

# SPACE-TIME AS A CAUSAL SET

by

LUCA BOMBELLI

Laurea, Università degli Studi di Milano,  
Milano, Italy, 1980

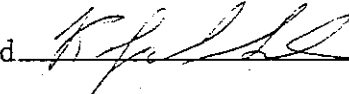
M.S., Syracuse University, 1983

## ABSTRACT OF DISSERTATION

Submitted in partial fulfillment of the requirements  
for the degree of Doctor of Philosophy in Physics  
in the Graduate School of Syracuse University  
December 1987

Approved

Date



19 Jan 88

This thesis describes a proposal for the structure of space-time at the smallest scales. The underlying new "substance" is what we call a causal set, a locally finite set of elements endowed with a partial order relation. It is conjectured that, when suitable causal sets on a large number of elements are considered, there exist unique lorentzian manifolds (up to small changes in the metric), in which the causal sets appear as uniformly distributed points, with metric-induced causal relations which agree with the partial order relation, and which are approximately flat on the length scales determined by the density of embedded points. These manifolds are free of causality violations and time-orientable, and provide a causal macroscopic interpretation of the partial order relation.

In an outline of the procedure for constructing the manifold associated with a causal set, we start by looking at small causal sets, thought of as embedded in the larger ones as subsets, which already contain the dimensionality information. We then propose ways of calculating effective dimensionalities for large causal sets, and of using the global structure of the causal set to determine the topology and other properties of the manifold, if it exists. For most causal sets, we expect this procedure not to yield any manifold, because no good embedding, in the sense described above, can be produced. In some cases, however, a suitable coarse-graining of the causal set can give a new causal set which does admit a good embedding, together with a set of "fields" on this causal set. The continuum approximation to the coarse-grained causal set will then consist of a manifold with metric and additional fields, and properties of the geometry and fields will depend on the degree of coarse-graining. In particular, the effective dimensionality can vary with length scale.

Dynamics is formulated in the sum over histories approach; which causal sets actually contribute most to the total amplitude, and whether these do have a well-defined continuum approximation, will then depend on the choice of basic amplitude for each history, and on how we define the class of histories we sum over. Provided such an approximation exists, a general argument is given, independently of the details of the dynamics, indicating that we can expect general relativity to be reproduced in the classical limit. A few possible choices for the quantities defining the dynamics are proposed.

*¿Para qué llamar caminos  
a los surcos del azar?*

Antonio Machado

## Acknowledgements

My first thoughts of gratitude go to the Syracuse relativity group, to which I owe more than I can say. In this group, I have found an uncommonly friendly atmosphere and a wealth of stimulating ideas.

I am especially grateful to the two faculty members I most closely worked with. Rafael Sorkin was not just a great advisor: working with him was a real pleasure, his thoughts have deeply influenced my view of physics, and my interaction with him, both as a physicist and as a person, was a continuous illuminating experience. Abhay Ashtekar made my previously hidden dream of working in relativity seem realizable; his style and approach to problems have since become a model for me, and he has been an invaluable source of encouragement.

Joshua Goldberg was always available and helpful, and provided many useful and interesting conversations. But not less important for me were the things I learned from, and the collaboration and friendship with, all other members of the group: Arvind Borde, Demetrios Christodoulou, Jerrold Garcia, Henry Kandrup, Rabinder Koul, Joohan Lee, Pawel Mazur, David Meyer.

I am very grateful to the Syracuse experimental high energy group for the use of the computer and of the  $\text{T}_{\text{E}}\text{X}$  program I used for this thesis.

## TABLE OF CONTENTS

Acknowledgements	iii
Table of contents	iv
1 INTRODUCTION	
1.1 Why look for a new fundamental structure of space-time	1
1.2 Quantum gravity and radical approaches to space-time structure	11
1.3 Our program	26
2 KINEMATICS OF CAUSAL SETS	
2.1 Posets: some definitions and properties	31
2.2 Metrics and causal structures on differentiable manifolds	44
2.3 Causal sets approximated by differentiable manifolds. I	52
2.4 Causal sets embedded in Minkowski space	61
2.5 Causal sets sprinkled in Minkowski space	72
2.6 Notions of dimensionality of a causal set	82
2.7 Causal sets sprinkled in arbitrary space-times	87
2.8 Coarse-grainings of causal sets	95
2.9 Causal sets approximated by differentiable manifolds. II	103
3 DYNAMICS OF CAUSAL SETS	
3.1 Causal sets and sums over histories	107
3.2 The amplitude: general considerations	116
3.3 The amplitude: possible expressions	122
3.4 Computer calculations	128
3.5 Conclusions	133
A APPENDICES	
A.1 Two-dimensional calculations with the "diamond lattice"	140
A.2 An application of simulated annealing	146
A.3 Sprinkling points in a given background	157
References	161

## 1. INTRODUCTION

### 1.1 Why look for a new fundamental structure of space-time

The idea that the basic structure underlying all of physics is that of a smooth 4-dimensional space-time manifold of simple topology has dominated the development of classical physics, quantum mechanics and classical general relativity, although there are big differences in the structures different theories have attributed to space-time *in addition* to that of smooth manifold (see, e.g., Penrose [15]). In fact, this development is largely due precisely to the mathematical tools of calculus offered by the continuum picture, and in turn it has strengthened our belief in the correctness of this picture. Certainly, our everyday life experience suggests a topological structure, and 4-dimensionality and trivial topology are the only possibilities for it, compatible with what we see at all scales in the range of our observations and experiments. The same cannot be said, however, about the manifold nature of space-time. The definition of a manifold is based on that of the real number continuum, and it has been remarked (see, e.g., Schrödinger [19], Penrose [16c]) that it is not only our everyday experience which has shaped the present concept of continuum, but also the particular form of calculus which was developed starting with Newton and Leibniz. Thus, we could consider as a cultural prejudice or "historical accident" the fact that we are using this notion of continuum and of manifold, based on a nonconstructive and nonintuitive definition of real numbers, as opposed to, e.g., using the so-called "non-standard analysis" as a basis for our theories; we accept the usual notions just because of

the familiarity we have with them.

How can our view of space-time structure be affected by an increased understanding of large- and small-scale physics? A better knowledge of the large-scale structure of space-time can only lead us to revise its topology. In fact, it is widely accepted that, on cosmological scales, the universe as a whole may very well be topologically nontrivial (in particular, it may be spatially closed). But this possibility can be, and has been, easily accommodated into our theories. What I wish to discuss here is the small-scale structure.

On experimental grounds, the conventional space-time picture holds well down to scales of the order of the smallest distances probed in high energy interactions, about  $10^{-16} \text{ cm}$ , and it is believed that it will continue to hold for many more orders of magnitude.<sup>†</sup> This picture, however, has been more and more often criticized on theoretical grounds in recent years, and many physicists now believe that it is only an approximation to a more fundamental structure. This new structure is expected to reveal itself in regimes where quantum gravitational effects become important, i.e., at length scales of the order of the Planck length,  $\ell_P = (G\hbar/c^3)^{1/2} \approx 1.6 \cdot 10^{-33} \text{ cm}$ , the only constant with dimensions of length one can construct using the fundamental physical constants  $\hbar$ ,  $c$  and  $G$ .

### Conceptual reasons

Conceptually, there are several reasons for believing in the existence of a more fundamental structure, although not all of them point along the same direction. First of all, in any attempt to unify the principles of quantum mechanics with

---

<sup>†</sup> We might notice, furthermore, that any possible observational indication that the smooth manifold picture is not adequate would be very indirect, and dependent on specific models for the interactions through which the effect was observed, since all the relevant observations themselves are made through a series of processes, all of which have length scales much greater than the possible manifold-breaking length, and would thus have a smoothing effect on the results.

general relativity, one is immediately led to the conclusion that, if space-time is indeed a lorentzian manifold, then the metric tensor must be subject to quantum fluctuations which become comparable to its expectation value at scales of the order of  $\ell_P$  (for a brief discussion, see, e.g., Misner, Thorne and Wheeler [50], §43.4). One qualitative way to see this, is the following well-known “uncertainty principle” argument. Suppose we wanted to make some measurement in a very small region of space-time. Then we would have to use probes with very small characteristic length and/or time scale, i.e., with very high momentum and/or energy. But this means having a large stress-energy tensor, which, in general relativity, produces a strong curvature, and thus a big distortion of space-time in the region of interest. We conclude that in a theory which includes both quantum mechanics and general relativity, the notion of a space-time point or of space-time metric will become fuzzy, signaling the breakdown of the conventional notion of space-time.

Second, we have Kaluza-Klein theories, motivated by the desire to find a unified theory of gravity and other interactions (originally, in the 20’s, electromagnetism, now all gauge theories). In these theories, space-time is a manifold  $\mathcal{K}$  of dimensionality greater than 4, which has (approximately) the structure of a fiber bundle over a 4-dimensional manifold  $\mathcal{M}$ . The size of the fiber (the “internal manifold”) is assumed to be of the order of  $\ell_P$ , thus, at large scales, the only structure we see is that of the 4-dimensional base space  $\mathcal{M}$ . If one starts with a dynamical theory of the metric alone in the full Kaluza-Klein space-time  $\mathcal{K}$ , described by an action of the same form as the Einstein-Hilbert action in general relativity, then in the low energy limit this induces on  $\mathcal{M}$ , a metric, and a set of scalar and gauge fields, the latter coupled to the metric in a similar way to the Yang-Mills couplings. Despite some phenomenological difficulties, it



is believed that these theories, or some of their modified versions, contain good indications of a geometrical origin of gauge interactions, and that, therefore, the 4-dimensionality and trivial topology of space-time is probably not correct at small scales.

Third, one could argue that, if we hope to eventually reduce physics to geometry (although, somehow, not everybody does), in order to reproduce the known physics, which is not scale invariant, the geometry has to include a fundamental length scale, and such a scale cannot arise from a smooth manifold (at least of relatively simple topology), with a dynamical metric, without any further structure.

Finally, the belief in the ultimate simplicity of nature makes a picture based on the continuum of real numbers, with lots of other structure added on, very suspicious: as Finkelstein [6] asks, why is space-time so complicated? We will see another class of conceptual reasons at the end of this section. The main practical reasons, however, that initially got physicists to start worrying about the correctness of the conventional description of space-time, have to do with difficulties in quantum field theory and general relativity.

### Technical reasons

In quantum field theory, calculations of amplitudes for physical processes yield infinities, which in some cases can be removed, by a combination of more or less ad hoc techniques, regularization of divergent integrals and renormalization of various quantities defined in the theory. The source of the problem in most cases lies in integrals in momentum space which extend to arbitrarily large values of the momenta, i.e., to arbitrarily small length scales. The simplest way to make the integrals converge is to introduce a momentum cutoff: a fundamental length scale. This scale could, but does not have to, be identified with "the

smallest possible length": it just indicates that something in the structure of the theory changes at small scales. For some theories, results of calculations may not depend crucially on how the change occurs, as long as we introduce in the parameters of the theory a suitable dependence on the above length scale (we call these renormalizable theories), whereas for some other theories this effective approach does not work: there is no fundamental reason for preferring one kind of theory over the other, and in both cases there is something about their small-scale behavior to be understood.

In general relativity, starting with the Schwarzschild solution for a black hole, an isolated object, and the Friedmann solutions in the cosmological context, singularities in the metric have been often appearing in the solutions of the field equations. While initially it was thought that these were a spurious consequence of the high degree of symmetry of the exactly known solutions, the singularity theorems of Hawking and Penrose proved that singularities will generically arise in many physically reasonable contexts. But a singularity signals the end of the region of applicability of Einstein's equation: the evolution of the gravitational field in its future light cone is completely undetermined. Thus, because of our reluctance to accept that points of infinite curvature can actually arise, and given that we would like to have, as an ultimate theory of gravity, a theory with a wider range of applicability, it has become important to look for ways to avoid singularities. From previous experience, we can expect a quantum theory to smooth out the singularities of its classical counterpart, just like it avoided the classical problem of an electron falling on the atomic nucleus by "spreading out" the electron to a wave function with support on an extended region around the nucleus. In our case, it seems natural to expect it to spread out the curvature that classically leads to a singularity over an extended region, and we are led to

look for a quantum theory of space-time geometry.

### First proposals

These difficulties were realized a long time ago. In particular, while a sufficient understanding of singularities in general relativity to appreciate the generality of this phenomenon in the theory was developed only in the late 60's, quantum field theory evolved rapidly enough that in the 30's physicists were already looking for fundamentally discrete structures for space-time,<sup>†</sup> which would help to solve the divergence problems (see, e.g., Ambarzumian and Iwanenko [2]; Silberstein [20]; Snyder [21]; Hellund and Tanaka [9]). These very first theories were not Lorentz invariant, since, e.g., in some cases space was made discrete, but time remained continuous, and soon new ideas came out which attempted to avoid this shortcoming. Schild [18] proposed a model in which space-time is a discrete set, obtained as a regular lattice in Minkowski space, which is left invariant by a discrete subgroup of the Lorentz group. The problem with this proposal is that one did not recover in some large scale view the invariance under (a good approximation to) the full Lorentz group. For example, in the simple case of a cubical lattice in 3+1 dimensions, the allowed discrete boost parameters are  $\beta = (n^2 - 1)^{1/2}/n$ , where  $n$  is a non-negative integer, so, in particular,  $\beta_{min} = \sqrt{3}/2$ . One does not see therefore how the Lorentz symmetry group can be recovered in a continuum approximation, although it is still possible to accommodate slowly moving particles in any spatial direction in the theory, since the spatial projections of the integer (null or) timelike vectors are dense in  $\mathbb{R}^3$ , and one can imagine a slow motion arising in a similar way to the Zitterbewegung of a Dirac particle.

---

<sup>†</sup> It is amusing to notice that it was thought that this discrete structure would show up at scales of the order of the nuclear dimensions.

Coish [4] proposed a more elaborate structure in which space-time points are specified by giving their cartesian coordinates, which however are not real-valued, but take their values in a very large but finite commutative field, i.e., the Galois field of prime order  $GF(p)$ . This proposal is not equivalent to an infinite cubic lattice in Minkowski space because of the periodic boundary conditions, but, for an adequately huge  $p$ , a sufficiently large number of points will approximate well some portion of Minkowski space: for  $p \sim 10^{10^{81}}$ , we can get a reasonable approximation to the usual geometry from  $10^{-13}$  cm to  $2 \cdot 10^9$  light years. All this is again rather artificial, but the interesting part about it is that, if we consider the "Lorentz" group of symmetries of the "metric form"  $-x_0^2 + x_1^2 + x_2^2 + x_3^2$  (remember, everything is  $GF(p)$ -valued), its representations include  $2 \times 2$  complex matrices, which have an action of a  $(p+1)$ -fold "covering" of this group, in a similar manner to  $SL(2, \mathbb{C})$  as a double covering of the ordinary Lorentz group. The generator of the extra transformations, which induce the identity of space-time, and can thus be considered as gauge, is identified by Coish with the electric charge. Although one should justify such an identification by looking at the couplings of the fields involved, the essential idea is that one can look for the origin of the internal symmetries observed in nature, in an alternative way to the Kaluza-Klein program: from a discrete fundamental structure, which gives these symmetries together with general space-time covariance in the continuum approximation.

Several other similar ideas at the same "pre-quantum gravity" level have been proposed. Some of them (e.g., Coxeter and Whitrow [5], Hill [10], Ahmavaara [1]) are roughly along the same lines as the ones described above, others (e.g., Takeno [25]) are more quantum mechanical. However, I will not go into their details here.

## Comments

The above proposals might rid us of the ultraviolet divergences of quantum field theory, although, from the point of view of gaining a deeper understanding of the physics involved, they do so in a rather trivial way. But, besides the question of whether they reproduce the correct “continuum limit”, the major conceptual criticism they are subject to is that they do not incorporate the spirit of general relativity. Taken in a restrictive sense, this means first of all that one should have tried to incorporate not just Lorentz invariance, but full diffeomorphism invariance in the discrete theory.<sup>†</sup> To implement this invariance, one could propose a theory of space-time as a lattice with a “random” structure, in which the distances between sites are not fixed, but dynamically determined, and which is not embeddable in a manifold with a metric fixed a priori. Theories of this kind, based on Regge calculus, a tool invented for doing calculations in a discretized version of general relativity, have been proposed, e.g., by Lee [74] (who introduced the concept of random lattice), Itzykson [72], and Lehto, Nielsen and Ninomiya [13]. One first chooses a “link structure” for the discrete set of points, which gives it enough rigidity that it already determines the topology of any continuum space-time manifold which will ultimately approximate it—in practice one usually makes it into a simplicial complex—, and then prescribes a dynamical theory which assigns lengths to all the links between points. Going over to the continuum, these lengths give the metric on the manifold.

---

<sup>†</sup> To put it another way, one can hardly expect to find the correct fundamental description of “pregeometry” by starting from symmetry considerations, just like one would not have found the description of physical geometry given by general relativity if one had followed, instead of the Riemannian approach to geometry, that of Klein’s Erlangen Programme, even though the latter had been fruitful in other areas, and it later proved itself useful again in the study of symmetries arising in the theory, and in particular in the use of fiber bundles. (One might say that general relativity really is a theory with the diffeomorphism group as symmetry group. But the concept of symmetry involved is not derived from known symmetries of euclidean geometry; rather, it is a generalization of the euclidean concept to a new structure.)

But there is a deeper sense in which one can try to follow the spirit of general relativity. General relativity taught us that the metric of space-time is not to be thought of as a fixed background on which physics takes place, but it participates in the dynamics. However, the space-time metric is just one mathematical structure sitting on top of a tower of constructions, and it becomes legitimate to ask why these other constructions should be taken as part of the background, fixed once and for all. One can thus ask a similar "tower" of questions: "Why is the space-time metric of lorentzian signature?", "Why does there exist a metric at all?", "Why can't the differentiable structure and topology of space-time be dynamical entities?", "Why is space-time four-dimensional?", "Why does it have a differentiable and topological structure in the first place?"

Once we start asking those questions, almost in the same spirit we might ask: "Why are the internal symmetries we observe in nature present?". Although this question might seem less geometrical than the previous ones, we have already had hints that the distinction may not be all that clear-cut. We might then continue with more "phenomenological" ones, like: "Why is the cosmological constant so small?", "Why is our universe so large?", "Why aren't there holes in space-time?", "Why are there (nearly massless) fermions?".

In the context of these questions, our search for a more fundamental structure for space-time takes on a whole new meaning, and we see why trying to base a theory on continuum language, on concepts like lattice embedded in Minkowski space or discrete subgroups of the Lorentz group, and in general keeping ideas borrowed from manifolds or 4-dimensionality in our models, won't do. What is needed instead is a "new substance", with which to build a theory of space-time. We use the term "substance" to stress the fact that this theory should make no reference, in its formulation, to facts that the questions above seek to

explain (except for motivational purposes), and should be logically independent of the concepts used in them. It can be thus compared to the introduction of new particles, with dynamics of their own, in a new fundamental description of matter.

The search for such a substance implicitly assumes that there is a real physical basis for the geometric structure we assign to space-time. This is contrary to the conventionalist view of geometry (see, e.g., Reichenbach [65], Grünbaum [64]), according to which we can assign any geometry to space-time (e.g., the metric does not need to have a lorentzian signature, or be related to the matter content by Einstein's equations, and some authors—e.g., Sextl [66]—make similar considerations even regarding the topology and dimension of space-time), as long as we choose appropriately our physical laws and prescriptions for associating geometrical quantities with measuring rods and clocks. In this view, choosing a geometric structure is something like choosing coordinates or a unit system, and the choice is just a matter of convention, although the one made in general relativity is the most convenient. If we can derive a geometric structure of the continuum in a natural way from a new substance, we will have a stronger argument than the continuum theory can give (see, e.g., Earman [63]) for the more empiricist view of physical geometry.

## 1.2 Quantum gravity and radical approaches to space-time structure

Questions along the lines of those posed at the end of the previous section had been asked long before the appearance of the physical motivation for them. Particularly suggestive for us is Riemann's query: "Why is there a spatial metric?". Riemann [19] attempted to start answering the question by remarking that if space were not a "continuous manifold" but a "discrete manifold", then there would be a natural notion of volume, or more precisely of ratio of volumes. Although one cannot construct a metric without additional structure just on a discrete set of points, the mere fact that space was discrete would explain the existence of a notion of volume: the number of points contained in a given region of space. Riemann's idea went unnoticed to physicists, still busy as they were in discussing whether even matter is discrete and made of atoms or not, and at the same time it preceded the development of the notion of space-time metric, without which, as we can see a posteriori, it could not develop into a prescription for deriving even just the spatial metric from a "discrete manifold".

Nowadays, a promising theory one can turn to, in the attempt to answer those questions, is quantum gravity, as we can see from the general remarks on quantum gravity and on singularities in general relativity I made in the previous section. In practically all the work on classical general relativity, and most of that in quantum gravity, however, there are still background elements which have not "joined the actors" in the play of dynamical evolution: the differentiable structure of the manifold, its topology, its dimensionality, its very manifold nature. This is largely due, in addition to the technical difficulties, to the rather different situation in which quantum gravity has found itself, as compared to most other developing physical theories. According to Taketani [70], the evolution of



a physical theory can be schematically divided into three stages: an initial stage in which the phenomenal effects of a particular "substance", or type of matter, are observed; a second stage in which the new substance in question is clearly recognized; and a final stage in which the comprehensive dynamics characterizing this substance is understood.<sup>†</sup> Quantum gravity, however, is forced to skip virtually all of the first stage and tackle the second and third stages simultaneously, hoping that the resulting theory will enable us to recognize with hindsight what features of already-known physics can serve as its "phenomenology".

In a sense the situation in which quantum gravity finds itself is a consequence of its lack of phenomenological basis, but the latter may not be merely a coincidence due to inadequate experimental techniques. Whereas some quantum gravitational effects (related, e.g., to black hole evaporation or properties of gravitational radiation) might in principle lead to observations, they are rather indirect and cannot point in a clear way to the "new substance" involved. Possibly, all characteristic effects of the latter will, even in principle, be observable only in terms of the "higher-level structures" used to describe macroscopic physics, and we might not be able to separate the identification of the new substance from the understanding of the mechanism by which it effectively gives rise to such structures, i.e., the understanding of its fundamental quantum dynamics and its "semiclassical approximation".

To some extent, the distinction between the second and third stages of Taketani's scheme corresponds to that between kinematics and dynamics of a theory, and the remark in the previous paragraph reminds of the suggestion that the latter distinction may also no longer be that sharp for quantum gravity, i.e., that the dynamics of quantum gravity may be built in, intimately woven

---

<sup>†</sup> 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.

into its kinematical structure. A theory which realizes this—not very precisely formulated—possibility has been looked for by physicists for some time already, attracted by the idea of bringing about such a conceptual unification in the principles of physics, but we are still far from a satisfactory proposal. As we will see shortly, even some of the most radical proposals for the fundamental space-time structure, including our own, have not been sufficiently developed yet, or have been constructed with an essential distinction between a kinematical part and a dynamical, quantum mechanical one.

The considerations above are not meant to detract from the value of those proposals, since, to a great extent, the possibility of making small modifications to one aspect of general relativity, while retaining enough of its usual structure to allow us to use many known results and techniques, is what has allowed us to get some clues about features that a full theory of quantum gravity should possess. Among the many illustrations of this fact, let us see a few which are directly relevant for us. The dynamics of Kaluza-Klein theories, including their supergravity variants, takes place in fixed background manifolds, but of very different topology from the usual one: although these theories are relatively “harmless” from the point of view analyzed here, as I remarked earlier they have introduced and “popularized” the idea that four may just be an effective dimensionality of space-time. Many studies have been done by now on topology change in general relativity (see, e.g., Geroch [33], Tipler [39], Yodzis [41], Sorkin [36], Borde [31] and other references therein). Most of the results in the subject are of a kinematical kind, i.e., they are topological conditions under which a space-time will exist which interpolates between two spacelike hypersurfaces of different topology, or they give restrictions on the curvature of the manifold, which, if we require the metric to satisfy in addition Einstein’s equation, become restrictions

on the matter stress-energy tensor. We have however obtained many indications about how topology changing processes—essentially quantum phenomena—will appear in the semiclassical limit, their connection to causality violations, and the possibility of using them to create particles out of space-time geometry, in the form of *geons*, “kinks” or “knots” in space (see, e.g., Sorkin [37]).<sup>†</sup> Several ideas about quantum geometry similar to those about geons were formulated by Wheeler [40], who gave fancy names to various effects expected to arise in his geometrodynamics vision. One of these is that of space-time *foam*, in which space-time appears smooth and nearly flat only at large scales, but is actually highly curved and has a complicated, continuously changing and irregular topology at Planck scales, although it is still described by a differentiable manifold. This idea was later developed further by Hawking [34], who showed that in the so-called euclideanized path integral approach to quantum gravity the dominant contribution to the amplitudes could come from manifolds “with about one unit of topology per Planck volume”.

Despite what was said above, some truly “backgroundless” ideas on how to deal with the question of a more fundamental structure for space-time have appeared. We now wish to describe some of these ideas. The starting question of course is: what feature of our conventional picture of space-time can be isolated as most fundamental and generalized, and what will this picture consequently be replaced by? The attempts to answer this question may differ, roughly, in two aspects. On the one hand, they may be more or less geometrical, as opposed to algebraic, in the sense that the elements of the basic structure they propose have a more or less direct correspondence with local elements of geometry. On

---

<sup>†</sup> Here by the term “geon” I mean a particle-like object made out of non-trivial spatial topology, as opposed to what Wheeler called a “geon”, which is essentially a lump of gravitational radiation, with no special topology involved.

the other hand, they may differ with respect to the amount of emphasis placed on incorporating into the structure parts of quantum mechanics from the very beginning, without formulating first the kinematics of the theory and then applying some quantization method. The first distinction probably has a mostly practical value, in the sense that a more geometrical proposal may be easier to work with (both because it is intuitively more immediate, and because one can see more directly how to implement in it the features one wishes to recover in their continuum approximation, like locality—more on this in §3.2), but it may not be as meaningful conceptually, since all proposals in the end are some kind of abstract mathematical structure, whose elements can never be identified with geometrical points. On the other hand, quantum mechanics by itself seems to be more of a logical-algebraic structure than a geometrical one, and there is thus a tendency for theories that do put in quantum mechanics at the beginning to be less geometric in character. A further potential distinction might be the extent to which “ingredients” of our continuum physics other than the purely geometrical ones, i.e., other fields and their symmetries and interactions, are seen to emerge from a unifying fundamental structure (in practice, though, none of the ideas we will see deals directly with this question). But all proposals agree in using simplicity of the fundamental structure as one of the criteria, and try to reduce all basic calculations to simple operations like counting: ultimately, physics could be reduced to combinatorics. Let us keep these aspects in mind while reviewing the proposals below.

### **Penrose’s spin networks**

One of the more fundamentally quantum mechanical ideas is that of Penrose’s [18] spin networks. Its main goal is to base all of quantum physics on simple—in principle—combinatorial rules, and free it from the notion of contin-

uum of real and complex numbers, even as regards the use of complex-valued probability amplitudes associated with physical processes or histories. The simple model described by Penrose is a first approximation to a full theory—still to be developed—of space-time geometry as emerging from a more fundamental structure, although some qualitative arguments can be given with regard to a more complete theory.

