) + H(S') \qquad \text{for all } S \text{ and } S', \text{ and} \tag{2.2}$$

$$H(S) = H(S+v) \qquad \text{for any vector } v \text{ in the lattice.} \tag{2.3}$$

It follows that

$$0 \le H(S) \le H_0\,|S| \tag{2.4}$$

for every non-vacuous S, where $|S|$ denotes the number of elements in S, and where the constant H_0 is defined to be the value of H on a set $\{\ell\}$ consisting of a single lattice point. We may as well assume also that the empty set $\emptyset$ satisfies $H(\emptyset) = 0$, so that (2.4) will be satisfied without exception.

Remarks. We will refer to $H(S)$ as the "information content" of the set S (although the term *entropy* of S would be closer to the usual usage in the literature). In practice, we will distinguish between topological and measure-theoretic information. By definition, the "topological" information associated with a choice between n alternatives is $\log(n)$. If each alternative is assigned a probability $p_i \ge 0$ with $p_1 + \cdots + p_n = 1$, then the measure-theoretic information (or more properly the "expected information") associated with such a choice is

$$-p_1 \log(p_1) - \cdots - p_n \log(p_n). \tag{2.5}$$

Here $0 \log(0) = 0$ by definition. Note that the expression (2.5) attains its minimum zero when all but one of the probabilities is zero, and attains its maximum of $\log(n)$ when $p_1 = \cdots = p_n = 1/n$.

Example 2.1. Topological information for subshifts. By definition, a *subshift* of K^L is a closed set $A \subset K^L$ which is invariant under lattice translations. By a *partial configuration* over a finite set $S \subset L$ will be meant any mapping $\alpha : S \to K$. Note that a subshift A is completely determined if we specify which partial configurations α can be extended to full configurations $\mathbf{a} : L \to K$ belonging to A.

To each subshift $A \subset K^L$ we associate the "topological information function" $S \mapsto H_A(S)$, where $H_A(S)$ is defined to be the logarithm of the number of distinct partial configurations $\alpha : S \to K$ which can be extended to full configurations belonging to A. In other words, $\log^{-1} H_A(S)$ is equal to the number of elements in the image of the restriction map $A \subset K^L \to K^S$. The inequalities (2.1) through (2.3) are easily verified. Since the number of partial configurations over S is equal to $k^{|S|}$, the inequality (2.4) can be replaced by

$$0 \le H_A(S) \le |S| \log(k). \tag{2.6}$$

Intuitively, $H_A(S)$ can be described as the quantity of information about an unknown configuration $\mathbf{a} \in A$ which we would obtain by knowing its restriction to S. As a matter of notational convenience, we will often move the subscripts onto the main line, writing $H(A; S)$ in place of $H_A(S)$.

Note. Depending on taste, the reader may use natural logarithms throughout, or logarithms to the base 2 so that information is measured in "bits", or even logarithms to the base k so that formula (2.6) takes the simpler form $0 \leq H_A(S) \leq |S|$. Because of this ambiguity, we write $\log^{-1}$ rather than exp for the inverse transformation.

Every CA-map $f : K^L \to K^L$ gives rise to an descending sequence of subshifts

$$K^L \supset f(K^L) \supset f^2(K^L) \supset f^3(K^L) \supset \cdots.$$

Here f^t stands for the t-fold composition $f \circ \cdots \circ f$. The intersection $\Lambda = \Lambda(f)$ of this sequence of successive images $f^t(K^L)$ is another subshift which is called the *limit set* or the *eventual image* of the cellular automaton. Thus a configuration $\mathbf{a}_0$ belongs to Λ if and only if it can be written as a t-fold image $\mathbf{a}_t \mapsto \mathbf{a}_{t-1} \mapsto \cdots \mapsto \mathbf{a}_1 \mapsto \mathbf{a}_0$ under f for every integer $t > 0$. Associated with this descending sequence of subshifts is a descending sequence of information functions

$$H(K^L; S) = |S| \log(k) \geq H(f(K^L); S) \geq H(f^2(K^L); S) \geq \cdots$$

which clearly converges to the limit function $H(\Lambda; S)$. In fact $H(\Lambda; S) = H(f^t(K^L); S)$ for t sufficiently large, depending on S.

Example 2.2. Complete histories. There is an important $(n+1)$-dimensional subshift associated with an n-dimensional cellular automaton, as follows. By a *complete history* for the CA-map f will be meant a bisequence

$$\cdots \mapsto \mathbf{a}_{-1} \mapsto \mathbf{a}_0 \mapsto \mathbf{a}_1 \mapsto \mathbf{a}_2 \mapsto \cdots$$

of configurations $\mathbf{a}_t \in K^L$ satisfying the identity $f(\mathbf{a}_t) = \mathbf{a}_{t+1}$ for every integer $t \in \mathbf{Z}$. Every such complete history can be considered as a function $(t, \ell) \mapsto a(t, \ell) = a_t(\ell)$, or in other words as a configuration on the $(n+1)$-dimensional lattice $\mathbf{Z} \times L$, satisfying the equation

$$\mathrm{a}(t+1, \ell) = \Phi(\mathrm{a}(t, \ell + v_1), \ldots, \mathrm{a}(t, \ell + v_r)) \tag{2.7}$$

for all t and ℓ. (Compare (1.1)). Thus we can identify the space of complete histories for f with the subshift $\mathcal{A} = \mathcal{A}(f) \subset K^{Z \times L}$ consisting of all configurations satisfying (2.7).

There is an associated topological information function $S \mapsto H_{\mathcal{A}}(S)$, where now S varies over finite subsets of the product lattice. Note that this construction subsumes the previous one in the following sense. If we restrict a complete history $\mathbf{a} \in \mathcal{A}$ to the n-dimensional sublattice $0 \times L$, then evidently we obtain a configuration $\mathbf{a}_0$ belonging to the limit set $\Lambda(f)$, and conversely it is not difficult to show that every configuration in Λ can be obtained in this way. Hence $H_{\mathcal{A}}(0 \times S) = H_\Lambda(S)$ for every finite set $S \subset L$.

Example 2.3. Measure-theoretic information. Let μ be a probability measure on K^L, or more precisely on the σ-ring of Borel subsets of K^L, which is invariant under lattice translations. Then μ gives rise to an information function $S \mapsto H_\mu(S)$ as follows. Define the *probability* p_α of a partial configuration $\alpha : S \to K$ to be the measure of the open and closed cylinder set E_α consisting of all configurations $\mathbf{a} \in K^L$ whose restriction to S is equal to α. Thus the sum of the $k^{|S|}$ probabilities p_α is equal to $\mu(K^L) = 1$. Set

$$H_\mu(S) = H(\mu;\, S) = \sum -p_\alpha \log(p_\alpha), \tag{2.8}$$

to be summed over all partial configurations $\alpha : S \to K$, where as usual $0\log(0) = 0$. Then the inequalities (2.1), (2.2) and (2.3) are easily verified. In fact (2.2) can be sharpened to

$$H_\mu(S \cup S') \le H_\mu(S) + H_\mu(S') - H_\mu(S \cap S'). \tag{2.9}$$

Equation (2.9) is equivalent to the statement that the "*conditional entropy*" $H(S/T) = H(S \cup T) - H(T)$ is subadditive, or to the statement that

$$H(S/T) \le H(S/T_0) \quad \text{whenever} \quad T \supset T_0. \tag{2.10}$$

Proofs of these inequalities may be found in [18, p. 81].

The corresponding sharpened inequalities for topological information are unfortunately false. As an example, consider the subshift $A \subset \{0,1,2\}^L$ consisting of the constant zero configuration together with all possible configurations of ones and twos. Then $H_A(S) = \log(1 + 2^{|S|})$ for S non-vacuous, and the inequality (2.9) is essentially never satisfied.

The relationship between measure-theoretic information and topological information can be described as follows. Define the *support* $|\mu|$ to be the smallest closed subset of K^L which has measure 1. Thus the complement of $|\mu|$ can be described as the union of all cylinder sets E_α which have probability $\mu(E_\alpha) = 0$; where now α is to vary over partial configurations which are defined on arbitrarily large finite sets $S \subset L$. Then $|\mu|$ is a subshift of K^L, and it is not difficult to check that $0 \le H_\mu(S) \le H_{|\mu|}(S)$.

Example 2.4. ($\mathbf{Z} \times L$)-invariant measures. Given a CA-map $f : K^L \to K^L$, we are interested in probability measures μ_0 which are not only translation-invariant but also f-invariant, in the sense that

$$\mu_0(f^{-1}(E_\alpha)) = \mu_0(E_\alpha)$$

for every cylinder set E_α, and hence for every measurable set $E \subset K^L$. If μ_0 is such a doubly-invariant measure, then clearly the support $|\mu_0|$ is contained in the limit set Λ, so that $0 \le H_{\mu_0}(S) \le H_{|\mu_0|}(S) \le H_\Lambda(S)$. We will show that μ_0 gives rise to a unique translation-invariant measure μ on $K^{Z\times L}$, with support $|\mu|$ contained in $\mathcal{A}(f)$.

Lemma 1. *If μ_0 is an f-invariant and translation-invariant measure on K^L, then there is one and only one translation-invariant measure μ on $K^{Z\times L}$ which has support contained in $\mathcal{A}(f)$ and which satisfies the condition that the μ-probability of any partial configuration on a subset $0 \times S \subset 0 \times L$ is equal to the μ_0-probability of this same partial configuration.*

Evidently the associated information function $S \mapsto H_\mu(S)$ satisfies the conditions that $H_\mu(S) \le H_{|\mu|}(S) \le H_{\mathcal{A}}(S)$ for $S \subset \mathbf{Z} \times L$, and that $H_\mu(0 \times S) = H_{\mu_0}(S)$ for $S \subset L$.

Proof of Lemma 1. Every configuration on the hyperplane $0 \times L$ gives rise to a unique "forward history" on the half-space $\{t \ge 0\} \times L$. Hence, if we are given μ_0, then we can assign a probability p_α to any partial configuration α which is defined over a finite subset of this half-space. These probabilities are shift-invariant, and hence can be defined for partial configurations over an arbitrary finite subset of $\mathbf{Z} \times L$. The extension to arbitrary measurable subsets of $K^{Z\times L}$ is now straightforward. ∎

3. The d-dimensional entropy set function $\eta_d(X)$

It will be convenient to embed the n-dimensional lattice L into the n-dimensional real vector space $V \cong L \otimes \mathbf{R}$, which is spanned by any basis for L. It will also be convenient to choose some Euclidean metric for this vector space V, so that distances are defined.

