SPACETIME AND CAUSAL SETS *

Rafael D. Sorkin
Department of Physics
Syracuse University
*Syracuse, NY 13244-1130 ***
and
Enrico Fermi Institute
University of Chicago
5640 S. Ellis Avenue
Chicago IL 60637

Abstract

The causal set is a possible discrete substratum for spacetime. This idea is introduced and some aspects of causal set kinematics are presented, most of them relating to the question of how the discrete order defining the causal set corresponds to the geometrical structures of continuous spacetime. In particular a statistical notion of "fractal dimension" is developed in some detail. The lectures conclude with some preliminary remarks on dynamics, and a speculation on the cosmological constant.

Since the structure I am going to propose as the basis of spacetime is discrete, let me take a moment to recall briefly why many people find an underlying discreteness more natural than the persistence of a continuum down to arbitrarily small sizes and short times. I think the main reasons can be summarized by referring to three self-contradictions or "infinities" which animate many attempts to go beyond the so-called "standard model" of current physics.

The first infinity can be symbolized by the equation $Z = \infty$. It arises in quantum field theory, is really a family of infinities, rather than a single one, and is traditionally dealt with via "renormalization". The second infinity, symbolized by $R_{abcd} = \infty$, arises in classical General Relativity, at the singularities where the

* Published in J.C. D'Olivo, E. Nahmad-Achar, M. Rosenbaum, M.P. Ryan, L.F. Urrutia and F. Zertuche (eds.), *Relativity and Gravitation: Classical and Quantum* (Proceedings of the *SILARG VII Conference*, held Cocoyoc, Mexico, December, 1990) (World Scientific, Singapore, 1991), pp. 150-173

** permanent address

curvature blows up. This infinity is also very familiar. The third infinity arises in quantum gravity proper, and can be expressed by the equation $S_{BH} = \infty$, S being the black-hole entropy. This third infinity, perhaps less widely appreciated than the other two, results when one tries to actually count the degrees of freedom of the horizon in any direct manner [1]. (In addition to contributing this new infinity, quantum gravity makes some of the old ones worse, of course, since it ruins the perturbative renormalizability of the so-called standard model.)

It thus seems that the application of quantum ideas to gravity is spawning new contradictions, rather than ameliorating the old ones, as had been hoped. A further example of this is the apparent impossibility of measuring the metric on sub-Planckian scales, without the apparatus collapsing into a black hole.

Since all of these contradictions involve distances around the Planck length or below, their solution might be found in the hypothesis that, at around 10^{-32}cm, the continuous manifold of General Relativity gives way to what Riemann called a 'discrete manifold', carrying only a finite number of degrees of freedom in any finite volume. And as a "bonus" there is the possibility (mooted already by Riemann) that the discrete structure might carry its metric relations within itself, which a continuous manifold like $\mathbf{R}^4$ can never do. At the very least, we could expect an explanation of why a fundamental length appears in physics, just as the atomic radius furnishes a fundamental scale of distance for ordinary matter.

Now the particular discrete manifold I am proposing is known as a *causal set*, which is a set of elements endowed with an order relation that one calls 'causal' in the sense of the Correspondence Principle.[2-4] The way in which this intrinsic order-structure is taken to relate to (or better give rise to) the metrical continuum of General Relativity is suggested by the two correspondences,

$$N \longleftrightarrow V$$

$$x \prec y \longleftrightarrow x \in \bar{J}^{-}(y).$$

The first correspondence means that any region of what we ordinarily think of as continuous spacetime is made up of only a finite number, N, of elements of the underlying causal set, and that number is equal to the macroscopic volume V of the region, when volume is measured in fundamental units. Or to say this the other way around, what we are actually doing when we measure spacetime volume is indirectly to count the number of causal set elements comprising the region, just as weighing a bar of copper is an indirect way to count the copper atoms comprising the bar. The second correspondence is equally straightforward, and just states that the macroscopic light-cone structure of spacetime directly reflects the order relation of the underlying causal set. An analogy here might be the way the fracture-planes of a crystal reflect the geometrical structure of its underlying atomic lattice.

Before going any further, I should say more precisely what mathematical structure I am going to be dealing with. Formally, a causal set is a relation satisfying axioms of transitivity, acyclicity, and local finiteness. In symbols these axioms say, respectively,

$$x \prec y \prec z \Longrightarrow x \prec z$$

$$x \prec y \text{ and } y \prec x \Longrightarrow x = y$$

$$|[x, y]| < \infty.$$

Here $\prec$ is the order relation defining the causal set, and $|[x, y]|$, the cardinality of the "Alexandrov set" or "interval" $[x, y]$, is the number of elements z falling between x and y in the sense that $x \prec z \prec y$. Taken together, these axioms just say that a causal set is what mathematicians would call a locally finite, partially ordered set. (The local finiteness is expressed by the third axiom, which may be a little stronger than necessary. In its present form it resembles the macroscopic condition that a Lorentzian manifold be globally hyperbolic.)

One obvious motivation for this particular choice of discrete structure is that it relates very directly to macroscopic causality, which, as many workers have sensed, may have a more fundamental character than other macroscopically apparent relations like length. But what is especially appealing about causal sets is that their discreteness is essential to their ability to reproduce macroscopic geometry. If an infinite number of elements were present locally then the correspondence $V = N$ would lose its meaning, and without it we could at best hope to recover the conformal metric, but not the volume-element needed to get from the latter to the full metric g_{ab}.

To my mind, this blending of order with discreteness is important evidence that we are on the right track in our choice of fundamental discrete structure. Another strong encouragement is the prospect of the unification that would accompany a successful theory based on causal sets. In such a theory, the single relationship $\prec$ would unite within itself the topology, the differential structure, the metrical geometry, and of course the causal structure of spacetime. In particular it would explain the Lorentzian signature of the metric (this being the only signature for which distinct past and future directions can be defined), whereas from most other points of view the presence of a minus sign in the signature appears to be mathematically unnatural.