In the model, the space-time notion that is recovered is that of spatial directions around various points. One thinks of the universe as a large *spin network*, which can be represented by a (not necessarily connected) graph with three edges or *units* per vertex (with some units terminating at two vertices and some having a free end), in which to each unit is associated an integer, the *spin number*, to be thought of as its total angular momentum in units of  $\hbar/2$ . When the spin of a given unit is large, we can think of this unit as defining some axis in space. Since we are using total spin, it does not give us, by itself, its direction, but we can give a prescription for calculating its angle with respect to another nearby large unit, with some uncertainty depending on properties of the interconnecting network. In this way, appropriately chosen local subnetworks, in which large units are connected in such a way as to give small uncertainties give rise to local notions of direction, seeds for the construction of a local notion of flat geometry. The way in which one subnetwork is related to another one tells us how the directions defined by one are related to those of the other. Roughly speaking, we can think of the units and the directions associated with them as geodesics, and a local set of directions, the “intersection” of geodesics, as a (fuzzy) space-time point, even if we cannot yet talk about its position. Typically, the angle between a given direction in a local geometry and one in another geometry will not be well-defined, but will appear as “fuzzy”. This effect can be attributed to curvature present

between the two regions, which manifests itself in the bending of geodesics.

How does dynamics come into the picture? There are rules which tell us what the probability is of having a certain value for the spin number of a free unit in a network, once all the other spin numbers are known. These rules, which essentially reproduce the quantum mechanical rules for combining angular momenta (e.g., for a network consisting of three units meeting at a vertex, they give the quantum mechanical rule for addition of angular momenta), assign different frequencies of occurrence to various subnetworks in the universe: they are the physical laws. The rules themselves are not very simple to apply, and the simplicity of principles one has obtained is paid for in complexity: to calculate the angle between two units of a given subnetwork or its probability of occurrence, one has to calculate so-called "values" and "norms" of parts of it, in which the combinatorics can get complicated, and for which Penrose has developed a special algebra of "binors" (for details, see [18d]).

Thus, the proposal claims to build quantum dynamics into the rules for making spin networks. There is still however a clear distinction between kinematics and dynamics, since the combinatorial probability rules are added on to the prescription for drawing allowed networks, rather than following naturally from them. Furthermore, the 3-dimensionality of space is also added by hand, both in the probability rules and—probably—in those defining the allowed networks.

The model was developed in this way to take advantage of the properties of angular momentum, chosen as the fundamental concept because, besides being intimately related to the geometry of directions, its values are integers, multiples of a fixed unit,  $\hbar/2$ , and calculations done with it even in ordinary quantum mechanics are of a combinatorial nature. As a consequence, the requirement that the theory should account for the existence of spin- $\frac{1}{2}$  objects—which, as Penrose

stresses, is essential for a fundamental theory—is satisfied from the start.

The main reason why this is only a first model is that it assumes the units are all at rest with respect to each other. A more complete, relativistic one, in which relative motion is allowed, would probably illustrate how to construct the notion of position and distance as well as that of direction. At the same time, such a model might clarify the reason for interpreting the “spin number” of a unit as its angular momentum, which presently, if one considers just the theory in its abstract form, looks put in by hand.

This more complete model has not been developed, however, and work on the above ideas has led Penrose to the concept of *twistors*, which play, with respect to the conformal group, a similar role to that played by spinors with respect to the Lorentz group, and could therefore in principle lend themselves to a basically combinatorial formulation of conformal geometry, similarly to the spin network formulation of riemannian geometry (although the twistor theory that was developed is essentially based on complex analytic manifolds, rather than on combinatorial ideas). For an attempt to use the idea of a spin network, see, e.g., Hasslacher and Perry [8].

### Geroch's Einstein algebras

This proposal [7] (of an algebraic, non fundamentally quantum mechanical type) is motivated by the idea that the very notion of event or space-time point will probably cease to be meaningful in quantum gravity, and therefore one can try to formulate a theory which does not mention space-time as a collection of points at all, except as a derived concept. The starting observation is the “well-known” fact that the differentiable structure of a manifold  $M$  can be recovered from the knowledge of the ring  $R$  of smooth real-valued functions on it: each point of the manifold can be identified with a maximal ideal in the ring, namely

the set of functions which vanish there. As Geroch points out, given this ring  $R$ , one can characterize in terms of it vector fields, tensor fields, derivative operators (provided  $R$  has a distinguished subring  $S$  isomorphic to the reals, which is to be thought of as the constant functions on  $M$ , and is used in defining derivative operators in terms of  $R$ ), and metrics (i.e., symmetric isomorphisms between the module  $D$  of derivations on  $R$  and its dual  $D^*$ , on which we want to impose some additional requirement, to ensure that the operation of index contraction on tensors be well-defined). This structure is called *Einstein algebra*.

Geroch's idea is to consider such an algebraic structure as more fundamental than that of smooth manifold. Given any commutative ring  $R$  with the structure above that makes it an Einstein algebra, we can write down a condition which translates the Einstein equation in terms of it: the (once) contracted Riemann tensor (defined in terms of commutations of derivative operators) vanishes, or equals some desired tensor. If the algebra is one that can be constructed from a manifold, then this is nothing more than a rewording of Einstein's equation. But this condition can be imposed on the algebra independently of whether it really represents the ring of functions on a smooth manifold, and one could hope to be able to thus generalize classical general relativity to a case in which there is no space-time manifold.

This approach was never developed beyond the classical "equation of motion" stage, but, before one tries to develop some quantum framework for it, what remains to be checked is whether one actually gains something by considering Einstein algebras as a fundamental structure. It is possible, in fact, that the conditions required in their definition are so strong that they force the existence of an underlying manifold (see in this regard some related work by Yodzis [29]). If this were the case, a way to push Geroch's program further might be



to relax these conditions in a suitable way (a possibility which seems worth exploring is whether they would be automatically relaxed upon “quantization”). A different possibility might be to apply Geroch’s “algebraization” procedure to a formulation of general relativity which, already at the geometrical level, lends itself to a generalization, like the “new variables” approach to classical and quantum gravity of Ashtekar [30]. In this formulation, the constraint and evolution equations of general relativity are recast in a form which uses, as variables, an  $SU(2)$  spinor connection and a soldering form between  $SU(2)$  spinors and (spatial) tangent vectors defined on a spacelike hypersurface. The new form of the equations is equivalent to the traditional one when the soldering form is actually an isomorphism between the two spaces, but it is still meaningful even when the soldering form becomes degenerate. In that case the geometric interpretation of the equations changes, and it is possible that they can be taken to represent a more general situation, like topology change.

### 't Hooft’s proposal

In the less quantum mechanical, more geometrical proposal of 't Hooft [27], the fundamental structure is a discrete set of points, endowed just with a dynamically determined causal structure. This comes in the form of a partial order relation, in which the statement that an element is “greater” than another element is taken to mean that it is causally influenced, i.e., it lies to the future, of the latter. For convenience, using our own terminology, I will call such a structure a *causal set*.

The basic motivational observation is that, in the continuum, a causal structure is equivalent to a conformal structure. i.e., all of the metric except one “component”, the local volume element or conformal factor, while a discrete set has, as Riemann had already remarked, a natural notion of volume, given by

simply counting elements. 't Hooft points out that in fact in Minkowski space the metric, in the form of the straight line distance between any two points, can be simply expressed directly in terms of the causal structure and volumes (I will come back to this in §2.5). Now those expressions are meaningful for any set where causal relations and volume element are defined, in particular for causal sets, although in general we do not expect the “distances” thus defined to correspond to any notion coming from a metric on a manifold. However, we might expect that the causal sets which are “dynamically preferred” will admit a continuum “limit”, in the sense that they are embeddable in 4-dimensional Minkowski space, the relations between the points agreeing with the causal relations induced on them by the Minkowski space metric  $\eta_{ab}$ , and the distances calculated using the number of embedded points as volume agreeing, to some degree of accuracy, with those calculated using the metric volume.

Although 't Hooft did not have a concrete proposal for causal set dynamics, he remarked that the action would have to be a nonlocal expression, since discretized “general covariance” would force to treat in the same way all pairs of points in the lattice. This remark is correct if we replace “all pairs of points in the lattice” with “all pairs of causal nearest neighbors” (general covariance does not prevent us from using the available causal structure, just like in the continuum we are allowed to use the metric when we write down the action!), which still include pairs of very “distant” points.

### Myrheim's statistical geometry

A very similar (but independent) proposal to 't Hooft's was made by Myrheim [16] (and by Sorkin [25]; Sorkin's idea is the one on which our framework is based, so I will not describe it separately here), motivated by the same observations and with the same basic structure. Of all the previous approaches to a fundamental

space-time structure, Myrheim's is the one that has been most developed along the lines of the framework I will discuss in the following. Although he also did not discuss dynamics directly at the causal set level, Myrheim developed further the "kinematical" aspects, giving a construction for recovering the differentiable structure and metric on a manifold from its causal structure and volume element (which can be hopefully extended to discrete causal sets), and a (necessary, but not sufficient) condition that can be used to determine whether a causal set is embeddable in a space-time of a given dimensionality. Implicit in this condition is the idea that space-time dimensionality may only be an approximately defined concept, which could vary with length scale, and/or not be well-defined at some scales. Simple counting arguments then tell us that it is unlikely for a causal set to be embeddable in only 1 dimension ("phase space" says it is unlikely for a set consisting of many elements to have precisely that causal structure which makes it totally ordered), or in a very large number of dimensions (which would require a very "intricate" causal structure, as we will start exploring in chapter 2), but the precise dimensionality will have to be determined by more precise dynamical considerations.

As regards dynamics, Myrheim recasts the continuum vacuum Einstein equation in terms of volumes and the causal structure. We will come back to the details of Myrheim's proposal in the relevant sections of chapters 2 and 3, after the necessary concepts will have been introduced.

### **Other proposals**

There are (at least) a couple of other approaches which would be worth describing, but to which I shall devote only a few words. The first is the program, which reminds somewhat of Wheeler's (see MTW [53], box 44.5) "pregeometry as propositional calculus", undertaken by Finkelstein (before Wheeler's idea) in

a series of papers [6]. Unfortunately, I do not fully understand these papers, but a summary of my understanding is the following. The central idea is to put some notion of quantum logic at the basis for the construction of space-time. The latter is made of fundamental pieces, "quanta of time" or *chronons*, each one of which is generated by the "previous" ones using a set of logical rules, the discovery of which is what "breaking the *space-time code*" consists in. These space-time units form a *quantum space*, a  $*$ -algebra  $Q$  of operators on a Hilbert space, and they come equipped with a partial order, which is just the order of generation: space-time is a *quantum causal space*, i.e., a quantum space with a causal relation, where a *causal relation* is a transitive quantum relation between objects of  $Q$ . The task is to first discover the abstract, logical skeleton of the space-time code, and then translate it into a procedure for building up operators in  $Q$ , in the algebraic quantum language. The method used in this translation is what Finkelstein calls *algebraic quantization*.

The second program is that of constructing a theory of strings without a background, and several proposals have been made along these lines (some of the first references are [12] and [9]; see also [13]). I cannot comment on them, however, since I am not familiar with them.

The germs for a different approach lie in some developments in the study of quantum gravity in the already mentioned new variables. Solutions of some of the quantum constraints of the theory have been found by Jacobson and Smolin [35] in the form of states which have distributional support on fields concentrated on loops in the reference 3-manifold, and it is speculated that geometry (at least, 3-geometry) could arise in some sense as the superposition of many of these "loop states". Obviously, this is not yet a precisely formulated idea, and a lot of work has to be done to construct states which satisfy all the constraints, to investigate

in what sense they might be the most general solutions of these, and how they could be characterized independently, without using a 3-manifold to construct them.

Finally, another attempt, besides Finkelstein's program, has been made by Szabo [26] to build a fundamentally quantum mechanical theory, based on a non-distributive generalization of the lattice of space-time subsets, with a notion of causal structure defined on it. What Szabo tries to incorporate in his proposal is the non-Boolean nature of the lattice of propositions about a physical system in quantum mechanics (see, e.g, Jauch [77]). For other ideas, see also various papers contained in the volumes edited by Castell, Drieschner and von Weizsäcker [74].

### Comments

When one looks at the ideas above, one sees that there are some recurring, overlapping themes. Let us summarize and elaborate slightly on some of these:

(a) Discrete sets seem to be natural candidates for underlying sets of space-time structure. One avoids first of all the fact that all regions of space-time have "the same number of points", as in the continuum: which is to say that discrete sets have an intrinsic notion of volume. The "intrinsic metrical amorphousness" of the continuum, besides being unsatisfactory by itself, has supported to some extent the conventionalist view of geometry (for a discussion, see, e.g., Torretti [71], §7.2). Furthermore, discrete sets are in principle simple to do calculations with;

(b) It seems possible to derive the notion of space-time, and its geometrical properties including the metric, from other structures, which a priori might not even be geometrical in nature;

(c) The space-time one starts with, or derives from some other structure, does not need to be 4-dimensional or have a simple topology, and indeed these attributes

don't need to be defined, as long as there is a procedure for recovering, as effective large-scale geometrical properties of space-time, those we are accustomed to;

(d) A fundamental theory has to account for the existence of spin- $\frac{1}{2}$  particles, as well as the other particles and their symmetries; it is natural in particular to view spinors in some sense as more fundamental than tensors, since tensors can be "broken up" into spinors, the latter are also closely related to the conformal structure, and spin-1/2 particles (especially massless ones) are hard to get from geometry;

(e) A causal structure of space-time is a good candidate for a more fundamental structure than the metric itself; in particular, in the continuum case, as we will see, it uniquely determines the topological, differential and conformal structure, and it is only compatible with a metric of lorentzian signature; furthermore, it is naturally related to spinors (the geometrical interpretation of a spinor—up to a sign—at a point  $x \in M$  is that of a null flag—i.e., a null line in the Minkowski space  $T_x M$ , together with a half-plane bounded by this line, at least in 4 dimensions), it is simpler than the metric, and philosophically more appealing;

(f) As regards quantum dynamics of the "new substance", Wheeler [72] put it this way: "Surely the Lord did not on Day One create geometry and on Day Two proceed to 'quantize' it. The quantum principle, rather, came on Day One and out of it something was built on Day Two which on first inspection looks like geometry but which on closer examination is at the same time simpler and more sophisticated."

### 1.3 Our program

I want to examine here a proposal we have developed for the fundamental structure of space-time (at level 2 of Taketani's scheme, with some discussion at the third level) which incorporates several of the ideas described in the previous section. In particular, we want to base the mathematical description of a fundamental theory on simple operations like counting, i.e., on combinatorics, and to consider the causal structure as the feature of macroscopic geometry most directly related to the basic structure. I will argue that, if the new structure has a discrete underlying set, in addition to satisfying our first wish, we *gain* information with respect to the situation in the continuum case: all of the metric of the space-time manifold which approximates our structure will be recovered, and not just the conformal metric. It will be one of our major tasks for the future to formulate precisely this conjecture and prove it.

We thus propose that the matter underlying space-time has the structure of what we will call a causal set: a discrete set of elements with a partial order relation, which will be considered the "dynamical variable" of our theory. We will not, strictly speaking, attach any interpretation to this partial order relation, since it is based on a *new* substance, which does not have any precise correspondence with any large-scale concept. Loosely speaking, however, we will identify the structure with a causal order, which specifies which elements of the causal set, if we think of them as "events", are to the past and to the future of any given element, and which are unrelated to it. (This is similar to what happens in the kinetic theory of gases, where on large scales we use the notion of temperature, arising from the microscopic motion of the gas molecules, but we do not call "temperature" the kinetic energy of single molecules.)

We can think of space-time as *being* a causal set, although a proper quantum

mechanical treatment of its dynamics will imply that what we observe as physical space-time is not just one causal set but a whole class of them, in analogy to the fact that we cannot associate a single trajectory to the motion of a particle in quantum mechanics. The reason why space-time appears to us as a continuum is that, to this collection of dynamically preferred causal sets is associated an approximating smooth manifold  $\mathcal{M}$ , in which the discrete elements of each causal set appear as points uniformly “sprinkled in” with Planck density (we will see that this statement should actually be modified to allow such an approximating manifold—notice that I will never talk of a continuum “limit”—to be defined after a suitable coarse-graining of the causal set has been performed). The causal set uniquely defines a(n approximate) lorentzian metric on  $\mathcal{M}$  compatible with it in the sense that the intrinsic order of the causal set agrees with the partial order induced by the metric on the sprinkled points. Since the length scales at which we perform all our measurements are much larger than the Planck length  $\ell_P$ , we can never directly notice this discreteness, and it is a good approximation to consider space-time as a manifold for all practical purposes.

When we talk of dynamically preferred causal sets, we have in mind a formalism for dynamics based on some form for the action or amplitude for individual causal sets, whose value will be calculated by using just combinatorics (everything is a pure number), and with no adjustable parameters. In this regard, let us remark on a question which frequently comes up when one starts thinking about this approach to space-time geometry. We saw that the manifold picture of space-time is expected to break down at scales of the order of the Planck length. It might seem therefore that the scale at which this happens is really a parameter of the theory, and, even though we might not be saying it, we must be sneaking in a length scale somewhere: after all, how else would we recover in the



continuum approximation a theory with dimensionful quantities, starting from a purely dimensionless one based just on counting? Similarly, one might wonder why, when we imagined the continuum as uniformly sprinkled with points of the causal set, we stopped sprinkling points at Planck density, rather than some other value. The question therefore is: what tells us how many *cm* (say) the fundamental length is?

But one can see that this is actually the wrong question to ask, motivated by the habit of thinking in terms of the continuum as the fundamental structure, with the causal set points sprinkled in a previously defined geometry. (Somewhat paradoxically, this habit may be strengthened by familiarity with lattice gauge theories.) The point, of course, is that a causal set which is well approximated by a manifold *defines* there the natural unit of length: before the causal set, there was no metric on the manifold (if indeed there had been a manifold!), no scale with which to compare the spacing of points. The causal set theory (generalized, if necessary, to incorporate non-gravitational matter) should ultimately predict sizes of things, like atomic radii, wavelengths of radiation emitted during certain transitions, and so on, in natural units, i.e., in spacings between points. But a *cm* is just a conventional name for, e.g., a certain number of wavelengths of a certain kind of radiation, so the real question one should ask is: how many natural units will a conventional one like a *cm* turn out to be? This is the question whose answer, on dimensional grounds, we expect to be that the natural unit is approximately  $\ell_P \approx 1.6 \cdot 10^{-33} \text{ cm}$ .

We cannot yet say whether it will be necessary to add further structure to causal sets to account for the presence of the matter fields we observe in the continuum, but I will discuss a mechanism that might produce these fields using only the causal structure.

As a word of caution, I will emphasize that, if we look for observational consequences of the theory, we must not try to attach any operational meaning to the individual relationships between elements in the causal set, nor relate the intrinsic discreteness of the theory with any notion of discreteness that may arise from ordinary experience or lab experiments: any observational consequences of this framework will have to come through properties of the continuum approximation to it, although it may be that some of these properties find their only justification in the existence of the new discrete substance underlying space-time.

### Preview

The rest of this thesis is organized as follows. Chapter 2 is devoted to the relationship between causal structures (in particular discrete ones, of course), and geometric properties of continuum space-times, without dynamical considerations, i.e., to a study of the “substance” we are proposing, at Taketani’s second stage. In the first two sections I give some background on causal sets as mathematical structures and on known results about continuum causal structures and geometry of space-time. I then start the description of our own work on how, and to what extent, a discrete causal set determines a continuum with its topology, differentiable structure and metric. Particular attention will be paid to the question of dimensionality.

In chapter 3, I will deal with aspects of “stage 3” for our proposal, giving a framework for dynamics of causal sets, formulated in terms of a sum over histories (“path integral”) approach to a quantum theory. I will discuss this choice of quantum framework (which in the previous chapter will have been implicitly assumed in the discussion), and consider various possible choices for the basic amplitude, and to what extent the theory will reproduce general relativity in the continuum approximation. In this connection one could say that kinematics

for us will be the study of how each *single* causal set determines a continuum, while dynamics describes how “cooperative effects” between many causal sets determine which of these continua will be our large-scale space-time.

Some of what I describe here is not just my work, but collects the results and ideas which have emerged in a collaboration with J. Lee, D. Meyer and R.D. Sorkin, and, for some aspects, P. Jain. Some sections are essentially an expanded version of our recent paper [3], while others include developments of our proposal in new directions, and/or new results.

## 2. KINEMATICS OF CAUSAL SETS

### 2.1 Posets: some definitions and properties

The causal sets we want to use as fundamental structures in our theory are well-known objects in combinatorial theory, since they belong to the category of partially ordered sets. Let me then introduce this mathematical notion, and some terminology associated with it, which will be useful in the following sections. (In this section, when a new term will be defined, I will normally use italics if it is part of the standard combinatorics terminology, and quotes when it is part of the physicists' terminology, or our term. Combinatorics literature on partially ordered sets—see, e.g., Rival [64], Fishburn [61], Stanley [65]—is filled with definitions: of those, I will try to give only a few useful or potentially useful ones for our purposes, and even so, they might appear as many; a few more will appear, when needed, in later sections.)

#### Basic definitions

A *partially ordered set* (or *poset* for short)  $P$  is a set with an order relation (for simplicity of notation, I will use the same symbol for the poset and the underlying set), i.e., a relation, indicated by  $p < q$ , which is:

- (1) reflexive:  $\forall p \in P, p < p$ ;
- (2) transitive:  $\forall p, q, r \in P, p < q < r \Rightarrow p < r$ ;
- (3) antisymmetric:  $\forall p, q \in P, p < q \Rightarrow q \not< p$ , unless  $p = q$ .

Notice that the third condition excludes the existence of “closed loops”. If  $p < q$ , we say that  $q$  follows  $p$ , or is to the future of  $p$ . The posets we will consider will

always be connected and locally finite. Before we define precisely the meaning of these terms, we need some more definitions.

The *interval* (or “Alexandrov set”, as we will call it in the context of causal sets)  $A(p, q)$  defined by two elements  $p, q \in P$ , with  $p \prec q$ , is the intersection of the future of  $p$  with the past of  $q$ :

$$A(p, q) := \{r \mid p \prec r \prec q\}. \quad (2.1.1)$$

An interval is a special case of (*induced*) *subposet*, a subset  $P' \subset P$  in which two elements are related,  $p \prec q$ , iff they are related in  $P$ : i.e.,  $P'$  has the order relation induced by  $P$ . An element of  $P$  is called *maximal* if there are no elements to its future, and *minimal* if there are no elements to its past (a poset in general has many maximal and minimal points). A subset of  $P$  in which any two elements are related (i.e., a totally ordered subset) is a *chain*.

There is in poset theory a notion similar to that of nearest neighbor for lattices embedded in manifolds with positive-definite metrics: we say that an element  $q$  *covers*  $p$ , or that there is a “link” between them, if  $p \prec q$  and there is no  $r$  (distinct from  $p$  and  $q$ ) such that  $p \prec r \prec q$ . Let us indicate this relation by  $p \prec\prec q$ . I should remark that, for a general poset, there is no metric meaning associated with this notion of closeness in the partial order, although, as we will see later, in some cases it is related to a notion of closeness in a lorentzian metric. The concept of a link, is, however, useful for the description of (locally finite) posets, since knowledge of all links is equivalent to knowledge of all relations among elements:  $p \prec q$  iff there are elements  $r_1, r_2, \dots, r_n$  such that  $p \prec\prec r_1 \prec\prec r_2 \prec\prec \dots \prec\prec r_n \prec\prec q$ . This property of links can be used first of all to draw simple pictures of posets, called *Hasse diagrams*, where elements are represented by points, and links by lines connecting these points, with  $q$  placed above  $p$  if  $p \prec q$ , to indicate the

direction of the lines. (We will see below several examples of Hasse diagrams.) Furthermore, the same property will make it easier for us to prove some results about posets, since it guarantees that if two posets have the same links, then they are equal (see, e.g., the “proof” of lemma 2, in §2.4).

We shall call “path” between  $p$  and  $q$  a *maximal chain* between these elements: any chain whose size can only be made to increase, by extending it to the past or the future, if possible; in other words, a chain made of links, like  $p \prec r_1 \prec r_2 \prec \dots \prec r_n \prec q$  in the previous paragraph. Two paths between  $p$  and  $q$  need not have the same *length* (number of links), and we will call “maximal path” one with the maximum length.

A *connected* poset is one whose diagram is connected, i.e., such that, given any two elements, there is a set of paths which can be followed (moving forward and backward with respect to the partial order) to go “continuously” from one to the other. A *locally finite* poset is one in which all Alexandrov sets are finite. From here on, when we will talk of a “causal set”, we will always mean a connected, locally finite poset.

### Examples; number of posets and their realizations

Let me now give some examples of posets (recall that we specify only the links between elements):

- (a) the totally, or linearly, ordered set with  $N$  elements,

$$P_N^l := \{a_i; i = 1, \dots, N \mid a_i \prec a_{i+1}; i = 1, \dots, N - 1\};$$

- (b) the “fence” poset of size  $2N$ ,

$$P_{2N}^f := \{a_i, b_i; i = 1, \dots, N \mid a_i \prec b_i \forall i; a_i \prec b_{i-1} \forall i \neq 1\};$$

- (c) the “crown” poset of size  $2N$ ,

$$P_{2N}^c := \{a_i, b_i; i = 1, \dots, N \mid a_i \prec b_i \forall i; a_i \prec b_{i-1} \forall i \neq 1; a_1 \prec b_N\}.$$

Some posets are more conveniently defined not by an abstract “presentation” of their structure as above, but by some realization, like in the following example:

- (d) the poset  $P_{2^n}$ , whose elements are the subsets of a set  $A$  of size  $|A| = n \geq 1$ , ordered by inclusion (the Hasse diagram of  $P_{2^n}$ —figure 2.1.4—looks like a cube in  $n$  dimensions);

We will see a few more examples of posets in later sections.