Let $S \mapsto H(S)$ be some fixed function satisfying the axioms (2.1) through (2.3). In many interesting cases, it turns out that the quantity $H(S)$ grows roughly like the d-th power of the diameter of S, for some integer $0 \le d \le n$. More precisely, let us define the *growth degree* of the function H to be the smallest integer d such that

$$H(S) \le c \text{ diam } (S)^d \tag{3.1}$$

for some constant c, independent of S, and for all S with at least two distinct points. (We must exclude single point sets here.)

Note that H has growth degree $d = 0$ if and only if it is bounded. In this case, we define the "0-dimensional entropy" η_0 to be simply the supremum of $H(S)$ as the finite set $S \subset L$ becomes arbitrarily large.

To study the more interesting case $d \ge 1$, we first extend the function H as follows. If B is an arbitrary bounded subset of the ambient vector space V, then we define $H(B)$ to be $H(B \cap L)$, the value of H on the set of lattice points in B.

Definition. Given any compact set $X \subset V$, the *d-dimensional entropy* $\eta_d(X)$ is defined as follows. For any bounded set $B \subset V$ and any large real number t, consider the stretched and thickened set $B + tX$, consisting of all sums $\mathbf{b} + t\mathbf{x}$ with $\mathbf{b} \in B$ and $\mathbf{x} \in X$. For each fixed B, form the *lim sup* of the ratio $H(B+tX)/t^d$ as $t \to \infty$. Then form the supremum as the bounded set B becomes arbitrarily large. That is, define

$$\eta_d(X) = \sup_B \limsup_{t\to\infty} H(B + tX)/t^d. \tag{3.2}$$

We will prove the following.

Theorem 1. *If H has growth degree $d \geq 1$, then $\eta_d(X)$ is defined and finite for every compact set $X \subset V$, and satisfies*

$$0 \leq \eta_d(X) \leq c \text{ diam } (X)^d. \tag{3.3}$$

This function $X \mapsto \eta_d(X)$ is monotone, subadditive, and translation-invariant, that is

$$\begin{aligned} 0 \leq \eta_d(X) \leq \eta_d(Y) &\quad \text{for } X \subset Y, \\ \eta_d(X \cup Y) \leq \eta_d(X) + \eta_d(Y) &\quad \text{for any compact } X \text{ and } Y, \text{ and} \\ \eta_d(X + \mathbf{v}) = \eta_d(X) &\quad \text{for any vector } \mathbf{v}, \end{aligned}$$

and furthermore

$$\eta_d(tX) = t^d \eta_d(X) \text{ for any } t \geq 0. \tag{3.4}$$

Proof. This is almost completely straightforward. The only point which may require comment is translation-invariance, since the information function $B \mapsto H(B \cap L)$ is invariant only under translation by vectors in the lattice L. To prove invariance under arbitrary translations, choose some ball B_0 which is large enough so that every translate $\mathbf{v} + B_0$ contains at least one point of the lattice L. Then $H(B + \mathbf{v}) \leq H(B + B_0)$ for every bounded set B and vector $\mathbf{v}$, hence

$$\lim \sup_t H(B + t(X + \mathbf{v}))/t^d \leq \lim \sup_t H((B + B_0) + tX)/t^d.$$

The rest of the argument will be left to the reader. ∎

Caution: It definitely is not stated that η_d is an additive set function, that $\eta_d(-X) = \eta_d(X)$, or that $\eta_d(X)$ depends continuously on X. (Compare §6.3)

Note that the function η_d can be of interest for only one value of d, namely the growth degree for the given information function H. If we use the corresponding definition with a larger (or smaller) value of d, then the resulting entropy will be zero (or infinite).

Here is one immediate consequence of Theorem 1.

Corollary 1. *The entropy $\eta_d(X)$ is less than or equal to some constant times the d-dimensional Hausdorff measure of X. In particular, if X has Hausdorff measure zero, then $\eta_d(X) = 0$.*

The proof is straightforward. For reasonable subsets of some d-dimensional plane, we will see that η_d is just a constant multiple of Lebesgue or Hausdorff measure. More precisely:

Theorem 2. *If H has growth degree d, then for compact polyhedral subsets X of some fixed d-dimensional plane $P \subset V$ the associated entropy $\eta_d(X)$ is proportional to the d-dimensional Lebesgue measure of X. In other words there is a proportionality constant $\mathcal{H}_d(P) \geq 0$, which depends only on P, so that*

$$\eta_d(X) = \mathcal{H}_d(P) \text{ volume}(X)$$

for every compact polyhedral $X \subset P$. Furthermore, for such X, the "lim sup" *in the definition of η_d can be replaced by a "lim inf", without changing the value of $\eta_d(X)$.*

The proof will be given at the end of this section.

It follows from (3.3) that these constants $\mathcal{H}_d(P)$, which measure information per unit d-dimensional volume near P, are uniformly bounded:

$$0 \leq \mathcal{H}_d(P) \leq \text{constant}$$

for all d-planes $P \subset V$. However, we will see in §6.3 that this bounded function $P \mapsto \mathcal{H}_d(P)$ need not be continuous. The ratio $\mathcal{H}_d(P) = \eta_d(X)/\text{volume}(X)$ depends of course on the particular choice of Euclidean metric, although the entropy $\eta_d(X)$ does not depend on the choice of metric. As long as we consider only polyhedral subsets of some fixed d-plane, it follows that the set function η_d is strictly additive,

$$\eta_d(X \cup Y) = \eta_d(X) + \eta_d(Y) - \eta_d(X \cap Y),$$

and symmetric, $\eta_d(-X) = \eta_d(X)$. (Compare §6.2.)

Definition. In the case $d = 1$ the number $h_1(\mathbf{v}) = \eta_1([0,1]\mathbf{v})$ is called the "*directional entropy*" in the direction $\mathbf{v}$. Here the notation $[0,1]\mathbf{v}$ stands for the line segment from the origin to the vector $\mathbf{v}$. More generally, for any d, we will use the notation $h_d(\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_d)$ for the value of η_d on the parallelepiped $[0,1]\mathbf{v}_1 + \cdots + [0,1]\mathbf{v}_d$ which is spanned by the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_d$. (Compare [13]). Note the identity

$$h_d(t\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_d) = |t| h_d(\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_d). \tag{3.5}$$

Evidently the function $P \mapsto \mathcal{H}_d(P)$ and the function

$$\mathbf{v}_1, \ldots, \mathbf{v}_d \mapsto h_d(\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_d)$$

determine each other uniquely, and either one might reasonably be called the "d-dimensional directional entropy". (See also §7.) Presumably this directional entropy could be identically zero even if η_d is not identically zero. For example, in the n-dimensional case, $n \geq 2$, if we define $H(S)$ to be $|S|^{1/n}$, then $\eta_1(X)$ is non-trivial, being equal to the n-th root of the n-dimensional volume of X, and yet $\mathcal{H}_1$ is identically zero. However, such examples don't seem to arise in practice.

In the application to cellular automata, the value of the one-dimensional entropy $h_1(\mathcal{A}; \mathbf{v})$ at a vector $\mathbf{v} = (t, \ell)$ in the spacetime lattice $\mathbf{Z} \times L$ can be identified with the topological entropy of the mapping f^t composed with a translation by the lattice element ℓ. In particular, if $\mathbf{v} = (1, 0)$ then we obtain the topological entropy of the mapping f itself, as defined by Adler, Konheim, and McAndrew. Similarly, if μ is an invariant measure supported by $\mathcal{A}$, then $h_1(\mu; (1,0))$ is the measure theoretic entropy of f, as defined by Kolmogorov and Sinai. In the case of a one-dimensional CA-map, these invariants are interesting and non-trivial. However, for higher-dimensional CA-maps the one-dimensional entropy is usually infinite, except in trivial cases where it is zero. In most cases, the appropriate tool for studying n-dimensional cellular automata is rather the n-dimensional entropy in the $(n+1)$-dimensional vector space $\mathbf{R} \times V$. (See §§4-7.)

As a consequence of the last statement in Theorem 2, we have the following.

Corollary 2. *In the measure-theoretic case, the sharper inequality*

$$\eta_d(X \cup Y) \leq \eta_d(X) + \eta_d(Y) - \eta_d(X \cap Y) \tag{3.6}$$

is valid whenever the intersection $X \cap Y$ is a polyhedral subset of a d-plane. Similarly

$$\eta_d(X/Y) \leq \eta_d(X/Y_0) \tag{3.7}$$

whenever the subset $Y_0 \subset Y$ is a polyhedral subset of a d-plane.

Proof (assuming Theorem 2). The inequality (3.7) can also be written as

$$\eta_d(X \cup Y) \leq \eta_d(Y) + \eta_d(X \cup Y_0) - \eta_d(Y_0).$$

Since we are in the measure-theoretic case, we know by (2.10) that the corresponding inequality

$$H(B + t(X \cup Y)) \leq H(B + tY) + H(B + t(X \cup Y_0)) - H(B + tY_0)$$

is indeed satisfied. Hence we can divide by t^d, pass to the *lim sup* as $t \to \infty$, and then take the supremum as the bounded set B becomes arbitrarily large. Since the last term has a negative sign, we must use Theorem 2 to complete the argument. The proof of (3.6) is similar. ∎

It would be interesting to know whether these inequalities (3.6) and (3.7) are true also in the topological case, and if they are true without the special hypothesis on $X \cap Y$ or Y_0 in the measure-theoretic case.

