PHYSICAL REVIEW
LETTERS

VOLUME 59 | 3 AUGUST 1987 | NUMBER 5

Space-Time as a Causal Set

Luca Bombelli,[1] Joohan Lee,[1] David Meyer,[2] and Rafael D. Sorkin[1,3]
[1]*Department of Physics, Syracuse University, Syracuse, New York 13244*
[2]*Department of Mathematics, Syracuse University, Syracuse, New York 13244*
[3]*Institute for Advanced Study, Princeton, New Jersey 08540*
(Received 8 April 1987)

We propose that space-time at the smallest scales is in reality a causal set: a locally finite set of elements endowed with a partial order corresponding to the macroscopic relation that defines past and future. We explore how a Lorentzian manifold can approximate a causal set, noting in particular that the thereby defined effective dimensionality of a given causal set can vary with length scale. Finally, we speculate briefly on the quantum dynamics of causal sets, indicating why an appropriate choice of action can reproduce general relativity in the classical limit.

PACS numbers: 04.60.+n

By comparison with most other scientific theories, "quantum gravity" is in a very special, and somewhat unhappy, situation. In the schematic evolution of Taketani and Sakata,[1] a physical theory passes through three stages: an initial stage in which a particular "substance", or type of matter, presents itself in a characteristic group of phenomena; a second stage in which the new substance in question is clearly discerned in relation to the phenomena; and a final stage in which the comprehensive dynamics characterizing this substance is understood.[2] In contrast, quantum gravity is forced to skip virtually the whole of the first stage, and tackle the second and third stages together, hoping that the resulting theory will enable us to recognize with hindsight what features of already-known physics can provide its "phenomenology."

In an attempt to guess what these features are, people have asked questions such as "Why are there four space-time dimensions?" "Why is the cosmological constant so small?" "Why is our universe so large?" "Why are there (nearly massless) fermions?" and "Why are there no holes in space-time?" and have hoped thereby to uncover important clues to the nature of the quantum physics underlying classical space-time. Among such questions the most suggestive for our purposes is Riemann's query,[3] "Why is there a spatial metric?" which, in the light of later developments in physics and different geometry, one must now amend and augment, changing "space" to "space-time", and asking in addition the two supplementary questions: "Why is the metric of Lorentzian signature?" and "Why does space-time itself have the topological and differentiable structures that allow a metric field to be defined on it?"

The beginning of an answer to this trio of Riemannian questions is suggested by the fact—insufficiently appreciated in our view—that a classical space-time's causal structure comes very close to determining its entire geometry. By the causal structure of a space-time, one means the relation P specifying which events lie to the future of which other events. Ordinarily, one thinks of a space-time as a topological manifold M, endowed with a differentiable structure S, with respect to which a metric g_{ab} is defined. Then the causal order P is regarded as derived from the light cones of g. However, one can also go the other way[4]: Given a space-time obeying suitable smoothness and causality conditions (and of dimensionality >2), let us retain from all its structure only the information embodied in the order relation P. Then we can recover from P not only the topology of M, but also its differentiable structure, and the conformal metric, $g_{ab}/|\det g|^{1/n}$. Now a partial ordering is a very simple thing, and it is natural to guess that in reality g_{ab} should be derived from P rather than the other way around. The problem with this is that P lacks the infor-

© 1987 The American Physical Society

mation needed to determine the conformal factor $|\det g|^{1/n}$. In a manner of speaking, we get from P the metric, but without its associated measure of length (or better, volume).

There seems to be no way to overcome this problem within the context of continuous space-time, but on the other hand there are many reasons to doubt that space-time is truly continuous, including of course the infinities of quantum field theory and the singularities of general relativity. If instead we postulate that a finite volume of space-time contains only a (large but) finite number of elements, then we can—as Riemann suggested[3]—measure its size by *counting*. If this is correct, then when we measure the volume of a region of space-time, we are merely indirectly counting the number of "point events" it contains. No attempt to "pack more points into the same volume" could change their density, because it would only increase the physical volume of the region in which they were placed.

We thus arrive at the view that the new "substance" (or better, structure) underlying space-time is what Riemann might have called an "ordered discrete manifold," but we will call a "causal set." In this view volume is number, and macroscopic causality reflects a deeper notion of order in terms of which all the "geometrical" structures of space-time must find their ultimate expression.

What recommends this idea to us is not so much its simplicity, but the way it is potentially able to address questions like those quoted above. Indeed it already tells us that the metric must be Lorentzian, because no other signature than $(-++\ldots+)$ has the two-napped light cones from which the continuum's causal order is derived.