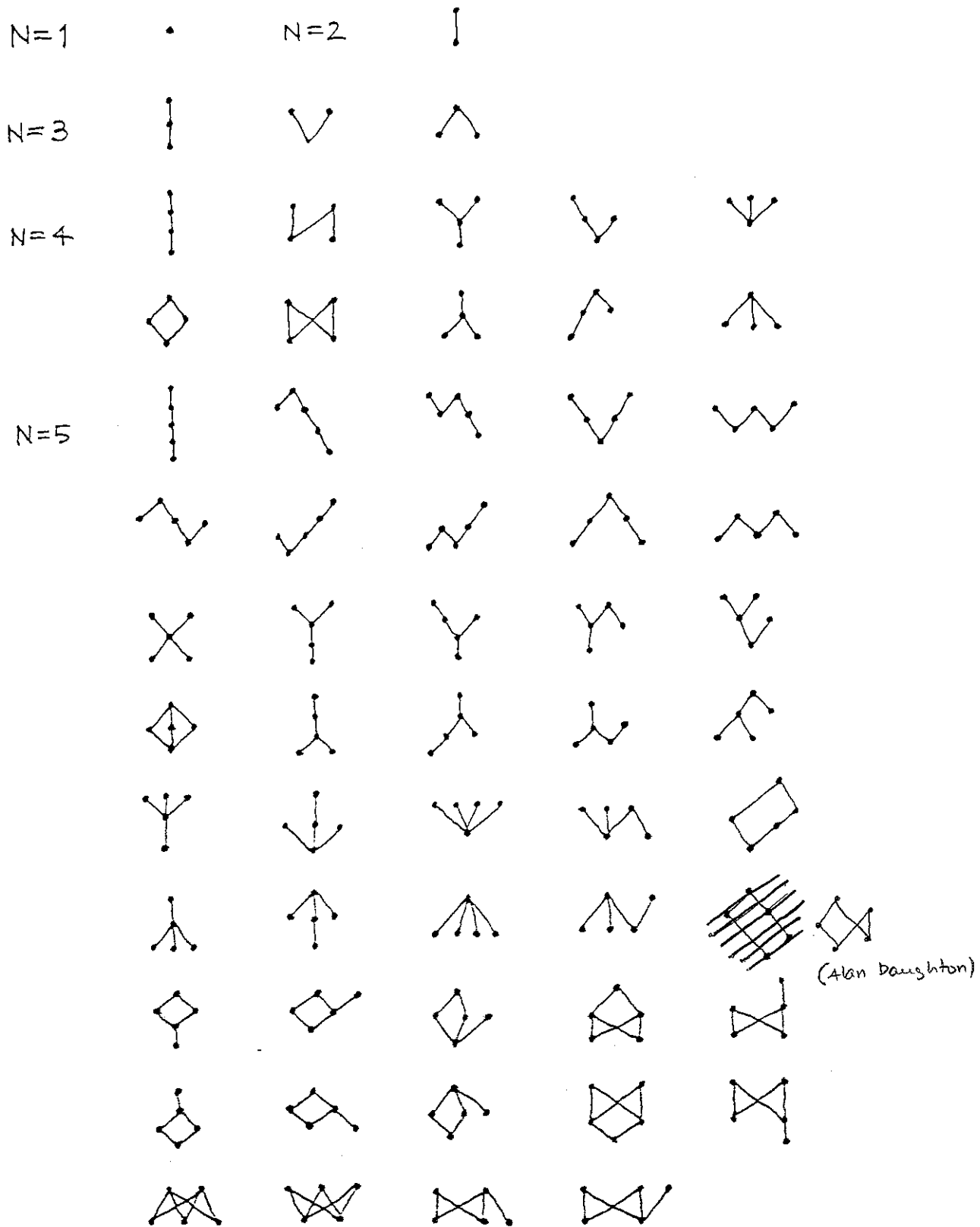
It is natural at this point to ask how many different posets one can construct with  $N$  elements, and to look for examples of posets. The first question turns out to be an unsolved problem in combinatorics: in fact the answer is known only for  $N < 8$ , although an asymptotic formula for large  $N$  is known. The table below gives, for  $N < 8$ , the number of inequivalent posets,  $D_N$ , and inequivalent connected posets,  $C_N$ , that can be built from  $N$  points. (By inequivalent we mean that we do not distinguish between posets which differ by a relabelling of elements, i.e., we use what are sometimes called *unlabelled* posets: in “physical” terms, we will always talk about equivalence classes of posets under the action of the permutation group, which is clearly irrelevant.)

$N$	1	2	3	4	5	6	7
$C_N$	1	1	3	10	44	238	
$D_N$	1	2	5	16	63	318	2045

(the connected posets up to  $N = 5$  are drawn in figure 2.1.1). For  $N \rightarrow \infty$ , the asymptotic behavior of the number of posets is

$$D_N \approx C 2^{N^2/4 + 3N/2} e^N N^{-N-1} = C 2^{N^2/4 - N \log_2 N + (3/2 + \log_2 e)N - \log_2 N}, \quad (2.1.2)$$

where  $C = \frac{2}{\pi} \sum_{i \geq 0} 2^{-i(i+1)} \simeq 0.8059$ , for  $N$  even, while a similar expression—with practically the same numerical value—gives  $C$  in the case of odd  $N$ . In



All connected posets with up to 5 elements

figure 2.1.1



the same limit  $N \rightarrow \infty$ , most posets (i.e., to leading order in  $N$ ) have diagrams consisting of 3 "layers" of points, the top and bottom ones with  $N/4$  points, the middle one with  $N/2$ , with each point being connected to about half of the points in its adjacent layers (Kleitman and Rothschild [63], and references therein).

Posets can be realized in different ways. In a loose sense we have already seen one way of representing posets when we introduced their Hasse diagrams, but we are now interested in showing that there are other (richer) mathematical structures in which poset structures can be identified, and which thus provide both applications of poset theory and means of visualizing some of their properties and definitions. In the framework of our causal sets, although the dynamical theory can be abstractly formulated without referring to any realization, as we will see, we often think of a poset  $P$  as realized in terms of points in a lorentzian manifold  $(\mathcal{M}, g_{ab})$ , with the causal relations induced by the metric reproducing the order relations in  $P$ : i.e., we have a map  $f : P \rightarrow \mathcal{M}$ , with  $f(p) \in J^-(f(q))$  iff  $p < q$  ( $J^-$  denotes the metrically defined "past" of a space-time point or region: it will be defined in §2.2). This is the reason why, in this context, we call (locally finite) posets causal sets; we will call the above a "causal realization" of  $P$ .

But the same poset could also be thought of, e.g., as a collection of subsets of some set  $X$ , with the inclusion relations between them reproducing the order relations: i.e., we could have a map  $f : P \rightarrow \mathcal{P}(X)$ , with  $f(p) \subset f(q)$  iff  $p < q$ , where  $\mathcal{P}(X)$  denotes the collection of subsets of  $X$ . In particular, if  $X$  is a topological space, the  $f(p)$ 's could be open subsets of  $X$ . Notice also that a causal realization is actually equivalent to one of the second kind: if we identify each point  $f(p) \in \mathcal{M}$  with the interior of its past light cone, we obtain a realization of  $P$  in terms of subsets of  $\mathcal{M}$  (light cones), ordered by inclusion.

As a third example, one of the common realizations of posets in combinatorics

is as points in some  $\mathbb{R}^n$ , in which the partial ordering induced by the coordinates reproduces the order relations in  $P$ : we have a map  $f : P \rightarrow \mathbb{R}^n$ , with  $f^i(p) \leq f^i(q) \forall i$  iff  $p < q$ . This gives a “linear realization” of  $P$ . There may be other physically interesting realizations of posets, besides the first two examples given here. One such possibility, related to the so-called “Bergmann manifolds”, will be briefly discussed in §3.5.

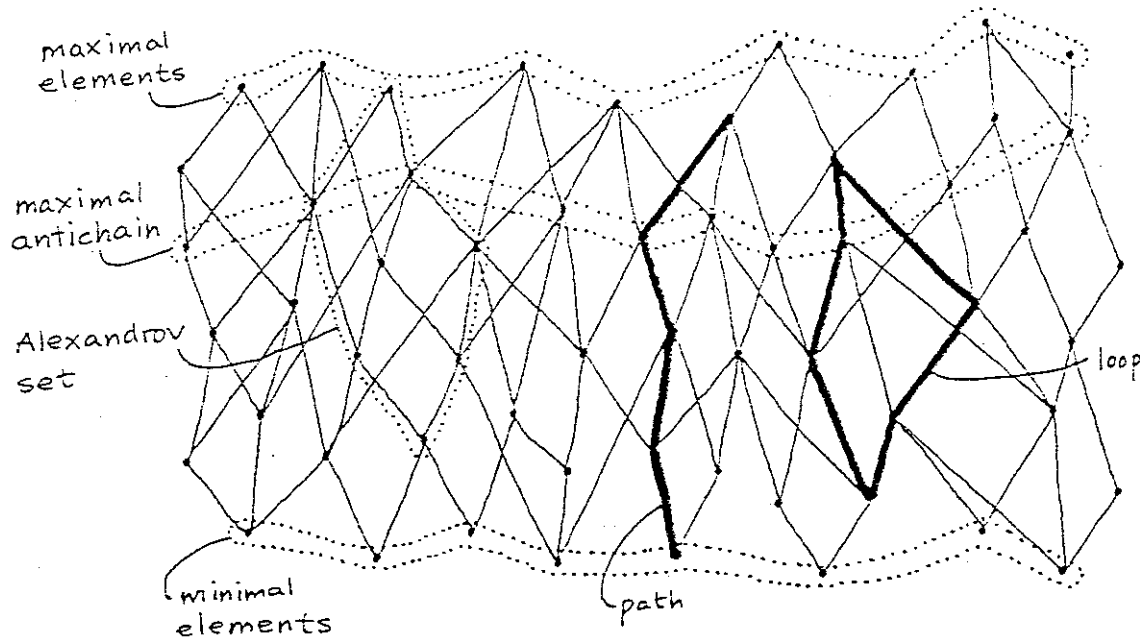
### Parameters of posets; important subsets

It is now time for some more definitions, before we start to see what can be done with posets. An *antichain* is any subset of a poset  $P$  in which no two elements are related (it is what in physics would be called an “acausal set”); whereas a *maximal antichain* is an antichain such that no further element can be added to it: all other elements of  $P$  are related to some element in it (a “spacelike hypersurface”). The *width* of a poset is the size of the largest (maximal) antichain, and its *height* the length of the longest path. A *join-independent* subset of  $P$  is a  $P' \subset P$  such that, for every  $p' \in P'$ , there is a  $p \in P$ , with  $p' \not\prec p$ , but  $p$  is to the future of every other element of  $P'$  (in particular, such a  $P'$  is always an antichain). The *breadth* of  $P$  is the size of its largest join-independent subset.

Although meaningful constructions with posets for us will only be the ones which can be defined independently of any labelling of their elements, it might be useful in some occasions to introduce labellings, just like it is often useful to introduce coordinates on manifolds. A partial labelling can be obtained by a function which slices a poset into antichains, a positive integer-valued *rank function*  $\rho$  on  $P$ , with (i)  $\rho(p) = 1$  if  $p$  is a minimal element, (ii)  $(p < q, p \neq q) \Rightarrow \rho(p) < \rho(q)$ , and (iii)  $\forall \bar{\rho} : 1 \leq \bar{\rho} \leq \max_{(q \in P)} \rho(q), \exists p \in P : \rho(p) = \bar{\rho}$  (often the stronger condition  $p < q \Rightarrow \rho(q) = \rho(p) + 1$  is required, but with it not all

posets are rankable, and we wish to allow ourselves more freedom in the ranking possibilities). One natural ranking defined for all posets is the following. Assign  $\rho(p) = 1$  to all minimal elements, and  $\rho(p) = i$  to any element which becomes minimal after removal of all elements of rank less than  $i$ .

We have so far identified "one-dimensional timelike" structures in a poset (paths), and "codimension-one spacelike" ones (maximal antichains). Another simple object we will need later is a "loop": a pair of paths between two points, which do not meet except at their endpoints. It is more difficult to identify other structures, but one definition worth giving is that of a "null path": a path which is also the Alexandrov set formed by its endpoints (in Minkowski space, the only points causally related to two null-related points are those on the null geodesic joining them). A few of these definitions are illustrated in figure 2.1.2.



Various kinds of subsets of a poset

figure 2.1.2

## Operations on posets

Now that we are acquainted with posets we can ask ourselves, what can be done, mathematically, once we have them: what operations can be performed on posets, and which other mathematical structures can be defined just by using partial orders? The second part will be postponed until §2.7, while, as regards operations, once again we will need only a few of the many that mathematicians have defined for posets (and we won't use all of them, some I just give here for possible future use). The simplest one is *duality*, which is what in physics terminology we would call "time-reversal":<sup>†</sup> it is defined by  $P \rightarrow P^*$ , where  $P^*$  has the same underlying set as  $P$ , but the arrows are reversed:

$$p < q \text{ in } P^* \quad \text{iff} \quad q < p \text{ in } P. \quad (2.1.3)$$

To combine two posets and produce a new one, we have several possibilities. In the (*cartesian*) *product* of  $P_1$  and  $P_2$ ,  $P_1 \times P_2$ ,

$$(p_1, p_2) < (q_1, q_2) \quad \text{iff} \quad p_1 < q_1 \wedge p_2 < q_2. \quad (2.1.4)$$

The *cardinal sum*  $P_1 + P_2$  of posets is just their disjoint union (and gives always a disconnected poset), and, together with the cartesian product, it makes the set of all posets into a commutative semi-ring (or into a ring, if one formally defines additive inverses). The *ordinal sum*  $P_1 \oplus P_2$  is " $P_2$  sitting on top of  $P_1$ ": its underlying set is  $P_1 \cup P_2$ , and

$$p <_{P_1 \oplus P_2} q \quad \text{iff} \quad (p <_{P_1} q) \vee (p <_{P_2} q) \vee (p \in P_1, q \in P_2). \quad (2.1.5)$$

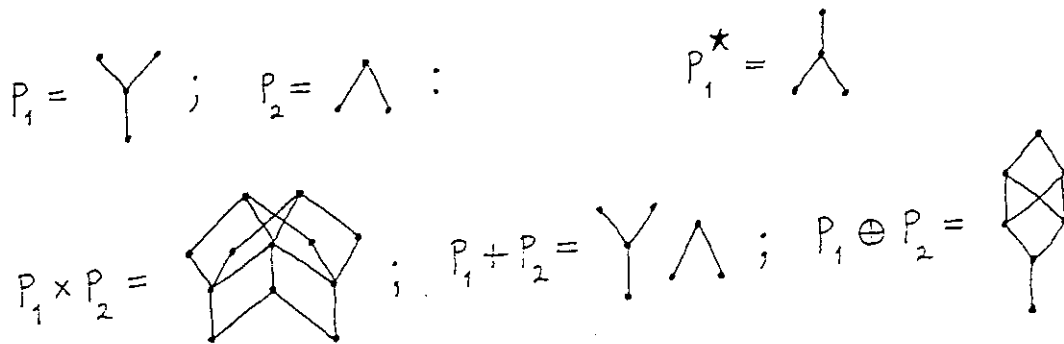
Of course, a trivial (but important in connection with the concept of coarse-graining) way of obtaining a poset from another is as a subposet, as described

---

<sup>†</sup> Although it might be more appropriate to call it *PT*, or *CT*, or *CPT*: we do not yet have a notion of charge conjugation or spatial orientation for posets. More on this in §3.5.

earlier. To visualize these operations, look at the examples represented by the Hasse diagrams in figure 2.1.3.

Finally, given two posets  $P$  and  $P'$  on the same underlying set one could define their "intersection" by saying that  $p < q$  in  $P \cap P'$  iff  $p < q$  in both  $P$  and  $P'$ . This operation, however, is only well-defined for labelled posets, since it depends on how we identify elements of  $P$  with elements of  $P'$ , and not just on the intrinsic poset structure. We are therefore not going to use it.



Operations on posets

figure 2.1.3

### Dimension of posets

The notion of dimension of a poset will become important in the following. A poset is not itself a topological space, but, as we will see in §2.7, there are ways of making a topological space out of it. It turns out, however, that the dimension one would define for this topological space is not very meaningful for the poset, and we will instead define several other notions of dimension: the useful ones are definitions in which the dimension in some sense is not a property of the poset only, but of the poset and some realization of it.

The first example is the notion mathematicians most commonly use, which we will call "linear dimension",  $ldim(P)$  (to distinguish it from the other notions

of dimension we will introduce later): it is the smallest  $n$  for which there exists a linear realization of the poset in  $\mathbb{R}^n$  as described above (the usual definition, although equivalent to this one, looks quite different, and it will not be necessary to give it here). This notion of dimension will be useful for us because of the number of results already known about it, but I must point out that it has little to do with the physically relevant dimension of a poset, which is the dimension of the lorentzian manifold which approximates well a large poset at large scales, in a sense which we will define in §2.3. In particular, the latter need not be defined for a given poset, whereas the linear dimension is always defined.

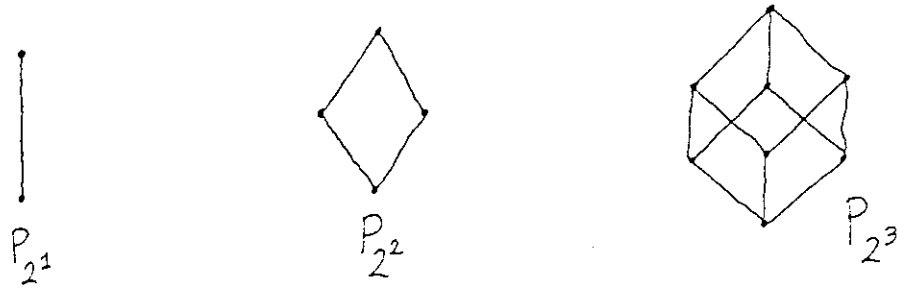
Let us give some known results on the linear dimension of a poset (see, e.g., Kelly and Trotter [62], Fishburn [61]). Upper bounds on  $ldim(P)$  are:

- (a)  $width(P \setminus E) + 1$ , where  $E$  is any set of maximal or minimal points;
- (b)  $|P|/2$ , i.e., half the size of the poset, if  $|P| \geq 4$ ;
- (c)  $ldim(P \setminus C) + 2$ , where  $C$  is any chain in  $P$ ;
- (d)  $\max(2, |P \setminus A|)$ , where  $A$  is any antichain;
- (e)  $width(P)$ ;
- (f)  $2 \cdot width(P \setminus A) + 1$ , where  $A$  is any antichain,

whereas a lower bound on the linear dimension of a poset is its breadth. Also, deleting one point from a poset can decrease its linear dimension by at most one, and the linear dimension of a cartesian product satisfies the relation

$$ldim(P \times Q) \leq ldim(P) + ldim(Q), \quad (2.1.6)$$

with equality, e.g., when  $P$  and  $Q$  are linearly ordered, or they each have unique maximal and minimal elements.



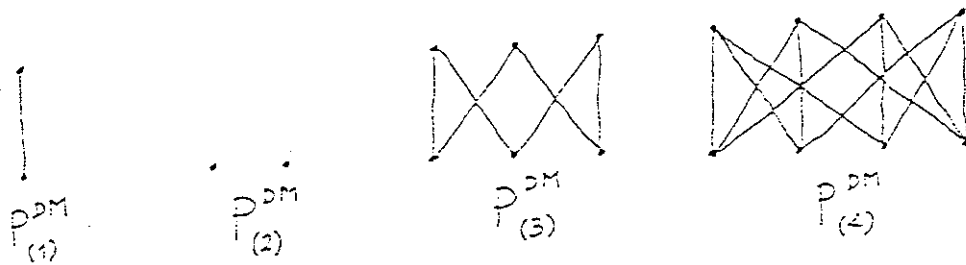
The first three hypercube posets

figure 2.1.4

The linear dimension of many small posets has been studied, and in particular there are two infinite families  $\{P_{(n)} \mid n = 1, 2, \dots\}$  of posets with  $ldim(P_{(n)}) = n$ :  $P_{(n)} = P_{2^n}$  (see example (d) above and figure 2.1.4), and

$$P_{(n)}^{DM} := \{\{a\} \mid a \in A\} \cup \{A \setminus \{a\} \mid a \in A\}, \quad (2.1.7)$$

where  $|A| = n$ , ordered by inclusion (Dushnik and Miller [60]). These latter posets saturate the upper bounds (b) and (c) and the lower bound on the linear dimension; the first four of them are shown in figure 2.1.5.



The first four Dushnik-Miller posets

figure 2.1.5

A second definition of dimension available in the literature is that of "multiposet dimension":  $mdim(P)$  is the minimum size of a collection of posets on the same underlying set as  $P$ , such that  $P$  itself belongs to the collection and such that, given any two elements  $p, q$  in the underlying set of  $P$ , there exists exactly one poset in the collection with respect to which  $p$  and  $q$  are related (with

either  $p < q$  or  $q < p$ ). At first this notion, like many others in combinatorics, might seem far-fetched, but it does capture some notion of complexity (see, e.g., Behrendt [58]):

- (a)  $mdim(P) = 1$  iff  $p$  is linearly ordered;
- (b)  $mdim(P) = 2$  iff  $ldim(P) = 2$  (Dushnik and Miller [60]);
- (c) if  $P'$  is a subposet of  $P$ , then  $mdim(P') \leq mdim(P)$ ;
- (d) for any  $P$ ,  $mdim(P) \leq ldim(P)$ ;
- (e) for any  $n \geq 3$ ,  $\exists P : ldim(P) = n \wedge mdim(P) = 3$  (namely the Dushnik and Miller posets (2.1.7)).



## 2.2 Metrics and causal structures on differentiable manifolds

The causal relations induced by lorentzian metrics on manifolds have been extensively studied, for the case of Minkowski space, starting just a few years after Minkowski's work (Robb [54]), and more recently have received an increased attention mainly because of their applications to "global methods" in general relativity, which use the behavior of families of geodesics to establish global topological and metric properties of space-times, in particular to prove theorems on the occurrence of singularities. Here, I will be interested in the results obtained regarding the amount of information that a causal structure by itself, abstracted from the other structures used in defining it, carries about the latter.

We start by defining some standard terminology related to causal structures (see, e.g., Kronheimer and Penrose [50]; Carter [43]; Hawking and Ellis [48]). Suppose  $(M, g_{ab})$  is a *space-time* (i.e., a differentiable (paracompact, Hausdorff) manifold with a smooth metric of lorentzian signature,  $(- + + \cdots +)$ —of any dimensionality), which is time-orientable (i.e., such that a future and a past light cone can be separately assigned in a continuous way to all points of  $M$ ). Then the *causal past*  $J^-(x)$  of a point  $x \in M$  is the set of all points in  $M$  which can be reached by a smooth past-directed causal (i.e., timelike or null) curve from  $x$ . The *chronological past*  $I^-(x)$  of  $x$  is the set of points in  $M$  which can be reached by a smooth, nondegenerate—we don't allow ourselves to stay always at  $x$  in this case—past-directed timelike curve. The *causal future* and the *chronological future* are defined similarly. All these "functions" can be extended to subsets  $S \subset M$  by taking the union over the points in them, e.g.,  $J^-(S) := \cup_{x \in S} J^-(x)$ . Chronological pasts and futures are always open in the manifold topology of  $M$ , whereas causal pasts and futures are not necessarily closed, but for most practical purposes they can be thought of as closed.

Space-times can be classified according to how “nice” their causal structure is, i.e., to how far it is from allowing a future-directed causal (timelike or null) curve to go back to its own past; mathematically a variety of *causality conditions* can be imposed, of varying strength, to prevent this. Since we will quote results which use them, let us give the definitions of some of the conditions one imposes. A space-time is *causal* if it contains no closed everywhere future-directed (or past-directed) curves. A space-time is *strongly causal* if for every point  $x$  in it and every open neighborhood  $\mathcal{U}$  of  $x$ , there is another open neighborhood  $\mathcal{V} \subset \mathcal{U}$  of  $x$  such that every future-directed smooth timelike curve which leaves  $\mathcal{V}$  never comes back to it. A space-time is *past and future distinguishing* if, for any two points  $x$  and  $y$ ,  $I^-(x) = I^-(y) \Rightarrow x = y$ , and similarly for  $I^+$ . The past and future distinguishing condition is equivalent to the statement defining the strong causality condition, but restricted to curves which pass through  $x$ , and is thus slightly weaker, but both are intended not to allow timelike curves to be even “almost closed”. Finally, a space-time is *stably causal* if its metric has a neighborhood (in a suitably defined notion of topology in the space of lorentzian metrics on  $\mathcal{M}$ —see, e.g., Hawking and Ellis [48]) every metric of which is causal. This is equivalent to the existence of a global *time function*  $t : \mathcal{M} \rightarrow \mathbb{R}$ , with  $\nabla_a t$  everywhere timelike.

Suppose a space-time  $(\mathcal{M}, g_{ab})$  is causal and time-orientable. Then the above notions define on  $\mathcal{M}$  a partial order, as we have already mentioned, the *causality relation*,  $x < y$  if  $x \in J^-(y)$ . Two other relations can be defined: the *chronological relation*,  $x \ll y$  if  $x \in I^-(y)$ , and the *horismos relation* or null relation,  $x \rightarrow y$  if  $x \in J^-(y) \setminus I^-(y)$ . These are not partial orders in the sense of §2.1 (e.g.,  $\ll$  is not reflexive), but, together with the causality relation, they satisfy other properties which, abstractly applied to a set, define a *causal space* (Kronheimer

and Penrose [50]). The expression "causal structure" refers collectively to these three relations, although, since in reasonably well-behaved cases they can all be defined in terms of any one of them, I occasionally use it to denote just the partial order  $\prec$ .

### Flat space-times

It has long been known that the causal structure of 4-dimensional Minkowski space contains all the information about its affine or linear (and therefore topological and differentiable) structure, and about the metric, up to a global scale factor. To show this, Robb [54] constructed these structures starting from a causal one satisfying a set of conditions, expressed purely in causal terms, which captured the fact that the causal structure had really been obtained from Minkowski space to begin with. Work on this subject was later done by Reichenbach [68], Mehlberg [52], Alexandrov [42], with similar results. The most complete and concise formulation of the result was given however by Zeeman [57]: the set of all one-to-one mappings  $f : M \rightarrow M$  of 4-dimensional Minkowski space such that  $x \prec y \Leftrightarrow f(x) \prec f(y)$  (such a map is called *causal isomorphism*) coincides with the Poincaré group together with dilatations.

I will not prove here these results, but I will just sketch a simple construction to illustrate how the 4-dimensional Minkowski space structure is recovered up to a global scale factor from the causal structure. Assume we have an infinite causal set whose structure is known to be compatible with 4-dimensional Minkowski space. Then the steps are the following:<sup>†</sup>

- (a) a null ray can be defined as a set of points  $\ell$  such that for any pair of points  $x, y \in \ell$  there is a unique causal path connecting them, i.e., a totally ordered set of points of the form  $\ell = \{x \mid x, z \in \ell \wedge x \prec y \prec z \Rightarrow y \in \ell\}$ ;

---

<sup>†</sup> This construction is due to R. Sorkin.

- (b) a null 3-plane is defined as the set of points which are spacelike related (i.e., unrelated) to a null ray, together with the ray itself:  $P := \{x \mid x \natural \ell\} \cup \ell$ , where the symbol  $\natural$  denotes the relation "spacelike related to";
- (c) a spacelike 2-plane is the intersection of two null 3-planes:  $S := P_1 \cap P_2$ ;
- (d) a spacelike line is the intersection of two spacelike 2-planes (which meet at more than one point):  $L := S_1 \cap S_2$ ;
- (e) a timelike 2-plane is the union of all (spacelike or null) lines joining points on two intersecting null lines;
- (f) a timelike line is the intersection of two timelike 2-planes (which meet at more than one point, and not at a null or spacelike line);
- (g) a parallelogram is formed by two pairs of parallel lines (coplanar lines which do not meet) intersecting at four points;
- (h) to get the linear structure fix now any point as the origin, and call vectors all other points, or equivalently all line segments with one end at the origin: addition of two vectors is defined by the diagonal of the parallelogram formed by these vectors; using the addition we can then easily define multiplication by a rational number, and from there multiplication by a real number;
- (i) what will become an orthonormal basis is chosen as follows: choose an arbitrary timelike vector  $T$  (here we actually make a choice for the global conformal freedom, besides picking a preferred 3+1 splitting); choose then any spacelike vector  $X$  such that both  $T + X$  and  $T - X$  are null (this says that  $X$  is normal to  $T$ ); choose a second spacelike vector  $Y$ , also satisfying the condition that  $T - Y$  and  $T + Y$  be null, and in addition that  $X + Y + \sqrt{2}T$  also be null (read:  $Y$  perpendicular to  $X$ ); finally, choose a third spacelike vector  $Z$ , with  $T - Z$ ,  $T + Z$ ,  $X + Z + \sqrt{2}T$  and  $Y + Z + \sqrt{2}T$

null vectors;

- (j) now we can impose the “meaning” of the conditions in (i) as a *definition* of the metric  $\eta$ : the four chosen vectors are linearly independent, and we set  $\eta(T, T) = -1$ ,  $\eta(X, X) = \eta(Y, Y) = \eta(Z, Z) = 1$ , and all the non-diagonal terms equal to zero; the definition is then extended to all other pairs of vectors by linearity.

