Discrete Kinetic Theory, Lattice Gas Dynamics and Foundations of Hydrodynamics
World Scientific (1989), 343–354.

Four topics in lattice gases: Ergodicity; Relativity; Information flow; and Rule compression for parallel lattice-gas machines *

Tommaso Toffoli

MIT Laboratory for Computer Science
Cambridge, MA 02139, USA

Abstract We briefly present four topics that have recently arisen in the study of lattices gases. (a) *Specific ergodicity* asks, for an invertible cellular automaton, what fraction of the total information needed to identify an individual state is devoted to specifying the position of this state on its orbit. We give empirical evidence that this question has a definite answer. (b) We prove that a wide class of lattice gases are *Lorentz invariant* in the limit as the lattice spacing goes to zero. (c) We present a nontrivial situation where information flows as a *locally additive* quantity, much as energy and momentum, and thus takes on a strikingly tangible aspect. (d) In the light of recent progress in identifying usable lattice gas rules for 3-D hydrodynamics, we discuss *rule compression trade-offs* for parallel lattice gas machines.

1 Specific ergodicity in cellular automata

1.1 Specific ergodicity...

Consider a finite set consisting of L points. The entropy of this set is

$$I_{\text{total}} = \log L, \tag{1}$$

in the sense that if you choose a point at random (with uniform probability over the set) it will take me $\log_2 L$ binary questions (on average, for large L, if I use an optimal strategy) to correctly guess your point. Consider now an arbitrary invertible dynamics (a permutation) τ on this set. Such a dynamics partitions the set into a collection of orbits, and for any point one can speak of the length ℓ of the orbit to which it belongs. The entropy of (the partition induced by) τ is

$$I_{\text{orbit}} = \sum_{\text{all orbits}} \left(-\frac{\ell}{L}\log\frac{\ell}{L}\right); \tag{2}$$

(this the number of questions needed to guess just which orbit the points belong to). On ther other hand, if I were told right away on which orbit the chosen point lies, I would have to ask on average only a number of questions equal to

$$I_{\text{phase}} = \sum_{\text{all points}} \frac{\log \ell}{L} \tag{3}$$

in order to precisely locate the point. Of course, $I_{\text{total}} = I_{\text{orbit}} + I_{\text{phase}}$.

The *specific ergodicity* η of the dynamics τ is defined as

$$\eta = \frac{I_{\text{phase}}}{I_{\text{total}}} = \frac{\langle \log \ell \rangle}{\log L}; \tag{4}$$

clearly, $0 \leq \eta \leq 1$. When $\eta = 1$, all the points lie on a single orbit (the system is ergodic); when $\eta = 0$, all orbits consist of a single point.

We would like to extend the above considerations to certain infinite systems, where $I_{\text{total}} = \infty$ and in general also $I_{\text{phase}} = \infty$ (so that definition (4) breaks down); in particular, to cellular automata.

*This research was supported in part by the Defense Advanced Research Projects Agency, in part by International Business Machines, and in part by the National Science Foundation.

1.2 ...in cellular automata

Let τ be an invertible cellular automaton. In the one-dimensional case, one can imagine "wrapping around" the system so that site i is identified with site $i+N$ (periodic boundary conditions), yielding a finite system τ_N distributed over N sites and consisting of $L = N^Q$ points (where Q is the number of states available to each site).

Let η_N be the specific ergodicity of τ_N. The specific ergodicity of τ itself is the limit, if it exists,

$$\eta = \lim_{N \to \infty} \eta_N. \tag{5}$$

Except for a few trivial cases (identity transformation, Bernoulli shifts, etc.) it isn't at all obvious that limit (5) should exist. Moreover, if the number of dimensions is greater than one, there are in general several wraparound aspect ratios (e.g., "squarish" vs "long and thin") that yield the same number N of sites, so that in defining the limit one may want to restrict one's attention to terms of the sequence that approximate a certain aspect ratio, or worry that the limit, if it exists, may depend on the aspect ratio itself.

1.3 Numerical evidence

We have numerically evaluated the first few terms of the sequence (5) for a number of simple cellular automata in one and two dimensions; this was done by randomly sampling the state space and explicitly following the orbit of each sample point until it closed on itself. In all cases, numerical convergence was rapid enough to suggest the existence of a definite limit; moreover, in the two-dimensional cases, the dependence on the aspect ratio was negligible. We'll present here the evidence for a few representative cases, and conclude with a brief discussion.

All the lattice gases considered here use indistinguishable particles (0 and 1 represent, respectively, the absence or the presence of a particle on a given space-time track). The one-dimensional ones use the lattice of Fig. 1a; the two-dimensional ones, that of 1b.