Kinematics and Dynamics

But how should one go about actually constructing a theory of causal sets? In the development of most physical theories one can distinguish two stages, corresponding roughly to what in mechanics are called kinematics and dynamics.

In the present case, the former stage would be concerned first of all with the fact that the main macroscopic properties we would like to make contact with are *emergent* in the sense that they become meaningful only when the causal set is configured in an appropriate way. Concepts such as length, topology, and dimension make little sense for a generic causal set; so it is necessary to understand in what circumstances they do emerge. We would like to be able to recognize when such circumstances prevail, and equally important, to learn how in such cases we can read information about the effective continuum geometry directly out from the underlying order itself. To appreciate the difficulty of this task, it is enough to glance at figure 1, which shows three fairly simple causal sets whose dimension is two in a sense we will make more precise below.

A second task of this kinematical stage of investigation would be just to gain familiarity with the new mathematics needed to describe causal sets, much of which belongs to the branch of combinatorial theory devoted to partial orders. Unfortunately this branch of mathematics is unfamiliar to most physicists, and it is also true that the questions mathematicians have concentrated on are not always those which are likely to be the most relevant physically. So a significant period of kinematical development is likely to be needed before we possess the concepts which will help us survey the different possibilities for constructing a dynamics for causal sets.

The dynamical stage of development of causal set theory would, of course, be the one in which we would understand their "laws of motion", laws which are presumably quantum in character. But because time itself is discrete in this case, we could not hope to write down some Hamiltonian generator of time-evolution for the causal set. Indeed it is not even clear what 'configuration space' could mean for causal sets, and therefore unclear what Hilbert space such a Hamiltonian could act on. The only available framework to work with thus appears to be that of the sum-over-histories. In this framework our job is to assign an amplitude to each causal set, and then figure out how to sum these amplitudes in order to obtain physically meaningful probabilities.

In looking for the correct amplitude function, I think, we have a couple of requirements to guide us. Most obviously, there is the requirement that the dynamics should possess a classical limit in which the causal set resembles a smooth manifold.

5

Figure 1. Three causal sets faithfully embedded in $\mathbf{M}^2$. Only the irreducible relations or "links" are represented.

Since a classical limit arises from the constructive interference of many histories, this presumably means that Lorentzian manifolds should be in some sense, "stationary phase" points of the causal set amplitude. Conversely, causal sets which are far from looking like manifolds (and they are the vast majority) should be points at which the amplitude is in some sense "extremely rapidly varying". (Notice, however, that the amplitude-function can have no actual derivative because a causal set cannot be varied continuously. This means that a classical dynamics for individual causal sets will probably not be definable at all, and the actual dynamics will not be able to be viewed as having "arisen via quantization"—just as one would expect for a truly fundamental theory.)

The other requirement, related in more than one way to the first, is what I would call "locality". By this I mean that the effective Action contributed by the family of all causal sets corresponding to a given Lorentzian manifold (M, g), should turn out to be the integral of a locally defined quantity in M. For reasons related to their inherent Lorentz invariance, this is much more difficult to achieve with causal sets than with ordinary sorts of lattices. On the other hand, if it *is* achieved then one can give simple dimensional reasons why the needed *form* of the resulting integrand (namely, $-\Lambda + \frac{1}{2\kappa}R$, where Λ is the cosmological constant, R the scalar curvature, and $\kappa = 8\pi G$) will follow almost of its own accord. Hence the thing to focus on in searching for the right amplitudes seems to be not the detailed structure of the Einstein equation, but just its local character in spacetime.

In the remainder of this lecture I would like to survey some of the progress that has been made in constructing a theory of causal sets along the lines I have just indicated. To be consistent with the historical scheme I advocated earlier, I ought to spend all of my time discussing kinematics; in fact I can not resist talking a little bit about dynamics as well. But perhaps that is more than excusable in the present case, because the lack of clearly relevant experimental or observational evidence to guide us, means that we really will have little indication whether we have been traveling in the right direction until we reach the realm of dynamics proper.

From Discrete to Continuous

Logically, the first task of causal set kinematics should really be to verify that it is possible, even in principle, for such an elementary structure to give rise to a Lorentzian manifold in some suitable approximation. Because the whole theory would be stillborn were it not true, we call this belief the "Hauptvermutung", after a conjecture that at one time was thought to be central to the theory of manifolds (and turned out to be completely false!). In order to see in what way we could formulate a conjecture of this type, let us return to the basic correspondences involving volume and causal precedence from which the causal set hypothesis sets off. In effect these correspondences are telling us what it should mean for a manifold M with Lorentzian metric g to approximate a given causal set C, and the criterion they embody might be expressed as follows:

> The manifold (M, g) "emerges from" the causal set C iff C "could have come from sprinkling points into M at unit density, and endowing the sprinkled points with the order they inherit from the light-cone structure of g."

(Here unit density means with respect to the fundamental unit of volume, which is unknown as yet, but expected to be around the "Planckian" value of $10^{-139}\text{cm}^3\text{sec.}$)

Actually, I probably should have attached the qualifier 'modulo coarse-graining' to the words 'could have come', because one probably would expect realistic causal sets to be rather chaotic on scales so small that only a relatively few elements are involved. Such a "microscopically rough" causal set would not be strictly embeddable in *any* continuum spacetime, but some appropriately "averaged" or "coarse-grained" version of it might be. In fact the notion of coarse-graining appears to be important in its own right; for example effective topology will not in general be preserved by coarse-graining, and this can give a precise meaning to otherwise elusive notions such as "scale-dependent dimensionality" and "spacetime foam." (We will encounter a small example of this later). For simplicity, however, let me ignore, henceforth, any need for coarse-graining in order to gain embeddability.