It is very easy to extend the construction to any other dimensionality, except 2: in 2 dimensions step (b) does not give anything (but the null rays themselves), and we cannot proceed to step 3. The above construction, although simple to follow, is misleading as far as the general metric case is concerned, since it uses the vector space structure of Minkowski space to define the metric and it leaves the role of the topology of the manifold at a very implicit level.

### Curved space-times

The above results might be a bit surprising, in view of the fact that an arbitrary conformal transformation does not change the causal structure of a space-time: it might seem that only an equivalence class of space-times, conformally related to a given Minkowski space, could be recovered. The subtlety is that in the reconstruction results one puts in the fact that the metric is flat, and this rules out all conformal transformations except the constant ones. For example, in the sketch of construction I described above, we were *looking* for a linear structure in the space, while in Zeeman’s theorem, the statement just refers to maps from Minkowski space to Minkowski space—and not, e.g., from Minkowski space to some other given conformally flat space-time.<sup>†</sup>

---

<sup>†</sup> It would clarify the issue considerably to spell out clearly what we mean by “reconstructing” the metric from the causal structure. Mathematically, construction arguments are given in functorial terms; in our case, we could thus formulate the questions we want to address in terms of functors from an appropriate category of causal spaces to one of Minkowski spaces or of conformally flat space-times.

In the case of a more general space-time, the analogous thing to knowing that the metric is flat is not a very natural requirement, if it can be at all worded. Thus, the result one obtains looks a bit different: the knowledge of the causal relations between all pairs of points allows us to determine uniquely a topology and differentiable structure on the manifold, and the metric up to a (local) conformal factor, i.e., to determine the conformal metric  $g_{ab}/(\det g)^{1/n}$ , where  $n$  is the dimension of the manifold. The most general proof of this fact was given by Malament [51], who showed the following: if  $(M, g_{ab})$  and  $(M', g'_{ab})$  are past and future distinguishing space-times, and  $f : M \rightarrow M'$  a causal isomorphism, then  $f$  is a homeomorphism, and it preserves future directed continuous timelike curves and null geodesics. This, together with Hawking's theorem (see [49]) that any homeomorphism preserving future directed continuous null geodesics is a smooth conformal isometry, proves the claim.

The remaining component of the metric can be recovered if one assigns a volume element to the manifold, as we will see explicitly in §§2.5 and 2.7. The construction for the metric in this case works for all dimensionalities, including  $n = 2$ .

### Topologies for space-time

An interesting aspect of the relationship between causal structures and geometry ("chronogeometry") in the continuum is that of the topologies for a space-time which can be defined in terms of its causal structure (the usual topology one gives to space-time comes with its differentiable structure: thus, it is there before any metric is put on it, and it cannot know anything about the causal structure). These alternative topologies are defined instead once the causal structure is given.

The most famous such topology is the *Alexandrov topology*: the coarsest topology in which all chronological pasts and futures  $I^-(x)$ ,  $I^+(x)$  are open. A

base for this topology consists of the Alexandrov neighborhoods: all open sets can be generated by taking unions of these.<sup>†</sup> In general the Alexandrov topology is coarser than the manifold topology, but the two are equivalent if the former is Hausdorff, or, equivalently, if the space-time is strongly causal (Kronheimer and Penrose [50]). (We remark here that if we applied this definition to a (discrete) causal set, we would in general obtain just the discrete topology on it, which is trivial).

Another topology for Minkowski space, which can be called *fine topology*, was defined by Zeeman [57b] as the finest one which induces on all straight timelike lines and all spacelike hyperplanes the same topology as the manifold topology. It is harder to work with than the Alexandrov topology or the manifold topology, but it is strictly finer than these, and it can be useful as a way to formulate questions about the causal structure in topological terms, since, e.g., the homeomorphisms in this topology coincide with the causal isomorphisms together with causal duality. This topology was later generalized by Göbel [46] to curved space-times.

But the most interesting one is the *path topology* of Hawking, King and McCarthy [49]: the finest topology which induces on all continuous timelike curves the same topology as the manifold topology. This definition may also not be too simple to work with, but it is also stronger than the manifold one, and it has a very interesting property: all homeomorphisms of the path topology, for any space-time, are smooth conformal isometries (Malament [51]). Furthermore, it might be considered as more physical than the Alexandrov topology (although, in the lat-

---

† If one wishes to use the Alexandrov topology for more general space-times than the ones covered by the definition I gave earlier, then the Alexandrov neighborhoods constitute a basis iff the following conditions are satisfied: (i)  $\forall x \in M, \exists y, z \in M : y \ll x \ll z$ ; and (ii)  $\forall x, y_1, y_2, z_1, z_2 \in M : y_i \ll x \ll z_i, i = 1, 2, \exists v, w \in M : y_i \ll v \ll x \ll w \ll z_i, i = 1, 2$ .

ter, open sets are defined using the causal structure, Alexandrov neighborhoods are still not very meaningful concepts operationally). A base for the path topology of  $\mathcal{M}$  is given by the sets of the form  $L_U(x, \epsilon) := B_U(x, \epsilon) \cap [I^+(x) \cup I^-(x) \cup \{x\}]$ , where  $B_U(x, \epsilon)$  is the image under the exponential map of a ball of radius  $\epsilon$  in the tangent space  $T_x\mathcal{M}$ , small enough to fit in a *convex normal neighborhood*  $U$  of  $x$ , a neighborhood such that any two points in it can be joined by a unique geodesic lying entirely within the neighborhood.

### Comment

Let me conclude this brief review with a remark of a more philosophical nature. The theorems quoted above, as we saw in connection with Robb's work, can hardly be said to actually reduce the space-time structure to that of a causal space, since they say: *if we know that a causal space really comes from a space-time manifold, then the latter is determined up to a local scale factor*. For flat space-time, where explicit conditions have been given on the causal structure to express this assumption, one could say (Torretti [71]) that we have "a set of geometric conditions with which the basic forms of physical causality must comply, if the principles of special relativity hold good". But in the general case we do not even have this: we don't know how to express the requirement that the causal space come from a manifold in terms of the causal space itself (some work related to this question was done, e.g., by Ehlers, Pirani and Schild [44] and by Woodhouse [56]). And, in neither case do we have a dynamical theory for the causal space.



### 2.3 Causal sets approximated by differentiable manifolds. I

We now start to consider to what extent the results of the previous section apply when the causal relations are defined for a discrete set of elements, i.e., to what extent a causal set determines the properties of an associated space-time manifold. From the concluding remark of §2.2, it would clearly be desirable to give a constructive procedure for obtaining such a manifold, if one exists, from the causal set. Only the outline of a procedure for doing so is available, however, and we will have to make a choice for what we mean by manifold “associated” with a causal set, and tackle the problem of showing, for such a choice, how unique the space-time manifold is. A very reasonable guess as to the kind of relationship there will ultimately be between the causal set  $P$  and the space-time manifold  $(\mathcal{M}, g_{ab})$  is that the general situation can be described by a mapping  $f : P \rightarrow \mathcal{M}$ . It has been said that we are putting in by hand an interpretation of the abstract elements of  $P$  as *points*, but there are a few arguments in favor of our choice. First and least, any other concrete possibility proposed (points  $\rightarrow$  strings, etc.) seemed more artificial. Second, the mentioned outline for a manifold construction procedure suggests the identification, since, in it, the causal set elements end up being vertices of a simplicial complex, which go over to points in a manifold of which the complex is a triangulation. Third, we expect to obtain only an approximate metric for the manifold, and part of the freedom left will be that of performing a small diffeomorphism, which, in one view, can be thought of as moving the embedded points around a little, thus making them look more like blurred points. Finally, if we invoke the “magic wand of quantum theory”, we can claim that in the end the complete theory, including quantum dynamics, will turn out to be equivalent to a formulation which does not use mappings of causal set elements to space-time points.

We will say that  $f$  is a *causal embedding* if it is a causal set isomorphism onto the image of  $P$ , i.e., if  $f(P)$  is a causal realization of  $P$ :

$$(1) f(p) \in J^-(f(q)) \text{ iff } p \prec q.$$

Notice that the existence of a causal embedding requires that  $(\mathcal{M}, g_{ab})$  be causal and time-orientable, or that any closed causal curves it might have be small enough not to be "seen" by the embedded points (non-time-orientability, on the contrary, logically cannot be confined to a small region of space-time). We will say that  $f$  is a *faithful embedding* with density  $\rho$  if, in addition, it satisfies:

- (2) the embedded points  $f(P)$  are distributed with uniform density  $\rho$ ; and
- (3) the characteristic length,  $\lambda$ , over which the continuous geometry varies appreciably is everywhere much greater than the mean spacing between embedded points.

Let us try to clarify the meaning of these conditions.

### Density

When we say that the embedded points are distributed uniformly we mean that their density has a constant value,  $\rho$ , according to the metric  $g_{ab}$ , in the following sense. If we take an arbitrary Alexandrov neighborhood of size  $V$  in  $\mathcal{M}$  the expected number of embedded points in it is  $\rho V$ , while the probability distribution for having  $N$  points in the neighborhood is a Poisson distribution,

$$P(N) = \frac{(\rho V)^N e^{-\rho V}}{N!}. \quad (2.3.1)$$

Since defining density in a lorentzian manifold is not as trivial as in a euclidean one, let me comment on this definition.

In euclidean geometry, we define density *at a point* by measuring the quantity of interest (in our case counting embedded points) inside invariantly defined

regions of known size, centered around the point in consideration, for sufficiently small regions, i.e., smaller than the scale over which the resulting density varies appreciably, but, in the case of a grainy structure like ours, large enough to contain many points. In a lorentzian manifold, the most natural substitutes for the spheres one uses in the euclidean case are Alexandrov neighborhoods, and, even when we consider Alexandrov neighborhoods of small size, these can extend between “far apart” regions in the manifold, if these are “approximately null-related” (the quotes stand for the fact that these expressions are actually meaningless: no timelike direction is “more null” than any other). Thus, if we wanted to produce a distribution of points with “non-uniform” density, in an invariant sense, we would at least have to make it uniform, for every  $p \in \mathcal{M}$ , in a thin neighborhood of its light cone. But because the light cones of all points in  $\mathcal{M}$  meet (this might not actually be true, but the argument can be made more precise), we are forced in the end to have a uniform density everywhere.

The reason for this situation can be traced to the following two related facts. First, euclidean spheres are invariant under the local symmetry group of euclidean space, the rotation group  $SO(n)$ , so, to define  $\rho_e$ , one only needs to consider a 1-parameter family of regions around  $p$ , parametrized by size, while Alexandrov neighborhoods are not Lorentz invariant, so one needs to consider many more regions around  $p$  to define  $\rho_l$  (similar remarks can be extended to the curved space-time case). Second, we have not specified what we mean by “slowly varying” lorentzian density; the slowness of the variation has to be measured in terms of the metric distance along the curve one chooses to follow, but if we want it to vary slowly along all directions, arbitrarily close to null ones, we end up with the problem we had before: the density cannot vary at all.

The last remark above, however, suggests one way out of the difficulty in

defining non-uniform densities in lorentzian manifolds, at the cost of losing full general covariance. Choose a timelike vector field  $t^a$  in  $M$ . Then we can define density with respect to  $t^a$  by using only Alexandrov neighborhoods whose endpoints are on the same integral curve of it. If, for a given  $t^a$ , suitable Alexandrov neighborhoods can be found (i.e., the density is statistically meaningful and slowly varying), then this density will coincide with the euclidean one we would have obtained from the riemannian metric  $h_{ab} := g_{ab} + 2\hat{t}_a\hat{t}_b$ , where  $\hat{t}_a := t_a/(t^m t_m)^{1/2}$ . In a given situation, there will be a range of timelike directions which can be used to define *the* density, and the latter can turn out to be non-uniform, while other directions will not give well-defined results.

The fact that the lorentzian density  $\rho_l$  coincides with the euclidean one  $\rho_e$  for two metrics related as above, is indicated by the following fact. Consider  $n$ -dimensional Minkowski space, with the metric expressed in the usual cartesian coordinates, i.e.,  $\eta_{\mu\nu} = \text{diag}(-1, +1, \dots, +1)$ , and use the same coordinates to construct a euclidean metric  $e_{\mu\nu} := \eta_{\mu\nu} + 2t_\mu t_\nu = \text{diag}(+1, +1, \dots, +1)$ . Then the volume elements of these two metrics in these coordinates are the same:

$$dV_\eta := \sqrt{-\eta} d^n x = dV_e := \sqrt{e} d^n x \equiv d^n x, \quad (2.3.2)$$

the euclidean and lorentzian values for the volume of any space-time region coincide, and a uniform density according to one metric is uniform according to the other as well, at least for the range of timelike vector fields for which  $\rho_l$  is well-defined. There is a (surprising) example of a distribution of points in Minkowski space which has a uniform  $\rho_e$ , but an obvious limit to the range of vector fields one can use to define  $\rho_l$ : the 2-dimensional diamond lattice described in appendix 1, one of the most regular distributions of points one can think of.

Let us now go back to the main point. Most of the above considerations, although interesting, can be bypassed in our case, where we are only interested in defining uniform density. The notion we defined by (2.3.1) is a “global” one, as opposed to that of density at single points, and more complete than the latter, in that it takes care of what happens in all Alexandrov neighborhoods, instead of requiring us to define which ones have the right size. It seems reasonable to think of (2.3.1), in particular, as imposing that any vector field can be used as possible direction for the pointwise density, because of the Lorentz invariant nature of the definition. Thus, according to (2.3.1), the diamond lattice of §A.1 does *not* have a uniform density, since too many Alexandrov neighborhoods, close to the null lines, are empty: the lattice is not random enough.

As for the reason why we did not set the constant density to be  $1 \ell_P^{-n}$ , or some other fixed  $\rho_0 = 1$  (*natural unit*) in condition (2), we will see later that the causal set we map into  $\mathcal{M}$  may not be the one we consider as the fundamental one, which defines the natural volume, but a coarse-grained version thereof.

### Characteristic lengths

Characteristic lengths over which the continuum geometry varies appreciably might arise as the following: topological-metrical scales, like the length of a closed spacelike geodesic or of one of the small closed causal curves allowed by condition (1), or, more generally, lengths associated with nontrivial generators of any homotopy group of the manifold; purely metrical scales, like  $R^{-1/2}$ ,  $(R_{ab}R^{ab})^{-1/4}$ , ..., and, in particular, length scales defined by geodesic focussing. In other words, the loosely stated condition (3) implies in particular that every point of  $\mathcal{M}$  should have an open convex normal neighborhood of size much greater than the spacing (e.g., if we call  $y^1, y^2, \dots, y^{(n-1)}$  the coordinates, the normal neigh-

neighborhood should contain a euclidean-type ball  $(y^0)^2 + (y^1)^2 + \dots + (y^{(n-1)})^2 = r^2$ , with  $r \gg 1$  in natural units). How much of the desired content of condition (3) escapes from such a way of stating it is not clear. Intuitively, the existence of these neighborhoods should exclude positive curvature, which makes geodesics converge, but even in a negative curvature space like anti-de Sitter space, time-like geodesics reconverge on a scale set by the cosmological constant, i.e., by the curvature itself, and thus such a metric would be excluded.

In any case, we could try to capture more of the meaning of condition (3) by requiring that each point have a suitable neighborhood, which is approximately flat. More precisely, each point would be required to have a neighborhood containing a Riemann normal coordinate ball of radius much greater than the spacing, as above, such that in these coordinates the components of the metric are close to  $\eta_{\mu\nu} = \text{diag}(-1, +1, \dots, +1)$ , and its derivatives are close to zero.

### Approximate isometry

The discussion above introduces the more general question of how to compare two metrics, to decide whether they are close to each other. In order to compare two metrics  $g_{ab}$  and  $g'_{ab}$  on two manifolds  $\mathcal{M}$  and  $\mathcal{M}'$ , we will first require that the manifolds be diffeomorphic. This in effect reduces the question to that of determining when two metrics  $g_{(1)ab}$  and  $g_{(2)ab}$  on a manifold  $\mathcal{M}$  are approximately equal:  $(\mathcal{M}, g_{ab})$  and  $(\mathcal{M}', g'_{ab})$  are approximately isometric if there exists a diffeomorphism  $\psi : \mathcal{M} \rightarrow \mathcal{M}'$  such that  $g_{ab}$  and the pullback  $\psi^* g'_{ab}$  of  $g'_{ab}$  by  $\psi$  are approximately equal.

To establish whether  $g_{(1)ab}$  and  $g_{(2)ab}$  are close to each other, we might be tempted to compare their values pointwise. But there is no way to give a good local definition of closeness for two lorentzian metrics: we obviously don't want

to compare components in some coordinate system, and any invariant “distance” we might define, like  $[(g_{(1)ab} - g_{(2)ab})(g_{(1)}^{ab} - g_{(2)}^{ab})]^{1/2}$ , cannot capture the notion of distance we want to define, since, e.g., it does not distinguish between the metrics  $g_{ab}$  and  $g_{ab} + u_a u_b$ , when  $u_a$  is a null covector with respect to  $g_{ab}$ . One way out is to introduce a reference positive-definite metric  $h_{ab}$  on  $\mathcal{M}$  (this can always be done), and consider  $h^{am} h^{bn} (g_{(1)ab} - g_{(2)ab})(g_{(1)mn} - g_{(2)mn})$  like a distance, which is effectively what has been done (Hawking [47], Geroch [45]) to define a topology on the space of all lorentzian metrics on  $\mathcal{M}$  (various topologies on this space can be obtained even without introducing an auxiliary metric, like the one I alluded to in §2.2, used to define stable causality).

We could give a non-invariant definition of the type given above for approximate flatness, and argue that it might be meaningful, but I will describe instead a “global” approach to the definition of closeness of two lorentzian metrics, in the sense of “global” that I used when discussing density. Such a definition comes again from thinking of the metrics as given by a volume element and a causal structure: we will say that two metrics are close to each other if their volume elements approximately agree,  $\det g_{(1)} \approx \det g_{(2)}$ , and also their light cones, in the following sense. Consider two points,  $x < y$ , in  $\mathcal{M}$ . The two metrics will assign different Alexandrov neighborhoods to them,  $A_1(x, y)$  and  $A_2(x, y)$ . The symmetric difference

$$\Delta A(x, y) := A_1 \Delta A_2 \equiv (A_1 \setminus A_2) \cup (A_2 \setminus A_1) \quad (2.3.3)$$

contains some information on how much the two causal structures disagree, at the scale set by the proper time between  $x$  and  $y$ . Thus, a quantity like

$$\delta(g_{(1)}, g_{(2)}) := \sup_{x < y} \frac{V[\Delta A(x, y)]}{V[A_1(x, y) \cup A_2(x, y)]} \quad (2.3.4)$$

could capture the difference between the two metrics, provided we make it into a well-defined quantity, by specifying which metric measures the volumes, and arrange for it to satisfy “transitivity of the relation  $\approx$ ”, i.e.,  $g_{(1)ab} \approx g_{(2)ab} \approx g_{(3)ab} \Rightarrow g_{(1)ab} \approx g_{(3)ab}$ . This would be guaranteed, e.g., if  $\delta$  satisfied the triangle inequality, at least for metrics sufficiently close to each other.

### Our main conjecture

We will require conditions (1), (2) and (3) to be satisfied for a manifold to be a reasonable continuum approximation to a causal set. Why we would want to impose the first two conditions should be clear. The reason for imposing (3) is not just that small lengths would not be meaningful, but also that (1) and (2) by themselves would be far from determining a unique approximating  $(\mathcal{M}, g_{ab})$ : given any manifold with the right causal structure, i.e., satisfying (1), we could always arrange the density to have a constant value by setting the conformal factor appropriately; but in doing so, we would in general introduce an unreasonably large curvature, or other small characteristic lengths. (It seems possible, however, 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 spacing between embedded points or smaller.)

But, taken together, these conditions are very strong: in general, for a given causal set  $P$ , there will be no manifold  $\mathcal{M}$  in which  $P$  can be faithfully embedded; in fact we expect almost all causal sets not to be faithfully embeddable anywhere. Even after imposing dynamics, there is no compelling reason to believe that the dynamically preferred causal sets will be faithfully embeddable. We will come back to this question, and to our proposal for obtaining nevertheless well-defined



continuum approximations, in §2.8: here, as I remarked earlier, we are interested instead in the uniqueness of such a continuum approximation, when it exists.

The question we posed at the beginning of this section then becomes: for a given causal set  $P$ , how unique are the topology, differentiable structure and metric of the manifold  $(M, g_{ab})$ , if we require that there exist a faithful embedding  $f : P \rightarrow M$ ? Our main conjecture (Hauptvermutung) is that the topology and differentiable structure are unique, and the metric is determined up to “small variations”, i.e., if there exist two such faithful embeddings, the two manifolds are approximately isometric.

In support of our expectation that a causal set has a structure rich enough to imply all the geometrical properties we attribute to continuous space-time, I will (almost) show that the dimensionality of this manifold is unique, and sketch the way in which one could prove that the topology and metric are also unique. The argument will rest on the fact that dimensional information is really contained in small subsets of the causal set, each of which, in the faithful embedding, covers an approximately flat region, while the information about topology and curvature is contained in the way these subsets are “put together”. We will thus take a long detour to study first properties of embeddings of causal sets in Minkowski space, in particular of small causal sets, and then other questions related to causal set kinematics, and come back to the general problem at the end of this chapter.

## 2.4 Causal sets embedded in Minkowski space

The question I want to consider in this section is: if we are given a causal set, how do we know in which, if any, Minkowski space it can be causally embedded (no uniform density requirement)? In particular, what is the minimal dimensionality of the latter? This minimal dimension will be called the *causal dimension* of the causal set. As I mentioned in the previous section, the results will in general be applied not to whole causal sets, but to small subsets of these which can be thought of as embedded in Minkowski space, since, because of their size, they do not see the curvature, and, because they are subsets of a larger causal set, they do not need to be uniformly embedded themselves. Suitable such subsets will contain the information on the physical dimension of the causal set, and part of what we want to know is which are those subsets and how do they contain dimensionality information.

On which features of the causal set does the causal dimension depend? It obviously depends on its size, but it also depends in a highly nontrivial way on the causal relations. It seems clear that, the larger the number of points, the higher the causal dimension one can force by using a suitable causal order. We will see below how to give an upper bound on the causal dimension for a fixed number of points, as well as which causal orders are more “efficient” for obtaining a large causal dimension, but let us first look at a few trivial examples:

- (a) For any  $N$ , the totally ordered set on  $N$  elements,  $P_N^l$ , can be embedded in 1-dimensional Minkowski space;
- (b) For any  $N$ , the “fence” causal set  $P_{2N}^f$  on  $2N$  elements has causal dimension 2, while the “crown” causal set  $P_{2N}^c$  has causal dimension 3; notice that removal of any single link transforms  $P_{2N}^c$  into  $P_{2N}^f$ .
- (c) For any  $N$ , the “brush” causal set  $P := \{a_i; i = 1, \dots, N \mid a_i < a_N; i = 1, \dots, N - 1\}$  has causal dimension 2;

(d) Any causal set  $P$  has the same causal dimension as its dual  $P^*$ .

The first remark we can make then is that just linking many points together will not necessarily increase the causal dimension. Rather, the causal dimension arises as a “nonlocal” combined effect from the whole structure of the causal set, as example (b) suggests, and it measures some aspect of its complexity.

### An upper bound on causal dimension

Before I give an upper bound on the causal dimension of a causal set, I will state three lemmas. The first one provides an alternative characterization of causal sets embedded in Minkowski space, which can be useful in a more general context than the proof of the theorem below. Either the second or the third can be used in the proof of the theorem on the upper bound (and they really incorporate all the difficulty of this proof) but both are given, because of their relevance to other sections below, and because neither of them has been proven, although the sketches of proofs I will give make us believe that they are true. Lemma 3, in particular, establishes a relationship between the causal dimension of a causal set, and the linear dimension we saw in §2.1. As a warmup for this relationship, I will give two (trivial) results: for any causal set  $P$ ,  $ldim(P) = 1 \Leftrightarrow cdim(P) = 1$  (the causal order on 1-dimensional Minkowski space coincides with the coordinate partial order on  $\mathbb{R}^1$ ), and  $ldim(P) = 2 \Leftrightarrow cdim(P) = 2$  (the causal order on 2-dimensional Minkowski space coincides with the coordinate partial order in  $\mathbb{R}^2$ , since one can take these coordinates to be the null ones,  $u = x^0 + x^1$  and  $v = x^0 - x^1$ ).

**Lemma 1.** If a realization of a finite causal set  $P$  is found in terms of balls in Euclidean space (a special case of the realization with open sets in a topological space, described in §2.1), another realization in terms of points embedded in

Minkowski space can immediately be constructed, and viceversa.

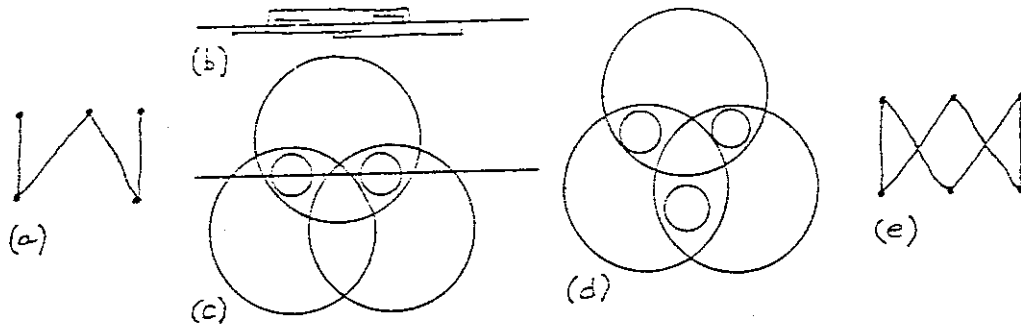
*Proof.* The proof is really very simple. Suppose first we have the realization with balls in  $n$ -dimensional Euclidean space. Then we can embed this Euclidean space in  $(n + 1)$ -dimensional Minkowski space, as, say, the  $t = 0$  hypersurface, and draw a set of past light cones such that the intersection of their interiors with the  $t = 0$  hypersurface coincides with the given balls. For each ball there is a unique such light cone, and it is easy to see that the vertices of these light cones provide a realization of the same causal set. Suppose, viceversa, that we are given a set of points in Minkowski space which realize a poset  $P$ . Then construct the past light cones of these points and consider their intersection with a flat hypersurface to the past of all the points. This defines a ball for each light cone, and the collection of these balls is the desired realization of the causal set. •

One potentially useful consequence of this lemma is that, since it reduces causal relations in Minkowski space to inclusion relations among balls in Euclidean space, we can avail ourselves of the known results on the number of regions that such a space can be divided into, when trying to establish whether a given embedding is possible. For example, the maximum number of regions  $\mathbb{R}^n$  can be divided into by  $k$   $n$ -balls is  $\binom{k-1}{n} + \sum_{i=0}^n \binom{k}{i}$  (see, e.g., Comtet [59]). There is one thing to be careful with, however, when using such results. The fact that in a given causal set some elements form an antichain does not mean that, in a causal embedding in Minkowski space, they will lie on the same spacelike hyperplane. Rather, there will exist hyperplanes on which some appear as points, while the light cones of the others intersect the hyperplanes in spheres, with no inclusion relations between them, nor with the points.