Fig. 2a shows a case (FREDS) where the specific ergodicity appears to be close to zero (and indeed can be proved to be zero). The collision rule embodies the three-input/three-output Fredkin gate[3], with the control line running along the middle track ("rest particle") of each node. This rule is particle-conserving.

Fig. 2b shows two cases where η_N goes up rapidly, and appears to have an asymptotic value close to 1.

Figure 1: Lattices used for the experiments, (a) in one dimension, and (b) in two dimensions.

FREDR is a variation of FREDS; after performing the Fredkin-gate operation, the three outputs are circularly permuted. RANDC is an invertible rule that was chosen at random, and is not particle-conserving.

A more interesting case is illustrated in Fig. 3,

Figure 2: Cases in which η appears to be close to 0 (a) or to 1 (b).

Figure 3: Here η_N seems likely to settle on a value close to 1/3.

where η seems likely to settle on a value ($\approx 1/3$) noticeably different from 0 or 1. In this rule, the three inputs are fed straight through or circularly permuted depending on their overall parity.

Fig. 4 illustrates the results for two two-dimensional cellular automata discussed in [14]. BBM implements the billiard-ball model of computation[3], and is particle-conserving; TRON implements a synchronization scheme for asynchronous computation: the four input signals are complemented if and only if they are all the same. In both cases, note how little η_N depends on the wraparound aspect ratio.

The above experiments where performed using a few AT clones full-time for several weeks, which is equivalent to several hours on a typical supercomputer. In view of the exponential complexity of the problem, even moderate improvements on the above estimates would require a drastic increase in computing power. For simple cellular automata, a speedup of $\approx 10,000$ is easily achieved with a dedicated, fully-parallel implementation consisting just of a few programmable-logic array chips—and we are planning to go that way.

Figure 4: Two-dimensional cases. (a) The BBM rule; (b) the TRON rule.

1.4 Discussion

To the extent that the limits suggested by the above evidence do indeed exist (in what follows we'll speak, for brevity, as if they existed) we are witnessing the emergence, at a macroscopic level, of an *intensive* quantity η (measured in "entropy per unit area"), associated with a cellular-automaton *rule* τ, out of the *extensive* quantity I_{phase} (a quantity having the dimension of entropy and associated with the microscopic structure of τ_N as a whole). In other words, η appears to be a property of a cellular automaton seen as a "material" (when boundary or finite-size effect can be neglected) rather than a "system" of a certain size and shape.

Confronted with an intensive quantity ostensively defined by *global* means, one wishes of course to find an alternative *local* definition for it (global conspiracies are hard to swallow). Are there any obvious connections between the conventional "local presentation" of a cellular automaton (e.g., via a *local map* specifying the next state of a site as a function of the current state of its neighbors) and the value of η?

As a starting point, consider a randomly chosen invertible dynamics on L points. Simple combinatorial arguments[9] show that, for a randomly chosen point, all possible orbit lengths, from 1 to L, are equally probable; thence, the expected specific ergodicity will depend on L as

$$\frac{\log L!}{L \log L}, \tag{6}$$

which goes to 1 as $L \to \infty$. Thus, unless we impose further structure, the specific ergodicity of a large invertible systems will typically be close to 1.

Next, consider *locally-additive conserved quantities* (LACQ), such as number of particles, energy or momentum, whose global conservation can be ascertained, in many lattice gases, by mere inspection of the local map. To each LACQ there corresponds a partition of the state space, of which the partition into orbits is just a refinement. Thus, for each size N, the existence of a LACQ sets an upper bound for the value of I_{phase} of the corresponding τ_N. Unfortunately, the number of classes that make up such a partition grows only as aN (with constant a); even if all the classes had approximately the same number of points (which is usually far from the truth—the typical class-size distribution being roughly multinomial), this knowledge would decrease a previous upper bound for I_{phase} by at most $\log(aN)$; and

$$\lim_{N \to \infty} \frac{\log(aN)}{\log Q^N} = 0, \tag{7}$$

so that the upper bound for η is not affected.

By the same token, even the much more numerous, physically "spurious" LACQ that arise in certain cellular automata (e.g., momentum conservation on individual row and columns as in the HPP lattice gas[7]) depress η by only a vanishing amount.