Now if we interpret sprinkling to mean random generation of points according to a Poisson process in M then we can actually prove a limiting version of the Hauptvermutung [5]. Namely let C_1 be the result of one such process and let C_2, C_3, etc. be the results of sprinkling in additional points at ever higher densities (i.e. we are imagining the fundamental length to go to zero). Then $C_1 \subseteq C_2 \subseteq C_3 \ldots$, there is a well defined sense in which the limit can be taken, and we have with probability one that

$$(M, g) = \lim_{j \to \infty} C_j.$$

This theorem is encouraging, but in reality no such limiting process can occur since the fundamental volume is small but not zero. Luca Bombelli has made some progress in this direction, but difficulties remain, even in knowing how to formulate the desired theorem precisely[6]. For now, let me just indicate the kind of result we would hope for. Let C be a causal set and let $f : C \to M$ and $f' : C \to M'$ be faithful embeddings (i.e. embeddings which are sprinkling-like according to some suitable criterion) of C into (M, g) and (M', g'). Then there exists an approximate isometry $h : M \to M'$ such that $f' = h \circ f$.

Assuming — as seems very likely — that causal sets do possess a structure rich enough to give us back a macroscopically smooth Lorentzian geometry, it is important to figure out how in practice one can extract geometrical information from an order relation. But before we can speak of a geometry we must have a manifold, and the most basic aspect of a manifold's topology is its dimension. So an obvious first question is whether there is a good way to recognize the effective continuum dimension of a causal set (or more precisely of a causal set that is sufficiently "manifold like" for the notion of its dimension to be meaningful). In fact *two* workable approaches exist, and I would like to describe each of them for you, concentrating, however, mainly on the second one.

Flat Conformal Dimension

The first approach relies on the existence of families of relatively small causal sets which can serve to characterize the dimension of any *Minkowski* spacetime via embedding. Since in this case, the number of embedded points is small, one cannot speak meaningfully of their density in the "target" Minkowski space, so we will require only that an embedding induce the correct order relations on the embedded points, omitting any requirement that they be "distributed uniformly with unit density".

To get an idea of what kind of causal sets can characterize dimension in such a manner, take a look at the prototypes for dimensions 1 through 3 shown in figure 2. It is obvious that the first one will embed in $\mathbf{M}^1$ but not in $\mathbf{M}^0$ (where $\mathbf{M}^d$ is Minkowski space of d-dimensions); and it is equally obvious that the second one will go into $\mathbf{M}^2$ but not into $\mathbf{M}^1$. Similarly the third one will pretty clearly embed in $\mathbf{M}^3$ (think of the diagram as a perspective drawing), and almost as clearly not embed in $\mathbf{M}^2$, as can be rigorously proven with a little trial and error. Now it turns out that there exist, for each spacetime-dimensionality d, analogous causal sets which will embed in $\mathbf{M}^d$ but not in $\mathbf{M}^{d-1}$, and I would like to take a moment to describe how this comes about.

Perhaps the most natural guess for such a causal set in dimension d would be the so-called "binomial poset" B_d, which can be realized as the set of all subsets of $\{1, 2, 3 \ldots, d\}$ ordered by inclusion. This poset has 2^d elements and appears, when depicted in the manner of figures 1 and 2, as a d-dimensional cube balanced on one corner. (In particular the causal sets of figure 2 are essentially B_1, B_2 and B_3, except that the maximum element has been left off of B_2 and both the maximum

Figure 2. Three simple dimension-characterizing causal sets

and minimum elements off of B_3.) Just as one might expect, it turns out that B_d will not embed in $\mathbf{M}^{d-1}$, but unfortunately it is not clear (for general d) that it will go into $\mathbf{M}^d$ either!

However, by deleting certain relations from B_d one can remove the impediments to this last embedding, and thereby acquire causal sets capable of characterizing arbitrarily great spacetime dimensions[7–9]. The particular family I will discuss [9] has for its representative in dimension d the poset P_d made by retaining from B_d only those relations embodied in the following rule:

$$S_1 \prec S_2 \Longleftrightarrow S_1 \subseteq S_2 \;\; \textbf{and} \;\; |S_1| \leq 1 \text{ or } |S_2| \geq d-1.$$

The proof that, indeed, P_d embeds in $\mathbf{M}^d$ but not in $\mathbf{M}^{d-1}$ relies on the "trick" of representing an embedded causal set by its "shadow" on a spacelike hyperplane H lying entirely to its past. These shadows (the intersections of the past lightcones with H) are spheres, and the order relation $\prec$ becomes the relation of inclusion for them. In order to show that P_d *will* embed in $\mathbf{M}^d$, one thus needs to find a corresponding set of spheres in $\mathbf{R}^{d-1}$, and such spheres can be constructed explicitly. To prove that such a spherical realization *does not* exist in $\mathbf{R}^{d-2}$, one uses Radon's lemma, which asserts that d points in $\mathbf{R}^{d-2}$ cannot be configured in such a way that the convex-hulls of disjoint subsets of the points remain disjoint in all cases. (For example, four points in $\mathbf{R}^2$ form a quadrilateral, and the two diagonal-segments must meet.) Now suppose that spheres realizing P_d could be found in $\mathbf{R}^{d-2}$. One shows that the minimal elements of P_d (i.e. the elements $\{1\}$, $\{2\}$, $\ldots$ $\{d\}$) can be realized as single points (spheres of radius zero), and that, given any disjoint subsets A, B of these points, there exist corresponding spheres S_A and S_B such that S_A includes A but excludes B, and vice versa. This implies that hull(A) is disjoint from hull(B) in all cases, contradicting Radon's lemma.

To my mind, the most surprising suggestion to emerge from the search for such "dimension-characterizing causal sets" is that the binomial poset B_6 (and therefore all subsequent B_n) may not be embeddable in $\mathbf{M}^d$ for any d at all. The evidence for this rests on the facts that, on one hand, embedding is either possible in $\mathbf{M}^6$ or impossible altogether[9], and on the other hand, symmetric embeddings in $\mathbf{M}^6$ already have been ruled out[7]. It thus seems possible that complete impossibility could be established without too much further effort.