For the extreme case, $d = n$, the construction of h_d simplifies somewhat. Since there is only one n-plane, namely V itself, the n-dimensional entropy can be described by a single real number $\mathcal{H}_n(V)$. In this case we will always adopt the convention that our Euclidean metric is chosen so that a fundamental domain for the lattice L has unit n-dimensional volume, or equivalently so that the quotient torus V/L satisfies

$$\text{volume}(V/L) = 1. \tag{3.8}$$

More explicitly, if $\mathbf{b}_1, \ldots, \mathbf{b}_n$ *is any basis for the lattice* L*, we assume that the parallelepiped* $[0,1]\mathbf{b}_1 + \cdots + [0,1]\mathbf{b}_n$ *has Euclidean volume equal to one, so that* $\mathcal{H}_n(V)$ *coincides with the directional entropy* $h_n(\mathbf{b}_1 \wedge \ldots \wedge \mathbf{b}_n)$. If we identify the lattice L with $\mathbf{Z}^n$, then in order to compute the number $\mathcal{H}_n(V)$ it suffices to consider large "discrete cubes" of the form $S_m = \{1, \ldots, m\} \times \cdots \times \{1, \ldots, m\}$ with $|S_m| = m^n$.

Lemma 2. *The ratio* $H(S_m)/|S_m|$ *converges to a limit* $\mathcal{H}_n$ *as* $m \to \infty$*. Furthermore* $H(S_m)/|S_m| \geq \mathcal{H}_n$ *for every integer* $m \geq 1$*. If the Euclidean metric is normalized so that (3.8) is satisfied, then this limit* $\mathcal{H}_n$ *can be identified with the ratio* $\mathcal{H}_n(V)$ *of Theorem 2.*

We will describe $\mathcal{H}_n$ as the average "*information per lattice point*" associated with the information function H. Note that this construction is only of interest when H has growth degree d equal to n. For if $d < n$, then the n-dimensional entropy h_n must be identically zero.

The proof of Lemma 2 can be sketched as follows. Fixing m, note that an arbitrarily large discrete cube S_N can be covered by $(1 + N/m)^n$ copies of S_m. Therefore

$$H(S_N) \leq (1 + N/m)^n H(S_m),$$

or in other words $H(S_N)/|S_N| \leq H(S_m)/|S_m| + \epsilon$, where it is not difficult to check that $\epsilon \to 0$ as $N \to \infty$ with m fixed. It follows that

$$\limsup H(S_N)/|S_N| \leq \inf H(S_m)/|S_m|.$$

But the inequality in the other direction is trivial, so equality must hold. ∎

To conclude this section, we must prove Theorem 2. Let us embed the compact polyhedral set $X \subset P$ in some large cube $I^d \subset P$ of edge length e, and express I^d as a union of N^d subcubes of edge length e/N, with N very large. Since the entropy $\eta_d(I^d)$ is precisely equal to the sum of the entropies of these subcubes by (3.4), it follows from subadditivity that the entropy of any partial union of subcubes is strictly proportional to their number. Since the proportion of subcubes which meet the boundary of X tends to zero as $N \to \infty$, the conclusion that

$$\eta_d(X)/\eta_d(I^d) = \text{volume}(X)/\text{volume}(I^d)$$

follows easily. More generally, this same argument works for any compact subset of P whose boundary has d-dimensional Lebesgue measure equal to zero. It may be conjectured that Theorem 2 remains true for arbitrary compact subsets of a d-plane.

To prove the second half of Theorem 2, first consider the statement for the cube I^d. An argument similar to the proof of Lemma 2 shows that

$$\eta_d(I^d) = \sup_B \inf_t H(B + tI^d)/t^d.$$

Now let Y be the closure of the complement of X in I^d, so that the volume of I^d is equal to the sum of the volumes of X and Y. If we divide the inequality

$$H(B+tX) \geq H(B+tI^d) - H(B+tY),$$

by t^d, take the lim inf as $t \to \infty$, and then the supremum over bounded sets B, we obtain the required inequality. ∎

4. Spatial entropies of CA-maps

We begin the discussion of entropies associated with CA-maps by discussing the "information per lattice point" $\mathcal{H}_n$ at some constant time. (Compare Lemma 2.) First consider the image $f(K^L)$ of an n-dimensional CA-map.

Lemma 3. *The topological entropy $\mathcal{H}_n = \mathcal{H}_n(f(K^L))$ satisfies $0 \leq \mathcal{H}_n \leq \log k$. Here the maximum value of $\log k$ is attained if and only if the map f is surjective, $f(K^L) = K^L$, and the minimum value of zero is attained if and only if $f(K^L)$ consists of a single constant configuration.*

Proof. If f is not surjective, then there must exist some partial configuration $\alpha : S \to K$ which does not extend to any configuration in $f(K^L)$. Choose some large discrete cube $S' \subset L$ which contains this set S. Then $H(f(K^L); S') < |S'| \log(k)$. But it follows from Lemma 2 that $H(f(K^L); S')/|S'|$ is an upper bound for $\mathcal{H}_n(f(K^L))$, which must therefore be strictly less than $\log(k)$.

Now suppose that f is not a constant map. The value of the configuration $f(\mathbf{a})$ at the origin depends on the values of $\mathbf{a}$ in some suitable discrete cube S. Tiling L by copies of S, we see easily that $\mathcal{H}_n(f(K^L)) > \log(2)/|S| > 0$. ∎

If $f : K^L \to K^L$ is any n-dimensional CA-map, then the sequence of closed translation-invariant sets

$$K^L \supset f(K^L) \supset f^2(K^L) \supset \cdots$$

with intersection Λ gives rise to a sequence of n-dimensional entropies

$$\log(k) \geq \mathcal{H}_n(f(K^L)) \geq \mathcal{H}_n(f^2(K^L)) \geq \cdots,$$

which clearly converges to the limit $\mathcal{H}_n(\Lambda) \geq 0$. This limit, which will play a central role in our discussions, will be called the "spatial" entropy of the limit set Λ, since it describes the distribution of information at some fixed time $t =$ constant. (In other words, it involves only the spatial lattice L and not the space-time lattice $\mathbf{Z} \times L$.) This spatial entropy is a very crude invariant, since it will not change if we replace f by some iterate $f \circ \cdots \circ f$, or compose it with an arbitrary lattice translation. It follows from Lemma 3 that $\mathcal{H}_n(\Lambda) = \log(k)$ if and only if f is surjective. However, we will see that $\mathcal{H}_n(\Lambda)$ may well be zero even if the limit set Λ is infinite.

Remarks on computability. Unfortunately, it must be admitted that this basic invariant $\mathcal{H}_n(\Lambda)$ tends to be very difficult to compute, even in the one-dimensional case. According to [7], there exists a one-dimensional CA-map $f : K^{\mathbf{Z}} \to K^{\mathbf{Z}}$ for which the collection of all finite substrings of bisequences from $\Lambda(f)$ is not recursively enumerable. *This means that there can be no effective procedure for computing the integers* $\log^{-1} H_\Lambda(S)$. If we work with some finitely iterated image $f^t(K^L)$, then in principle any $H(f^t(K^L); S)$ can be computed in finite time. However, the length of this computation tends to increase exponentially with the size of S (and with t^n), so that the $\mathcal{H}_n(f^t(K^L))$ may be difficult to compute, and $\mathcal{H}_n(\Lambda)$ may be impossible. We have seen that $\mathcal{H}_n(f^t(K^L))$ is zero only if the t-fold image $f^t(K^L)$ consists of a single constant configuration; however a decision as to whether or not $\mathcal{H}_n(\Lambda) = 0$ seems very difficult. In the one-dimensional case, it is at least possible to decide effectively whether or not any given CA-map is surjective, so that $\mathcal{H}_n(\Lambda) = \log(k)$. (See for example [1].) However, in higher dimensions it seems unlikely that any algorithm exists for deciding in general whether f is surjective.

If we are given a translation-invariant measure μ_0 on the limit set Λ, then the numbers $H(\mu_0; S)$ can be effectively computed, but again it seems very difficult to decide whether or not the entropy $\mathcal{H}_n(\mu_0)$ is zero. Furthermore, in saying that μ_0 is "given" we have avoided the major difficulty of effectively constructing interesting invariant measures on Λ.

A basic theorem of Dinaberg and Goodman asserts that the topological entropy of any map on a compact metric space is the supremum of the measure-theoretic entropies of its invariant measures. (Compare [18].) In the case $n = 1$, we can apply this result to the shift map on Λ, and conclude that $\mathcal{H}_n(\Lambda)$ is equal to the supremum of $\mathcal{H}_n(\mu_0)$ as μ_0 varies over all translation-invariant measures supported by Λ. I don't know whether the corresponding statement is true for $n > 1$. Even in the one-dimensional case, if we insist that our measures μ_0 be not only translation-invariant but also f-invariant, then I don't know whether the supremum of $\mathcal{H}_n(\mu_0)$ is still $\mathcal{H}_n(\Lambda)$. (Compare §6.3.) Furthermore, I don't know whether this supremum is actually realized by some particular μ_0.

One very special case is well understood. **Definition:** By the *standard measure* σ on K^L is meant the measure for which each partial configuration $\alpha : S \to K$ has probability

$$p_\alpha = \sigma(E_\alpha) = 1/k^{|S|}. \tag{4.1}$$

Evidently this is the unique translation-invariant measure on K^L with maximal spatial entropy $\mathcal{H}_n(\sigma) = \log k$. A theorem which was proved by Blankenship and Rothaus in the one-dimensional case and by Maruoka and Kimura in the higher-dimensional case asserts that this standard measure is f-invariant if and only if f is surjective. (See [6,12,1].) *Thus, in the surjective case, there is a unique measure which is both f-invariant and translation-invariant, and which has maximal spatial entropy, equal to* $\log k$.

5. Causality

In this section we begin to study the dynamics of an n-dimensional CA-map $f : K^L \to K^L$. The presentation will be based on the following idea, suggested by Lind and Smillie. Recall that the "space-time" lattice $\mathbf{Z} \times L$ is embedded in a vector space $\mathbf{R} \times V$, and that $\mathcal{A} \subset K^{\mathbf{Z}\times L}$ is the $(n+1)$-dimensional subshift consisting of all complete histories for f.

Definition. By a *causal cone* for the subshift $\mathcal{A} \subset K^{\mathbf{Z}\times L}$ will be meant any convex polyhedral cone $C \subset \mathbf{R} \times V$ which is spanned by some finite set $S \subset \mathbf{Z} \times L$ having the following two properties:

1. The value of a complete history $\mathbf{a} \in \mathcal{A}$ at the origin must be uniquely determined by its values at the points of S. Hence the value of $\mathbf{a}$ at any lattice point (t, ℓ) is determined by its values at the points of $(t, \ell) + S$.