**Conjectural lemma 2.** Adding one maximal or minimal point to a causal set can increase its causal dimension at most by 1.

*Sketch of proof.* Suppose a causal set  $P_{N-1}$  of size  $N - 1$  has causal dimension  $n$ . Then, by lemma 1, it can be realized as a collection of  $N - 1$   $(n - 1)$ -balls  $B_i^{n-1}$  in  $(n - 1)$ -dimensional Euclidean space  $E^{n-1}$ . Embed this space in  $E^n$  as, say, the  $x^n = 0$  hypersurface. Suppose we want to add a minimal element  $x_N$  to the causal set, with  $p_N \prec p_{i_k}$ , for some subcollection of elements  $\{p_{i_k} | i_k = 1, \dots, M < N - 1\}$ . Then we want to show that from the  $B_i^{n-1}$  we can construct  $n$ -balls  $B_i^n$  such that the inclusion relations among them are preserved, and that we can find an additional  $n$ -ball  $B_N^n$  with  $B_N^n \subset B_{i_k}^n$  for all  $i_k$ , but not contained in the others. It seems reasonable to expect that one can do so.

An illustration of what I mean by this construction can be found in figure 2.4.1. We start with a 5-element causal set of causal dimension 2. A causal embedding in 2-dimensional Minkowski space yields the 1-balls shown, as intersections of the past light cones of the embedded points with a spacelike line. After promoting these 1-balls to a set of 2-balls in a plane, with the same inclusion relations, an additional 2-ball can be added, to yield a realization of the 6-element crown, which, by example (b) above, has causal dimension 3.



Adding a minimal element

figure 2.4.1

**Conjectural lemma 3.** Given a linear realization of a causal set in  $\mathbb{R}^n$ , one can obtain a causal one in  $(n + 1)$ -dimensional Minkowski space.

*Sketch of proof.* The idea is the following. A linear realization of a causal set can be thought of one in a space-time with “square (or maybe better triangular) light cones”. It seems not possible to obtain a map from  $\mathbb{R}^n$  to  $n$ -dimensional Minkowski space which takes all “square” light cones into round ones, but it may be possible to produce one into a suitable region of  $(n + 1)$ -dimensional Minkowski space. We start by mapping  $\mathbb{R}^n$  linearly onto itself, squeezing all “square” light cones into thin wedges. This can be done by a map which takes the  $i$ -th standard basis vector  $e_i$  into  $e_i + \alpha \sum_{j=1}^n e_j$  (normalization is not important), for some  $\alpha > 0$ , extended to all vectors by linearity. Now compose this map with one that takes  $\mathbb{R}^n$  onto one of the subspaces of the tangent space of the origin in  $(n + 1)$ -dimensional Minkowski space which are tangent to its light cone, with the axis of the “square” light cone of  $0 \in \mathbb{R}^n$  mapped to the line along which they are tangent. The original points of the linear realization in  $\mathbb{R}^n$  are now all spacelike related to each other, but “almost null-related”. The final step is to project them to each other’s light cones, to obtain the correct relations. To do this, start from the minimal points, and project them (say, along radial directions on each  $t = \text{const}$  hyperplanes) on the light cone of the origin; then proceed by projecting the points in the second layer onto the intersection of the light cones of the first layer points they are linked to, and so on. If  $\alpha$  is large enough, i.e., the “square” light cones squeezed down enough, this projection should displace the points by an “arbitrarily small” amount, which should ensure that only the desired relations between them are established in the process, unless somehow the “errors” accumulate.

I will proceed to the “theorem” on the upper bound for the causal dimension, as if either of these lemmas had been proved. From now on all results whose proof depends on the validity of the lemmas will be indicated by the symbol  $\star$ .

\* **Theorem 1.** A causal set of size  $N$  can always be embedded in  $(N - 1)$ -dimensional Minkowski space (and has thus causal dimension  $< N$ ).

*Proof.* It is very simple to prove the theorem by induction. The (only) 2-element causal set  $P_2^l$  obviously can be embedded in 1-dimensional Minkowski space. Thus, for  $N = 2$ , the claim is true. Now, any causal set  $P_N$  of size  $N$  can be thought of as constructed from a causal set  $P_{N-1}$  of size  $N - 1$  by adding a maximal (or minimal) point. But, if lemma 2 holds, we know that by adding such a point the causal dimension can increase at most by 1. Therefore, if the statement of the theorem is true for  $P_{N-1}$ , it is also true for  $P_N$ . If lemma 3 holds, then the bound can be tightened: since  $ldim(P) \leq |P|/2$  (§2.1),  $cdim(P) \leq |P|/2 + 1$ ; other bounds can be obtained using the other known inequalities for  $ldim(P)$ . •

#### Small causal sets and dimension

We now want to study the question of what is the minimal structure a causal set must have in order to tell us something about the manifold. Let us call *n-irreducible* a causal set whose causal dimension is  $n$ , and decreases upon removal of any point in the causal set (presumably by one). Then we will call a *pixie* for  $n$  dimensions a “minimal”  $n$ -irreducible causal set, i.e., one with the minimum size.<sup>†</sup> We do not know yet whether  $n$ -pixies exist for all  $n$  (because we don't know that for all  $n$ , causal sets with  $cdim(P) = n$  exist), nor if they are unique, but we can say something about their size (indicate an  $n$ -pixie by  $\mathcal{P}_n$ ):

\* **Lemma 4.** The size of an  $n$ -pixie is at least  $n + 1$ , and  $n > m \Rightarrow |\mathcal{P}_n| > |\mathcal{P}_m|$ .

*Proof.* The first part follows immediately from theorem 1. The second part from

<sup>†</sup> It will be useful to assume, at a later point, that pixies can always be embedded so that no two points in them are null-related. We will assume here that this is always possible—if in a particular embedding, two points end up null-related, move one of them slightly inside the light cone of the other—but it will have to be checked once pixies are known.

lemma 2: suppose  $|\mathcal{P}_n| < |\mathcal{P}_m|$ ; if we remove any maximal or minimal element of  $\mathcal{P}_n$ , we get a causal set of causal dimension  $n - 1$ ; if we keep removing elements, at some point we will have a causal set of causal dimension  $m$  and size less than  $|\mathcal{P}_n| < |\mathcal{P}_m|$ . But then  $\mathcal{P}_m$  cannot be an  $m$ -pixie, contrary to our assumption. •

Notice that the same result holds for linear dimension, since it just depends on lemma 2, which for linear dimension is known (not just conjectured) to be true.

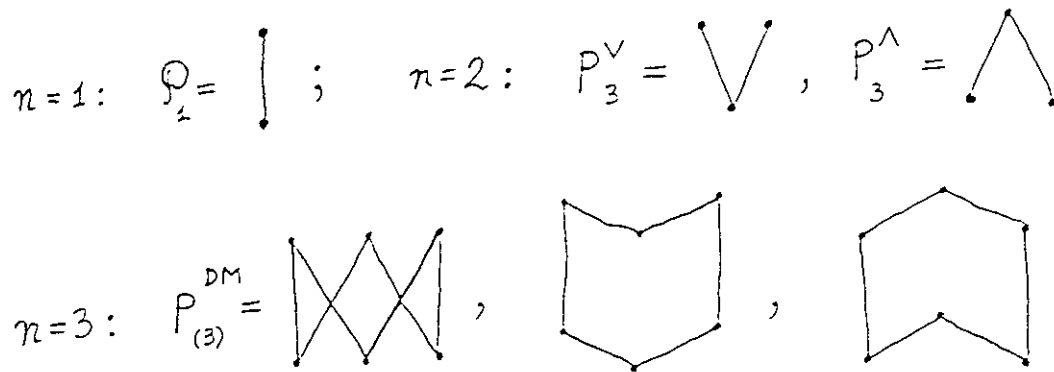
Let us look explicitly at the lowest dimensional cases (figure 2.4.2). In dimension 1 there is only one pixie,  $\mathcal{P}_1 = \{a, b \mid a < b\}$ ; in two dimensions, two 3-element pixies,  $\mathcal{P}_3^\vee$  and  $\mathcal{P}_3^\wedge$  (duals of each other); and in three dimensions, three 6-element pixies (one self-dual,  $\mathcal{P}_{(3)}^{DM}$ , and the other two taken into each other under duality), as one can see by trial and error. But in fact one can also use the knowledge available for pixies in the linear dimension sense (see, e.g., Kelly and Trotter [62]) and show that

**Lemma 5.** The 3-dimensional causal pixies coincide with the 3-dimensional linear pixies (i.e., the three 6-element pixies in figure 2.4.2).

*Proof.* First of all, any  $\mathcal{P}_3$  has 6 elements: if it had less than 6, it would have linear dimension, and thus causal dimension, less than 3; but all three 6-element linear pixies are by inspection irreducible in the causal sense, and they must be pixies. Consider now any causal 3-pixie  $\mathcal{P}_3$ . It cannot have linear dimension 2, since then its causal dimension would also be 2; it cannot have linear dimension greater than 3, because of lemma 4. Thus it must have linear dimension 3, and, since it has 6 elements, it must be a pixie. •

It is simple to convince oneself that the three 6-elements causal sets in figure 2.4.2 are really pixies, by removing any one element and verifying explicitly that





Pixies for 1, 2, and 3 dimensions

figure 2.4.2

the resulting 5-element causal set can be embedded in 2-dimensional Minkowski space (it may be useful to look at figure 2.1.1, where all 5-element causal sets are shown). Also, I should remark that a result like lemma 5 cannot be shown if one replaces “pixies” by “3-dimensional irreducible causal sets”.

Finding pixies for higher dimensions becomes quickly difficult, but some facts can already be inferred, like the fact that  $n$ -pixies, if they exist, are not unique, and, if  $\mathcal{P}_n$  is one, then so is  $\mathcal{P}_n^*$ . Theoretically, it would be very helpful if we had a lower bound on the causal dimension of a causal set. Since this is not available, we can try to guess what pixies might look like. A computer program, described in appendix 2, has been written to help in this task, by telling us whether a given causal set  $P$  can be (causally) embedded in a given dimension  $n$ . However, this program (interesting as it is by itself) will by no means solve our pixie problem: it can show that  $P$  can be embedded in  $n$  dimensions, but it can never guarantee that it can't be embedded.

Almost as useful as knowing a family of pixies for all dimensions, would be to know some family of causal sets containing (at least) one causal set for each dimension, just like we had in the case of linear dimension (see equation (2.1.3)).

One possible such family is the “hypercube” causal sets  $P_{2^n}$  (figure 2.1.4). This has the property that

**Lemma 6.** Any (connected) causal set  $P$  of size  $N$  can be embedded as a sub-causal set in  $P_{2^{N-1}}$ .

*Proof.*<sup>†</sup> (This proof has nothing to do with causal realizations: it is just combinatorics.) From the definition of  $P_{2^N}$ , we see that it can be considered as the set of all strings of  $N$  bits (e.g., (01101...1001)), where  $p < q$  iff the string  $q$  contains at least all the 1's present in the string  $p$ . The proof will now proceed by induction. The 2-element causal set  $\{(0), (1)\}$  is  $P_{2^1}$ , so a fortiori it is contained in it. Suppose now that  $P$ , of size  $N$ , is embeddable in  $P_{2^{N-1}}$ , and consider a causal set  $Q$ , of size  $N + 1$ , obtained by adding a minimal element  $p_{N+1}$  to  $P$  (without loss of generality, as remarked in the proof of theorem 1). Then  $P$  is a subset of all possible  $(N - 1)$ -strings of bits. Construct a collection of  $N$ -strings of bits as follows. To each string/element  $p_i \in P$ , append one bit which is 1 if  $p_i$  is to the future of  $p_{N+1}$ , and 0 otherwise; finally add an extra  $N$ -string to the collection, given by (00...01). It is simple to check that this collection of  $N$ -strings realizes  $Q$ , and it is by construction embedded in  $P_{2^N}$ . •

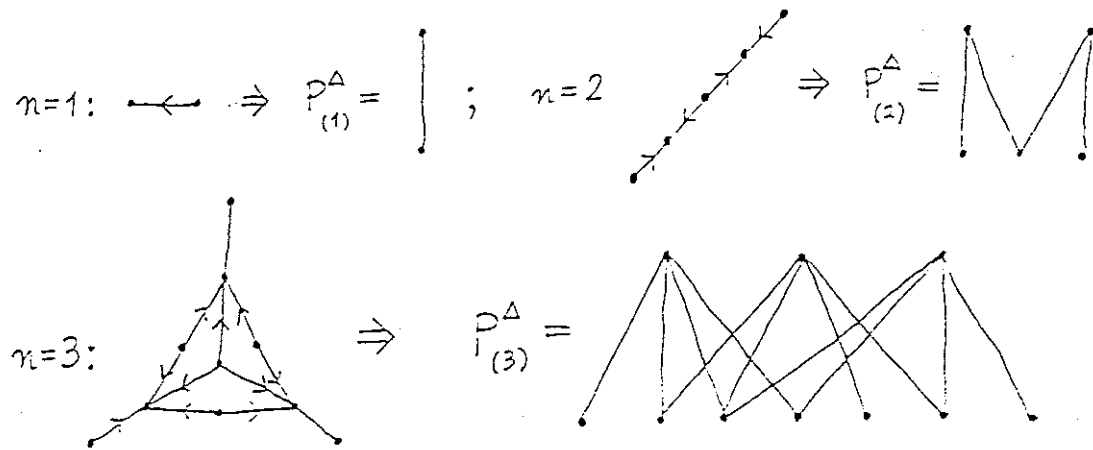
Another possible family in which to look for causal sets of all dimensions is the collection of causal sets  $P_{(n)}^\Delta$  defined by the following construction. Take an  $(n - 1)$ -simplex and join its barycenter to its  $n$  vertices, and the barycenters of all its 1-, 2-, ...,  $(n - 2)$ -faces to the vertices of the corresponding faces. Finally, add  $n$  external points and join one of them to each vertex. Then the causal set is represented by all these  $n + 2^n - 1$  points (vertices, barycenters and external points), with the above links defining the order relation, each vertex following

---

<sup>†</sup> Proof provided by D. Meyer.

(or preceding, if desired) all points it is joined to.

One can easily see that  $P_{(1)}^\Delta = P_2^I$ ,  $P_{(2)}^\Delta \supset P_3^V$ , and  $P_{(3)}^\Delta \supset P_{(3)}^{DM}$ , and they are embeddable, respectively, in 1, 2, and 3 dimensions. However,  $P_{(4)}^\Delta$  is also embeddable in 3 dimensions, as can be shown by producing the corresponding 2-ball diagram (not reproduced here because it is not trivial, and looks a bit messy—especially in black and white). Since  $l\dim(P_{(4)}^\Delta) \geq 4$ , because  $P_{(4)}^\Delta \supset P_{(4)}^{DM}$ , this fact shows that linear dimension does not always coincide with causal dimension, and cannot be used as a physical definition of dimension of a causal set.



The first three  $P_{(n)}^\Delta$ 's

figure 2.4.3

### Other aspects of causal dimension

I conclude with a simple result on the behavior of causal dimension under the operations defined for causal sets:

**Lemma 7.** For finite causal sets  $P$  and  $Q$ , under the operations of disjoint (cardinal) and ordinal sum,

$$cdim(P \div Q) = cdim(P \oplus Q) = \max[cdim(P), cdim(Q)]. \quad (2.4.1)$$

*Proof.* The proof is obvious: suppose  $\text{cdim}(P) = m$ ,  $\text{cdim}(Q) = n$ ; then there exist separate causal embeddings  $f : P \rightarrow \mathbb{R}^{\max(m,n)}$  and  $g : Q \rightarrow \mathbb{R}^{\max(m,n)}$ , with all embedded points contained in compact subsets of  $\mathbb{R}^{\max(m,n)}$ , respectively  $C \supset f(P)$  and  $D \supset g(Q)$ ; but then an embedding of  $P + Q$  is obtained by embedding  $P$  and  $Q$  simultaneously, with all points  $f(P)$  spacelike related to  $g(Q)$ , and one of  $P \oplus Q$  with all of  $f(P)$  to the past of  $g(Q)$ . •

Less trivial, and much more immediately useful, would be to show a similar result to lemma 7, but for the cartesian product, analogous to the known result for linear dimension. Even showing that  $\text{cdim}(P \times P_2^l) \leq \text{cdim}(P) + 1$  for all  $P$  would be extremely useful, since it would immediately imply that  $P_{2^n} \equiv (P_2^l)^n$  has causal dimension at most  $n$ . From the fact that a causal set  $P$  of size  $N$  is embeddable in  $P_{2^{N-1}}$ , as we saw, it would then follow that  $P$  has causal dimension at most  $N - 1$ , i.e., the result of theorem 1, without having to prove lemmas 2 or 3.

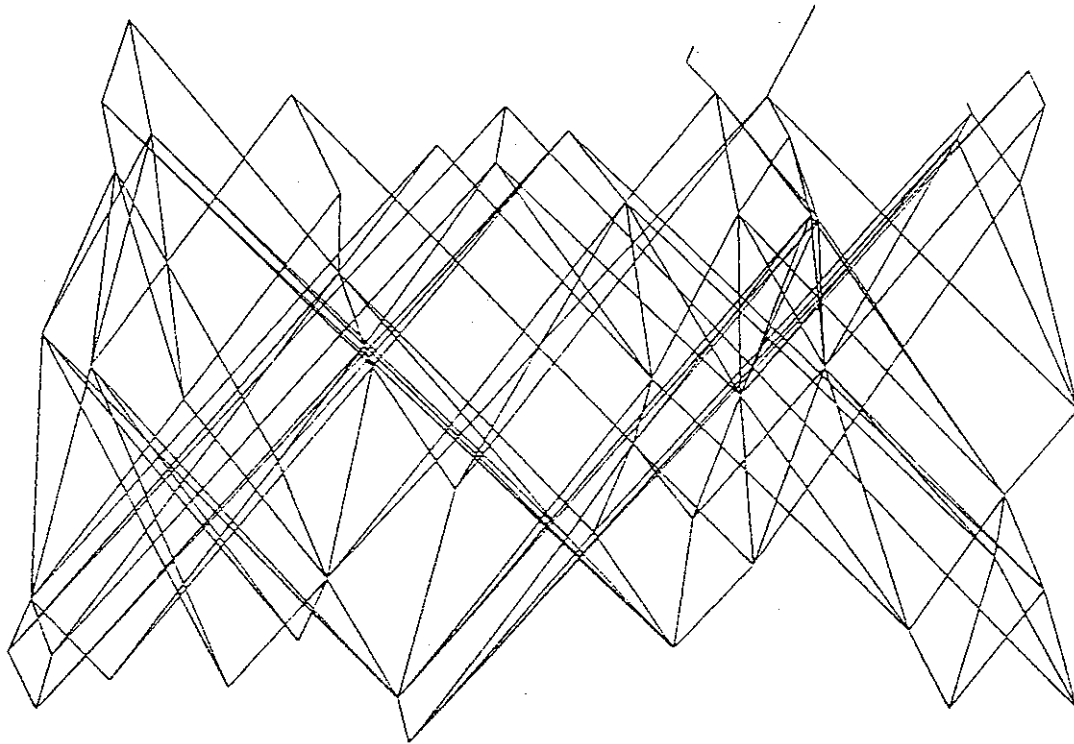
## 2.5 Causal sets sprinkled in Minkowski space

In this section we will mainly address the following question: if we start with Minkowski space, and sprinkle points in it randomly (i.e., with uniform density), what properties do we expect the resulting causal set to have? In particular, the properties we will be interested in are of a “statistical” nature: how many relations will there be, on the average, between all points in a given space-time volume? About how many links will each point have? How many paths will there be between any two of them?

The motivation for considering this kind of question is twofold. On the one hand, we will use these results in the next section to define notions of dimensionality for general causal sets, based on the dependence of these “statistical” or “global” quantities on dimension found here. These notions will be meaningful when looking at properties of large collections of elements, e.g., those belonging to some large Alexandrov set, as opposed to looking for specific causal relations among few elements, like searching for pixies. On the other hand, as we will see in §2.7, the results given here will form the basis for a study of what happens to various properties of causal sets when the embedding manifold is curved.

Consider thus  $n$ -dimensional Minkowski space, with a fixed uniform random sprinkling of points in it. This, as we saw, means that an Alexandrov neighborhood of size  $V$  will contain on the average  $\rho V$  points, with a Poisson probability distribution of its containing some number  $N$  of points, and that a lorentzian random sprinkling of points looks exactly the same as a euclidean random sprinkling. This is very useful, since it means we can carry over some of the techniques developed in euclidean space for the treatment of random or poissonian lattices (Santalo [83]; Itzykson [76]; Lee [79]).

But in a very important respect lorentzian random sprinklings are different from euclidean ones. Looking at an example of such a sprinkling, for  $n = 2$ , shown in figure 2.5.1, we can notice the high number of links per point, in particular along “almost null directions”, which makes this diagram look quite different from a euclidean 2-dimensional random lattice, where each point has 6 links on the average (no matter how large the lattice is), or from, say, figure 2.1.2, more closely corresponding to one’s naive expectations. In fact, it is easy to see that Lorentz invariance of a random sprinkling implies that, if the figure showed an infinite region, each point would have an infinite number of links.



A causal set sprinkled in  $n = 2$

figure 2.5.1

To find out how fast this number grows with volume, we can pick an Alexandrov neighborhood  $A(x, y)$  in  $M$  (with  $x$  an *embedded* point), and calculate the

expected number of links  $x$  will have with other points in  $A$ , by writing the expected number of links with a *manifold* point  $x'$  contained in a volume element  $dV'$  of  $A$ , and integrating with respect to the latter over  $A$ . We start with the following observation. Since for an Alexandrov neighborhood of volume  $V$ , the probability that there will be exactly  $N$  embedded points inside it is  $(\rho V)^N e^{-\rho V}/N!$  (equation 2.3.1), in particular, the probability of having no points is  $e^{-\rho V}$  and that of having exactly one point in an infinitesimal volume element  $dV$  is just  $\rho dV$ . Thus, the probability of having a link between  $x$  and  $x'$  is

$$P(dx \prec dx') = e^{-\rho V|A(x,x')|} (x \prec x') \rho dV'. \quad (2.5.1)$$

But this is also the expected number of such links: for  $x$  and  $x'$  fixed, let us call  $\chi(x, x')$  the variable whose value is 1 if there is a link between these points, and 0 otherwise. Then its expectation value is

$$\langle \chi(x, x') \rangle = 0 \cdot P(\chi=0) + 1 \cdot P(\chi=1) = P(\chi=1). \quad (2.5.2)$$

But  $P(\chi=1)$  is just (2.5.1), and therefore the expected total number of links  $x$  has is

$$\langle N_l(x) \rangle = \int_{J^+(x)} e^{-\rho V(x,x')} \rho dV'. \quad (2.5.3)$$

The result is that  $\langle N_l \rangle$  grows logarithmically with volume in 2 dimensions, linearly in 3, and so on.

Although, for definiteness, one can read the formulae in this section setting  $\rho = 1$  (in natural or Planck units), we will leave  $\rho$  (uniform but) arbitrary in them. By doing this, we obtain, at very little cost, an increased “dimensional readability” of the formulae, and a greater generality we might want to make use of, in cases where we have to “coarse-grain” the causal set (see §2.8).

### Expected number of paths

Let us now see how to calculate the expected number of paths between two timelike related points  $x$  and  $y$ ,  $x \prec y$ , a distance  $l$  apart in  $n$ -dimensional Minkowski space (see also §3.4). By a similar argument to the one that led to (2.5.1) and (2.5.3), we obtain that the probability of having a  $k$ -link path between two *embedded* points  $x$  and  $y$ , with intermediate points  $x_1, x_2, \dots, x_{k-1}$  (in this order), contained respectively in the small regions  $dV_1, dV_2, \dots, dV_{k-1}$  is

$$\begin{aligned} P(x \prec dx_1 \prec \dots \prec dx_{k-1} \prec y) \\ = \rho^{k-1} dV_1 dV_2 \dots dV_{k-1} \nu(x - x_1) \nu(x_1 - x_2) \dots \nu(x_{k-1} - y), \end{aligned} \quad (2.5.4)$$

where we have defined  $\nu(x - x') := e^{-\rho V(x, x')}(x \prec x')$ , and the expected total number of  $k$ -paths from  $x$  to  $y$  is just (2.5.4), summed over all locations of the  $k - 1$  intermediate points:

$$\langle N_k \rangle = \rho^{k-1} \int dx_1 \int dx_2 \dots \int dx_{k-1} \nu(x - x_1) \nu(x_1 - x_2) \dots \nu(x_{k-1} - y), \quad (2.5.5)$$

where all integrals extend over  $A(x, y)$ . Using  $\langle N_k \rangle$ , we can write down expressions for other quantities of interest, like the expected total number of paths between  $x$  and  $y$ ,

$$\langle N_p(x, y) \rangle = \sum_{k=1}^{\infty} \langle N_k(x, y) \rangle, \quad (2.5.6)$$

or the average number of links (length) of one such path,

$$\langle N_l(x, y) \rangle = \frac{1}{\langle N_p(x, y) \rangle} \sum_{k=1}^{\infty} k \langle N_k(x, y) \rangle, \quad (2.5.7)$$

which can all in principle be calculated, giving an  $n$ -dependent relationship between  $V(x, y)$  and  $\langle N_k \rangle$ ,  $\langle N_p \rangle$  or  $\langle N_l \rangle$ .



In practice, these calculations are not simple, but I will briefly describe a way to get, e.g., the asymptotic value for  $\langle N_p \rangle$  in 2 dimensions, as  $V(x, y) \rightarrow \infty$ .<sup>†</sup> We start by noticing that (2.5.5) is in the form of a convolution,

$$\langle N_k(x, y) \rangle = (\nu * \nu * \dots * \nu)(x, y) = \nu^k(x, y) \quad (2.5.8)$$

(from now on, for simplicity of notation, we will set  $\rho = 1$ ), from which it follows that

$$\langle N_p(x, y) \rangle = \sum_{k=1}^{\infty} \nu^k(x, y) = \left( \frac{1}{1 - \nu} - 1 \right) (x, y), \quad (2.5.9)$$

in a sense I will now specify.