In the opposite direction, it is an interesting question whether one can also derive dimension-characterizing causal sets from B_n by adding relations instead of removing them. In particular one can map the cube representing B_n into $\mathbf{M}^n$ in such a way that all the edges become null, and add in the new relations thereby induced to obtain a new causal set D_n. If, as I would conjecture, D_n will not embed in $\mathbf{M}^{n-1}$, then it has the advantage over P_n of admitting a much more regular embedding into $\mathbf{M}^n$, making it more natural to use as a dimension indicator in the non-flat case (see next paragraph). In addition, this ability to embed regularly fits in nicely with a potential application which I will just mention, since it has hardly been explored: the possibility of constructing an analog of cubical homology for causal sets by using the D_n as building blocks.

The notion of dimension that results from considering embeddability in Minkowski space may be called "flat conformal dimension", because it would be insensitive to arbitrary conformal changes in the flat metric on $\mathbf{M}^d$. It does not directly apply to the kind of very large causal set we are typically interested in, because we certainly do not want to limit ourselves to the case where the continuum approximation to the causal set in question is free of curvature. On the other hand it is true by definition that any continuum manifold is flat on small scales, so it would be natural to try to extract the desired dimensional information by looking at suitable small subsets of our causal set. If among them we found instances of (say) P_d with all values up to some maximum d, then we could conclude that the effective continuum dimension of our causal set was this maximum d.

Statistical Dimension

The second method of recognizing dimensionality that I referred to has a statistical character; and, since it yields dimension values which need not be integers, I will call it 'fractal'. Unlike the method just discussed, this one uses volume information as well as conformal information, and seems to be somewhat more efficient for that reason. Actually, one can imagine several distinct methods of this ilk, all of which involve counting suitable substructures of a given Alexandrov subset of our causal set. The specific one I will describe (suggested independently by Myrheim and Meyer) is the only one which has been fully analyzed analytically, as far as I know.

Given an Alexandrov subset $[a, b]$, the simplest thing to count is the number N of its elements, and the next simplest is probably the number R of (nontrivial) *relations* it contains, i.e. the number of pairs of (unequal) elements $u, v \in [a, b]$ such that $u \prec v$. If the full causal set can be faithfully embedded in $\mathbf{M}^d$ then N and R can be estimated by forming expectation values with respect to a Poisson process taking place within $\mathbf{M}^d$. I will describe this in a moment, but first let me summarize the situation.

First notice that, by Lorentz invariance, the expectation values of R and N can depend only on the volume in $\mathbf{M}^d$ of the Alexandrov neighborhood ("double light cone") bounded by a and b, and, of course, on the dimension d itself. Forming a ratio to make the volume drop out gives us

$$\frac{R}{N^2} \sim \frac{3}{4} \frac{d!\,(d/2)!}{(3d/2)!}$$

Because the right-hand-side is a monotonically decreasing function of d, this equation can be inverted to associate a unique "Myrheim-Meyer fractal dimension" to each value of the ratio R/N^2. For suitable small Alexandrov subsets of a "manifold-like" causal set, this fractal dimension should average out to the true continuum dimension. For larger ones it would be necessary to take the curvature into account (c.f. [10]).

Now let me go into more detail about how the mean and variance of R may be found analytically. To that end, consider a Poisson sprinkling of points into $\mathbf{M}^d$, and let us select a pair of sprinkled points, a, b, such that the corresponding interval or "Alexandrov subset" $[a, b]$ contains precisely N elements. To find the mean value $< R >$ we may imagine dividing $A := J^+(a) \cap J^-(b) \subseteq \mathbf{M}^d$ into very small subregions dx, and defining for each pair of such subregions, the "Boolean" random variable $\chi(dx, dy)$ which equals one if and only if $x \prec y$ and both dx and

dy are occupied by sprinkled points. (In the limit of infinitesimal subregions, we may ignore the possibility that either dx or dy holds more than one point.) With these definitions we have (in the limit)

$$R = \sum_{dx,dy} \chi(dx, dy),$$

whose expectation-value yields immediately

$$< R > = \sum_{dx,dy} < \chi(dx, dy) >,$$

a result which owes its simplicity to the fact that summation commutes with expectation-value, even in the presence of correlations.

Now the expectation of $\chi(dx, dy)$ is just the probability that it equals 1, which in turn is just the product of the probability $P(dx \text{ occupied})$ with the conditional probability $P(dy \text{ occupied} \mid dx \text{ occupied})$; hence

$$< \chi(dx, dy) >= \frac{N}{V} dx \cdot \frac{N-1}{V} dy = \frac{N(N-1)}{V^2} dx dy,$$

whence, setting $\phi = R/N(N-1)$, we have

$$< \phi > = \frac{< R >}{N(N-1)} = \int_{x \prec y} \frac{dx}{V} \frac{dy}{V} =: I_2.$$

Thus the computation of $< \phi >$ reduces to the evaluation of the integral I_2, which, from the above derivation, may be described as (half) the probability that a randomly chosen pair of points in A are related to each other. Evaluating I_2 then leads to the result quoted above:

$$< \phi >= \frac{< R >}{N(N-1)} = \frac{3}{4} \binom{3d/2}{d}^{-1} =: I_2,$$

where the parentheses denote a binomial coefficient. (See [7] for the evaluation of a more general class of integrals of this type, two of which we will use below.) Notice that $I_2(d)$ is indeed a monotone function of d, allowing us to associate a unique (but in general fractional) d to each value of the ratio ϕ.

But how reliable, for a given N, can we expect this estimate of dimension to be? To answer this question we need the standard deviation of ϕ. In fact, it is more convenient to work with $\ln \phi$ and its standard deviation $\Delta \ln \phi \sim \Delta\phi/\phi = \Delta R/R$.