The picture of space-time as a causal set is by no means new, but to our knowledge, previous proposals along this direction[5] either have remained undeveloped[6] or, if pursued further, have led to formulations[7] in which the issues we would like to address here were not dealt with: How is this picture related to our ordinary one of space-time as a smooth manifold, and, eventually, will one be able to recover general relativity in the classical limit? In trying to deal with these issues, practical difficulties soon come up, but we will see that it is possible to formulate some reasonably detailed conjectures, some of whose proofs will appear elsewhere.[8]

Before proceeding further, let us put the notion of causal set into mathematically precise language. A *partially ordered set* (or "poset" for short) is a set C provided with an order relation, $<$, which is transitive (i.e., $x<y<z \Longrightarrow x<z$) and noncircular (i.e., $x<y<x \neq y$ is excluded: no "closed timelike curves"). It is also customary to adopt the convention that $x<x$, in which case the ordering $<$ is called reflexive. A partial ordering is *locally finite* if every "Alexandroff set" $A(x,y)$ contains a finite number of elements, where $A(x,y):=\{z\,|\,x<z<y\}$ is the intersection of the future of x with the past of y. A *causal set* is then by definition a locally finite, partially ordered set.

Let us now take up this view that space-time is a causal set, and try to relate it to the picture of space-time as a continuum. Suppose that we have an arbitrary causal set C containing many elements—at least 10^{130}, say—and we want to see which manifold-with-metric, if any, it looks like at large scales. Then we should seek a manifold M (with a time-oriented Lorentzian metric g_{ab} free of closed timelike or null curves), and an embedding $f{:}C \rightarrow M$ of the causal set into the manifold, such that the following conditions are satisfied: (1) The causal relations induced by the embedding agree with those of C itself, i.e., $f(x) \in J^-(f(y))$ iff $x<y$, where $J^-(p)$ denotes the set of points of M to the causal past of p; (2) the embedded points are distributed uniformly with unit density[9]; and (3) the characteristic length λ over which the continuous geometry varies appreciably is everywhere much greater than the mean spacing between embedded points.[10]

In specifying in condition (2) that the points be embedded with *unit* density we have assumed that the metric g is expressed in natural units, i.e., ones in which $\int(-g)^{1/2}dx$ directly counts elements of C. Ultimately the theory (generalized, if necessary, to incorporate nongravitational matter) should predict atomic radii in natural units, and then one will be able to say precisely how many elements of C make up a conventionally defined space-time volume, such as 1 $\mathrm{cm}^3\cdot\mathrm{sec}$. On dimensional grounds, one expects this number to be near the Planck density of $7\times10^{138}\ \mathrm{cm}^{-3}\cdot\mathrm{sec}^{-1}$.

Let us call an embedding which satisfies the above three conditions "faithful." It is clear that an (M,g) in which we can embed C faithfully need not exist at all (in fact, it probably almost never exists); but if it does, then our discussion up to now leads us to expect that it is essentially unique. In other words, we can expect that any pair of faithful embeddings, $f_1{:}C \rightarrow (M_1,g_1)$, $f_2{:}C \rightarrow (M_2,g_2)$, are related by a C-preserving diffeomorphism $h{:}M_1 \rightarrow M_2$, which is an approximate isometry of g_1 to g_2. (By C-preserving we mean $f_2 = h \circ f_1$). A precise formulation and proof of this statement would establish rigorously that the continuum approximation is well defined, and therefore that a causal set has a structure rich enough to imply all the geometrical properties we attribute to continuous space-times.

As a start in proving such a statement, we will now sketch an argument that the space-time dimensionality must be the same for all faithful embeddings of a given causal set C. To begin with, let $f_1{:}C \rightarrow (M_1,g_1)$ be a fixed faithful embedding of dimension n_1, and let $A=A(x,y)$ be an Alexandroff subset of C, which is "small" in the sense that its image by f_1 has a linear size of order unity with respect to g_1. Then the corresponding Alexandroff neighborhood A_1 and M_1 will also be small in this sense; whence it must be approximately isometric to an Alexandroff neighborhood in n_1-dimensional Minkowski space, since otherwise the geometry in it would induce length scales incompatible with condition (3). {Here A_1 is the Alexandroff neighborhood in M_1 between $f_1(x)$ and $f_1(y)$, not the finite set $f_1[A]$, of course.} Now one can show[8] using condition (2) that, with great likelihood, some such "small" A will have the property that it cannot be embedded in fewer than n_1 flat dimensions. But since for any other faithful embedding, f_2, the corresponding neighborhood A_2 will also be approximately flat, we must have $n_2 \geq n_1$; and therefore $n_1 = n_2$ by symmetry.