Let us use the null coordinates  $u = x^0 + x^1$ ,  $v = x^0 - x^1$ , and choose the Alexandrov neighborhood defined by  $u \in [0, l]$ ,  $v \in [0, l]$  (there is no loss of generality, since the result has to be Lorentz invariant, i.e., it depends only on the product  $uv$ ). Then, functions  $f(x)$  become functions  $f(u, v)$  defined on  $[0, \infty) \times [0, \infty)$ , and we can define their Laplace transform

$$\tilde{f}(\alpha, \beta) := \iint_0^{\infty} \frac{du dv}{2} e^{-\alpha u - \beta v} f(u, v), \quad (2.5.10)$$

and its inverse

$$f(u, v) = \frac{1}{(2\pi i)^2} \iint_{c-i\infty}^{c+i\infty} 2 d\alpha d\beta e^{\alpha u + \beta v} \tilde{f}(\alpha, \beta), \quad (2.5.11)$$

where  $c$  is any real number such that the line over which we integrate in (2.5.11) is to the right of all singularities of  $\tilde{f}(\alpha, \beta)$ . It can be easily checked that  $\widetilde{f * g} = \tilde{f} \tilde{g}$ . Thus, what I mean by the right hand side of (2.5.9) is the inverse Laplace

---

<sup>†</sup> Calculation by R. Sorkin.

transform of  $[1 - \tilde{\nu}(\alpha, \beta)]^{-1} - 1$ , and one can check that

$$\tilde{\nu}(\alpha, \beta) = \int dy \frac{e^{-2\alpha\beta y}}{1+y} =: F(2\alpha\beta). \quad (2.5.12)$$

Note that  $F$  satisfies  $F'(x) = F(x) - 1/x$ .

We can use the saddle point approximation to estimate the first term (while the second doesn't contribute, because the inverse transform of 1 is  $2\delta(u)\delta(v)$ ),

$$\frac{1}{(2\pi i)^2} \iint 2 d\alpha d\beta \frac{e^{\alpha u + \beta v}}{1 - \tilde{\nu}(\alpha, \beta)} = -\frac{1}{2\pi^2} \iint d\alpha d\beta e^{\Phi}, \quad (2.5.13)$$

where  $\Phi(\alpha, \beta) := \alpha u + \beta v - \ln(1 - F(2\alpha\beta))$ , by approximating the integrand with a gaussian around the maximum of  $\Phi$ . The result is that

$$\langle N_p(x, y) \rangle \approx \frac{e}{2\pi} \frac{\gamma^2}{2 - \gamma^2} \frac{e^{\gamma l}}{\sqrt{\gamma l}}, \quad (2.5.14)$$

where  $\gamma$  is defined by  $F(\gamma^2/2) = 1$ , and turns out to have the value  $\gamma \approx 0.93254288$ . Therefore

$$\langle N_p(x, y) \rangle \approx 0.33283897 \frac{e^{\gamma l}}{\sqrt{\gamma l}}. \quad (2.5.15)$$

Tests of the validity of (2.5.14) (in particular of the approximations involved in deriving it) have been done. These involved checking whether it satisfied the recursion relations derived from considering the paths contributing to  $N_p$  as being extensions of different classes of paths in various Alexandrov neighborhoods contained inside  $A(x, y)$ . Whereas some results are in agreement with (2.5.14), others are in slight disagreement (on the value of the proportionality constant in (2.5.15)), and the question is not completely settled.

**Width of the distributions**

One of our concerns, in the application of formulae like (2.5.5) and (2.5.6), will be how close do we expect the value of  $N_k$  or  $N_p$  to be to  $\langle N_k \rangle$  or  $\langle N_p \rangle$ , respectively, for a given Alexandrov set of volume  $V(x, y)$ . In other words, we would like to know the width of the probability distributions for  $N_k$  and  $N_p$ , a much harder task than finding their mean values, and one which we will only start dealing with. The squares of these widths, the variances of the distributions, are of course

$$\begin{aligned}\sigma_{N_k}^2 &:= \langle (N_k - \langle N_k \rangle)^2 \rangle = \langle N_k^2 \rangle - \langle N_k \rangle^2 \\ \sigma_{N_p}^2 &:= \langle (N_p - \langle N_p \rangle)^2 \rangle = \langle N_p^2 \rangle - \langle N_p \rangle^2 \\ &= \sum_{i,j} \langle N_i N_j \rangle - \sum_{i,j} \langle N_i \rangle \langle N_j \rangle.\end{aligned}\tag{2.5.16}$$

The second term can be written using (2.5.5). For the first term, we notice the following. Call  $N_{i,j}(x, y)$  the number of self-intersecting  $(i, j)$ -loops between  $x$  and  $y$ , i.e., pairs of paths from  $x$  to  $y$ , one with  $i$  links and the other with  $k$ , which are allowed to overlap by any amount. Then

$$N_{i,j} = \begin{cases} N_i N_j & \text{for } i \neq j \\ \frac{1}{2}(N_i N_j + N_i) & \text{for } i = j, \end{cases}\tag{2.5.17}$$

so that

$$\sum_{i,j} \langle N_i N_j \rangle = 2 \sum_{i \leq j} \langle N_{i,j} \rangle + \sum_i \langle N_i \rangle.\tag{2.5.18}$$

To describe an  $(i, j)$ -loop we use two sets of intermediate points  $x_1, x_2, \dots, x_{i-1}$  and  $y_1, y_2, \dots, y_{j-1}$ , but  $\langle N_{i,j} \rangle$  is not simply the product of two expressions like (2.5.6) (this would be  $\langle N_i \rangle \langle N_j \rangle$ ), because now the probability of a link  $x_a \prec x_{a+1}$  is not independent of the probability of a link  $y_b \prec y_{b+1}$ , if  $A(x_a, x_{a+1}) \cap A(y_b, y_{b+1}) \neq \emptyset$ . However, it is still possible that a generalization of the argument that led to (2.5.5) will give an expression for  $\langle N_{i,j}(x, y) \rangle$ .

### Expected total number of relations

As the dimension of the Minkowski space in which we sprinkle our points increases, the causal structure induced on them becomes more intricate. As a result, we expect that, for fixed  $N$ , the total number of related pairs of points out of the possible  $\binom{N}{2}$  increases with  $n$ . This is in fact true, and Myrheim [16] showed that, if we call  $f(n) := \langle N_{rel} \rangle / \binom{N}{2}$  the “ordering fraction”, i.e., the expected number of related pairs of elements  $N_{rel}$  in the causal set over the total number of pairs, its values for causal sets uniformly embedded in Minkowski spaces of low dimensionalities are

$n$	1	2	3	4
$f(n)$	1	$\frac{1}{2}$	$\frac{8}{35}$	$\frac{1}{10}$

To obtain these results, we may proceed as follows. Consider  $N$  points sprinkled uniformly in an Alexandrov neighborhood  $A$  (of volume  $V = N\rho^{-1}$ ) in  $n$ -dimensional Minkowski space. Then the expected number of elements to the future of  $x \in A$  is  $\rho V[J^+(x) \cap A]$ , and the expected number of relations in  $A$

$$\langle N_{rel} \rangle = \int_A \rho^2 V[J^+(x) \cap A] dV. \quad (2.5.19)$$

### Number of inequivalent causal sets

I will now give a rough estimate of the number of causal sets on  $N$  elements which can be faithfully embedded in  $n$ -dimensional Minkowski space. Consider one such causal set and a faithful embedding of it in  $n$ -dimensional Minkowski space, with mean distance  $l$  between embedded points and density  $\rho (= l^{-n})$ , so that the causal set occupies a total volume  $V = Nl^n = N\rho^{-1} = L^n$ , where  $L$  is thus a characteristic linear size of the causal set. Each embedded point  $p$  will be free to move in a small region around its position, without altering its causal

relations to other points. What is the size of this region? Suppose we move  $p$  in a timelike direction. We can keep moving it while it does not change any causal relation, i.e., by a distance  $s$  such that the volume  $V(s) = O(sL^{n-1})$  swept by its light cone is less than  $\rho^{-1}$  (we assume here that the causal set is connected, and its image in Minkowski space has a fairly regular shape, with a small “surface to volume ratio”). Therefore,  $s = O(\rho^{-1}L^{1-n})$ , and our estimate for the volume available to each point is  $s^n = O(\rho^{-n}L^{n-n^2}) = O(l^n N^{1-n})$ . We thus imagine the total volume  $V$  divided into cells of size  $s^n$ , of which there are a total of  $L^n/s^n = O(Nl^n/N^{1-n}l^n) = O(N^n)$  (many of them unoccupied: there are many more cells than embedded points), and estimate the number of different causal sets we can faithfully embed in  $V$  as the number of inequivalent ways of picking  $N$  cells out of the  $N^n$ , i.e.,  $\binom{N^n}{N} \sim 2^{(n-1)N \log_2 N}$ .

We can compare this with the asymptotic estimate for the total number of  $N$ -element causal sets (2.1.2) (assuming that the connectedness requirement does not reduce this estimate significantly). We see that, for large  $N$ , only a vanishingly small fraction of causal sets can be faithfully embedded in any given dimension (at least in flat space). This suggests that, up to modifications due to possible embeddings in curved spaces, as  $N$  increases, the dimension in which most causal sets are faithfully embeddable can be expected to increase as well, and, since we know that causal sets have more and more a three-layer structure (§2.1), that not only pixies exist for all dimensions, but that two- or three-layer pixies exist for all dimensions.

### Obtaining the metric

To conclude this section, I will illustrate, in this special case of flat space-time, how the metric distance between two embedded points can be explicitly expressed using only the causal set itself. 't Hooft [27] showed that, for any two

timelike related points  $x \prec y$  in 4-dimensional Minkowski space, the volume  $V$  of the Alexandrov neighborhood  $A(x, y)$  is related to the proper time  $l$  between  $x$  and  $y$  by

$$V = \frac{\pi l^4}{24}, \quad \text{i.e.,} \quad l(x, y) = \left( \frac{24V}{\pi} \right)^{1/4}, \quad (2.5.20)$$

while, for  $x$  and  $y$  spacelike related, the proper distance is

$$s(x, y) = \min_{u, v} l(u, v), \quad (2.5.21)$$

where  $u$  is to the future and  $v$  to the past of both  $x$  and  $y$ . But these expressions can be directly used for causal set points: e.g., the "expected value" for the timelike distance between two causal set elements  $x$  and  $y$  after embedding in Minkowski space is (2.5.20), where  $V$  is the number of elements in  $A(x, y)$ .

These equations are immediately generalizable to any dimensionality. Equation (2.5.21) does not change, while to generalize (2.5.20), we just notice that in  $n$ -dimensional Minkowski space, the volume of the Alexandrov neighborhood defined by  $x$  and  $y$  is twice that of a cone with base an  $(n-1)$ -ball of radius  $l/2$  and height  $l/2$ :

$$V = \frac{2}{n} \frac{2\pi^{(n-1)/2}}{(n-1)\Gamma[(n-1)/2]} \left( \frac{l}{2} \right)^n = \frac{\pi^{(n-1)/2}}{2^{n-2}n(n-1)\Gamma[(n-1)/2]} l^n, \quad (2.5.22)$$

which can again be easily solved for  $l$  in terms of  $V$ .

Notice that we could have tried to find  $l$  just by counting the number of links in the longest path from  $x$  to  $y$ . This latter method however is subject to larger statistical fluctuations, and we do not know the coefficient of proportionality between  $l$  and the number of links, which, roughly speaking, depends on how "straight" the longest path is, and could be a function of  $n$  (and of  $l$  itself). Furthermore, we will need the two methods together shortly.

## 2.6 Notions of dimensionality of a causal set

Now that we have defined the conditions under which we have a satisfactory embedding of a causal set in a curved lorentzian manifold, it might seem natural to define the dimension of the causal set to be that of the manifold in which we can faithfully embed it (provided we can show it to be unique). We will in fact call this the *physical dimension* of the causal set. Such a notion, however, is not entirely satisfactory from the mathematical point of view, since it is not well-defined for all causal sets—actually for almost none—, nor from a practical point of view, since it does not lend itself to simple computation methods, and it forces us to solve the full embedding problem for the causal set in order to find its dimension, instead of giving us a tool that can help us find one aspect of the continuum approximation. It would be desirable to define a dimension for a causal set which reduces to the physical dimension when a faithful embedding exists, is always well-defined, and is such that a computer can be taught how to calculate it.

As we remarked earlier, the notion one would define by constructing a topological space from the causal set in the standard way we will describe in §2.7, and the ones used in combinatorics (at least the linear dimension) are useless in this respect, since they are not directly related to the physical dimension. The causal dimension defined using embeddings in Minkowski space also won't do, and in fact it seems that there can be no *global* notion of dimension that satisfies our requirements. We can instead define (several variations of) a *quasilocal* dimension for a causal set, which we will call *fractal dimension*, associated not with the whole causal set, but with different regions in it.

Consider first  $n$ -dimensional Minkowski space, and a region of finite volume

$V$  of arbitrary shape in it. If we rescale the region to a similar one of volume  $V' = \alpha V$ , then all linear dimensions in it will be rescaled by a factor  $\alpha^{1/n}$ , all two-surfaces by  $\alpha^{2/n}$ , etc., and, similarly, all other quantities related to the region will change in a computable way, in terms of  $\alpha$  and  $n$ .

This suggests that we proceed in the following way, in the spirit of the calculations of dimensionality for fractals. Given two elements  $p, q$  in any causal set  $P$ , count the number of elements  $V$  contained in  $A(p, q)$ , and calculate one of the many quantities defined by the elements in  $A(p, q)$ , whose theoretical (statistical) dependence on the volume of the Alexandrov set and on the dimension  $n$  is known for causal sets *uniformly* (and randomly) embedded in Minkowski space (as discussed in §2.5, or in the examples below). Then, by inverting the above relation, one can in principle obtain an effective dimensionality for this Alexandrov set. This effective dimensionality  $n_{eff}$  is meaningful if there exists a large region  $R$  of  $P$ , covered by Alexandrov sets all of roughly the same size  $V$ , and all yielding the same  $n_{eff}$ . In general, however, if the calculation of  $n_{eff}$  is repeated for different pairs of points, one will get different values, which just reflects the fact that we are applying the method to a causal set not faithfully embeddable in Minkowski space. We expect nevertheless the effective dimensionalities to agree, if the causal set we consider is faithfully embeddable in a curved manifold, provided they are derived from Alexandrov sets small enough that the result is not significantly affected by curvature. Thus, if  $n_{eff}$  is constant for all Alexandrov sets contained in  $R$ , for a range of volumes  $V$  down to volumes of a few units, then we will say that  $n_{eff}$  is the *fractal dimension* of the region  $R$  in  $P$ .

Which quantities can we actually use to calculate the effective dimension of the Alexandrov set  $A(p, q)$ ? I will give some examples.



### Length of the longest path

In  $n$ -dimensional Minkowski space, the volume  $V(x, y)$  of  $A(x, y)$  is related to the proper time along the geodesic between  $x$  and  $y$  by (2.5.22). Therefore, if  $p$  and  $q$  are points in a uniformly embedded  $P$ , a similar relationship will hold between the number of elements in  $A(p, q)$  and the length of the longest path between  $p$  and  $q$ , up to statistical fluctuations, and, by counting these two quantities for an Alexandrov set and inverting (2.5.22), we can estimate its  $n_{eff}$ .

Two things have to be taken into account, however. First, in the relationship between the size of  $A(p, q)$  and the length of the longest path there is an unknown coefficient, which could depend both on dimension and on the size of the neighborhood, and should be calculated theoretically, or found from properties of computer-generated sprinklings of points in Minkowski space. The theoretical problem is difficult, however, and large enough sprinklings are difficult to produce (and do calculations with); the numerical results obtained so far are not very satisfactory.

Second, the reliability of this method depends of how big the statistical fluctuations are. We understand the fluctuations in the number of points in  $A(p, q)$  from the definition of uniform sprinkling: they go like the square root of this number, which, for large  $A(p, q)$ , is a small fractional error. The crucial factor is then the fluctuation in the length of the longest path. An estimate of it could be obtained from a knowledge of the shape of the probability distribution for  $N_l$ , the number of links in a path from  $p$  to  $q$ , whose expectation value is given by (2.5.7): in particular, from the tail of this distribution for large  $N_l$ .

Besides statistical fluctuations, a limit to the straightforward applicability of this method (and all other methods) appears when  $A(p, q)$  becomes large enough not to be negligible with respect to the radius of curvature of space-time. We

will see in §2.7 how to turn this to our advantage, and in fact use it to estimate the curvature.

### Total number of relations

Another possibility is to consider a large  $A(p, q)$  and count the number  $N$  of points in it and the number of relations  $N_{rel}$  between them. Computing the “ordering fraction”  $N_{rel}/\binom{N}{2}$ , and comparing the result with the expectation value  $f(n)$  for various dimensions, given by the entries in the table in §2.5 (suitably extended to higher dimensions if necessary), gives an estimate of  $n_{eff}$ . This method seems potentially more reliable than the first one as regards statistical fluctuations, because it is based on larger numbers, and  $N_{rel}$  will not fluctuate much, since it is a sum of a large number of variables, each associated with the existence of a relation between two elements of  $P$ , and the central limit theorem tells us that the fluctuations will be of the order of  $\sqrt{N_{rel}}$ . In comparing it with other methods, we should also take into account to what extent the small fluctuations are offset by a weaker dependence of  $N_{rel}$  on  $n$ , particularly for high  $n$ .

Numerical results have been obtained,<sup>†</sup> by uniformly sprinkling 300 points in an Alexandrov neighborhood of 2-dimensional Minkowski space (for sprinkling algorithms, see appendix 3). The average  $n_{eff}$  for all Alexandrov sets of size 0-30 was 2.1, and similarly for other ranges of sizes, with the best value, 2.007, for sizes 150-180. In 3 and 4 dimensions, only smaller sprinklings have been produced.

### Midpoint method

A method which also uses large numbers is that of comparing  $N$  with the number of points  $N_{1/2}$  in  $A(p, r)$ , where  $r$  is the “midpoint” between  $p$  and  $q$ . This corresponds to rescaling linear sizes by a factor  $\alpha = 1/2$ , and thus volumes by a factor  $N_{1/2}/N = (1/2)^n$ : this ratio depends exponentially on  $n$ , for fixed

---

<sup>†</sup> Results by D. Meyer.

$N$ , so this method is very good for large  $n$ . How do we define the midpoint  $r$ ? The most useful definition, among several possible ones, seems to be the following. All points  $s$  in the “equatorial plane” of the Alexandrov set are such that  $V(p, s) = V(s, q)$ . Among these,  $r$  can be distinguished by the fact that, subject to the above equality,  $V(p, s)$  is maximum for  $s = r$ .

One of the main advantages of the midpoint method is that it is very simple to use, and it looks very promising for computer simulations.

## 2.7 Causal sets sprinkled in arbitrary space-times

Until now, whereas part of the discussion in this chapter referred to general causal sets, which could be faithfully embeddable only in a curved manifold, or not at all, the quantitative results quoted were obtained from causal sets embedded in Minkowski space. Such a restriction did not allow us to go very far in causal set kinematics: the only interesting questions we could ask were continuum oriented and about dimensionality; but, when talking about dynamics, it would limit even more our possibility of understanding the relationship between causal sets and continuum approximation, since we know that in general relativity the action associated with a region of space-time is calculated from its curvature. For a causal set, the ingredients we can use to construct an action are the various elements of structure we have been identifying along this chapter (relations, links, paths, loops, pixies), or other possible subsets that we might identify as “seeds” of structure. We thus want to know how to extract topological and metrical information from a causal set, and what effects topology and curvature have on these elements of structure: how the results of §2.5 have to be modified in a more general manifold.

### Obtaining the topology

We have seen some results on how causal sets contain dimensionality information. Can we get further topological information, i.e., does the causal set  $P$  encode the topology of the manifold  $M$  it is sprinkled in? The way in which we can hope to extract the “germs” of a notion of incidence, around which topology is essentially built, from the causal set, is by constructing from  $P$  a finite topological space, or a finite simplicial complex—a set of triangles, tetrahedra and analogous higher-dimensional simplices, which in simple cases can be considered

as a triangulation of a manifold—or a, similar but more general, finite cell complex. I have to stress that these constructions must depend only on the structure of  $P$  itself, and not on, e.g., any mapping  $f : P \rightarrow M$ . The standard ones are the following (see, e.g., Stanley [65] and Sorkin [24b]).

To obtain a simplicial complex  $\Delta(P)$  from  $P$ , associate with every element a vertex, and with every chain of length  $k$  a  $k$ -simplex, whose vertices are those associated with the  $k + 1$  elements in the chain. Thus, if a  $j$ -chain is contained in a  $k$ -chain ( $j < k$ ), then the corresponding  $j$ -simplex is a face of the  $k$ -simplex. The resulting  $\Delta(P)$  is called the *order complex* of  $P$ , and, obviously, its dimension is the height of the causal set. (Notice that not every simplicial complex is the order simplex of some causal set.)

To obtain from  $P$  a finite topological space  $\mathcal{T}(P)$ —essentially equivalent to the order complex, which can be also obtained from  $\mathcal{T}(P)$ —, use the same underlying set for  $\mathcal{T}(P)$  as for  $P$ , and associate an open set with every “past set” (i.e., a subset of  $P$  containing the past of all its own elements—this is called in combinatorics an *order ideal*). Obviously, a base for this topology is the set of all “elementary past sets”, pasts of single elements of  $P$  (or *principal order ideals*). Notice that this is not just any old finite topological space, it is  $T_0$ : any two distinct elements have distinct sets of neighborhoods (although all neighborhoods of one may very well contain the other—this topology in general is not discrete). The mapping from finite causal sets to  $T_0$  finite topological spaces is a bijection, and it is simple to give a procedure for going back from any such topological space to its associated causal set.

It might seem at first that a simplicial complex would be more desirable than a finite topological space. However, as we remarked regarding the dimension of the simplicial complex, some of its topological properties are not related to

those of the topological space we were hoping to get. The space  $\mathcal{T}(P)$  defined above has the same problem, but finite topological spaces have the following nice feature.

Consider a manifold  $\mathcal{M}$  and a finite open cover  $\mathcal{C}_1$  thereof. This defines a finite  $T_0$  topological space  $\mathcal{T}_1$  in which each point is one of the regions in which the cover divides  $\mathcal{M}$ , and the open sets in  $\mathcal{C}_1$  form a sub-base for the topology. (In turn, we get from  $\mathcal{T}_1$  a poset  $P_1$ ,<sup>†</sup> which, if  $\mathcal{C}_1$  is not a degenerate kind of cover, has a unique rank function: the one associating to each  $p$  the number of open sets of  $\mathcal{C}_1$  whose intersection defines it.) If we now consider an enlarged cover  $\mathcal{C}_2$  of which  $\mathcal{C}_1$  is a subcover, we obtain, by the same procedure, a new finite  $T_0$  topological space  $\mathcal{T}_2$ , with a natural projection  $\pi_1 : \mathcal{T}_2 \rightarrow \mathcal{T}_1$ , which takes  $p_2 \in \mathcal{T}_2$  to the  $p_1 \in \mathcal{T}_1$  such that  $p_2 \subset p_1$ . Proceeding in this way, we obtain a sequence of spaces  $\mathcal{T}_1 \leftarrow \mathcal{T}_2 \leftarrow \mathcal{T}_3 \leftarrow \dots$ , with projections  $\pi_i : \mathcal{T}_{i+1} \rightarrow \mathcal{T}_i$ . But such a sequence, in the category of topological spaces, admits an *inverse limit*  $\mathcal{T}_\infty$ . This inverse limit in general is not a Hausdorff space, but in a mild way: if we define an equivalence relation for in  $\mathcal{T}_\infty$  by  $x \sim x'$  iff they are not “Hausdorff separated” by the topology, i.e., there are no two open neighborhoods  $\mathcal{U} \ni x$ ,  $\mathcal{U}' \ni x'$ , with  $\mathcal{U} \cap \mathcal{U}' = \emptyset$ , then the quotient space  $\mathcal{T}_\infty / \sim$  is nontrivial, and, furthermore, it is homeomorphic to  $\mathcal{M}$ , the original manifold (alternatively,  $\mathcal{M}$  is the set of closed points in  $\mathcal{T}_\infty$ ).

In our case of a causal set sprinkled in  $\mathcal{M}$ , we can also define a cover of  $\mathcal{M}$  by choosing a collection of Alexandrov neighborhoods which cover it. Therefore, if we kept sprinkling points, up to an infinite density, the causal structure on the limiting sprinkling would contain the topological information on the manifold (this is not surprising, given the continuum theorems quoted in §2.2). But,

---

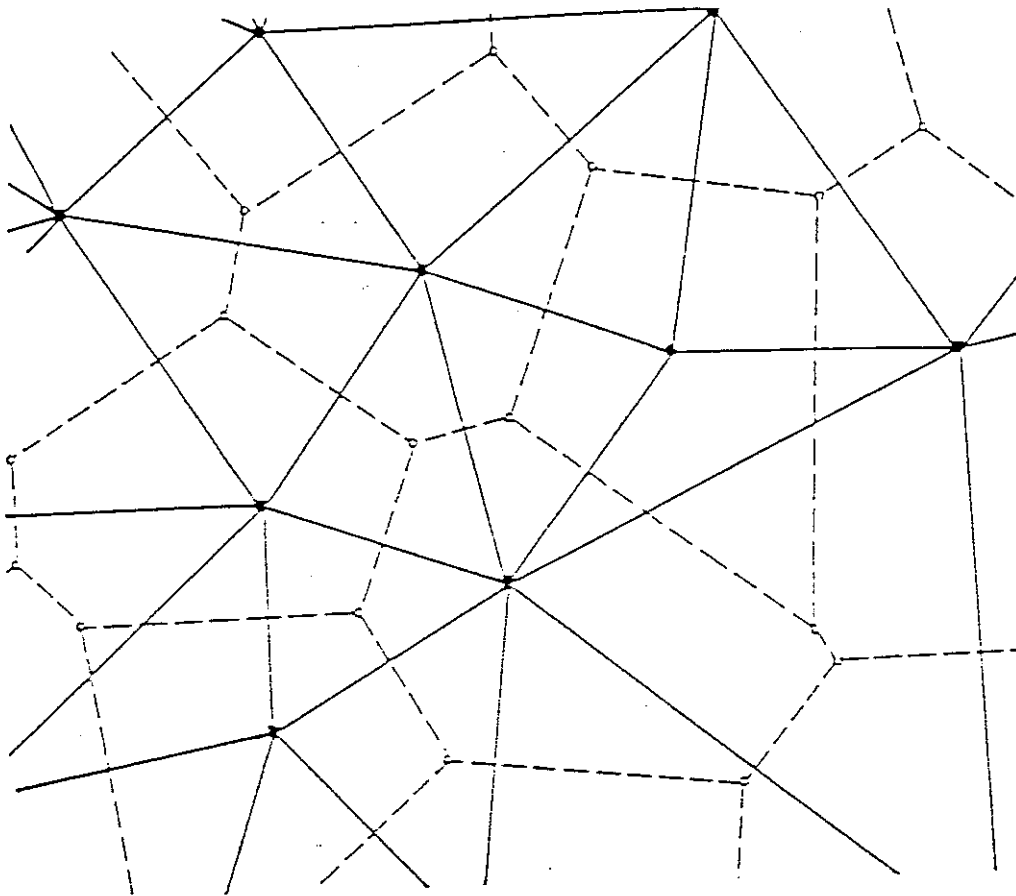
<sup>†</sup> I don't call it “causal set” because the expression would be misleading in this context; furthermore, this poset is not just locally finite: all its “future” and “past” sets are finite.

because of our condition on the length scales of  $\mathcal{M}$ , we expect this information to be available even at a finite sprinkling stage. To be able to use this approach, we should state precisely and prove this last sentence, which includes showing that all the intersection properties of the Alexandrov neighborhoods which define the cover of  $\mathcal{M}$  are fixed by the structure of  $P$ .