Then, in terms of the derivative of $\ln I_2$ with respect to d (call this derivative $f(d)$ for short), we can write the uncertainty in the dimension estimator as

$$\Delta d \sim \frac{\Delta \ln \phi}{f(d)} \sim \frac{\Delta R/R}{f(d)} \sim \frac{\Delta R}{R},$$

where I have used the easily verified fact that $f(d)$ is of order unity (varying in fact between -0.55 near $d=1$ and -0.955 for $d \to \infty$). Our estimated d can thus be regarded as reliable when the standard deviation of R is much smaller than R itself: $\Delta R \ll R$.

In order to evaluate ΔR we may once again express R as the sum

$$R = \sum \chi(dx, dy) =: a + b + c + \cdots,$$

where each of the letters a, b, $c \ldots$ stands for one of the Boolean variables χ. It follows immediately that

$$R - <R> = (a - <a>) + (b - <b>) + \cdots,$$

whence

$$(R - <R>)^2 = \sum_a (a - <a>)^2 + \sum_{a \neq b} (a - <a>)(b - <b>).$$

The first term's expectation-value (which just corresponds to the ordinary "$\sqrt{n}$-type fluctuations") is easily evaluated:

$$<(a - <a>)^2> = <a^2> - <a>^2 = <a> - <a>^2,$$

which implies

$$<\sum_a (a - <a>)^2> = \sum <a> - \sum <a>^2 = <R>.$$

Here I have used in the upper line that $a^2 = a$ (since $a = 1$ or 0) and in the lower line that $<a>^2$ is small of quadratic order, whence neglectable.

Most of the contribution to ΔR comes from the $a \neq b$ sum, however, which registers the enhanced fluctuations engendered by the mutual correlations among the variables $\chi(dx, dy)$. Its expectation-value

$$\sum_{a \neq b} [<ab> - <a><b>]$$

contains three types of contributions, depending on how the subregions involved in $a = \chi(dx, dy)$ and $b = \chi(dz, dw)$ overlap. In the "generic" case (i) x, y, z, w are all distinct, but they can also (ii) form a 3-chain (if $y = z$ or $x = w$), or (iii) form a "V" (if $x = z$) or a "Λ" (if $y = w$). Take for illustration the third case, with $x = z$: $a = \chi(dx, dy)$, $b = \chi(dx, dw)$. For the mean value of ab we then find

$$< ab >= P(dx, dy, dw \text{ all occupied}) = \frac{N}{V}\frac{N-1}{V}\frac{N-2}{V}dxdydw,$$

while for $< a >< b >$ we have (as earlier)

$$< a >< b >= (\frac{N}{V}\frac{N-1}{V}dxdy)(\frac{N}{V}\frac{N-1}{V}dxdw),$$

which is proportional to dx^2dydw and therefore negligible in the limit $dx \to 0$. The net contribution of this case to ΔR^2 is therefore

$$\sum_{dx,dy,dw} < \chi(dx, dy)\chi(dx, dw) >= \frac{N(N-1)(N-2)}{V^3} J_3,$$

where J_3 is the integral

$$J_3 = \int \frac{dx}{V}\frac{dy}{V}\frac{dz}{V}$$

taken over all triples of points such that $x \prec y \prec z$. Handling cases (i) and (ii) similarly, and combining results, we find for the variance of R

$$\Delta R^2 = N(N-1)[(2N-4)(J_3 + I_3) - (4N-6)I_2^2 + I_2],$$

where I_3 is the triple integral formed like J_3 but with x, y, z forming a 3-chain instead of a "V": $x \prec y \prec z$. (Incidentally, the evaluation of the integrals I_3 and J_3 reveals a peculiar identity between them, which—if it could be generalized—might aid in evaluating other integrals of the same sort. The identity is $I_3 = I_2 J_3$.)

For large N, ΔR thus goes like $N^{3/2}$, with a coefficient depending on dimension. If $d \gg 1$ as well, we can estimate the integrals I_2, I_3, J_3 using Stirling's approximation (all three integrals being generalized binomial coefficients) to obtain

$$\Delta R^2 \sim \frac{4}{3}\sqrt{\pi d}\, 4^{-d} N^3,$$

from which (dropping a prefactor proportional to $d^{-1/4}$)

$$\frac{\Delta R}{R} \sim \frac{1}{\sqrt{N}}\left(\frac{27}{16}\right)^{d/2}.$$

Hence

$$\Delta R \ll < R > \quad \text{if} \quad N \gg \left(\frac{27}{16}\right)^d .$$

For large d this is a big number, but not as big as the 2^d elements required to characterize flat conformal dimension by means of the special causal sets described earlier.

Computer tests of this method of recognizing dimension [10] show that it works reasonably well in low dimensions ($d = 2, 3, 4$), with the computed dimension converging rapidly to the true one as the number of sprinkled points becomes larger than about 300-400, the results being in general agreement with the error analysis given above. Figure 3 shows a typical case, in which a thousand points were sprinkled randomly into an Alexandrov neighborhood in $\mathbf{M}^3$, and a value of d computed for each interval within the resulting causal set. The plot depicts the distribution of the resulting values of d as a function of the size of the interval. Many similar pictures may be found in reference [10], including an especially interesting example in which the starting spacetime was a "Kaluza-Klein cylinder" of 1+1-dimensions. In that example one sees clearly how the effective spacetime dimension falls gradually from 2 down to 1 as the size of the interval in question increases. In effect, one sees how coarse-graining can induce "dimensional reduction", and how such scale-dependent dimensionality (and more generally, topology) becomes a perfectly well-defined concept in the context of causal sets.