On the other hand, in a space-distributed system such as a cellular automaton there often exist *localized invariants*[10], i.e., nonconstant functions that (a) are defined over a *finite* patch of space and (b) are left constant by the dynamics for at least some of the possible intial states of this patch (a trivial example is given by the "mirrors" in the BBM model[14]); each localized invariant will appear with a definite density (expected number of instances per unit area) in a random initial configuration. As stressed by Norman Margolus, localized invariants "factor out information," so to speak, in an amount measured in *bits per unit area*. Note that localized invariants can be enumerated in a systematic way by *local means*, that is, by inspecting the local map and its iterates; starting with a default upper bound of 1 for η, each new localized invariant discovered in this enumeration knocks off a finite amount from the current upper bound, so that the corresponding decreasing sequence of upper bounds has a definite limit. (The situation here is analogous to that encountered in computing by local means the *topological entropy* of a dynamics[1].)

However, it is not clear whether the quantity defined by this converging procedure coincides with "specific ergodicity" or only provides an upper bound for it.

1.5 Conclusions

We hope the above numerical results and discussion will stimulate a theoretical investigation of the subject.

2 Lorentz invariance in lattice gases

2.1 Orientation

Cellular automata have recently contributed a new tool to investigations in hydrodynamics, namely, lattice gases whose individual interactions directly reflect (although in a stylized way) basic principles of microscopic physics, and whose macroscopic behavior converges, for low flow speeds, to the Navier-Stokes equation. At higher flow speed these models depart from Galilean invariance (cf. [8]); this is usually viewed as a *blemish* (the Navier-Stokes equation *is* Galilean invariant), and much effort is being devoted to the synthesis of lattice-gas models in which Galilean invariance is retained to a higher order of approximation.[1]

Leaving practical hydrodynamics aside, we ask whether this alleged blemish can be seen instead as a *feature*; specifically, whether, by interpreting the speed of particles as comparable to the speed of light, the deviations from Galilean invariance are consistent with

[1]In a lattice gas, the maximum speed of microscopic propagation of information through the system is the speed of the particles themselves. Since in the above models of hydrodynamics the speed of the lattice-gas particles is treated as representative of the average speed of real-gas particles—rather than of the speed of light—violation of Galilean invariance is bound to occur at sufficiently high flow speeds.

Lorentz invariance. After all, cellular automata have on more than one occasion shown that they intrinsically "know" more physics than one would tend to credit them with; it is not inconceivable that lattice gases that depart from Galilean invariance may approximate real physics, in some respects, better than the Navier-Stokes equation itself.

Figure 5: On an intuitive distance scale, a lattice gas may fail to well approximate the Navier-Stokes equation, and yet be closer to real physics than the Navier-Stokes equation itself.

Figure 6: In the two-dimensional random walk on a lattice, circular symmetry naturally emerges at the macroscopic level out of a square-symmetry microscopic law.

To begin with, is it at all plausible that a dynamics defined of a *regular spacetime lattice*, and thus having a strongly distiguished inertial reference frame, could be Lorentz-invariant? We'll start by pointing at some circumstantial evidence.

Consider the two-dimensional random walk on the x, y lattice. At the microscopic level, this dynamics is not rotation-invariant (except for multiples of a quarter-turn rotation); however, invariance under the continuous group of rotations *emerges* at the macroscopic level (Fig. 6). In fact, for $r^2 = x^2 + y^2 < t$ and in the limit as $t \to \infty$, the probability distribution $P(x, y; t)$ for a particle started at the origin converges to

$$\frac{1}{2\pi t} e^{-\frac{1}{2}\frac{x^2+y^2}{t}}, \tag{8}$$

i.e., depends on x and y only through $x^2 + y^2 = r^2$.

Now, there is a strict formal analogy between a *circular rotation* by an angle ϕ in the x, y plane and a Lorentz transformation with velocity β in the t, x plane—which can be written as a *hyperbolic rotation* by a rapidity $\theta = \tanh^{-1}\beta$:

$$\begin{bmatrix} t \\ x \end{bmatrix} = \begin{bmatrix} \cosh\theta & \sinh\theta \\ \sinh\theta & \cosh\theta \end{bmatrix} \begin{bmatrix} t' \\ x' \end{bmatrix} \tag{9}$$

Riding on this analogy, one may hope to find a microscopic dynamics on the t, x lattice for which Lorentz invariance (which is out of the question at the microscopic level) would emerge at the macroscopic level.

Let's look at the one-dimensional random walk on a lattice, with probability p of moving to the right and $q = 1 - p$ of moving to the left. For $p = q = 1/2$, the evolution of the resulting binomial distribution is characterized, macroscopically, by a mean $\mu = 0$ and a standard deviation $\sigma = \sqrt{t/4}$ (Figure 7a).