2. S must lie strictly to one side of some hyperplane through the origin. Thus, if $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$, then C consists of all linear combinations $t_1\mathbf{v}_1 + \cdots + t_r\mathbf{v}_r$ with $t_i \geq 0$; and no such linear combination can be equal to the zero vector unless $t_1 = \cdots = t_r = 0$.

Every CA-map has at least one causal cone. For if $f(\mathbf{a}_t) = \mathbf{a}_{t+1}$ is given by

$$\mathbf{a}(t+1, \ell) = \Phi(\mathbf{a}(t, \ell + v_1), \ldots, \mathbf{a}(t, \ell + v_r)),$$

as in (2.7), then evidently the value of $\mathbf{a}$ at the origin is determined by its values at the points $\mathbf{v}_1 = (-1, v_1), \ldots, \mathbf{v}_r = (-1, v_r)$, which lie in the open half-space $t < 0$.

Let us assume that the function Φ really depends on every one of its r arguments, so that this collection of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is uniquely determined.

Definition. With this hypothesis, the causal cone spanned by the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_r$ in the half-space $t < 0$ is called the *1-step backward cone* or *past cone* $-F_1$ for the CA-map f. The opposite cone F_1, spanned by the vectors $-\mathbf{v}_1, \ldots, -\mathbf{v}_r$, is called the *1-step forward cone* or *future cone.*

Note. In the trivial case of a constant function f, we adopt the convention that $r = 0$, and that F_1 consists only of the zero vector. In practice, since this case is completely uninteresting, it will often be excluded.

Note also that F_1 can be characterized as the smallest convex cone in $\mathbf{R} \times V$ with the following property: If two initial configurations $\ell \mapsto \mathbf{a}(0, \ell)$ and $\ell \mapsto \mathbf{b}(0, \ell)$ on the spatial lattice $0 \times L$ are identical except at the origin, and if we define $\mathbf{a}(t, \ell)$ and $\mathbf{b}(t, \ell)$ for $t = 1, 2, 3, \cdots$ by (2.7), then the two resulting configurations on the half-space $t \geq 0$ must coincide everywhere outside of the cone F_1.

More generally, let m be any positive integer. If we replace the CA-map f by its m-fold composition with itself, then we are led to the following construction. The *m-step forward cone* is defined to be the smallest convex cone $F_m \subset \mathbf{R} \times V$ such that, if two initial configurations $\mathbf{a}(0, \ell)$ and $\mathbf{b}(0, \ell)$ coincide except at the origin, then the corresponding configurations $\mathbf{a}(m, \ell)$ and $\mathbf{b}(m, \ell)$ at time m must coincide everywhere on $m \times L$ outside of the intersection $F_m \cap (m \times L)$.

It follows from this definition that $-F_m$ is a causal cone satisfying $F_1 \supset F_m$, and more generally $F_m \supset F_{mp}$. Thus, we are led to define the *limiting forward cone* F to be the intersection of the nested sequence

$$F_1 \supset F_2 \supset F_6 \supset F_{24} \supset \cdots .$$

(In fact, it is not difficult to check that the successive F_m actually converge to this limit cone F in the sense that for any open neighborhood $U \supset F$, all but finitely many of the F_m, intersected with the unit ball, lie inside U.) This intersection F is clearly a closed convex cone in the $(n+1)$-dimensional spacetime $\mathbf{R} \times V$. (I don't know whether the negative $-F$ is necessarily a causal cone, or even a polyhedral cone.)

One elementary but useful invariant is the *dimension* of this forward cone, where $0 \leq \dim(F) \leq n+1$. In the case of a one-dimensional CA-map, the possible dimensions 0, 1, and 2 correspond very roughly to Wolfram's cellular automata of Class 1, 2, or 3. Thus if $\dim(F) = 0$, then some iterated image $f^t(K^{\mathbf{Z}})$ consists of a single constant configuration, and there is nothing to study. If $\dim(F) = 1$, then the cone is a half-line consisting of pairs (t, x) with $t \geq 0$ and $x = ct$. Intuitively, the constant c measures the speed at which information travels. (Compare §6.) Finally, if $\dim(F) = 2$, then F consists of all (t, x) with $t \geq 0$ and $c_1 t \leq x \leq c_2 t$, so that in this case F is described by an entire interval $[c_1, c_2]$ of speeds. In the n-dimensional case, the ratios $\mathbf{x}/t$ with $(t, \mathbf{x}) \in F$ range over some compact convex subset of n-space.

We will say that a vector $\mathbf{v} \in \mathbf{R} \times V$ is *timelike*, or belongs to the *timelike cone* $F \cup (-F)$ if either $\mathbf{v} \in F$ or $-\mathbf{v} \in F$. The remaining vectors, which do not belong to $F \cup (-F)$ are said to be *spacelike*. Note that non-zero vectors in $0 \times V$ are necessarily spacelike. However, depending on the particular CA-map being studied, vectors along the time-axis $\mathbf{R} \times 0$ may be either timelike or spacelike. In the one-dimensional case, the open set of spacelike vectors splits into two components, hence we can further distinguish between "right spacelike" and "left spacelike" vectors.

Let us fix some causal cone C (for example the m-step backward cone $-F_m$).

Definition. By the *umbra* of a compact set $X \subset \mathbf{R} \times V$ with respect to C will be meant the closed set consisting of all points $\mathbf{x} \in \mathbf{R} \times V$ which are totally shadowed by X in the following sense. Any piecewise smooth path $\tau \mapsto \mathbf{p}(\tau)$ which starts at the point $\mathbf{p}(0) = \mathbf{x}$ and has velocity vectors $d\mathbf{p}(\tau)/d\tau$ in the cone C must eventually meet the set X. (Compare figure 1.)

Let B be a ball, centered at the origin, which is large enough so that the value of a complete history $\mathbf{a}$ at the origin is uniquely determined by its values at the other lattice points in $B \cap C$.

Lemma 4. *With C and B as above, for an arbitrary compact X, the values of a complete history $\mathbf{a} \in \mathcal{A}$ at all lattice points $\mathbf{x} = (t, \ell)$ in the C-umbra*

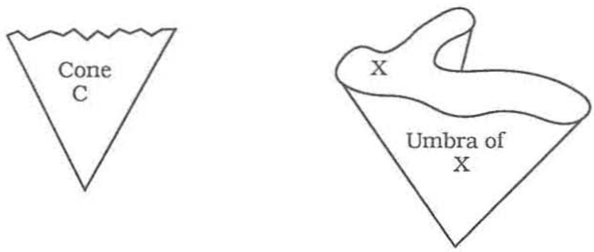

Figure 1.

of X are uniquely determined by the values of $\mathbf{a}$ *at the lattice points in the neighborhood* $X + B$.

Proof. Since $C - \{0\}$ lies in an open half-space, we can choose a linear function $\tau : \mathbf{R} \times V \to \mathbf{R}$ which is strictly negative on $C - \{0\}$. After approximating by a linear map with rational coefficients and clearing denominators, we may assume that τ takes integer values on the lattice $\mathbf{Z} \times L$. The proof will be by induction on the integer $\tau(\mathbf{x})$. (In the special case where C is the backward cone $-F_m$, we can of course simply take $\tau = t$.) To begin the induction, the lemma is trivially true for $\tau(\mathbf{x})$ near $-\infty$, since no such point can lie in the umbra of X. For a given lattice point $\mathbf{x}$ in this umbra there are two possibilities. If each lattice point $\mathbf{x}' \in \mathbf{x} + (B \cap C)$ also lies in the C-umbra of X, then by induction we can compute the value of $\mathbf{a}$ at these points, and hence at $\mathbf{x}$. On the other hand, if one of these points $\mathbf{x}'$ is not in the umbra of X, then the line segment from $\mathbf{x}$ to $\mathbf{x}'$ must cross X, and therefore $\mathbf{x}' \in X + B$. ∎

As an immediate consequence of Lemma 4, we have the following.

Lemma 5. *The information function $H = H_{\mathcal{A}}$ associated with the subshift of complete histories for an n-dimensional cellular automaton has growth degree at most n. Hence the corresponding n-dimensional entropy function η_n is defined and satisfies $0 \le \eta_n(X) \le c\,\mathrm{diam}(X)^n$ for any compact $X \subset \mathbf{R} \times V$.*

Proof. Apply Lemma 4 to the boundary of a large cube I^{n+1}. Clearly the entire cube is in the umbra of its boundary ∂I^{n+1}. Since the number of lattice points in $\partial I^{n+1} + B$ grows like the n-th power of the edge length, it follows that $H(I^{n+1})$ is less than some polynomial function of degree n in the edge length. ∎

It follows that the same statement is true for any subshift of $\mathcal{A}$, or for any translation-invariant measure supported by $\mathcal{A}$. More generally, the same argument applies to *any* subshift or translation-invariant measure which possesses a causal cone.

Here is another consequence of Lemma 4.

Theorem 3. *Let C be a causal cone for an n-dimensional cellular automaton, and let X and Y be compact subsets of $\mathbf{R} \times V$ with Y contained in the C-umbra of X. Then $\eta_n(Y) \leq \eta_n(X)$.*

In particular, if the umbra of X is compact, it follows that $\eta_n(\text{Umbra}(X)) = \eta_n(X)$. The proof is straightforward. ∎

Corollary 3. *If the spatial entropy $\mathcal{H}_n(\Lambda)$ of §4 is zero, then the n-dimensional entropy $\eta_n(X)$ is zero for every compact set $X \subset \mathbf{R} \times V$. More generally, if $\mathcal{H}_n(P) = 0$ for any n-plane which intersects a causal cone C only at the origin, then η_n is identically zero.*

Proof. The hyperplane P cuts the space $\mathbf{R} \times V$ into two half-spaces, one of which contains the cone C. Let $\mathbf{x}$ be any point in the interior of the other half-space $\mathbf{H}$. Then the translated cone $\mathbf{x} + C$ intersects $\mathbf{H}$ in a compact region with non-vacuous interior which lies in the C-umbra of its intersection with P. The hypothesis that $\mathcal{H}_n(P) = 0$ implies that the entropy of this region is zero. It follows easily that the entropy of any compact set is zero. ∎

6. Examples

This section will describe some explicit examples of one-dimensional cellular automata.