Before leaving the topic of dimension, I want to mention three of the other possible "statistical" methods I alluded to earlier. So consider once again our Alexandrov neighborhood A in $\mathbf{M}^d$ of height T and volume $V = [(\pi/4)^{\frac{d-1}{2}}/d\,(\frac{d-1}{2})!]T^d$, and containing N randomly sprinkled points comprising an interval $[a, b]$ in the larger causal set. The "midpoint dimension" of $[a, b]$ (perhaps the notion closest in spirit to the Hausdorff dimension of a metric space) would be defined as $d = \log_2(N/N_1)$, where $N_1 = |[a, z]|$ (say) is the number of elements contained between a and a "midpoint element" z, which one might take to be the middle element in a longest chain from a to b, or alternatively (cf. [11]) an element that maximizes $\min\{|[a, z]|, |[z, b]|\}$ (with '$|\cdot|$' denoting cardinality, as earlier). One might also try to define a dimension by relating N (which is effectively an estimate of the volume of A) to T, as estimated, perhaps, in terms of the length of a longest chain from a to b (cf. below). Finally there is the intriguingly simple definition $d = \ln N/\ln\ln P$, where P is the total number of chains (or of paths) from a to b. There is good reason to believe that P has the large N behavior $P \sim e^{aT}$ [where $a = a(d)$], which if true would make this last formula an asymptotically valid estimator for large N (inasmuch as $N \propto T^d$

Figure 3. Scatter plot of the fractal dimensions of the intervals in a thousand-element causal set sprinkled into $\mathbf{M}^3$.

for all N). To put any of these methods on a sound analytic basis, one would need better control over random variables like the length of the longest chain from a to b, the total number of such chains, and the total number of paths from a to b (a path being defined as an unrefineable chain).

Length

Beyond recognizing dimensionality, there is the problem of extracting information about global connectivity (i.e. topology), and ultimately about the metric tensor itself. Concerning the latter, one may ask in particular how to estimate the geodesic distance between two causal set elements (assuming as always that the causal set resembles a continuum geometry (M, g).) Thus, let x and y be elements of the causal set C for which $x \prec y$. The most obvious way to define a "distance" from x to y is just to count the number of elements in the longest chain joining them, where a joining chain is by definition a succession of elements, $x \prec z_1 \prec z_2 \prec z_3 \ldots \prec y$. Clearly a maximal path in this sense is analogous to a timelike geodesic, which maximizes the proper-time between its endpoints.

Now, how does length defined in this way compare to geodesic length? Once again, most of the work has been on causal sets sprinkled into Minkowski space, but the results there are very encouraging. For maximal chains involving up to about twenty links, David Meyer found empirically that there seems to be a quite good proportionality between the Minkowski separation T of two sprinkled points, and the number of links L in the longest chain joining these points. What's more[11], in the asymptotic limit of large T, this proportionality becomes exact: $L = cT$. The constant c depends on the spacetime dimension and is not known exactly, although fairly tight bounds on its value are available.[11] If proportionality holds up when curvature is present, we will have a way to extract timelike distances directly from the order relation $\prec$. For spacelike distances, no equally direct method appears to exist, but there would seem to be a good prospect of getting them indirectly from the ensemble of timelike ones.

There are also other ways to deduce geometrical information from order information. For example, the type of counting technique we used to define fractal dimension will, in certain situations, yield the value of the scalar curvature, and in other situations the radius of the "internal circle" of a Kaluza-Klein vacuum-metric[7,10]. However all these methods have been developed only on the assumption that the causal set is faithfully embeddable in *some* continuum spacetime. What we still lack is a way to judge directly whether or not this is the case; i.e. whether the causal set can give rise to any Lorentzian manifold at all.

Counting Problems

Before leaving kinematics and turning briefly to the question of dynamics, I would like to describe some progress that has been made on what one can call "counting problems". We have already seen how geometrical knowledge can be gleaned just by counting suitable substructures, but I think such enumeration is likely to be important for another reason. I think it furnishes numbers in terms of which the amplitude entering into the sum-over-histories (or better, over causal sets) could be defined. In any case let me mention some counting problems whose solution is known, as well as some which are clearly important, but which have not been solved so far.

We have already encountered the notion of a chain in connection with the definition of timelike length. A natural question is how many chains exist between two specified causal set elements. For a causal set sprinkled into *two-dimensional* Minkowski space it is possible to evaluate in closed form the expected number of chains joining two sprinkled points separated by the time lapse T, and one finds <chains> $= \mathrm{I}_0(\sqrt{2}T)$, where $\mathrm{I}_0(x) = \mathrm{J}_0(ix)$ is a so-called modified Bessel function of the first kind. Notice in connection with this formula that mathematicians have defined 'chain' to mean *any* totally ordered sequence of elements, even if the sequence "skips over" intermediate elements. Thus, every subset of a chain is also a chain, and only a *maximal* chain need consist of "links" in the sense of *unrefineable* two-element chains.

One can also ask for the number of chains of a *fixed* length occurring as subsets of a causal set sprinkled into an Alexandrov neighborhood of height T in $\mathbf{M}^d$. In this case the expected number $< C_k >$ of k-element chains is known as a function of T for all dimensions d. Also known is a generalization of this formula to the case of constant spacetime curvature[7,10].

In this connection, I might mention some mysterious identities that have been found to relate certain expectation values of the sort we have been considering[7,10]. In $\mathbf{M}^d$ we have that $< V >< R >=< C_3 >< N >$, where 'V' stands for the number of three-element subsets such that one element precedes the other two, and 'C_3' stands for the number of 3-element chains. This identity (equivalent to the relation $J_3 I_2 = I_3$ mentioned earlier) relates the numbers of certain specified sub-causal-sets of a sprinkled causal set, and is rigorously true in any dimension. A similar, but so far only empirical identity has turned up in $\mathbf{M}^2$ in computer work by Jorge Pullin and Eric Woolgar; it states that $\langle L\rangle^2 = 2\langle W\rangle\langle N\rangle$, where now L is the number of links (unrefinable two-element chains) and W the number of "V's" each of whose "arms" is a link. (The letter 'W' stands for 'wedge', and such "wedges" will be mentioned again later.)