In general, $\mu = (p - q)t$. If we shift the parameter p away from its center value of $1/2$, the center of mass of the distribution will start moving at a uniform velocity $\beta = p - q$. Let's try to offset this motion by a Galilean transformation

$$x = x' + \beta t'. \tag{10}$$

Macroscopically, the new system will evolve, in the new frame, just as the old system did in the old frame—except that now $\sigma = \sqrt{pqt} = \sqrt{(1-\beta^2)t/4}$, so that the diffusion will have slowed down by a factor $1 - \beta^2$ (Fig. 7b).

Intuitively, as some of the resources of the "random walk computer" are shifted toward producing coherent macroscopic motion (uniform motion of the center of mass), fewer resources will remain available for the task of producing incoherent motion (diffusion). Not only do we get something analogous to the Lorentz-Fitzgerald "time expansion;" we also get a slow-down

Figure 7: (a) Symmetric random walk ($p = 1/2$). (b) Asymmetric random walk ($p = 3/4$); note that, as the center of mass picks up a speed $\beta = p - q$, the rate of spread goes down by a factor $1 - \beta^2$.

factor $(1 - \beta^2)$ that is reminiscent of, though different from, the correct relativistic factor $\sqrt{1 - \beta^2}$.

2.2 A deterministic model of diffusion

In the above example we tried to offset by a *Galilean* coordinate transformation the consequences of a trasformation of the dynamical parameter p, and noticed that as we tamed μ we lost control of σ. A *Lorentz* transformation, though closer to the goal, would still not have yielded invariance. We could have tried other coordinate transformations, and eventually succeded.[2]

However, with dynamical *parameters* we are barking up the wrong tree. What relativity says is that a *certain* spacetime transformation (Lorentz transformations) can always be offset by an appropriate transformation of the dynamical *variables*. We shall now present a lattice-gas model of diffusion that has the same macroscopic phenomenology as the random walk, but is microscopically deterministic and reversible. Unlike the random walk, changes in the macroscopic parameters μ and σ arise in this model from changes in the initial distribution of microscopic *states*, rather than by tampering with the microscopic *laws*. This model is exactly Lorentz-invariant in the continuum limit, i.e., as the lattice spacing λ goes to zero.

Let us consider a cellular automaton that is similar to the HPP lattice gas[7], though operating on a *one-dimensional* lattice of spacing λ (sites per unit length), which will have the form of Fig. 8a when "unfolded" over time. There are two kinds of particle, i.e., left-moving and right-moving, and no more than one particle of each kind can be at any one site. When two particles collide, each reverses its direction (Fig. 8b). As long as particles are identical (say, all black), this reversal is indistinguishable from no interaction (Fig. 8c).

Figure 8: (a) One-dimensional lattice, unfolded over time. The tracks, with slope ± 1, indicate potential particle paths; the nodes indicate potential collision loci. (b) Bouncing collision. (c) No-interaction collision.

Now let us paint just one particle red (in which case the reversal *does* make a difference), and study the evolution of its probability distribution $\rho(x;t)$ when both right- and left-going particles are uniformly and independently distributed with linear density (particles per unit length) $s = n/\lambda$, where n is the lattice

[2] Indeed, the desired transformation here is a Lorentz transformation followed by a scaling of both axes by a further factor $\sqrt{1 - \beta^2}$.

occupation density (particles per site).

For fixed s, as $\lambda \rightarrow 0$ (continuum limit) $\rho(x;t)$ converges to the solution of the *telegrapher's equation*[6]

$$\partial_{tt}\rho = \partial_{xx}\rho - 2s\partial_t\rho, \tag{11}$$

(in this discussion, we are deliberatly omitting details; the reason will be obvious by the end of Section 2.3). The latter distribution, in turn, converges to the solution of the diffusion equation

$$\partial_t\rho = \frac{1}{2s}\partial_{xx}\rho \tag{12}$$

in the same circumstances ($t \rightarrow \infty$, $|x| < \sqrt{t}$) as the binomial distribution does. We shall now introduce the freedom to independently vary the densities s_+, s_- of, respectively, right- and left-going particles; as a consequence, the red particle's distribution's center of mass will drift, and the diffusion rate will be affected too—much as in the asymmetric random walk case. However, this time we have strict Lorentz-invariance (in the continuum limit): to every Lorentz transformation of the coordinates, $t, x \mapsto t', x'$, there corresponds a transformation of the initial conditions, $s_+, s_- \mapsto s'_+, s'_-$, that leaves the form of ρ invariant.