Example 6.1. The simplest block maps. Suppose that our alphabet has only two elements, $K = \{0, 1\}$, and that $f(\mathbf{a}) = \mathbf{a}'$ with

$$\mathbf{a}'(\ell) = \Phi(\mathbf{a}(\ell - 1), \mathbf{a}(\ell)), \tag{6.1}$$

so that our block map Φ is a function of just two variables. Then there are sixteen different possibilities for f, corresponding to the sixteen distinct maps from $K \times K$ to K. These sixteen CA-maps can be divided into six groups, as indicated in the following table.

Block map		$\mathcal{H}_1(\Lambda)/\log 2$	**Speed**	**Class**
constant	(2 cases)	0	—	1
Max or Min	(2 cases)	0	[0, 1]	1
1 - Max or 1 - Min	(2 cases)	.81137	1/2	2
i+j+constant (mod 2)	(2 cases)	1	[0, 1]	3
function of one variable	(4 cases)	1	0 or 1	2
other non-symmetric	(4 cases)	.69424	0 or 1	2

Here the "speed" is the set of ratios x/t in the forward cone F, and can be described intuitively as the rate at which information is moved to the right as we apply f. (Compare §5.) The "class", loosely following Wolfram [19], can be defined as 1 if the entropy is zero, 2 if information remains localized, traveling with constant speed, and 3 if information spreads out linearly with time, so as to affect a larger and larger region of the lattice. (Compare §7. We will not attempt to discuss Wolfram's Class 4, which is beyond the scope of this paper.)

Suppose for example that $\Phi(i,j) = \mathrm{Max}(i,j)$. Then the image $f(K^{\mathbf{Z}})$ consists of all bisequences which contain no string of the form 010 . Similarly, $f^2(K^{\mathbf{Z}})$ consists of all bisequences which contain no 010 or 0110 , and so on. The entropies $\mathcal{H}_1(f^m(K^{\mathbf{Z}}))/\log 2$ with $m \geq 0$ form a sequence

$$1 > .811 > .694 > .613 > .551 > .504 > .465 > .433 > \cdots$$

which converges very slowly to zero. The limit set Λ can be described as follows. A finite string of 0's and 1's can be embedded in a bisequence belonging to Λ if and only if it has the form $1^*\ 0^*\ 1^*$, that is a string of 1's followed by 0's followed by 1's (where any of these three substrings may be empty). Since the number of such strings of length m grows only quadratically with m, it follows that $\mathcal{H}_1(\Lambda)$, the rate of exponential growth, is zero.

Next consider the block map $\Phi(i,j) = 1-\mathrm{Max}(i,j)$. In this case, it follows immediately from the discussion above that the image $f(K^{\mathbf{Z}})$ consists of all bisequences which contain no isolated zeros. Using a standard procedure for computing the entropy of a subshift of finite type, we see that $\mathcal{H}_1(f(K^{\mathbf{Z}}))$ is equal to $\log(\omega)$ where $\omega = 1.7548776662$ is the real root of the equation $x^3 - 2x^2 + x - 1 = 0$ (see [9].) But in this case, inspection shows that the second iterate f^2 acts on this image by a single shift to the right. For any string of the form $*1*$ is replaced by something of the form $*00$ under f, and hence by $**1$ under f^2, and similarly each string of the form $00*$ or $*00$ maps to $**0$ under f^2. It follows that the limit set Λ is equal to $f(K^{\mathbf{Z}})$.

The remaining cases are all straightforward. For example in the last row of the table the limit set is equal to $f(K^{\mathbf{Z}})$, and consists either of all bisequences with no double 1's or all bisequences with no double 0's, with entropy equal to $\log((1+\sqrt{5})/2)$.

In the four cases which are described as of Class 1, the entropy function $X \mapsto \eta_1(X)$ is of course zero. In the ten Class 2 examples, it is not difficult to show that $\eta_1(X)$, for polyhedral X, is just the length of the image of X under a suitable projection from $\mathbf{R} \times V$ to the real line. The two Class 3 examples can be analyzed as follows.

Example 6.2. Left and right permutive block maps. Here is a larger class of examples for which it is possible to almost completely compute the function η_1. Let f be the one-dimensional CA-map which is associated with the block map

$$\mathbf{a}'(\ell) = \Phi(\mathbf{a}(\ell+p), \mathbf{a}(\ell+p+1), \ldots, \mathbf{a}(\ell+q)), \tag{6.2}$$

with $p \leq q$. By definition, such a map is *left permutive* if for every fixed $\mathbf{a}(\ell+p+1), \ldots, \mathbf{a}(\ell+q)$ the correspondence

$$\mathbf{a}(\ell+p) \mapsto \Phi(\mathbf{a}(\ell+p), \mathbf{a}(\ell+p+1), \ldots, \mathbf{a}(\ell+q))$$

is a permutation of K. Every left permutive CA-map is surjective, $f(K^L) = K^L$, and hence has maximal spatial entropy $\mathcal{H}_1(\Lambda) = \log k$. *Right permutivity* is defined similarly. As an example, the block maps $(i,j) \mapsto i+j+\mathrm{constant}(\mathrm{mod} 2)$ of Example 6.1 are both left and right permutive. If Φ

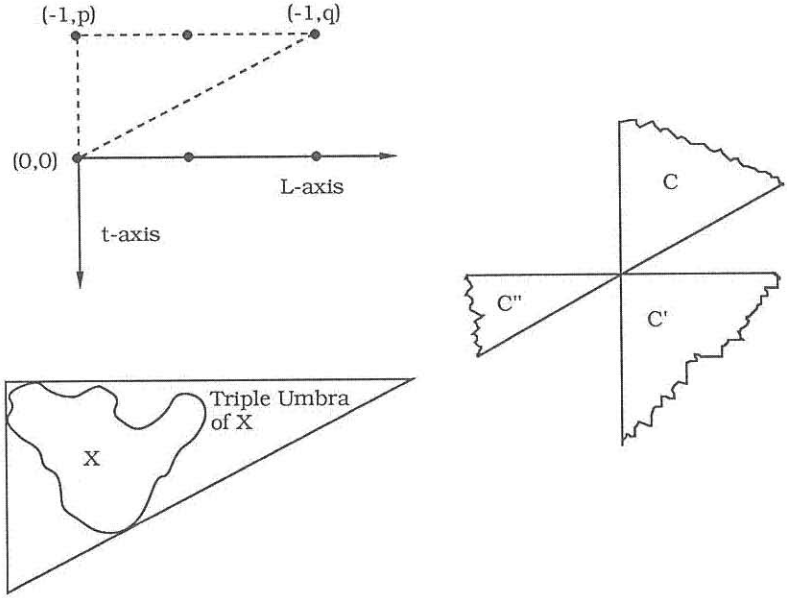

Figure 2.

is left permutive, then inspection shows that its space $\mathcal{A}$ of complete histories admits not only the usual causal cone $C = -F_1$, but also an additional causal cone C', as shown in figure 2. Similarly, if Φ is right permutive, then there is an additional causal cone C''. Thus if Φ is both left and right permutive, then any compact $X \subset \mathbf{R} \times V$ has three distinct umbras. A straightforward extension of Theorem 3 shows that the union $\tilde{X}$ of these three umbras satisfies $\eta_1(\tilde{X}) = \eta_1(X)$. But if X is connected, then it is not hard to see that $\tilde{X}$ is a triangle T, as illustrated in figure 2. Using (3.4), we see that $\eta_1(X)$ is proportional to the edge length of this triangle $T = \tilde{X}$.

As an example, let us compute the entropy of the reflected triangle $-T$. Evidently, $\widetilde{(-T)}$ has twice the edge length of T, so $\eta_1(-T) = 2\eta_1(T)$. As another example, note that the directional entropy $h_1(\mathbf{v}) = \eta_1([0,1]\mathbf{v})$ is given by

$$h_1(t,x)/h_1(0,1) = \mathrm{Max}(|x+pt|, |x+qt|, (q-p)|t|).$$

This same formula is true whether we are working with topological or measure-theoretic entropy.

We will next describe two examples, due to Lind and Smillie, to show that the topological entropy $\hat{h}(\mathbf{v})$ need not depend continuously on $\mathbf{v}$, and to show that topological entropy need not be the supremum of measure-theoretic entropies if we insist on using shift-invariant measures.

Example 6.3. Gliders and Walls. Let $n = 1$, and let the alphabet $K = \{o, w, r, l\}$ consist of the following four elements: There is a neutral background state o, and a stationary particle w which never moves and which forms an absorbing wall. Finally, there is a right-moving particle r and a left-moving particle l which travel with unit speed until they are absorbed by a wall or until two of them come together. A right and left particle which approach each other and arrive at two sites which have distance two [or one] are transformed, at the next time step, into a wall [or a pair of contiguous walls]. To be more precise, the associated block map $\Phi : K^3 \to K$ maps triples of the form $*w*$ or $rl*$ or $r*l$ or $*rl$ to w. It maps all other triples or the form $r**$ to r, and all other triples of the form $**l$ to l. Finally, any triple not covered by these descriptions is mapped to o. We will show that the topological directional entropy of the resulting space $\mathcal{A}$ is given by

$$h_1(t, x) = (|t| + |x|)\log(2) \quad \text{for} \quad x \neq 0, \quad \text{but} \tag{6.3}$$

$$h_1(t, 0) = 2|t|\log(2). \tag{6.4}$$

Thus entropy is discontinuous as $x \to 0$. In fact the discontinuity as $x \to 0$ is precisely the largest possible one. (Compare (7.5).)

Remark. According to Lind and Smillie, the directional entropy is always continuous in spacelike directions. (See the Remark following Lemma 6.) I don't know whether measure-theoretic directional entropy is always continuous. In the one-dimensional case, Sinai proves that it converges to a well-defined limit as we approximate any irrational direction by rational directions.