It would be easy to pose many similar counting problems holding equal interest to those just indicated, but I'm sure several such will occur to you without my mentioning them explicitly. However, I *would* like to mention here a slightly different enumeration problem, one with obvious bearing on the question of how, in the "classical limit of causal set theory", a smooth geometry emerges at large length scales from the underlying discrete order. That is the problem of determining the total number of distinct causal sets which can result from sprinkling N points into a given Lorentzian manifold (M, g). Clearly this is related to the question of how far a given sprinkling can be perturbed without changing the induced order, and it is possible to base a "guesstimate" of the number of sprinklings on this kind of consideration. Still needed, is a better justification of this, or some other estimate, and a refinement of it which would clarify how the number of distinct sprinklings depends on geometrical parameters of the region in question, like its volume, its curvature, and the shape of its boundary.

Amplitudes

There are some other kinematical issues which I would have liked to discuss had there been time, especially coarse-graining and the related possibilities of scale-dependent dimensionality, emergent matter fields, and the like. Rather than spend my remaining minutes on such topics, however, I would like to report on some intriguing results and preliminary work concerning the dynamical stage of causal set theory.

As I said earlier, the sum-over-histories framework appears to be the most natural point of departure for anyone trying to develop a plausible "quantum equation of motion" for causal sets. In this framework each causal set C would carry an amplitude $A(C)$; and dynamics would be contained in the amplitude-function $A(\cdot)$, together with the combining rules telling us how to use the amplitudes to construct meaningful probabilities. Without yet possessing these latter rules in their final form, we can still investigate some of the consequences of particular choices of $A(\cdot)$, and thereby try to gain insight into what features an adequate choice would have to have.

Consider, for example an amplitude of the form $A = \exp(i\beta R)$, R being the total number of relations, as before. This looks like a familiar path-integral amplitude (with trivial "measure" and) with R and β playing the roles of "classical Action" and "coupling constant" respectively. [Rescaling β, one might say that the (fractal) dimension was being used as the Action, in this case!] To prevent the amplitude-sums from diverging, one can take the total number of causal set elements to be a fixed integer N. Now, if we were to go over to the corresponding "statistical mechanics" problem by continuing β to imaginary values, then we would be dealing with a random causal set of N elements, and with probability-weight given by the "Boltzmann factor" $\exp(-\beta R)$. It happens that just this problem has been studied in connection with a certain "lattice-gas" model [12].

Actually what was studied was not exactly the "Gibbs ensemble" given by the "Boltzmann factor" $e^{-\beta R}$, but instead the corresponding "micro-canonical ensemble", in which R is fixed and every causal set with that R (and with N elements) is weighted equally. The first result of interest is that, in the "thermodynamic limit", $N \to \infty$, at least two, and probably an infinite number, of phase transitions occur as R/N^2 is varied (corresponding to varying βN^2 in the "canonical ensemble"). For small values of this parameter, the most probable causal sets possess only two "layers" (i.e. no chain has more than three elements), and the phase transitions mark thresholds at which successively greater numbers of layers begin to contribute. In some very general sense the causal set is thus becoming more manifold-like with each such transition, but so far there is no evidence that genuine manifold behavior sets in for any value of the parameter.

One noteworthy event that accompanies the 2-level to 3-level phase transition is a spontaneous breaking of time-reversal symmetry. In the 2-level phase the most probable configurations look similar to their T- (or better CPT-) reversals, but in the initial 3-level phase the causal sets of high-probability have very unequal numbers of elements in their top and bottom layers. I obviously would not want to claim that this effect was at the root of the cosmological time-asymmetry, but it does demonstrate the possibility that something of the sort could ultimately emerge from a better understanding of causal set dynamics.

The other preliminary results I want to report on here concern not the causal set's own dynamics, but rather that of a "'scalar field" living in a fixed (or "background") causal set. Such additional degrees of freedom might or might not be necessary to incorporate "matter", but in either case, their study should serve to clarify issues like locality which I would expect to influence critically the choice of amplitude for the causal set itself. So let C be a given causal set, and let ϕ be a real-valued function on C, or in other words a "real scalar field". One would like to discover an Action $S(\phi)$, defined purely in terms of ϕ and the order $\prec$, which in appropriate situations will reproduce the known behavior of a quantum scalar field in curved spacetime.

In what we have done so far[13], we have limited ourselves to causal sets sprinkled into 2-dimensional Minkowski space $\mathbf{M}^2$, and searched for a function $S(\cdot)$ which would approximately reproduce the Action functional of a free scalar field. To that end, we considered within C both links and "wedges". (Recall that the latter are triples of elements x_0, x_1, x_2 such that the pairs $x_0 \prec x_1$ and $x_0 \prec x_2$ are both links, a structure I referred to earlier in connection with counting problems.) Letting ϕ_1 denote $\phi(x_1)$, etc., I can form the quadratic expression $S_W = -\sum(\phi_1-\phi_0)(\phi_2-\phi_0)$ as a sum over all wedges, and in the same way $S_L = -\sum(\phi_1-\phi_0)^2$ as a sum over all links. We chose the first expression because it resembles the square of the gradient expressed in u-v-coordinates, if you take the two links of the wedge as rightward-null and leftward-null. The second expression also resembles the continuum Lagrangian to some extent, but clearly can't be correct by itself since it can only have one sign. Now one can argue that there should exist *some* linear combination of S_W and S_L that will yield (on average) the correct Action for arbitrary *linear* ϕ. What is surprising, however, is that the simple difference $S = S_W - S_L$ appears to be the combination that works (modulo an overall normalization depending on N). The tests are still unfinished, but for the three test fields $\phi \equiv t, \phi \equiv x$, and $\phi \equiv t + x$, we find errors in $< S >$ of about 15% for 500 points and 6 runs, and of about 1% for 10,000 points and 10 runs.

For some nonlinear functions, including $\phi \equiv x^2$ and $\phi \equiv \sin(t)$, the results are again not too bad, but for more rapidly oscillating ones like $\phi \equiv \sin(5t)$ they are decidedly not too good, and they don't get noticeably better as the number of sprinkled points is increased. (Here the points are sprinkled into an Alexandrov neighborhood whose height in the t-direction is unity.)