2.3 No statistics involved

In the two-dimensional random walk of Section 2.1, macroscopic rotational invariance emerged for *statistical* reasons: though there was no one-to-one correspondence between the microscopic walks in the original frame and those in a rotated frame, approximately as many walks terminated in a given area element $dx\,dy$ of one frame as in the corresponding element $dx'dy'$ of the other. What is surprising is that no statistics is involved in the emergence of Lorentz invariance in the example of the previous section; the distribution $\rho(x;t)$ is Lorentz-invariant because *every individual microscopic history* of red and black particles is matched by an identical history in the Lorentz-transformed frame. This feature is independent of the details of the dynamics, provided that the lattice gas conforms to the following general "format"

1. The spacetime lattice group (Fig. 8a), as generated by the particles' elementary movements (in one time step), is *free Abelian*.

2. The dynamics is invertible.

3. The only interactions that are allowed are *collisions*; i.e., an isolated particle cannot change state (and thus "sense" when it is crossing a lattice track in the absence of a particle on that track).

We'll sketch the proof. Consider an arbitrary history of an arbitrary lattice gas obeying the above format, as in Fig. 9a (the "color" of the particles is irrelevant). Let us now Lorentz-transform the coordinates, yielding (for $\theta = \log 2$) the lattice of 9b. Can we specify initial conditions yielding, in the new frame, a history identical to 9a? This history would have to be as in 9c, but then a particle path might fall outside of an available lattice track, as indicated. Let's "borrow" for such a path an adjacent track (assuming it's not occupied), yielding history 9d, with a displacement error of no more than half the lattice spacing; this error does not diverge with time. A conflict would arise if the track to be "borrowed" were already occupied by another particle—but the probability of such a conflict is vanishing for vanishing particle densities. Thus, for any assigned particle density ρ, as the lattice occupation density n and the lattice spacing λ both go to zero at the same rate so as to leave ρ constant (continuum limit), *all microscopic histories* (except a vanishing fraction) can be reproduced in *microscopic detail* (with a vanishing error) in the Lorentz-transformed frame by a suitable transformation of the initial conditions. QED

For one spatial dimension, a Lorentz transformation concides with a scaling of the two "lightspeed" axes c_+, c_- by factors μ_+, μ_- such that $\mu_+\mu_- = 1$ (i.e., spacetime volume is invariant). For lattice gases conforming to the above format, an analogous scaling transformation can be introduced for any number of spatial dimensions, and is the only reasonable generalization of the Lorentz transformation when the speed of the particles themselves is interpreted as the "speed of light." With this reading of the term "Lorentz transformation," the above argument on Lorentz invariance at the level of individual microscopic histories applies verbatim to any number of spatial dimensions.

3 Information as an additive conserved quantity

We wish to bring attention to a situation where information flow strictly obeys the continuity equation,

Figure 9: (a) Original history. (b) Lorentz-transformed lattice. (c) Attempt to make an identical history in the new frame. (d) The best we can achieve with the available spacetime tracks.

and thus takes on a strikingly tangible aspect.

In many spatially-extended dynamical systems governed by short-range interactions—such as an ordinary fluid—one encounters locally additive conserved quantities (e.g., energy, electric charge). In an invertible system, information (or "fine-grained entropy") is also conserved. However, information is not, in general, additive. Because of spatial correlations, if the volume occupied by the system is partitioned into several pieces and the information from all the individual pieces is added up, the result is not the total information of the current macroscopic state—which does not change with time as the system evolves—but an overestimate which may widely fluctuate with the passage of time. In other words, information, though conserved in a global sense, is not localized, and one cannot write for it transport equations of the kind that are familiar for energy, momentum, etc.

It turns out that information becomes a *locally additive* conserved quantity when the following four conditions are met:

1. Small perturbations
2. from equilibrium
3. of particle-conserving
4. invertible cellular automata

Though local additivity of information may arise in more general situations, the above conditions allow us to arrive at an exact proof in a number of interesting cases (including systems that support a full-featured hydrodynamics and others that are known to be computation-universal). In particular, one can take advantage of the fact that a wide class of invariant distributions for systems obeying conditions 3 and 4 are explicitly computable.

3.1 Equilibrium

We shall condense the argument (see [15] for more details), using the HPP lattice gas[7] as an example. Because of invertibility and particle conservation, such system a admits of equilibrium distributions that are *spatially homogeneous* and *wholly uncorrelated*: for each of the four possible directions of motion, $h = x_+, x_-, y_+, y_-$, the mass-density (expected occupation number) ρ_h is site-independent, and the ρ's are uncorrelated both within sites and between sites.[3] These equilibria are characterized by the *balance condition*

$$\rho_{x_+}\rho_{x_-}\overline{\rho}_{y_+}\overline{\rho}_{y_-} = \overline{\rho}_{x_+}\overline{\rho}_{x_-}\rho_{y_+}\rho_{y_-} \equiv \sigma, \qquad (13)$$

where $\overline{p}$ denotes $1-p$.