A cover of a topological space also defines a simplicial complex, either by first constructing the poset  $P_1$  and then taking its order complex, or by constructing the so-called *nerve* of the cover. In the nerve, to each element of the cover we associate a vertex, and to each non-empty intersection of  $k$  elements, a  $(k-1)$ -simplex, spanned by the  $k$  vertices corresponding to the  $k$  elements. We might hope, therefore, that the nerve of an appropriately constructed cover of  $\mathcal{M}$  by Alexandrov neighborhoods would be a simplicial complex with the right topology to be a triangulation of  $\mathcal{M}$ . This hope is supported by the fact that it is possible to construct coverings of an  $n$ -dimensional  $\mathcal{M}$  with open sets such that only intersections of up to  $n+1$  of them are non-empty (which in fact is the basis for the definition of the *covering dimension* of a topological space), and thus such coverings will have simplicial complexes of dimension  $n$  as nerves. The problem in our case is that we don't know whether such a "good" cover can always be made out of Alexandrov neighborhoods, and, just like in the "truncated inverse limit" approach above, we don't know whether it is possible to choose a cover whose intersection properties are expressible just in terms of  $P$ .

Can one construct a simplicial complex in a more direct way? For random lattices in  $n$ -dimensional Euclidean space there are two procedures. The first is the well-known Dirichlet-Voronoi (or Wigner-Seitz) construction, which always yields a simplicial complex and an associated dual cell complex (see, e.g., Itzykson [76]). To each lattice site we associate one  $n$ -cell, composed of all points which

are closer to this site than to any other one, and is shaped like a polyhedron, whose faces are  $(n-1)$ -cells of the complex. Each of the latter belongs to the hyperplane bisecting, and perpendicular to, the straight line segment between the site in consideration and another nearby site. When two such sites are associated to a common  $(n-1)$ -cell, the line between them is a link in the simplicial complex. If two sites linked to a common site correspond to adjacent  $(n-1)$ -cells, the three sites together define a triangle, and so on (see figure 2.7.1, from [79]).



Dirichlet-Voronoi complex in 2 dimensions

figure 2.7.1

The second construction is a generalization of the first one, and the generalization might be useful for its implementation on a causal set, although it requires some additional structure to be chosen in the latter. I will give it for a



given sprinkling in a lorentzian manifold, following closely what Lee [79] does for euclidean space. (Lee's construction uses, instead of Alexandrov neighborhoods aligned with a vector field, arbitrary convex regions which differ from each other by a translation and/or dilation. When applied to spheres, it is equivalent to the Dirichlet-Voronoi construction.)

Suppose  $(M, g_{ab})$  is a stably causal space-time with a sprinkling of points  $f(P)$  satisfying our faithful embedding conditions. Then every choice of global time function  $t : M \rightarrow \mathbb{R}$  (always possible because of stable causality) determines a unique triangulation of  $M$ . The time function  $t$  gives a timelike vector field  $t^a := g^{ab} \nabla_b t$  defined everywhere on  $M$ . Among all possible Alexandrov neighborhoods  $A(x, y)$  in  $M$  we can now isolate the ones, which are "along  $t^a$ ", in the sense that their endpoints  $x$  and  $y$  belong to the same integral curve of  $t^a$ , and denote them  $A_t(x, y)$ . Let us call "cluster" a set of  $n + 1$  embedded points such that there exists an  $A_t$  on whose boundary they all lie, and which does not contain any other embedded point. Because of our embedding condition (3), each cluster will be contained in a convex normal neighborhood, and any two points of it are connected by a unique geodesic (contained within  $A_t$ ). We can then make the points of each cluster into a simplex (if more than two points lie on the same geodesic, a degenerate situation which will happen with negligible probability, the simplex will be a singular one; another degenerate situation happens when more than  $n + 1$  points lie on the boundary of an  $A_t$ ). The main claim is that the collection of all these simplices forms a triangulation of  $M$ : the proof is essentially the same as the one given by Lee [79] for flat euclidean space.

Of course, once we have the vector field  $t^a$ , we can also define a riemannian metric  $h_{ab} := g_{ab} + 2t_a t_b$ , and follow the construction using spheres with respect to  $h_{ab}$ . This possibility is attractive because, as we saw, it gives both a simplicial

complex and a cell complex. The two methods are not equivalent, and which one will be more convenient to use will depend on how simple it is to characterize the “clusters” in each case, just using the causal set structure, since the construction, as given, heavily depends on the embedding. Furthermore, although the assumption that the space-time be stably causal is not unreasonable for manifolds arising from faithful embeddings, one question to resolve, in trying to recast the construction in causal set terms, is: how can we say “choose a time function” in terms of  $P$ ? Since a time function really chooses a slicing of  $\mathcal{M}$  into spacelike hypersurfaces, the corresponding notion for a causal set is that of rank function, defined in §2.1, but this still doesn’t tell us how to choose clusters of points.

For a proposal for deriving topological concepts from posets, which however does not seem suited to their interpretation as causal sets, see also Marlow [80].

### Obtaining the metric

As we saw, causal sets will always be embedded in such a way that any point  $x_0$  will have an open convex normal neighborhood of volume much greater than one. We thus choose Riemann normal coordinates in this neighborhood for the description of the metric and the curvature, such that at  $x_0$  the metric has the Minkowski form  $g_{\mu\nu}(x_0) = \eta_{\mu\nu}$ . Then, the volume of the Alexandrov neighborhood defined by  $x_1 = (l/2, 0, 0, 0)$  and  $x_2 = (-l/2, 0, 0, 0)$  (i.e.,  $x_1$  and  $x_2$  lie on a timelike geodesic through  $x_0$  and are at a distance  $l/2$  before and after  $x_0$ , respectively), is (Myrheim [16]):

$$V = \frac{\pi l^4}{24} \left\{ 1 + l^2 \left[ \frac{1}{180} R(x_0) + \frac{1}{30} R_{00}(x_0) \right] + \mathcal{O}(l^3) \right\}, \quad (2.7.1)$$

which generalizes (2.5.20). We saw in §2.6 how to use (2.5.22), i.e., (2.5.20) generalized to higher dimensions, to calculate the fractal dimension of  $P$ . This suggests that we generalize (2.7.1) to higher dimensions and use it, for a particular

$P$  faithfully embedded in  $(M, g_{ab})$ , to obtain a “statistical” quantity that we can think of as related to the curvature of  $P$ , once the dimensionality of  $M$  has been obtained from looking at small Alexandrov sets, and we can look at larger ones for deviations from the flat space relationships.

It is important as well to derive the dependence of causal set properties, other than the relationship between volume and length, on curvature, and to check the results with computer-generated sprinklings of points in curved manifolds. The main motivation, in this case, is to find out which quantities are more suitable to enter the definition of the action (or amplitude), given the fact that the latter, in general relativity, is just  $\int R dV$ .

## 2.8 Coarse-grainings of causal sets

As we have already remarked, we expect most causal sets not to be faithfully embeddable in any manifold  $(M, g_{ab})$ . If we believed that in the final theory the dynamically preferred causal sets, i.e., those whose amplitudes interfere constructively (in the path integral language we will adopt for the description of dynamics), are such that they do admit a faithful embedding, then this remark would not be a problem. Such a picture, however, is contrary to our expectations, for at least two (related) reasons.

In the first place, recall the successes of the many Kaluza-Klein type theories, in which space-time is a manifold, but its topology at Planck scales is very different from the effective large-scale one. Such a situation could not arise after faithfully embedding a causal set, because of our requirement on the length scales defined by the geometry. Now, it is conceivable that our preferred causal sets turned out to be faithfully embedded in Kaluza-Klein-like manifolds, with internal manifolds of size much larger than the spacing between points, but it is more appealing to believe that, although the above possibility may be partly correct, the right way of looking at Kaluza-Klein theories is that they give an indication that our large-scale manifold is only an effective approximation to something which at small scales does not have a well-defined dimensionality or topology, and that the (non-gravitational) fields we study on this manifold are just remnants of a structure without a geometrical meaning.

Secondly, on quantum mechanical grounds, we are led to expect that the microscopic structure will be characterized by fluctuations, or described by the superposition of many different causal sets, and we certainly do not expect that, if a causal set  $P$  is faithfully embeddable, any small variation of the causal structure of  $P$  will lead to another faithfully embeddable causal set, since we could be,

e.g., adding a link which creates one higher-dimensional pixie. Furthermore, if the space-time foam picture has a physical significance, it implies in our terms that most causal sets which contribute to a (euclidean??) “path integral” are not faithfully embeddable. It could be that they contain “one unit of topology per Planck volume”, but they could equally well have one (pixie-like) unit of dimension per Planck volume, looking roughly like Kaluza-Klein space-times with different “internal” manifolds.

Finally, I might add, as an internal reason to this theory for wishing that not all physically relevant causal sets be faithfully embeddable in a manifold, that, if they were, not only would life be more boring, but we would lose our main candidate mechanism for making new effective fields appear, in a purely “causal” fundamental theory.

We must thus start to handle the more general case, and associate a manifold to (certain) non faithfully embeddable causal sets. Since our arguments indicate that the trouble lies in the small-scale structure of the causal sets, we introduce the notion of *coarse-graining*, to smooth out this structure. This will take us one step further in the study of kinematics of causal sets in their own right, and will enable us to talk of a causal set representing a larger scale view of another.

A priori, one could think of (at least) two ways of defining a coarse-graining  $P'$  for a causal set  $P$ . One of them is the analog of “lumping together” or identifying elements in some set  $X$  into equivalence classes, and quotienting out by this equivalence relation, inducing somehow a structure among the equivalence classes (this is natural when dealing with topological spaces, or, like in Kaluza-Klein theories, when a group action defines a natural equivalence relation). In our context, an equivalence relation is not so natural, and furthermore care should be taken that no inconsistencies be introduced in the causal relationships among

classes of elements, however defined; but a consistent definition is the following. Consider a collection of Alexandrov sets  $A_i = A(p_i, q_i) \subset P$  such that  $\cup_i A_i = P$  and the volume of all the  $A_i$ 's is roughly equal to some value  $V$ , and induce on it the partial order  $A_i \prec A_j$  iff  $p_i \prec p_j$  and  $q_i \prec q_j$ . This makes  $\{A_i\}$  into a causal set, which we can call a *cover coarse-graining* of  $P$ . A parameter  $p \in [0, 1]$  that characterizes the amount of coarse-graining performed, with  $p = 1$  corresponding to no coarse-graining, can be defined as  $p := V^{-1}$  or  $p := (N - V)/(N - 1)$ .

The second approach to coarse-graining is provided by the notion of subcausal set. A *subset coarse-graining* of a causal set  $P$  will be a  $P' \subset P$ , with the induced partial order, satisfying a condition intended to ensure that it represents a larger-scale view of  $P$ : we require that there exist a parameter  $p \in [0, 1]$  such that the fraction of all  $n$ -element Alexandrov sets of  $P$  which contain  $k$  elements of  $P'$  is approximately  $\binom{n}{k} p^k (1 - p)^{n-k}$ . This definition is similar to (2.3.1) in its "global" character, and, since it implies we can think of  $P'$  as having been obtained by picking at random a fraction  $p$  of the elements of  $P$ , it makes  $P'$  appear sprinkled with uniform density in  $P$ .

Both definitions are meant to lead to coarse-grainings which can be interpreted as preserving only those features of  $P$  with characteristic volume-scale larger than  $1/p$ . In some respects, a subset coarse-graining is simpler to handle than a cover coarse-graining, and it is certainly easier to think about. We will however consider this definition of subset coarse-graining as a tentative one, since it may be forcing too much randomness into  $P'$ . In particular, we would accept a situation in which each Alexandrov set of  $P$  contains *exactly*  $pn$  elements of  $P'$ , if this is possible, and we would like to have the possibility of picking the points systematically in some cases, to avoid situations in which the elements of the subset are just as non-embeddable as  $P$ , e.g., because we are left with a

high-dimensional pixie somewhere.

As regards the cover coarse-graining, one potential problem with it is that we might pick a few  $A_i$ 's which are strongly boosted with respect to the others (i.e., they look very long, along null directions), and would have few relations, of a kind that may make the resulting  $P'$  complicated in an undesired way. One obvious (by now) way to prevent this is to align the  $A_i$ 's using some rank function in  $P$ . The additional structure that this presupposes on  $P$  may however affect properties of  $P'$ , in a way that should be investigated. To balance these difficulties—coming in some sense from a mismatch between the prescription for obtaining a cover coarse-graining and the way we associate a manifold with a causal set, by a map that picks out *points* in the manifold—, it seems easier to induce additional structure besides the causal one on a cover coarse-graining than on a subset coarse-graining, as we will see below.

### Effect on geometry

From our motivation for the introduction of coarse-grainings, it is clear that the most important question to address is that of the effect of coarse-graining on the embedding of a causal set in a manifold. First of all, whichever definition we choose, since the natural notion of volume in a causal set applies to the original  $P$ , before coarse-graining, the density of embedded points in a faithful embedding (if it exists) of the coarse-grained  $P'$  will not be unit, but  $p$ .

As regards dimensionality, it is trivial to show that

**Lemma 8.** If a causal set  $P$  has causal dimension  $n$ , then any subset coarse-graining  $P'$  of  $P$  has causal dimension  $n' \leq n$ .

*Proof.* Obviously, if  $P$  is embeddable in  $n$ -dimensional Minkowski space, then so is any  $P'$  obtained by removing elements from  $P$ . •

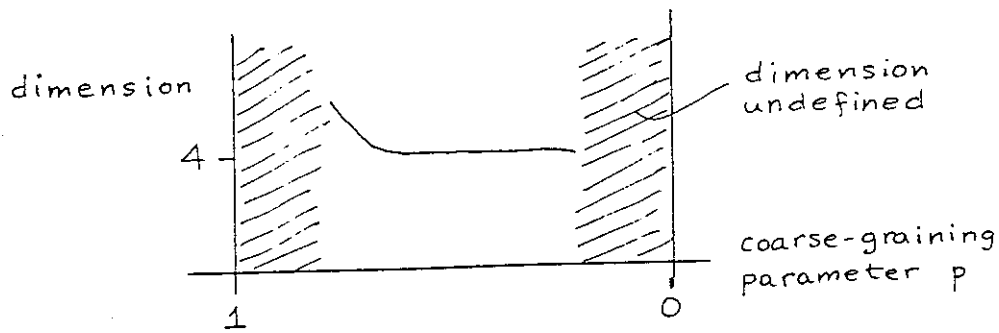
It is very simple to construct examples of subset coarse-grainings that do not affect the causal dimension of a causal set; in particular, if  $P$  is a very large faithfully embeddable causal set, e.g., one obtained by sprinkling points in a manifold, then coarse-graining it with a factor  $p$  not too far from 1 should not make any difference in its dimension. On the other hand, consider the extreme case of a coarse-graining of  $P$ , with  $\text{cdim}(P) = n$ , down to a set with fewer elements than those in an  $n$ -pixie (in the limit, just 1): this will necessarily imply  $\text{cdim}(P') < \text{cdim}(P)$ .

For the cover coarse-graining, similar arguments are harder to make. A result like lemma 8 would hold, e.g., if one could choose in each set of the cover an element  $r_i \in A_i$ , such that  $r_i < r_j$  iff  $A_i < A_j$ , since then the causal set  $\{A_i\}$  would be equivalent to the subset coarse-graining  $\{r_i\}$ , but it seems unlikely that such points exist in general. It cannot be excluded that, if the cover is “bad” in the sense described above, new higher-dimensional pixies are created by the links of the “long” Alexandrov sets, thus increasing the dimension.

The effect of either coarse-graining on other notions of dimension (fractal, physical) is also hard to establish, but we can conjecture that, even for physical dimension, the general effect on a large causal set will be similar to that implied in the remarks above, which can be qualitatively represented in a curve like that of figure 2.8.1. After some degree of coarse-graining, an initially non-embeddable causal set can acquire a well-defined physical dimension, which might however correspond to a Kaluza-Klein type manifold, and need not be the “macroscopic” dimensionality of the causal set, obtained by further coarse-graining.

As regards topology, we will assume that we can start talking in a meaningful way of the topology of a causal set when we can produce from it a simplicial complex and/or cell complex using the Dirichlet-Voronoi construction. At this stage,





Behavior of dimension under coarse-graining?

figure 2.8.1

the topology might be too complicated for a faithful embedding to exist. After a suitable amount of coarse-graining, if such an embedding becomes possible, the topology might still be non-trivial, either globally, because of a possible Kaluza-Klein nature of the manifold, or because of the presence of localized structures, e.g., handles, in the manifold. Further coarse-graining (if desired) could then wash out this remaining nontrivial topology.

### Emergence of structure

As we mentioned above, the notion of coarse-graining is our prime candidate mechanism for producing new effective fields in the continuum approximation, starting from a purely “causal” fundamental theory. Let us see how this comes about. After we perform a coarse-graining, the resulting causal set  $P'$  will obviously contain less information than the original one: it contains less elements, and, just by looking at  $P'$ , we cannot tell how many extra elements there were, and how they were related. Are we interested in the extra information (since, after all, it is only  $P'$  which will be faithfully embedded and give rise to our continuum geometry)? Certainly yes, since, dynamically, the history we will associate an amplitude with is necessarily that represented by the original  $P$ , while the amplitude we would calculate for  $P'$  using the same procedure will in general be very different, leading to a different interference with other histories, and thus

to a different selection of dynamically preferred causal sets, which are the ones that determine the classical limit. From our experience with effective theories, it seems plausible, however, that the amplitude for  $P$  can be expressed in terms of  $P'$  and the “extra information” I mentioned.

But how can we preserve the information on the original causal set in the coarse-grained one? We can do this by attaching numbers to various elements of structure in  $P'$ , which tell us where the additional elements were and how they were linked. As an illustration, consider the following examples. If  $P'$  is a subset coarse-graining of  $P$ , to each element  $p' \in P'$  we can associate the number of links it had in  $P$  (or extra links), to each pair  $p' < q'$  the volume of  $A(p', q')$  in  $P$ . If  $P'$  is a cover coarse-graining, to each element  $p'_i = A_i \in P'$  we can associate the number of elements in the original  $A(p_i, q_i) \subset P$ , to each pair  $A_i, A_j$  the number of elements in  $A(p_i, q_i) \cap A(p_j, q_j)$ , ..., to each collection  $A_i, A_j, \dots, A_k$  the number of elements in  $A(p_i, q_i) \cap A(p_j, q_j) \cap \dots \cap A(p_k, q_k)$ . This last possibility is suggestive, since we expect the  $A_i$ 's with non-empty intersections to be “nearby”; in particular, if we had defined a simplicial complex from  $P'$ , a collection of  $k$  nearby  $A_i$ 's would define a  $k$ -simplex in the complex. One can envisage the possibility of associating a different “field” to each  $k$ -skeleton of the complex, which then, by its nature, would correspond to a different realization of the local symmetry group in the continuum approximation, i.e., to a different kind of tensor field (including “internal indices”?), and, if we had available the cells of a dual complex, like in the Dirichlet-Voronoi construction, “fields” associated with  $(n-k)$ -cells could have a natural interpretation as  $k$ -tensor densities, dual to  $k$ -forms defined on simplices (this is a standard association—see, e.g., Itzykson [76]).

It is therefore natural to think that, assuming we had found a way to encode

in  $P'$  the “extra information”, the “fields” thus defined would be something like a lattice version of the ordinary fields we deal with in particle physics. It would be nice and simple if this were the right picture, but we should bear in mind that between the length scales at which particles appear pointlike to us today and Planck scales there is a difference of 17 orders of magnitude, and thus what we call elementary particle might be an enormously complicated object in our description (we don't expect to be forced to coarse-grain a causal set by a factor close to  $10^{17}$ !). On the other hand, the grand unification (GUT) scales are only about a factor of  $10^3$  larger than Planck scales, so it might appear that GUT physics is describable in relatively simple terms.

But there is room for other, intermediate structure to manifest itself between the causal set level of description and the one in terms of fields in a smooth 4-dimensional background. Recalling the remarks we made about the effect of coarse-graining on topology, we see that the structure which appears as fields on a macroscopic manifold may appear as geons, wormholes, foam-like structure or “internal manifolds” at a smaller scale, and be “purely causal” at the smallest ones. In the transition from each stage to the next, there will probably be an averaging process, which will make fundamentally different structures appear similar. The level of description we will use for a given phenomenon will then depend on the simplicity of the description at each level and our desire to account for it in the most fundamental terms.

## 2.9 Causal sets approximated by differentiable manifolds. II

To wrap up the discussion of causal set kinematics, I will now put together the various concepts introduced along this chapter, and outline the qualitative picture that is emerging, on how individual causal sets can determine a continuum with its additional structure. The main result we would like to prove in this context is our “main conjecture” of §2.3, for which I will only be able to provide the sketch of a proof, and summarize the various leads that have appeared towards a possible proof.

For the case of causal sets which are known to come from a continuum sprinkling, if we assume the existence of  $n$ -pixies, finite causal sets that can force a given manifold dimensionality  $n$ , then, using lemma 8, we can show the following:

**Theorem 2.** If we randomly sprinkle points in any finite region of a Minkowski space of arbitrary dimension  $n$  with increasing density, we will certainly end up after a finite number of steps with a set of points whose causal relations define a causal set  $P$  with causal dimension  $n$ .

*Proof.* Suppose  $n$ -pixies have size  $|\mathcal{P}_n| = N$ . Then define  $N$  regions  $S_i$  of  $n$ -dimensional Minkowski space in the following way. Embed an  $n$ -pixie  $\mathcal{P}_n$  in this space and take a neighborhood of arbitrary (small enough) fixed volume  $V$  around each of its points  $p_i$ , such that, if we moved each  $p_i$  arbitrarily within its corresponding neighborhood, we would not affect the causal relations between them (this is always possible, if none of the  $p_i$ 's are null related—see the first footnote of §2.4). Once we have these regions  $S_i$ , discard the original  $\mathcal{P}_n$  and start sprinkling points at random in the space. After a certain finite number of steps, we will have at least one point in each  $S_i$  (more precisely, the probability that a given  $S_i$  will contain no points when the density is  $\rho$  is given by  $e^{-\rho V(S_i)}$ ). At

this stage, the sprinkled points define a causal set  $P$ . This causal set has causal dimension  $n$ . To see this, consider a sub-causal set  $P'$  obtained by choosing one point inside each  $S_i$  and discarding all other points.  $P'$  is then equivalent to  $\mathcal{P}_n$  by construction, thus its causal dimension is  $n$ . Then, by lemma 8, the causal dimension of  $P$  has to be at least  $n$  and, since it cannot be greater because it is embedded in  $n$ -dimensional Minkowski space, it is precisely  $n$ . •

We will assume that this theorem still holds in any curved space-time in which all length scales defined by the metric are bounded below by some constant, for the following reason. For the theorem to hold, we really need only sprinklings in a compact region of Minkowski space, since all we use in the argument is a region containing all the  $S_i$ 's, and such a region can certainly be taken to be compact. Therefore, if a curved space-time  $\mathcal{M}$  satisfies the assumption above, when the sprinkling density has become very large compared to the density defined by the length constant, the local regions of  $\mathcal{M}$  which are approximately flat contain many embedded points, and we are locally in a situation in which we can apply theorem 2. The only implicit assumption we are making about causal sets in this generalization of theorem 2 is that  $n$ -pixies are such that their size does not increase too rapidly with  $n$ —at most like  $2^n$ , a very generous bound, with their height being roughly constant—, so that we have a bound on the size, in natural units, of the compact region we need  $\mathcal{M}$  to be approximately flat in.

Our “main conjecture” was that any pair of faithful embeddings,  $f : P \rightarrow (\mathcal{M}, g_{ab})$  and  $f' : P \rightarrow (\mathcal{M}', g'_{ab})$ , are related by a  $P$ -preserving diffeomorphism  $\psi : \mathcal{M} \rightarrow \mathcal{M}'$  which is an approximate isometry of  $g_{ab}$  to  $g'_{ab}$ . The following argument is a start in proving such a statement.

To begin with, let  $f : P \rightarrow (\mathcal{M}, g_{ab})$  be a fixed faithful embedding of dimension  $n$ . From the point of view of  $\mathcal{M}$ ,  $f(P)$  is a sprinkling of points to which (the

generalized) theorem 2 applies. Therefore,  $P$  must contain "small"  $n$ -pixies as subsets. But the mapping  $f'$  is also a faithful embedding of  $P$ , hence the existence of  $n$ -pixies in  $P$  forces the dimension of  $M'$  to be at least  $n$ . Since the same could have been said interchanging the roles of  $M$  and  $M'$ , we have  $n' \geq n$  and  $n \geq n'$ , or  $n = n'$ .

Let us call "small" an Alexandrov set  $A(p, q)$  of  $P$  whose height is a few units. Then the corresponding Alexandrov neighborhoods  $A(x, y)$  in  $M$ , where  $x = f(p)$  and  $y = f(q)$ , and similarly in  $M'$ , will also have a small volume in this sense, as measured in the respective metrics; whence they must be approximately isometric to Alexandrov neighborhoods in  $n$ -dimensional Minkowski space, since otherwise the geometry in them would induce length scales incompatible with condition (3). Certainly  $P$  can be "covered" by such small Alexandrov sets. We thus have a covering of  $P$  by sets whose pairwise intersection properties are fixed by the causal relations (elements in the intersections are characterized by their causal relations with the maximal and minimal elements of the Alexandrov sets involved). This goes over, in each manifold, to a covering made of neighborhoods which are approximately isometric to the corresponding ones in the other manifold, and with fixed pairwise intersection properties. This makes it plausible that the two manifolds will be globally approximately isometric.

If our main conjecture were proven along the lines of the argument above, we would be only at the stage the continuum theory of causal spaces is: we would know that, if a faithful embedding exists, then it is (essentially) unique, but we still would not know how restrictive the assumption is for a causal sets; in fact we would not know how to characterize it from the causal set point of view. To understand such a characterization we would need a constructive proof of the uniqueness conjecture, which used, e.g., some of the simplicial complex

constructions mentioned earlier. Such a proof would, along the way, deal with the issue of when the simplicial complex actually is the triangulation of a manifold, and, perhaps, how many different causal sets give different triangulations of the same manifold.

### 3. DYNAMICS OF CAUSAL SETS

#### 3.1 Causal sets and sums over histories

We must now discuss how to handle dynamics with causal sets. As I remarked earlier, there are reasons to look for a fundamental structure of space-time which is truly unified and self-contained, with quantum structure, including dynamics, built into it, so that one will not have to apply some quantization procedure to a previously defined classical structure. This means more than just skipping the formulation of a classical version of dynamics when building up the theory, and going directly to quantum dynamics. However, for lack of a deeper understanding of the relationship between quantum theory and the various levels of description of space-time geometry, the latter is precisely what we shall do.