In fact the difficulty was to be expected as a manifestation of a very general conflict among locality, discreteness, and Lorentz invariance that I alluded to earlier. What happens in this case is that the links and wedges, which we would like to be rather "small", can instead be "very long and skinny". For example, a link which in one reference-frame looks to be purely timelike and of small size, will in a highly-boosted frame appear to be very stretched out and almost null. By including such links (and the analogous wedges) in our expression for $S(\phi)$ we make S depend on very large finite differences instead of only on small ones that could furnish a good approximation to the gradient of ϕ. In fact S_W alone would in some sense already have been a good approximation to S had not this problem been present, and one can probably understand the effect of subtracting S_L as a partial cancellation of the contribution of the "long skinny wedges". What is needed, then, is a full cancellation of their contribution, leaving behind an expression for the Action that one could expect to be accurate for a much wider range for test functions.

In fact, there is at the moment, the prospect of obtaining such a cancellation by using "diamonds" instead of "wedges"; and there is also a very promising simpler approach, which, however, works most naturally only in in dimensions two and four. But the main point I want to make here is that locality in this sense might be the lodestar leading us to an appropriate dynamics for the causal set itself, and not just for whatever fields may inhabit it. By achieving locality for "matter" fields like ϕ we could of course acquire a new type of "random lattice" approximation to quantum field theory in curved spacetimes (including as a special case Minkowski space itself of course). But more importantly for our present purposes, we might then begin to understand how to build local amplitudes for the causal set itself. Indeed it could even turn out that the resulting ϕ-dynamics would directly yield a useful effective causal set amplitude in the manner of "induced gravity", that is, via "integrating out" the ϕ-fluctuations in the sum-over-histories.

That is about all it seems appropriate to say on dynamics for now, but before concluding, I want to throw out a quite speculative idea that illustrates how the assumption of a discrete substratum for spacetime can suggest otherwise unexpected routes to the resolution of basic puzzles like the smallness of the observed cosmological constant. Suppose, in fact, that this smallness of Λ were a *statistical effect* due to the very large number of elements making up the relevant portion of the universe. Then we might expect whatever cancellations were responsible for this smallness to take place not exactly, but just with some statistically imposed accuracy, and this suggests the formula

$$\Lambda \sim \frac{1}{\sqrt{N}},$$

N being the number of elements in question. Taking the fundamental volume to be of Planck magnitude, and estimating the 4-volume of the observable universe between the "big-bang" and the present epoch, yields an N of the order of 10^{240}. Then Λ would be somewhere around 10^{-120} in natural units, which (coincidentally?) is currently the largest value not yet ruled out by astronomical data.

This research was partly supported by NSF grants PHY 9005790 and 8918388.

References

1. Bombelli, L., Koul, R.K., Lee, J. and Sorkin, R.D., "A Quantum Source of Entropy for Black Holes", *Phys. Rev.* **D34**, 373-383 (1986); 't Hooft, G., "On the quantum structure of a black hole", *Nuclear Phys. B* **256**:727-745 (1985).

2. Bombelli, L., Lee J., Meyer, D. and Sorkin, R.D., "Spacetime as a causal set", *Phys. Rev. Lett.* **59**, 521 (1987); Moore, C., "Comment on 'Space-time as a causal set'," *Phys. Rev. Lett.* **60**, 655 (1988); Bombelli, L., Lee J., Meyer, D. and Sorkin, R.D., reply to C. Moore in *Phys. Rev. Lett.* **60**, 656 (1988); Bombelli, L. and Meyer, D.A., "The origin of Lorentzian geometry," *Physics Lett.* **A141**, 226 (1989).

3. Bombelli, L., *Space-time as a Causal Set*, Ph.D. thesis, Syracuse University (1987);
Sorkin, R.D., "A Specimen of Theory Construction from Quantum Gravity", in *How Theories are Constructed: The Methodology of Scientific Creativity* (Proceedings of the Thirteenth Annual Symposium in Philosophy, held Greensboro, North Carolina, March, 1989) (to appear).

4. Finkelstein, D., "The space-time code", *Phys. Rev.* **184**, 1261 (1969);
Myrheim, J., "Statistical geometry," CERN preprint TH-2538 (1978);
't Hooft, G., "Quantum gravity: a fundamental problem and some radical ideas,", in *Recent Developments in Gravitation* (Proceedings of the 1978 Cargese Summer Institute) edited by M. Levy and S. Deser (Plenum, 1979).

5. Meyer, D.A. and Sorkin, R.D., "On Poset Completion", in preparation

6. Bombelli, L. and Sorkin, R.D., "When are Two Lorentzian Metrics close?", in preparation

7. Meyer, D.A., *The Dimension of Causal Sets.* Ph.D. thesis, M.I.T. (1988).

8. Meyer, D.A., "The Dimension of Causal Sets I: Minkowski dimension," Syracuse University preprint (1988);
"Spherical containment and the Minkowski dimension of partial orders," Syracuse University preprint (1989).

9. Brightwell, G. and Winkler, P., "Sphere Orders", *Order* **6**, 235 (1989).

10. Meyer, D.A., "The Dimension of Causal Sets II: Hausdorff dimension," Syracuse University preprint (1988).

11. Brightwell, G. and Gregory, R., "The Structure of Random Discrete Spacetime", *Phys. Rev. Lett.* **66**:260-263 (1991).

12. Dhar, D., "Entropy and phase transitions in partially ordered sets", *J. Math. Phys* **19**:1711-1713 (1978); and "On phase transitions in posets", *Pacific J. Math.* **90**: 299-305 (1980);
Kleitman, D.J. and Rothschild, B.L., "A Phase Transition on Partial Orders", *Physica* **96A**: 254-259 (1979).

13. Daughton, A., Pullin, J., Sorkin, R.D., and Woolgar, E., work in progress.