To prove (6.3) and (6.4), we start by describing the space $\mathcal{A}$ of complete histories. Given $\mathbf{a} \in \mathcal{A}$, suppose to fix our ideas that $\mathbf{a}$ involves particles of all three kinds. Then it is not difficult to check that the "wall" particles are concentrated within a vertical strip, with all right moving particles to the left of it and all left moving particles to the right of it. From this information, we can compute the topological entropy $\eta_1(\mathcal{A}, X)$ for any compact polyhedral X as follows. Let S_0 be an arbitrary closed vertical strip, and let S_{-1} and S_1 be the closed half-planes to the left and right of S_0. For v equal to ± 1 or zero, let X_v be the image of $X \cap S_v$ under the projection $(t, x) \mapsto x - vt$. hen $\eta_1(X)/\log 2$ is equal to the supremum, over all such strips S_0, of the sum $\text{length}(X_{-1}) + \text{length}(X_0) + \text{length}(X_1)$. The proof is not difficult. Applying this recipe to the case of a closed interval we obtain the required formulas (6.3) and (6.4). ■

Now let us look at measure-theoretic entropies. If μ is any shift invariant measure on the limit set $\Lambda \subset K^L$ then we claim that μ can be written as a linear combination of three measures μ_{-1}, μ_0, and μ_1 with the property that each μ_v admits only particles with speed v, with probability one. First note that, with μ-probability one, it is impossible for any two of the particles r, l and w to exist in the same bisequence. To see this, let s be any finite string in the given alphabet, of length m, which occurs with μ-probability greater than $1/p$. Then it is not difficult to show that any string of length greater than pm must contain two disjoint copies of s with positive μ-probability. Thus, if s contains say an r and an l, then we obtain a contradiction since we cannot have any l to the left of an r in any bisequence belonging to Λ. (I am indebted to Hurd for this argument.) Now if τ_v is the probability that a bisequence $\mathbf{a} \in \Lambda$ contains at least one particle with speed v, then it follows easily that μ splits as

$$\mu = \tau_{-1}\mu_{-1} + \tau_0\mu_0 + \tau_1\mu_1 + (1 - \tau_{-1} - \tau_0 - \tau_1)\mu',$$

where μ' admits only the background state o, with probability one.

Every such measure μ extends to a measure supported by $\mathcal{A}$. (Lemma 1.) It follows that the corresponding directional entropy has the form

$$h_1(\mu;\ (t, x)) = \alpha_{-1}|x - t| + \alpha_0|x| + \alpha_1|x + t|,$$

where the coefficients $\alpha_v \geq 0$ satisfy $\alpha_{-1} + \alpha_0 + \alpha_1 \leq \log 2$. In particular, the measure-theoretic entropy in the timelike direction $(1, 0)$ is at most equal to $\log 2$. *Note that this is only half of the topological entropy in this direction, as given by* (6.4). (Of course, if we are willing to allow f-invariant measures which are not shift invariant, then we can achieve $2\log 2$ as a measure-theoretic entropy.)

Example 6.4. Just gliders. If we simplify this example so that two gliders simply annihilate each other when they meet, and there are no walls, then the analysis becomes more difficult. Again any shift invariant measure on the limit set Λ has the property that r's and l's cannot coexist, so measure theoretic entropy has the form $h_1(t, x) = \alpha_{-1}|x - t| + \alpha_1|x + t|$. But, in the case of topological entropy in a timelike direction, there is the additional possibility that left and right moving particles exist in carefully balanced proportions so as to annihilate each other precisely within a long narrow strip. The conclusion is that

$$h_1(t, x) = (|t| + |x|)\log(2) \qquad \text{for} \qquad |x|/|t| \geq .546\ldots,$$

but

$$h_1(t, x) = |t + x|\log(|2t/(t + x)|) \;+\; |t - x|\log(|2t/(t - x)|)$$

otherwise. (Compare [17].) In the later case, topological entropy is strictly larger than any (shift-invariant) measure-theoretic entropy. In this example, topological entropy is continuous, but not subadditive or piecewise-linear.

Example 6.5. Reflecting walls. First consider an invertible cellular automaton with alphabet $K = \{o, r, l, b\}$ consisting of a neutral background state and two particles r and l which travel in opposite directions with unit speed. When these two particles meet, they travel through each other without interacting. Here the symbol b stands for the state in which two opposite particles land on the same lattice point at the same time. The directional entropy, either topological or measure-theoretic based on the standard measure σ of (4.1), is easily computed as

$$h_1(t,x)/\log 2 = |t+x| + |t-x|. \tag{6.5}$$

(Thus entropy is non-zero in all directions, even though this example is completely trivial from a computational point of view.) Now suppose that we add a fifth element w which is a stationary reflecting wall. Thus a right moving particle which reaches the lattice point to the left of w gets converted on the next move to a left moving particle at the same location. For this enlarged system, it is not difficult to check that the measure-theoretic entropy, using the $(k=5)$ standard measure, is

$$h_1(t,x) = |x| \log 5.$$

Thus entropy has increased along the x-axis, as compared with (6.5), *but has decreased to zero along the the t-axis.* In particular, the standard measure is not the measure of greatest entropy in the t-direction.

More generally, if we assign probability p to the alphabet element w and probabilities $(1-p)/4$ to the other four elements, the values at different points of the lattice L being independent random variables, then one can check that the measure-theoretic entropy $h_1(1,0)$ is zero whenever $p > 0$, but is $\log 4$ when $p = 0$. *Thus measure-theoretic entropy does not depend continuously on the measure.*

Quite similar behavior occurs for the $k = 2$ left permutive CA-map $\mathbf{a} \mapsto \mathbf{a}'$ given by

$$\mathbf{a}'(\ell) \equiv \mathbf{a}(\ell) + (1 - \mathbf{a}(\ell+1))\mathbf{a}(\ell+2) \pmod 2, \tag{6.6}$$

as studied by Coven. If we consider only histories for which all odd lattice sites are off, $\mathbf{a}(2\ell+1) = 0$, then this reduces to the additive rule

$$\mathbf{a}'(2\ell) \equiv \mathbf{a}(2\ell) + \mathbf{a}(2\ell+2) \pmod 2.$$

Hence topological entropy is non-zero in all directions. Yet for σ-almost-every initial configuration we have infinitely many odd and even sites which are on, and the behavior is rather different. Fixing some bisequence $\mathbf{a}$, let us say that the lattice point ℓ is "distinguished" if $\mathbf{a}(\ell) = 1$, and if the next lattice point to the right which is on, $\mathbf{a}(\ell') = 1$, has odd distance, $\ell' - \ell \equiv 1(\mathrm{mod} 2)$. Then it is not hard to check that every distinguished site will remain on forever, and that in the interval between two distinguished sites the configuration will repeat periodically with period equal to the largest power of 2 which is less than or equal to the interval length. Thus, as far as the measure-theoretic entropy $\eta_1(\sigma; X)$ is concerned, we are dealing with a CA-map of Class 2.

7. Directional entropy and the dual cone

Now let us look at the n-dimensional entropies associated with an n-dimensional CA-map. We will work either with the topological entropy associated with the space $\mathcal{A} \subset K^{Z \times L}$ of complete histories, or with the measure-theoretic entropy associated with some translation invariant measure μ supported by $\mathcal{A}$. Thus, to each n-dimensional hyperplane $P \subset \mathbf{R} \times V$ there is associated an entropy $\mathcal{H}_n(P)$, which measures information per unit n-dimensional volume. If $\mathbf{u}$ is a unit vector orthogonal to P, then it will be convenient to use the notation $\hat{h}(\mathbf{u}) = \mathcal{H}_n(P)$. More generally, if $\mathbf{p}$ is any non-zero vector orthogonal to P, we will set

$$\hat{h}(\mathbf{p}) = \|\mathbf{p}\|\mathcal{H}_n(P) = \|\mathbf{p}\|\mathcal{H}_n(\mathbf{p}^\perp), \tag{7.1}$$

where $\|\mathbf{p}\|$ is the Euclidean norm. In the case $n = 1$, this amounts to setting $\hat{h}(\mathbf{p}) = h_1(J\mathbf{p})$, where J is a 90° rotation of the plane. For $n > 1$ it amounts to setting $\hat{h}(\mathbf{v}_1 \times \cdots \times \mathbf{v}_n) = h_n(\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_n)$, using a suitably defined n-fold cross product operation.

Now let C be any causal cone for the space $\mathcal{A}$. By definition, the *dual cone* $\hat{C}$ consists of all vectors $\mathbf{p}$ with the property that $\mathbf{p} \cdot \mathbf{v} \le 0$ for all vectors $\mathbf{v} \in C$. The geometric meaning of this construction is the following. A vector $\mathbf{p}$ belongs to the interior of $\hat{C}$, or of $-\hat{C}$, if and only if every non-zero vector in the orthogonal complement $\mathbf{p}^\perp$ is spacelike. In the trivial case $\dim(C) = 0$, note that $\hat{C}$ is the entire space $\mathbf{R} \times V$. In the case $\dim(C) = 1$, clearly $\hat{C}$ is an entire closed half-space. In all other cases, $\hat{C}$ is strictly smaller than a half-space.

In the special case $C = -F$ there is one unit vector $\mathbf{u}_0 \in -\hat{F}$ of particular interest, namely the unit normal vector to the spatial hyperplane $0 \times V$. Note that $\mathbf{u}_0$ always lies in the interior of $-\hat{F}$. By definition, $\hat{h}(\mathbf{u}_0)$ coincides with the spatial entropy $\mathcal{H}_n(0 \times V)$, as studied in §4.

Clearly the n-dimensional entropy $\hat{h}(\mathbf{u})$ is always defined, and is bounded on unit vectors. In other words there is a finite constant c so that

$$0 \le \hat{h}(\mathbf{p}) \le c\|\mathbf{p}\| \tag{7.2}$$

for all vectors $\mathbf{p}$. If the spatial entropy $\hat{h}(\mathbf{u}_0)$ is zero (or more generally if $\hat{h}$ vanishes anywhere in the interior of any $\hat{C}$), then it follows from Corollary 3 that the n-dimensional entropy $\hat{h}$ is identically zero.

We will prove the following two basic results.