From the point of view of a continuum (M,g) in which C is faithfully embedded, the latter resembles a "random lattice"[11] obtained by "sprinkling in points" until Planckian density is reached. If one took such a picture too seriously, the question might seem to arise of why we stop sprinkling at a specific density, rather than continuing to some higher, or even infinitely high, value, but of course such a question would be meaningless because M is only an approximation to C, and not vice versa.

Although we have so far based our notion of continuum approximation on embedding a causal set into the approximating manifold, one might anticipate that small-scale fluctuations in the causal order will prevent physically realistic sets C from being embeddable in any manifold, (M,g). To handle this more general possibility, it is useful to introduce a notion of coarse graining. A coarse graining C' of a causal set C will be a subset of C, endowed with the ordering obtained by our restricting to the subset the order relation of C. But since we wish to use coarse graining to obtain a larger-scale view of the original causal set C, not any subset C' is appropriate. Rather C' should be "representative" of C, as it would be, for example, if its elements had been selected randomly from those of C with probability p. Such a coarse graining can be interpreted as preserving only those features of C which have characteristic volume scale larger than $1/p$. Thus, it can happen that a causal set which cannot be embedded faithfully in any manifold may become embeddable when coarse grained. This could happen, for example, if the original causal set contained different regions looking like Kaluza-Klein space-times with different "internal" dimensionalities, and the coarse graining were on a scale bigger than that characterizing the internal metrics. Notice in this connection that, even if C *can* be faithfully embedded in some manifold, coarse graining with a small enough p can wash out any small-scale structures which lend C a specific effective topology, leaving C' with a simpler effective topology than C, including perhaps a lower space-time dimensionality.

We see, then, that the dimensionality of a causal set can vary with scale in elaborate but, nonetheless, well-defined ways. For example, a causal set which looks four dimensional at large scales might, in precisely the same sense, look eight dimensional at smaller scales. Under a still lesser degree of coarse graining, it might comprise three large domains of dimensionalities 10, 11, and 26, while in itself (i.e., with no coarse graining at all) it might not even have any meaningful dimensionality. In the same way, other topological features (holes, handles, geons,[12] "foam," etc.) might appear at one degree of coarse graining, without necessarily being present at either larger or smaller scales. This is a kind of behavior that has often been imagined, but is difficult or impossible to describe with use of only the concepts of manifold and metric.

If a causal set really does underlie the continuum, then it will no doubt come into its own in situations where the continuum description *already* seems to break down, such as inside a black hole, in the very early universe, or in topology-changing processes. It is easy to see that, here also, causal sets possess the necessary kinematic flexibility; but no further conclusions can be drawn at the level of Taketani's "stage 2," on which our reflections so far have remained. Let us conclude, then, with a few thoughts on a dynamical (or "stage 3") question: how—and for what choice of action—can ordinary general relativity be expected to arise in the classical limit?

To suggest an answer is difficult without having first specified a form of quantum dynamics appropriate to causal sets. For the present discussion, let us merely assume that a quantum amplitude is defined for every causal set C, and that the classical limit will result in the familiar way from the condition of constructive interference among the amplitudes of those causal sets contributing to a given continuum geometry (M,g). Now let us begin with such a geometry and consider the collection of causal sets C which can be faithfully embedded in it. Suppose that the amplitude of a typical such C is "multiplicative," in the sense that it is approximately the product of the separate amplitudes of any (still macroscopic) regions into which C may be divided. In that case the action (by which we mean of course the *logarithm* of the amplitude) must take the form in the continuum approximation of an integral over M of a locally defined scalar (density). But since the quantum "sum over histories" washes out the anisotropy associated with any single embedded C, the resulting effective action density can only depend on the metric g_{ab} itself (in other language, the result must be "generally covariant").[11] Therefore it can be expanded as

$$L_{eff} = L_0 + L_2 + L_4 + \dots,$$

where L_0 is just a constant (the effective cosmological constant), L_2 is a multiple of the scalar curvature, L_4 is a sum of curvature-squared terms, etc. But since the action itself must have the dimensions of $\hbar$, the coefficients occurring in the term L_{2k} would be "expected" to be of the order of $\hbar l_{\text{Planck}}^{2k-n}$ in ordinary units, and for a continuum of n space-time dimensions. Then neglecting the relatively miniscule terms $L_4 + L_6 + \dots$ leaves us with $L_0 + L_2$, or precisely the usual general relativity Lagrangean *with* cosmological term, L_0.