Figure 10: (a) The HPP lattice. (b) The HPP node as a four-input, four-output function.

[3]Note that such a distribution is *not* an equilibrium one for, say, an Ising spin model such as Q2R[16, 14], which is invertible but not spin-conserving.

3.2 Evolution of mass-density

With reference to Fig. 10, where the ρ_h denote input densities to a site and the ρ'_h the corresponding output densities at a given interaction step, the rate of response to a small perturbation from equilibrium is given by the Jacobian matrix

$$\mathbf{J}^{\text{mass}} = \left[\frac{\partial \rho'_h}{\partial \rho_k}\right] = \mathbf{I} + \sigma \mathbf{A}, \tag{14}$$

where $\mathbf{I}$ is the identity matrix, and σ, defined in (13), determines the overall intensity of scattering of the perturbation. The *normalized scattering matrix* $\mathbf{A}$ is of the form

$$\mathbf{A} = \begin{bmatrix} u_{x+} & u_{x-} & -u_{y+} & -u_{y-} \\ u_{x+} & u_{x-} & -u_{y+} & -u_{y-} \\ -u_{x+} & -u_{x-} & u_{y+} & u_{y-} \\ -u_{x+} & -u_{x-} & u_{y+} & u_{y-} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix} \begin{bmatrix} u_{x+} \\ u_{x-} \\ -u_{y+} \\ -u_{y-} \end{bmatrix}^T, \tag{15}$$

where $u_h = -1/\rho_h\overline{\rho}_h$. Since the elements of the first vector of (15) add up to zero, it is clear that the sum of the mass-density perturbations at the output of a site equals that at the input (this is, of course, a direct consequence of particle-conservation).

3.3 Evolution of information-density

A perturbation (with respect to the equilibrium value) of mass-density ρ_h at a given site yields a corresponding perturbation of the accompanying information-density $s_h = -(s_h \log s_h + \overline{s}_h \log \overline{s}_h)$; the ratio r_h between these two perturbations is

$$r_h = \frac{ds_h}{d\rho_h} = -\log \frac{\rho_h}{\overline{\rho}_h}. \tag{16}$$

We can thus write a Jacobian matrix for the propagation of information-density perturbations

$$\mathbf{J}^{\text{info}} = \left[\frac{\partial s'_h}{\partial s_k}\right] = \left[\frac{r_h}{r_k} J^{\text{mass}}_{hk}\right] = \mathbf{I} + \sigma \mathbf{B}, \tag{17}$$

where

$$\mathbf{B} = \begin{bmatrix} r_{x+} \\ r_{x-} \\ -r_{y+} \\ -r_{y-} \end{bmatrix} \begin{bmatrix} u_{x+}/r_{x+} \\ u_{x-}/r_{x-} \\ -u_{y+}/r_{y+} \\ -u_{y-}/r_{y-} \end{bmatrix}^T. \tag{18}$$

Note that the elements of the first vector in (18) are not equal in magnitude as in (15). However they still add up to zero—because of the balance condition (13). Thus, in the limit for small perturbations, the amount of information-density perturbation is conserved across an interaction, which means that no correlations are introduced—and the perturbed distribution remains (in the present linear approximation) wholly uncorrelated during the evolution.

3.4 Conclusions

We have exibited a nontrivial system in which information is transported as an additive conserved quantity. The evolution of the system is completely characterized by the site-by-site values of the mass-density perturbations, since correlations across sites vanish.

In this situation one has direct computational access to the dynamics of the perturbed distribution, since the amount of work needed to compute such evolution is proportional to the number N of sites (rather than to the number of microstates, which increases exponentially with N). Fig. 11 shows two stages of the evolution of a localized perturbation in the HPP gas (with $\rho_{x+} = 0.933$, $\rho_{x-} = 0.500$, and $\rho_{y+} = \rho_{y-} = 0.789$); this figure plots the deviations of the ρ's from their equilibrium values. The corresponding deviations of the s's are proportional to those of the ρ's—with different proportionality coefficients for the four directions of travel, as given by (16)—and yield a qualitatively similar plot.

Symbolic treatment is also fully-accessible, since one is confronted with a set of coupled linear finite-difference equations.