Lemma 6. *If $\hat{C}$ is the dual of a causal cone, then the following inequalities are valid:*

$$\hat{h}(\mathbf{p}+\mathbf{q}) \le \hat{h}(\mathbf{p}) + \hat{h}(\mathbf{q}) \quad \text{whenever } \mathbf{p}+\mathbf{q} \in \hat{C}, \tag{7.3}$$

$$\hat{h}(\mathbf{p}+\mathbf{q}) \le \hat{h}(\mathbf{p}) + 2\hat{h}(\mathbf{q}) \quad \text{whenever } \mathbf{p} \in \hat{C}, \tag{7.4}$$

$$\hat{h}(\mathbf{p}+\mathbf{q}) \le 2\hat{h}(\mathbf{p}) + 2\hat{h}(\mathbf{q}) \quad \text{in all cases.} \tag{7.5}$$

Remark. As one immediate consequence of (7.3), together with (7.2), we see that the function $\hat{h}$ is continuous throughout any $\hat{C}$. (According to §6.3, the function $\hat{h}$ is not always continuous everywhere.) In particular, the directional entropy $\mathcal{H}_n(P)$ [or equivalently $h_n(\mathbf{v}_1, \ldots, \mathbf{v}_n)$] as long as the plane P spanned by the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is spacelike. In the special case of a left or right permutive block map (Example 6.2), Lind and Smillie point out that the dual cones $\pm\hat{C}$ and $\pm\hat{C}'$ cover the vector space. Hence $\hat{h}$ is subadditive and continuous everywhere.

Here is another useful consequence. Define the *null space* $N \subset \mathbf{R} \times V$ to be the set of all vectors $\mathbf{n}$ such that $\hat{h}(\mathbf{n}) = 0$. Then it follows from (7.5) that N is a vector space. Furthermore, if $\mathbf{p} \in \hat{C}$, then it follows from (7.3) and (7.4) that the entropy function $\hat{h}$ is constant throughout the plane $\mathbf{p} + N$ parallel to N. Note that the orthogonal complement $N^\perp$ can be described as the intersection of all n-dimensional planes W with entropy $\mathcal{H}_n(W) = 0$. If some causal cone C has dimension strictly less than $n+1$, we will prove that N is non-trivial.

Lemma 7. *Every n-dimensional plane W which contains a causal cone C has entropy $\mathcal{H}_n(W) = 0$. Hence $0 \le \dim(N^\perp) \le \dim(C)$.*

The dimension of $N^\perp$ provides a very crude measure of the complexity of the system. Thus the CA-maps with $\dim(N^\perp) = 0$ have n-dimensional entropy identically zero, and can perhaps be identified with Wolfram's cellular automata "of Class 1". If $\dim(N^\perp) = 1$, then it is not difficult to show that there is a unique *entropy vector* $\mathbf{h} \in \hat{F}$ with the property that

$$\hat{h}(\mathbf{p}) = |\mathbf{p} \cdot \mathbf{h}| \tag{7.6}$$

for every vector $\mathbf{p}$. This entropy vector is non-zero, and lies in the line $N^\perp$. These cases, in which information remains localized and travels with constant speed, should perhaps be identified with Wolfram's cellular automata "of Class 2". (Compare Example 6.1.)

Proof of Lemma 6. In the case $n = 1$ the three different cases are illustrated in figure 3. (The hypothesis that a vector $\mathbf{p}$ belongs to $\hat{C}$ means that the cone C lies completely to one side of the complementary hyperplane $\mathbf{p}^\perp$.) The conclusion then follows from Theorem 3. But these same figures can be used to illustrate the n-dimensional case: We need only form the cartesian product of one of these figures with a very large $(n-1)$-dimensional cube I^{n-1} which lies in the orthogonal complement of this 2-plane. This yields all possible configurations described by Lemma 6, since any two vectors in $\mathbf{R} \times V$ lie in some two-dimensional subspace. There is an additional complication in the higher-dimensional case since we must add the product of the shaded region $\mathcal{R}$ in the plane with the boundary ∂I^{n-1} in order to completely shade the region $\mathcal{R} \times I^{n-1}$. However, the n-dimensional volume of this boundary becomes negligible in comparison as the cube I^{n-1} gets bigger. ∎

Proof of Lemma 7. This follows immediately from Theorem 3, since any polyhedron P which lies in a hyperplane W containing C is in the C-umbra of its boundary ∂P, which has dimension $n-1$. ∎

Here are two further statements which can be proved by the same method.

Lemma 8. *If* $\mathbf{v}$ *and* $\mathbf{w}$ *belong to* $\hat{C}$, *then*

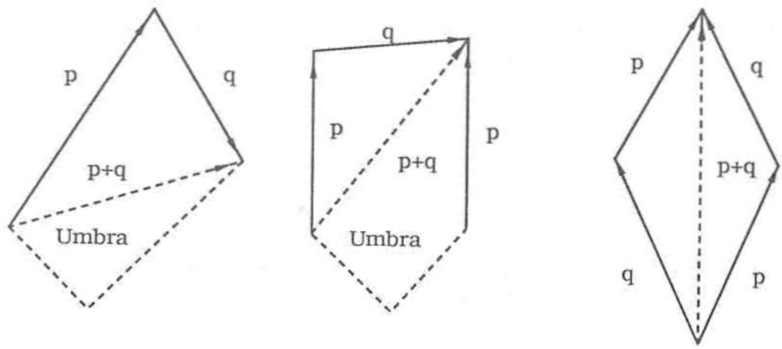

Figure 3.

$$0 \leq \hat{h}(\mathbf{v}) \leq \hat{h}(\mathbf{v}+\mathbf{w}),$$

with strict inequalities if $\hat{h}$ is not identically zero, and if these vectors belong to the interior of $\hat{C}$.

Finally, a sharper version of Corollary 3. Let $\mathbf{u}$ be any unit vector in the interior of $\hat{C}$ and let

$$\psi(\mathbf{v}) = \inf\{t \geq 0 \quad \text{suchthat} \quad \mathbf{v} + t\mathbf{u} \in \hat{C}\}$$

be the distance from $\mathbf{v}$ to $\hat{C}$ along lines parallel to $\mathbf{u}$.

Lemma 9. *Then $\hat{h}(\mathbf{v}) \leq \hat{h}(\mathbf{u})(\psi(\mathbf{v}) + \psi(-\mathbf{v}))$.*

The proof, in each case, is an application of Theorem 3, together with an appropriate two-dimensional diagram. Details will be left to the reader. ∎

Let $f : K^L \to K^L$ be a CA-map which is bijective, or in other words has a well defined inverse CA-map. Then clearly in addition to the usual causal cone $C = -F \subset \mathbf{R} \times V$ for the set of complete histories under f, which lies in the half-space $t \leq 0$, there will be an addition causal cone C' which is contained in the half-space $t \geq 0$. Note that both cones have only the zero vector in common with the hyperplane $t = 0$.

More generally, it will be convenient to say that a CA-map is "quasi-invertible" if it ts space of histories has two distinct causal cones C and C' which are strictly separated by some hyperplane. In other words, they lie in opposite half-spaces, and each one has only the zero vector in the common boundary hyperplane. An equivalent condition is that the dual cones $\hat{C}$ and $-\hat{C}'$ have a common interior point. In this case, evidently all of the Assertions of §5 are valid not only for the dual cone $\hat{C}$ but also for the dual cone $\hat{C}'$.

As an example, in the one-dimensional case, clearly every block map which is left or right permutive gives rise to a quasi-invertible CA-map.

We will prove that the entropy function $\hat{h}$ is linear on $\hat{C} \cap (-\hat{C}')$, at least in the measure-theoretic case. (Compare Boyle and Krieger [3].) Intuitively, this means that the function $W \mapsto \mathcal{H}_n(W)$ is "linear" on the set of separating hyperplanes. Suppose that we are given some fixed $(\mathbf{Z} \times L)$-invariant measure μ supported by $\mathcal{A}$.

Theorem 4. *With C and C' as above, the measure-theoretic entropy $\hat{h} = \hat{h}_\mu$ is linear throughout the intersection $\hat{C} \cap (-\hat{C}')$. More precisely, there exists an "entropy vector" $\mathbf{h} \in C \cap (-C')$ with the property that $\hat{h}(\mathbf{v}) = \mathbf{h} \cdot \mathbf{v}$ for every $\mathbf{v} \in \hat{C} \cap (-\hat{C}')$. Further*

$$\hat{h}(\mathbf{v}) \geq |\mathbf{h} \cdot \mathbf{v}|$$

for every vector $\mathbf{v}$.

In particular, it follows that the intersection $C \cap (-C')$ is non-trivial. It follows also that the entropy $\hat{h}(\mathbf{v})$ is strictly positive except possibly in directions orthogonal to $\mathbf{h}$.

Proof of Theorem 4. First consider the one-dimensional case. Consider a triangle whose edges are all "spacelike", in the sense that they are parallel to lines strictly separating C from C'. As illustrated in figure 4, we will assume that the edge E_3 is in the C-umbra of the union $E_1 \cup E_2$ of the other two edges, and where both E_1 and E_2 are in the C'-umbra of E_3. Then clearly $\eta_1(E_3) = \eta_1(E_1 \cup E_2)$ by Theorem 3. We will show that $\eta_1(E_1 \cup E_2) = \eta_1(E_1) + \eta_1(E_2)$, and hence that

$$\eta_1(E_3) = \eta_1(E_1) + \eta_1(E_2). \tag{7.7}$$

More generally, for any two compact sets X and Y, let us define the "entropy-correlation" to be the difference

$$\eta_1(X) + \eta_1(Y) - \eta_1(X \cup Y) = \eta_1(X) - \eta_1(X/Y) \geq 0. \tag{7.8}$$

Since we are working with measure-theoretic entropy, we know by Corollary 2 that this number can only decrease if we replace either X or Y by a subset which is closed line segment. *In particular, if the entropy-correlation between X and Y is zero, then the same is true for any subsets of X and Y which are closed line segments.*

Consider the situation of figure 5. Here we have two collinear edges E_1' and E_2' which shadow the regions $\mathcal{R}_1$ and $\mathcal{R}_2$ respectively (taking the union of the C-umbra and the C'-umbra). Let us assume that $E_1 \subset \mathcal{R}_1$ and $E_2 \subset \mathcal{R}_2$. Then $\eta_1(E_1' \cup E_2') = \eta_1(E_1') + \eta_1(E_2')$ since η_1 is a measure within each line, and $\eta_1(E_1') = \eta_1(\mathcal{R}_1)$, $\eta_1(E_2') = \eta_1(\mathcal{R}_2)$ by a umbra argument, hence the entropy-correlation

$$\eta_1(\mathcal{R}_1) + \eta_1(\mathcal{R}_2) - \eta_1(\mathcal{R}_1 \cup \mathcal{R}_2)$$