The reason why L_0 does not by the same token dominate L_2, rendering it negligible as well, cannot be understood purely from the above considerations, which apply, *mutatis mutandis,* in many theories for which Einstein gravity represents a low-energy approximation to some more complete dynamics (induced gravity,[13] higher derivative gravity,[14] or string theory[15]). For us here the important point is that it seems to be easy to suggest amplitudes that have the necessary multiplicativity, and therefore that can be expected to yield something very close to general relativity in the classical limit. In another place we will propose a specific form for such an amplitude, and examine the approximate large-scale action to which it gives rise. On the basis of the result—and of a certain interpretation of the quantum sum over histories for gravity—we will then try to derive the smallness of the cosmological constant from the effect on this sum of order fluctuations of the kind referred to earlier: fluctuations altering the effective dimensionality or leading to causal sets for which the notion of effective dimension does not even make sense.

We would like to thank A. P. Balachandran, Nicola Cabibbo, the Syracuse University Relativity group, and, especially, Pankaj Jain for stimulating discussions. This research was supported in part by the National Science Foundation Grant No. PHY-8700651.

[1]See articles by M. Taketani and S. Sakata, in Suppl. Progr. Theor. Phys. **50**, (1971).

[2]A good recent example is the theory of hadrons, with the three stages being respectively the observation of resonances, the discovery of quarks, and the formulation of QCD.

[3]G. F. B. Riemann, in *Über die Hypothesen, welche der Geometrie zugrunde liegen,* edited by H. Weyl (Springer-Verlag, Berlin, 1919).

[4]S. W. Hawking, A. R. King, and P. J. McCarthy, J. Math. Phys. **17**, 174–181 (1976); D. Malament, J. Math. Phys. **18**, 1399–1404 (1977); for a general discussion of causal structures of space-times, see E. H. Kronheimer and R. Penrose, Proc. Cambridge Philos. Soc. **63**, 481–501 (1967).

[5]There have also been proposals in which a partial ordering has been used as fundamental structure, but with a directly topological meaning, rather than a causal one. See, e.g., R. D. Sorkin, in *General Relativity and Gravitation, Vol. 1,* edited by B. Bertotti, F. de Felice, and A. Pascolini (Consiglio Nazionale delle Ricerche, Roma, 1983), and to be published.

[6]G. 't Hooft, in *Recent Developments in Gravitation, Cargèse 1978,* edited by M. Levy and S. Deser (Plenum, New York, 1979); J. Myrheim, CERN Report No. TH-2538, 1978 (unpublished).

[7]D. Finkelstein, Phys. Rev. **184**, 1261–1271 (1969).

[8]L. Bombelli, P. Jain, J. Lee, D. Meyer, and R. D. Sorkin, to be published.

[9]The notion of "uniformly distributed" can be slightly tricky to define, but, roughly, we mean that every Alexandroff neighborhood $J^-(p) \cap J^+(p)$ in the manifold contains a number of points of C equal to its volume, within the Poisson-type fluctuations which could be expected from a random "sprinkling" of points.

[10]Notice that condition (2) on the density would not help us to determine a unique approximate metric if we did not also have condition (3) on the characteristic length λ: Given any manifold with the right causal structure, i.e., conformal metric, we could always arrange the density to be unity by setting the conformal factor appropriately; but in doing so, we would in general introduce an unreasonably large curvature, or other small characteristic lengths. However, it seems plausible that conditions (1) and (2) alone determine the continuum geometry "up to arbitrary variations on small scales, and small variations on arbitrary scales" (where small scale means size unity or smaller).

[11]T. D. Lee, in *How Far Are We from the Gauge Forces?,* edited by A. Zichichi (Plenum, New York, 1985).

[12]R. D. Sorkin, in *Topological Properties and Global Structure of Space-Time,* edited by P. G. Bergmann and V. de Sabbata (Plenum, New York, 1986).

[13]S. L. Adler, Rev. Mod. Phys. **54**, 729–766 (1982), and Philos. Trans. Roy. Soc. London A **310**, 273–278 (1983); A. D. Sakharov, Dokl. Akad. Nauk. SSSR **177**, 70–71 (1967) [Sov. Phys. Dokl. **12**, 1040–1041 (1968)].

[14]T. Tomboulis, Phys. Lett. **70B**, 361–364 (1977); K. S. Stelle, Phys. Rev. D **16**, 953–969 (1977).

[15]A. Sen, Phys. Rev. Lett. **55**, 1846–1849 (1985); C. G. Callan, D. Friedan, E. J. Martinec, and M. J. Perry, Nucl. Phys. **B262**, 593–609 (1985); R. I. Nepomechie, Phys. Rev. D **32**, 3201–3207 (1985); W. Fischler and L. Susskind, Phys. Lett. B **171**, 383–389 (1986), and **173**, 262–264 (1986).