4 Rule compression for lattice-gas machines

We conclude this paper with some thoughts suggested by the most visible feature of this conference, namely, the wide consensus there appears to have grown over a viable class of lattice-gas model for 3-D hydrodynamics.[4] These models have brought to light a paradoxical situation.

Lattice-gas hydrodynamics had started out as a champion of fine-grained reductionism:

[4]This is an informal and tentative discussion. I have tried to extrapolate from ideas and data that I gathered from discussions with Doolen, Frisch, Hasslacher, Hénon, Margolus, Rivet, Sellers, and others.

Figure 11: Evolution of mass-density for a small perturbation from equilibrium, for the HPP gas, shown at times $t = 16$ and $t = 128$.

"Let's replace a calculus based on continuum constructs (density, pressure, flow rate, etc.) by a large number of discrete interactions between stylized particles, so as to directly reduce macroscopic phenomenology to microscopic first principles. Because of this, our model will speak with more sincerity and authority.

"For comparable results, it's true, we'll need many more operations—but also much simpler ones. And today's technology allows us to casually contemplate moving from a few bulky general-purpose floating-point units to a dedicated array of millions of elementary 'collision' processors."

Indeed, the original HPP rule[7], with four bits' worth of particles at each site and a single type of interaction, could be realized by a collision processor consisting of just a few gates. The FHP rules[4] were a major improvement over the HPP model—and going from four particles to six or perhaps seven didn't seem to rock the boat much. However, one could already suspect that a 3-D version would require a substantial amount of machinery for the collision processor—certainly more gates for the latter than for the storage of a dozen or so bits of state variables.

Now, it turns out that one has to go all the way up to a rule involving at least 24 bits per site in a pseudo 4-D interconnection scheme[5, 13] before performance (in terms of Reynolds number) flattens out. A fast, brute-force implementation of such a rule requires a *billion* bits' worth of lookup table—definitely way beyond the size of the fanciest floating-point processor. Has the dream of a thousand processors per chip been replaced by the nightmare of a thousand-chip processor? Shall we loose faith in cellular automata's promise of massive parallelism? Shall we go and warn the neural-network community that they too may be bound for trouble?

Shall the next cellular automata machine consist of a gram of silicon labeled "state variables" and a ton of silicon labeled "processors?" Or shall one be forced to take one architectural route for dedicated *lattice-gas machines* and a quite different one general-purpose *cellular automata machines*?

At a more fundamental level: Feynman was worried by the fact that, according to QED, an electron seems to have to carry around a trunk full of floating-point processors in order to decide where to go next[?]. Shall a more digitally-oriented elementary particle have to carry around the equivalent of ten phone books in order to figure out what to do on meeting another particle?

The actual information contents of the above lookup table for 3-D hydrodynamics is obviously a small fraction of a billion bits. After all, workers in the field can state the recipe in perhaps a thousand words. But if I ask pointblank for a definite entry of the table, a lot of thinking (a few hours of computing time) will be needed to extract this entry from the recipe. What are the most favorable *storage space* vs *access time* tradeoffs? The answer depends not only on the nature of the tables themselves, but also on the computing resources one has in mind.

Originally, CRAY and Connection Machine users took advantage of their machines' natural parallelism

(64 bits in the CPU path, in one case, and 64K atomic processors in the other) to process many sites at once in a SIMD scheme. The broadcast "single instruction" was one step of an algorithm that described the collision rule in Boolean terms—just a few dozen steps for the HPP rule. But, much as the size of the lookup table, also the time of the algorithm grows exponentially with the number of bits at a site, and this scheme soon became too slow to be useful.

Today, CRAY users do not attempt to take advantage of the CPU's potential 64-bit parallelism for doing the site-updating. They employ most of the machine as a single neighbor-gathering pipeline, and do the collision part by brute-force lookup, using the machine's fast memory[12, 13]. As long as memory limits are not exceeded, it doesn't pay at all to compress the lookup table, since any unused portion of the memory would be wasted.

Architectures like the proposed CAM-8[11] (which are perfectly viable in a single-processor mode) are intrinsically amenable to a high degree of multiprocessor parallelism, since the "data-flywheel" that stores state-variables and moves them around as needed is realized in a much less expensive way than with a conventional supercomputer. In CAM-8, a lookup table of, say, a million bits for each processor approximately matches the cost of the associated data-flywheel; with lookup tables of this size, an assembly on *several thousand processors* is technologically and economically feasible. With much larger lookup tables, however, the cost of the table itself would overwhelm all other costs, and the number of processors one could practically put together would be inversely proportional to the size of the lookup table. In this situation, any amount of compression of the lookup table would yield a proportionate increase in the number of processors.