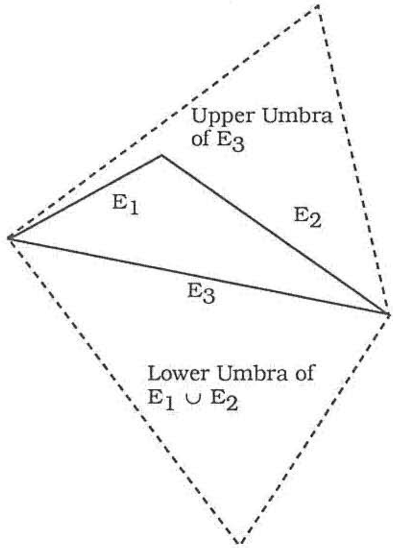

Figure 4.

is zero. Replacing $\mathcal{R}_1$ and $\mathcal{R}_2$ by their subsets E_1 and E_2, it follows that the entropy-correlation $\eta_1(E_1) + \eta_1(E_2) - \eta_1(E_1 \cup E_2)$ is also zero, which completes the proof of linearity in the one-dimensional case. Passing to the dual notation, it follows that $\hat{h}(\mathbf{v}) = \mathbf{h} \cdot \mathbf{v}$ for all $\mathbf{v} \in \hat{C} \cup (-\hat{C}')$, and for some fixed vector $\mathbf{h}$. This vector must lie in C, for otherwise we could choose a vector $\mathbf{w}$ in the interior of $\hat{C}$ orthogonal to $\mathbf{h}$. The equation $\hat{h}(\mathbf{v} + \epsilon\mathbf{w}) = \hat{h}(\mathbf{v})$ for all $\mathbf{v}$ in the interior of $\hat{C} \cap (-\hat{C}')$ and all small ϵ would then contradict Lemma 8. Finally, the inequality $\hat{h}(\mathbf{v}) \geq |\mathbf{h} \cdot \mathbf{v}|$ for every vector $\mathbf{v}$ follows using (7.3). In fact, if we express $\mathbf{v}$ as the sum of a vector $\mathbf{p}$ which belongs say to $\hat{C}$ and a vector $\mathbf{q}$ orthogonal to $\mathbf{h}$, then we have

$$\hat{h}(p + \epsilon\mathbf{v}) = (1 + \epsilon)\hat{h}(\mathbf{p}) \leq h(\mathbf{p}) + \hat{h}(\epsilon\mathbf{v})$$

for small ϵ, and the required inequality follows.

Just as in the proof of Lemma 6, this one-dimensional proof extends easily to higher-dimensional cases. ∎

The case in which f is actually an invertible CA-map is of particular interest. We can then use the standard invariant measure σ of (4.1), and the "entropy vector" will have the form $\mathbf{h} = (t, \mathbf{x})$ with $t = \log k > 0$. The ratio $\mathbf{x}/t \in V$ can then be described as the "mean shift" associated with the invertible CA-map. In the one-dimensional case, Boyle and Krieger [3] show that the mean shift function $f \mapsto \mathbf{x}/t$ is a homomorphism from the group $\mathrm{Aut}(K^L)$ of all invertible CA-maps to the additive group of the (one-dimensional) ambient vector space $V = L \times \mathbf{R}$. In particular, it follows that

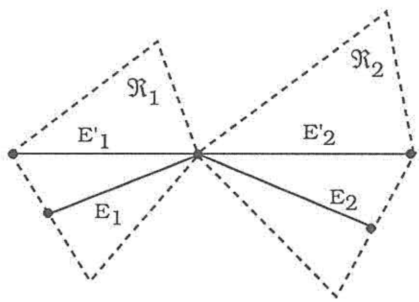

Figure 5.

the kernel of this homomorphism forms a subgroup of $\mathrm{Aut}(K^L)$, which they call the group of "shiftless" automorphisms. I have no idea whether these conclusions remain true in the higher-dimensional case.

I also do not know whether the assertion of Theorem 4 remains true for topological entropy. In one special case the proof does go through. *If the CA-map f is actually invertible, and if C and C' are the causal cones associated with f and its inverse, then Theorem 4 is true in the topological case also.* For in this case we can choose $\mathcal{R}_1$ and $\mathcal{R}_2$ so that the information-correlation

$$H(t\mathcal{R}_1) + H((t\mathcal{R}_2) - H(t\mathcal{R}_1 \cup t\mathcal{R}_2)$$

is not just asymptotically small, but is precisely zero. Hence it remains zero when we pass to appropriate subsets. ∎

Appendix A. Commuting maps

The constructions of §2 and §3 extend naturally to a more general case of a space Y and a collection of commuting maps $f_1, \ldots, f_n$ from Y to itself.

First suppose that Y is a compact metric space, with distance function $d(x, y)$, and suppose that the f_i are commuting continuous maps from Y to itself. Let L be the lattice $\mathbf{Z}^n$ with standard basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$, and let $\mathcal{A} \subset Y^L$ be the space of "complete histories" for the collection of maps f_i, that is the space of functions $\mathbf{a} : L \to Y$ satisfying the identity $f_i(\mathbf{a}(\ell)) = \mathbf{a}(\ell + \mathbf{e}_i)$. Given a finite subset $S \subset L$, define the *S-distance* $d_S(\mathbf{a}, \mathbf{b})$ between two complete histories as the maximum of $d(\mathbf{a}(s), \mathbf{b}(s))$ as s varies over S. Then for each $\epsilon > 0$ we can define the information content $H(\epsilon; S)$ to be the logarithm of the minimum number of sets with d_S-diameter less than ϵ which are needed to cover $\mathcal{A}$. The axioms of §2 are easily checked. If these information functions $S \mapsto H(\epsilon; S)$ all have growth degree at most d, then the d-dimensional entropy $\eta_d(\epsilon; X)$ is defined for every $\epsilon > 0$ and every compact $X \subset L \otimes \mathbf{R}$. Passing to the limit as $\epsilon \to 0$, we obtain a topologically invariant d-dimensional entropy function $0 \le \eta_d(\mathcal{A}, X) \le \infty$ (where this limit can be infinite).

As an example, if we have two commuting d-dimensional CA-maps, say f and g mapping $Y = K^{L'}$ to itself, then these two maps together with the lattice translations give us $d+2$ commuting maps on a the compact space Y. The associated d-dimensional entropy function $X \mapsto \eta_d(X)$, for $X \subset (\mathbf{Z} \times \mathbf{Z} \times L') \otimes \mathbf{R}$, is always well defined and finite.

Similarly, if we are given n commuting measure preserving transformations from a probability measure space Y to itself, then measure-theoretic entropies are defined. Choose a finite partition of Y, or equivalently a measurable map $\mathcal{P}$ from Y onto a finite set K. Then every complete history in $\mathcal{A} \subset Y^L$ corresponds under $\mathcal{P}$ to a configuration in K^L. Each partial configuration $L \supset S \to K$ has an associated probability, so the measure-theoretic information content $H(\mathcal{P}; S)$ is defined. Again, if this information function has growth degree at most d, it follows that the associated d-dimensional entropy $\eta_d(\mathcal{P}; X)$ is defined and finite. Passing to the limit as the partition becomes arbitrarily fine, we obtain an invariant d-dimensional entropy function $\eta_d(\mu; X)$.

References

[1] S. Amoroso and Y. N. Patt, "Decision procedures for surjectivity and injectivity of parallel maps for tessellation structures", *J. Comp. Syst. Sci.*, **6** (1972) 448–464.

[2] R. Bowen, "Topological entropy and axiom A", *Global Analysis*, Proc. Symp. Pure Math. **14**, AMS 1970, 23–41; R. Bowen and O. Lanford, "Zeta functions of restrictions of the shift transformation", *Global Analysis*, Proc. Symp. Pure Math. **14**, AMS 1970, 43–49.

[3] M. Boyle and W. Krieger, "Periodic points and automorphisms of the shift", *Trans. A.M.S.*, **302** (1987) 125–149.

[4] E. M. Coven, "Topological Entropy of Block Maps," *Proc. A.M.S.*, **78** (1980) 590–594.

[5] L. W. Goodwyn, "Some Counter-examples in Topological Entropy", *Topology*, **11** (1972) 377–385.

[6] G. A. Hedlund, "Transformations commuting with the shift", in *Topological Dynamics*, J. Auslander and W. G. Gottschalk, eds. (Benjamin 1968) 259–289; "Endomorphisms and automorphisms of the shift dynamical system", *Math. Syst. Theor.*, **3** (1969) 320–375.

[7] L. Hurd, "Formal language characterizations of cellular automaton limit sets", *Complex Systems*, **1** (1987) 69–80. (See also: "The application of formal language theory to the dynamical behavior of cellular automata", in preparation.)

[8] B. Kitchens and K. Schmidt, in preparation.

[9] D. Lind, "Applications of ergodic theory and sofic systems to cellular automata", *Physica*, **10D** (1984) 36–44.

[10] D. Lind, "The entropy of topological Markov shifts and a related class of algebraic integers", *Erg. Th. and Dyn. Sys.*, 4 (1984) 283–300.

[11] D. Lind, "Entropies of automorphisms of a topological Markov shift", *Proc. A.M.S.*, **99** (1987) 589–595.

[12] A. Maruoka and M. Kimura, "Conditions for injectivity of global maps for tessellation automata", *Info. and Control*, **32** (1976) 158–162.

[13] J. Milnor, "Directional entropies of cellular automaton-maps", of *Disordered Systems and Biological Organization*, E. Bienenstock et al., (Springer, 1986) 113–115.

[14] N. Packard, "Complexity of growing patterns in cellular automata", in *Dynamical Systems and Cellular Automata*, J. Demongeot, E. Goles, and M. Tcheunte, eds. (Academic Press, 1985).

[15] C. Shannon and W. Weaver, *The Mathematical Theory of Communication*, (U. Ill. Press 1949, 1963).

[16] Ya. Sinai, "An answer to a question of J. Milnor", *Comment. Math. Helv.*, **60** (1985) 173–178

[17] J. Smillie, "Properties of the directional entropy function for cellular automata", to appear.

[18] P. Walters, *An Introduction to Ergodic Theory*, (Springer 1982).

[19] S. Wolfram, "Universality and complexity in cellular automata", *Physica*, **10D** (1984) 1–35. (Reprinted in *Theory and Applications of Cellular Automata* (World Scientific, 1986). See also: "Cellular automata as models of complexity", *Nature*, **311** (1984) 419–424.)