Taking advantage of the spatial symmetries of the collision rule (rotation-invariance and other isometries) would allow one to compress the table by about two orders of magnitude (tens of megabits rather than gigabits); however, a matching decompression algorithm would add several thousand Boolean operations to the lookup proper. In a CRAY environment, this would be prohibitively time-consuming. On the other hand, in a CAM-8 environment it would be rather easy to replace the large table by the smaller one complemented by a hardwired Boolean network for the decompression algorithm. In order to achieve a reasonably small number of logic levels (say, a few dozen) so as to facilitate pipelined operation,[5] the network would have to be of a constant shallow depth, and this may bring the actual gate count into the tens of thousands rather than thousands. Still, the silicon area taken by the decompressing network would not exceed that of the compressed table itself.

One may conjecture that, if the CAM-8 architecture with uncompressed lookup tables should yield a performance gain of one to two orders of magnitude over a CRAY-XMP in running current 3-D lattice-gas schemes for hydrodynamics, the same architecture *with* a compression scheme would easily yield *two to four* orders of magnitude.

In conclusion, there seems to be a strong incentive in devising a compact "collision unit" for lattice gas machines, which would play a specialized role somewhat analogous to that of the floating-point unit in more conventional computers.

One challenge is, of course, to design this collision unit as a suitable combination of programmable lookup tables and *programmable* logic, so as to yield not only the desired *efficiency*, as discussed above, but also a good amount of *flexibility* in running a wide range of rules—in particular, many cellular automata rules of physical interest besides hydrodyanamics (cf. [14]). This appears to be a good ground for tight interaction between mathematical physics, information mechanics, and technology.

References

[1] ADLER, R. L., A. G. KONHEIM, and M. H. MCANDREW, "Topological entropy," *Trans. Amer. Math. Soc.* **114** (1965), 309–319.

[2] FEYNMAN, Richard, *The Character of Physical Law*, MIT Press (1965).

[3] FREDKIN Edward, and Tommaso TOFFOLI, "Conservative logic," *Int. J. Theor. Phys.* **21** (1982), 219–253.

[4] FRISCH, Uriel, Brosl HASSLACHER, and Yves POMEAU, "Lattice-Gas Automata for the Navier-Stokes Equation," *Phys. Rev. Lett.* **56** (1986), 1505–1508.

[5] Even though the CAM-8 architecture is compatible with very extensive pipelining (hundreds of thousands of steps), there are a number of practical of reasons for striving to keep the pipelines short.

[5] FRISCH, Uriel, et al., "Lattice gas hydrodynamics in two and three dimensions," *Complex Systems* **1**:4 (1987), 633-707.

[6] GOLDSTEIN, S., "On diffusion by discontinuous movements, and on the telegraph equation," *Quart. J. Mech. Appl. Math.* **IV.2** (1951), 128-156.

[7] HARDY, J., O. DE PAZZIS, and Yves POMEAU, "Molecular dynamics of a classical lattice gas: Transport properties and time correlation functions," *Phys. Rev.* **A13** (1976), 1949–1960.

[8] HAYOT, Fernand, "The effect of Galilean non-invariance in lattice gas automaton one-dimensional flow." *Complex Systems* **1** (1987), 753–761.

[9] LOVÁSZ, László, *Combinatorial Problems and Exercises*, Akadémiai Kiadó, Budapest (1979).

[10] MARGOLUS, Norman "Physics-like models of computation," *Physica* **10D** (1984), 81–95.

[11] MARGOLUS, Norman, Tommaso TOFFOLI, "Cellular Automata Machines," to appear in *Large Nonlinear Systems*, Gary DOOLEN ed. (1988), as a revised version of a paper by the same title appeared in *Complex Systems* **1**, 967–993 (1987).

[12] RIVET, Jean-Pierre, "Hydrodynamique par la methode des gaz sur réseaux," Doctoral Thesis, Université de Nice (France), Observatoire de Nice (June 1988).

[13] SHIMOMURA, T., G. DOOLEN, B. HASSLACHER, and Castor FU, "Calculations Using Lattice Gas Techniques," *Los Alamos Science*, special issue no. 15 (1987), 201–210.

[14] TOFFOLI, Tommaso, and Norman MARGOLUS, *Cellular Automata Machines—a new environment for modeling*, MIT Press (1987).

[15] TOFFOLI, Tommaso, "Information transport obeying the continuity equation," *IBM J. Res. Develop.* **32** (1988), 29–36.

[16] VICHNIAC, Gérard, "Simulating physics with cellular automata," *Physica* **10D** (1984), 96–115.