We will follow the sum over histories framework familiar from path integrals, and our main task will then be that of finding a reasonable expression for the causal set action (or amplitude), to be used in selecting whole classes of causal sets as contributing to making up a continuum, and to understand what consequences these collective effects might have on this continuum, possibly of a kind that would not follow from collective effects due to a superposition of continuum geometries. The approach can be thus criticized for putting quantum mechanics in—including the continuum of complex numbers with which we define the basic amplitudes—“by hand”, but our hope is to produce a reasonable theory, which can be argued to reproduce general relativity in the continuum approximation, and which might give us some insight into a deeper level, at which quantum



dynamics appears naturally.

The sum over histories quantization is a more natural procedure for handling causal sets than, say, canonical quantization, since the latter uses in a much more essential way structures induced on hypersurfaces, which are much less natural for causal sets than they are for general relativity. The reason for this is that a causal set does not induce anything natural on a maximal antichain—the analog, as we saw earlier, of a spacelike hypersurface—which can then be evolved; in fact, by their own nature, causal set elements do not persist in time and there is no set on which one knows how to define a configuration space (we will make, despite this, an attempt to formulate a notion of evolution for causal sets below).

Causal sets are not the only approach to quantum gravity in which defining configuration variables and momenta is difficult, since already in topology changing situations in the continuum we face similar difficulties. Besides, we may not wish anyway to formulate a fundamental theory of quantum gravity in terms of states and measurements on a surface, but give instead more importance to histories and more realistic measurements distributed in space-time: after all, using quantum fluctuation-type arguments, we should not expect to be able to say what a spacelike hypersurface is if we can't say what a space-time point is. Finally, it is difficult to define a quantum theory based on canonical quantization, which implements the full 4-dimensional diffeomorphism invariance, so important in classical general relativity. Similarly to what I said regarding the notion of continuum, we could take the point of view that it is our cultural prejudice, formalized starting with Newton, which makes us talk of the evolution of states defined at a certain time, and distinguish physics into initial conditions and evolution equations (the need to overcome this distinction has been stressed, e.g., by Wheeler and by Wigner).

Sums over histories, however, still use spacelike hypersurfaces; in this sense, we could compare the role of the path integral formulation in quantum dynamics to that of general relativity in the continuum question: just like general relativity did not have anything new to say directly about the use of the real numbers as the thing space-time is “made of”, but contained the seeds for a deeper understanding of the question, so path integrals still make use of spacelike hypersurfaces and configuration spaces to allow us to calculate quantities of physical interest, although the concept of “state of a system” is not so extensively used, and the “initial conditions” can be replaced by the more covariant notion of choice of a subset of the set of all possible histories of the system. In fact it seems natural to try to question the existence of various structures, like the apparatus of states and observables, in the conventional quantum theory, similarly to the questions we were asking two chapters ago about space-time (following many authors, among which Penrose [18]), and to speculate on the existence of a more fundamental quantum theory lacking this apparatus, making again the conjecture that this theory will naturally merge with a fundamental theory of space-time.

#### Path integrals without hypersurfaces?

I will mention here an interesting idea of Sorkin [38] which goes a long way towards freeing the path integral formulation of quantum theory from the use of spacelike hypersurfaces, starting with a restatement of one of the basic facts, how the usual path integral formulation expresses the quantum mechanical laws of probability, in a perhaps unconventional way, which is intended to be as general as possible, in terms of the kinds of questions one asks from a quantum theory. In the path integral formalism one typically partitions the set of all histories into subclasses  $E_1, E_2, \dots$ , chosen, e.g., by some conditions on the histories in some compact region  $S$  of space-time, and calculates relative probabilities associated

with different subclasses. To calculate these probabilities, we proceed as follows. Choose a spacelike hypersurface  $\Sigma$ , defined by the two conditions that “time”, defined in an appropriate way, has a constant value  $t_c$  on it, and such that  $S$  is to the past of it. Then, for a given subclass  $E$ ,

$$P(E; t_c) = \int_{q \in C} dq \left| \int_{E_{q, t_c}} [d\gamma] e^{iS[\gamma]} \right|^2, \quad (3.1.1)$$

where  $C$  is the configuration space on  $\Sigma$ , and the paths  $\gamma$  we integrate over—here not causal sets necessarily—for each  $q \in C$  are  $E_{q, t_c} := \{\gamma \mid \gamma(t_c) = q\} \cap E$ .

Thus, in this formulation, we don't need to ask questions that include fixing  $\gamma(t)$  for some  $t$ —although conventionally one fixes  $\gamma$  at a final time  $t$ , in which case one can identify  $t_c$  with  $t$  and avoid having to perform the integral over  $dq$  in (3.1.1)—, but nevertheless we are forced to define a spacelike hypersurface  $\Sigma$  and a configuration space on it. This  $\Sigma$  is just a mathematical tool and plays apparently no physical role; indeed we are free to choose any  $\Sigma$  subject to the two conditions stated above (3.1.1); unitarity of the evolution of the physical system is what guarantees the answer in (3.1.1) to be independent of  $t_c$ .

Following Schwinger, let us now expand the right hand side of (3.1.1):

$$\begin{aligned} P(E; t_c) &= \int_{q \in C} dq \left( \int_{E_{q, t_c}} [d\gamma] e^{iS[\gamma]} \right) \left( \int_{E_{q, t_c}} [d\gamma] e^{iS[\gamma]} \right)^* \\ &= \int_{q \in C} dq \left( \int_{E_{q, t_c}} [d\gamma] e^{iS[\gamma]} \int_{E_{q, t_c}} [d\gamma'] e^{-iS[\gamma']} \right). \end{aligned} \quad (3.1.2)$$

Suppose that  $S[\gamma]$  is such that, when the time-orientation of the path  $\gamma$  is reversed, it changes sign:  $S[\gamma^*] = -S[\gamma]$ . Then the second integral over  $E_{q, t_c}$  in (3.1.2) can be thought of as an integral over histories  $\gamma^*$  which, from  $q$ , go

backwards in time (and meet again the requirements  $E$  on the way back):

$$P(E; t_c) = \int_{q \in C} dq \int_{\tilde{E}_{q, t_c}} [d(\gamma \circ \gamma'^*)] e^{i(S[\gamma] + S[\gamma'^*])}, \quad (3.1.3)$$

where  $\tilde{E}_{q, t_c}$  is the set of composite paths  $\gamma \circ \gamma'^*$  such that both  $\gamma$  and  $\gamma'$  are in  $E_{q, t_c}$ . But, looking at this expression, it is natural to consider  $\gamma \circ \gamma'^*$  as a single round-trip path  $\tilde{\gamma}$  and, if  $S$  is additive, to write

$$\begin{aligned} P(E; t_c) &= \int_{q \in C} dq \int_{\tilde{E}_{q, t_c}} [d\tilde{\gamma}] e^{iS[\tilde{\gamma}]} \\ &= \int_{\tilde{E}_{t_c}} [d\tilde{\gamma}] e^{iS[\tilde{\gamma}]}, \end{aligned} \quad (3.1.4)$$

(notice that this is a real quantity, even though it looks complex), where  $\tilde{E}_{t_c} := \cup_{q \in C} \tilde{E}_{q, t_c}$ , the collection of all paths that go forward up to a time  $t_c$ , and then backward in time. Equation (3.1.4) mentions spacelike hypersurfaces only indirectly, in the  $t_c$  requirement, and it makes no explicit reference to a configuration space. We will call this the “closed time” formulation of the sum over histories.

Following the steps above for causal sets, we would thus have to come up with an action  $S[P]$  (or better an amplitude-with-measure, “[ $dP$ ]  $e^{iS[P]}$ ”), and a prescription for saying what  $t = t_c$  means in causal set terms; furthermore, either  $S[P]$  changes sign under time-reversal, or we need to define a configuration space  $C$  (with perhaps an associated boundary term in the action, expressed in terms of elements of  $C$ , which makes this second possibility even less attractive). But, as we will now see, causal sets offer the possibility of eliminating all trace of  $\Sigma$ , by a suitable generalization of what we mean by “time”.

## Time

A definition of time is essential for dynamics, as we have seen. It provides part of the conditions used to select the classes of paths which contribute to sums over histories and variations of a path, and its choice, although probably simpler than that of the amplitude, is also crucial in determining whether the dynamics satisfies the desired properties (see §3.2 below).

In a proposal by Sorkin [38], time is taken to be defined in terms of the volume of “pasts sets”, in a manifold or in a causal set. One of the simplest possibilities is to define  $t$  as the volume of the past of a single element  $p \in P$  (for a causal set), and the spacelike hypersurfaces used in equations (3.1.1) to (3.1.4) are those in which this  $t$  is constant. In the continuum implementation of this proposal, we obtain a “modified” path integral version of quantum gravity—in the standard one, no time restriction is placed on the paths. But results obtained in this version of quantum gravity, with a simplified, finite-dimensional configuration space, indicate that it may lead to non-unitary evolution. In particular, the classical evolution equations of the theory, obtained by extremizing the action with the constraint that all variations preserve the total time associated with the metric about which we perturb, are non-local in time. This was to be expected, because time itself is defined in a non-local way, such that the time increment between two nearby hypersurfaces  $\Sigma_t$  and  $\Sigma_{t+\Delta t}$  can be calculated only by looking at the whole past of the hypersurfaces, and not just at the region between them. In this case, a consequence of this fact is that a metric which is a solution of the classical equations for a time  $t$  cannot be obtained by extending a solution valid for a shorter time.

We have therefore considered a second possibility for  $t$ , the total past volume of a hypersurface, with the condition that all elements, or points, of it have the

same past volume. In this way, the classical evolution equations—and their corresponding Schrödinger-like, or Wheeler-De Witt-like equations—are more likely to be local in time.

The most interesting possibility, however, comes from relaxing one condition in our second definition of time: we can choose histories by requiring *only* that their total past volume have a fixed value, and call this the time. In connection with the closed time formulation of the sum over histories, this implies that  $\tilde{E}_t$ , in (3.1.4) is defined just as the set of paths with total volume  $2t_c$ , which meet both the conditions defining the subclass  $E$  and their time-reversal: no mention whatsoever is made of spacelike hypersurfaces (except for the fact that the paths have to “turn around” *somewhere*).

For any choice of time, when we perform the variations of the histories to find the classical field equations, we will impose that all variations satisfy the time constraint that the final hypersurface be given by  $t = \text{const} = t_c$ . This is equivalent to adding a Lagrange multiplier term to the action and varying freely the resulting modified action. Thus, if we choose as time the total past volume, without additional conditions, the modified action is  $S + \lambda(\int dV - t_c)$ , while, if time is defined by the past volume of individual points on  $\Sigma$ ,  $\lambda$  becomes a field over  $\Sigma$ .

### Evolution of a causal set?

One of the remarkable features of causal sets is their essentially covariant, space-time nature. As we discussed, there is no such thing as a “3+1 formulation” for causal sets. Nevertheless, in the continuum approximation to the theory we want to recover general relativity, which does admit a 3 + 1 formulation, where one considers a spatial metric on a hypersurface and its second fundamental form (or some other equivalent tensor field) as the dynamical fields, and solves

an initial-value problem to find the evolution of these fields once their values at some initial “time” are given (together with some prescription for what kind of foliation into constant time surfaces one wants the resulting space-time to come out with). We could thus wonder whether there is any notion of evolution that can be formulated already at the causal set level, and goes over to the standard one in the continuum approximation.

A positive answer would not just be potentially useful for calculations with causal sets, but it would also throw further light on the relationship between discrete and continuum descriptions of space-time. Another, perhaps more important in practice, reason why we would like to know how much can be said about  $3 + 1$  splittings of causal sets is that they might be necessary to describe causal set dynamics in the path integral formulation, as indicated above, unless its closed-time version can be used. Finally, some approaches to fundamental structures that people have thought about try to build in quantum dynamics by generating elements of structure in an evolution-type picture (like Finkelstein’s idea that I mentioned in §1.2, and other approaches that resemble cellular automata).

But evolution of a causal set does not have to be equated, as in the continuum, with “path in phase space”. We can consider the question we have at hand as similar to the one we discussed in the context of coarse-grainings: how can we encode, in one piece of a causal set, information that will enable us to recover the missing, time-evolved, piece? In some cases this question has an answer.

Suppose we have a causal set  $P$  and we interpret it as representing all the evolution of a larger causal set, up to a certain “time”, represented by the set of maximal elements of  $P$ . We can add one more layer to  $P$  by giving an arbitrary collection  $\{R_i\}$  of antichains of  $P$ , and interpreting each  $R_i$  as one element of the

new layer, identified by the elements of  $P$  which are linked to it and to its past:  $P' = P \cup \{R_i\}$ , and two elements of  $P$  are related in  $P'$  iff they were related in  $P$ , while  $p \prec_{P'} R_i$  iff  $p \in R_i$ , and all  $R_i$ 's are unrelated to each other. Then we can add a second layer by assigning a collection of subsets  $\{R'_i\}$  of  $P'$  to define  $P'' = P' \cup \{R'_i\}$ , and so on.

Of course, having added  $k$  layers to a causal set does not help to know what the  $(k+1)$ -th layer will be, but this is just like in a manifold, where differential geometry by itself gives us a formalism for evolving data on a hypersurface, without telling us what the actual evolution will be: the prescription for evolving comes from a space+time splitting of Einstein's equation. A question to worry about is instead whether this evolution for causal sets will be local. As we have seen, and figure 2.5.1 illustrates, elements in the top layer of a causal set are often linked to elements many layers below. Whether this will be a problem or not can be learned only after more detailed studies are done of the link patterns in large causal sets, faithfully embedded in a manifold (only for those do we really need a local evolution).



### 3.2 The amplitude: general considerations

Now that we have discussed the general framework we want to use for causal set dynamics, including the choice of time, we should look at the other ingredient which goes into the sums over histories: the basic amplitude, a functional  $A$  which assigns to each causal set  $P$  a complex number  $A[P]$  (neglecting for now the possibility that we might have to define a configuration space as well). The combined choice of time and amplitude is really the crucial question in our approach to causal set dynamics. However, unlike the case in which one is constructing, e.g., some field theory for a new kind of field, we cannot be directly guided, in this choice, by similarity with other theories, but we have to “start from scratch”, and use only indirectly what we have learned from studying other physical theories. It will be useful then to remind ourselves of the conditions we would like our dynamics to satisfy, even before we try to find the “correct” one.

I should emphasize, furthermore, that we cannot use analogies with lattice gauge theories, in which the basic variables are continuous-valued fields defined at discrete points (or links between them), nor with Regge calculus, the discrete formulation of general relativity, in which the basic variables are again real-valued variables attached to links between points. Only after coarse-graining and embedding of the causal set in a manifold can we start thinking of it as a lattice, but the action must be defined for an abstract causal set.

The biggest “responsibility” as regards properties of the dynamics is usually given to the amplitude, but it is clear that different choices of time can make dynamics satisfy or not the basic conditions I will state below. For simplicity of discussion, however, I will talk mainly in terms of the amplitude, or the action  $S[P]$  such that  $A[P] = e^{iS[P]}$ .<sup>†</sup> After pointing out that those conditions go a long

---

<sup>†</sup> Thus we would tend to say that the amplitude must be a unimodular complex number. We

way towards ensuring that in the continuum approximation we recover general relativity, with a possible cosmological constant, I will sketch an argument to show that the cosmological constant is likely to vanish, while in the next section we will see a few general forms for  $S$ , and one possible explicit expression for it, some of whose properties will be studied later on.

### First conditions

The general criteria for choosing one amplitude over another come from requirements imposed by the continuum approximation. The properties we want dynamics to satisfy, even before looking at the specific theory it reduces to in the continuum approximation, are:

- (1) **existence of a classical limit:** one should be able to formulate a variational principle for the theory, asserting that the histories (in our case, causal sets) which contribute most to the amplitude for a given "process"  $E$  are a set of histories "close to each other", which all have approximately the same amplitude (to "first order in the variation" the amplitude is the same); the classical limit will be this whole set of histories (there won't be a unique classical one, since there won't be a "field equation" to select it for us);
- (2) **existence of a continuum approximation:** the histories which are "close to each other" according to (1) should have similar geometric properties (insofar as these can be defined, using, e.g., simplicial complexes), and admit approximations by space-time manifolds which are also close to each other: the latter will define *the* classical solution in the continuum view;
- (3) **(approximate) locality, or additivity of the action (or multiplicativity of the amplitude):** if the causal set  $P$  is divided into two separate regions, i.e.,  $P = P_1 \cup P_2$  (with  $P$  connected, but  $P_1 \cap P_2 = \emptyset$ , as point

---

will assume this, although it is difficult to see how, in the continuum limit, the amplitude associated with each geometry, resulting from the superposition of many causal sets, will also turn out to be of unit magnitude.

sets), then  $S[P] \approx S[P_1] + S[P_2]$ , i.e.,  $A[P] \approx A[P_1] \cdot A[P_2]$ , up to "boundary terms";

- (4) **independence of the probabilities of  $t_c$** : the values of expressions like (3.1.4) should not depend on the value of the time parameter, as an internal consistency requirement on the theory. In usual continuum theories, this follows from unitarity, but here we will consider it as a more fundamental condition. Although certainly related to independence of  $t_c$ , unitarity will be viewed more as a part of "phenomenology".

The first two conditions usually are not even mentioned, since the second is satisfied from the start, and the first is also automatically satisfied when one writes the action for a continuum-valued theory as a functional of differentiable tensor fields defined over the space-time region in consideration. But they are essential here: existence of the classical limit, which is not guaranteed for all expressions, as we will see, becomes a necessary condition for being able to write down the continuum action as a functional of differentiable tensor fields. Locality is not satisfied by all continuum theories, but it is simple to ensure, since it is equivalent to the action being the integral of a locally defined quantity. Independence of  $t_c$  and unitarity is the first condition which is not trivially satisfied even by continuum theories.

Finally, in view of the "closed time" approach to sums over histories described in the previous section, I will add another condition to our list of desiderata:

- (5) **change of sign under time-reversal**: if we replace a causal set  $P$  by its dual  $P^*$ ,  $S[P^*] = -S[P]$ , or  $A[P^*] = A[P]^*$ .

## Relationship with general relativity

Let us now assume that conditions (1) and (2) are satisfied. Then we might ask ourselves which expression for  $S[P]$  is actually the “correct” one. It can be shown by a simple argument that any expression which satisfies the first three conditions above will give, in the continuum approximation, general relativity with a cosmological constant, up to small corrections. This argument does not rely on any specific form for  $S[P]$ , nor in fact on very specific features of causal set theory: it applies, *mutatis mutandis*, to many theories, for which general relativity represents a low energy approximation to some more complete dynamics (like induced gravity, higher derivative gravity, or string theory).

Let us thus consider any assignment of an action  $S[P]$  (or an amplitude) to every causal set  $P$ , which admits a continuum classical limit and is additive (multiplicative). In that case, the action must take, in the continuum approximation, the form of an integral over  $M$  of a locally defined scalar (density). This scalar can be constructed from any tensor fields available in the manifold, but since the quantum “sum over histories” washes out the anisotropy associated with any single embedded  $P$ , the resulting effective action density can in fact only depend on the metric  $g_{ab}$  itself (in other language, the result must be “generally covariant”),<sup>†</sup> and it can therefore be expanded as

$$L_{eff} = L_0 + L_2 + L_4 + \dots, \quad (3.2.1)$$

where  $L_0$  is just a constant (the effective cosmological constant),  $L_2$  is a multiple of the scalar curvature,  $L_4$  is the sum of curvature-squared terms, etc. But since

---

<sup>†</sup> Strictly speaking, we should take into account the other fields on  $M$  defined by the causal set coarse-grainings as well. Thus, in our argument we are in effect “integrating out” these fields, and using the amplitudes to ask questions only about the metric.

the action itself must have the dimensions of  $\hbar$ , the coefficient occurring in the term  $L_{2k}$  would be expected to be of order  $\hbar \ell_P^{2k-n}$  in ordinary units, and for a continuum of  $n$  space-time dimensions. Then, neglecting the relative minuscule terms  $L_4 + L_6 + \dots$  leaves us with  $L_0 + L_2$ , or precisely the usual general relativity lagrangian 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, but has to come from a more detailed analysis that uses explicitly the form of  $S[P]$  and of  $t_c$ .

### The cosmological constant problem

So far, I have tried to stress the connections between causal set dynamics and aspects of continuum theories, in particular general relativity. I would now like to point out a special feature arising in causal set dynamics, which does not have any direct correspondence in the continuum approximation, but may have important consequences for continuum dynamics.

As we just saw, we can expect the action, in the continuum approximation, neglecting higher powers of the curvature, to be of the standard form

$$S[g_{ab}] = (2\kappa)^{-1} \int R d^n V - \Lambda \int d^n V, \quad (3.2.2)$$

where  $\kappa$  can be used to define the gravitational constant (e.g., in 4 dimensions we set  $2\kappa = 16\pi G$ ), and  $\Lambda$  is the cosmological constant. Not only the continuum action, but also what to do with it, e.g., how to determine “classical paths”, should be derived from the fundamental causal set dynamics. In the latter, we perform “small variations” in the causal set and check how much  $S[P]$  changes. Defining “small variations” of  $P$  is hard, and to do it properly we would need some notion

of distance or topology on the set of all causal sets, or at least on those which admit some continuum approximation. This, in fact, is the causal set version of the approximate isometry question we were dealing with in §2.3; the two things should be related by the fact that when two causal sets are “nearly isomorphic”, in their continuum approximations the metrics will be “nearly isometric”, and similarly for the other continuum fields defined along the way.

For the sake of the following argument, let us neglect all fields other than  $g_{ab}$ , induced by  $P$  in the continuum (e.g., assume that they vanish because  $P$  was faithfully embeddable from the beginning). Then, among the possible small changes of  $P$ , there will be some which only induce a small change in  $g_{ab}$ , and some which are not representable as changes in the continuum geometry, because, say, they are fluctuations in dimension or topology. For  $g_{ab}$  to be a classical solution, the expression (3.2.2) (with the addition of the Lagrange multiplier term) should be invariant under all these changes. Suppose now that one of the “small changes” we can make on  $P$  is, e.g., to transform it into  $P \times P_4^\diamond$ , where  $P_4^\diamond$  is the diamond-shaped 4-element causal set ( $P_{(2)}^{DM}$ ). This will have the effect of multiplying the total volume by 4—it is not *really* a small change—and, if we think of  $P_4^\diamond$  as a little circle, the manifold corresponding to the new causal set will be metrically  $\mathcal{M} \times S^1$  (in this qualitative argument, we will also neglect the fact that such a small  $S^1$  would violate our condition (3) on faithful embeddings), so  $S - \lambda V$  will just scale with the radius of the circle. Thus, if its variation is to vanish,  $S - \lambda V$  must vanish itself. On the other hand, we know from the variation of the metric  $g_{ab}$  that the Einstein equation with effective cosmological constant  $\Lambda + \lambda$ ,  $R_{ab} - (\Lambda + \lambda)g_{ab} = 0$ , must be satisfied. These two conditions together imply that both  $R$  and the effective cosmological constant  $\Lambda + \lambda$  must vanish separately.

### 3.3 The amplitude: possible expressions

One tempting way to go about constructing an action would be to find a relationship involving the continuum action density, i.e., the scalar curvature  $R$ , and the quantities defined for a causal set, i.e., volume and causal structure, and take this to be a defining relationship, in some sense, for the discrete case. Such a relationship was already found in (2.7.1). After deriving this equality, Myrheim [16] also noted that, not only does Einstein's equation for  $g_{\mu\nu}$  guarantee that the  $l^2$  terms in (2.9.1) vanish, but the converse is also true: if for  $x_1$  and  $x_2$  along *any* timelike geodesic through  $x_0$  we have

$$V = \frac{\pi l^4}{24} [1 + O(l^3)], \quad (3.3.1)$$

then Einstein's equation is satisfied at  $x_0$ . But, while (3.3.1) could be taken as a basis for some hypothetical "classical" theory of causal sets which are known to be embeddable in some fixed dimensionality, it is clearly not what we are looking for: it relates volumes and lengths—which can be expressed, although not too simply, in terms of causal sets—not to the scalar curvature, but to some combination of the components of the Ricci tensor; and, most importantly, it is based on manifolds, and it is far from clear how and with what justification we could extend it to general causal sets.

Instead, we want to define an action in terms of counting substructures in a causal set. Several such substructures have been identified in chapter 2, and different kinds of actions can be formed with them.

First of all, let us see why an apparently obvious choice doesn't work. If we define  $S[P]$  to be the total number of links in  $P$  (up to a factor 2, the number of "nearest neighbors" of each point, summed over all points), it would be an

integer, and an integer action—or any action which is an integer multiple of some constant—cannot satisfy condition (1) (unless this constant is an extremely small number, but then it would be hard to justify why we effectively introduce such a small parameter in the theory). Furthermore, it would be badly non-local, since, as we know, most of the links of a given element are to elements which are very far from it. As pointed out by 't Hooft [27], at first sight it is difficult to see how the necessary (approximate) locality of the theory can be obtained, from some natural definition of causal set dynamics.

### Multiloop action

We define *multiloop* in a causal set  $P$  as any set of loops in  $P$ , no two of which have any element in common. Then, if  $M$  is the number of distinct multiloops in  $P$ , the *multiloop action* of  $P$  is

$$S_m[P] := \ln M. \quad (3.3.2)$$

This action satisfies (1), at least to the extent that for sufficiently large  $P$ ,  $S_m$  can be made to vary by an arbitrarily small amount, but it certainly does not satisfy condition (5), since  $S_m[P^*] = S_m[P]$ . Its intuitive motivation is the fact that loops capture curvature, and we use multiloops instead of single loops because the latter don't increase fast enough in number to yield an additive action. Let us see how additivity works out with  $S_m$ . Consider again  $P = P_1 \cup P_2$ . Then

$$M[P] = M[P_1] \cdot M[P_2] + \{\text{multiloops with one or more loops shared by } P_1 \text{ and } P_2\}. \quad (3.3.3)$$

But most of the contribution to  $M[P]$  will come from multiloops made of small loops—there are many more of these, since they have to be non-intersecting—so



the second term in (3.3.3) will effectively be of the same magnitude as the first term, with an actual value depending on the causal structure “near the boundary between  $P_1$  and  $P_2$ ”. Thus, upon taking the logarithm,

$$\begin{aligned} S_m[P] &= \ln M[P] \approx \ln(2 M[P_1] M[P_2]) \\ &\approx \ln M[P_1] + \ln M[P_2] = S_m[P_1] + S_m[P_2], \end{aligned} \quad (3.3.4)$$

for large causal sets. This argument does not take into account possible surface terms in the definition (3.3.2) of  $S_m[P]$ , which, from our experience with continuum theories, could well be present. What form such terms would have for causal sets is not obvious. If they are present, they do not have to go over to the surface terms of the continuum approximation: the latter can be a combination of causal set volume and surface terms.

Our argument for the locality of  $S_m[P]$  is a handwaving one, and it would be nice to see an analogy with a different physical theory, equally in danger of being non-local a priori, where there is however a better argument to show its locality. This analogy is provided by the 2-dimensional Ising model.

The Ising model is a square lattice, to each site  $i$  of which we associate a spin  $s_i = \pm 1$ , and with a total energy  $H = -\sum_{\langle ij \rangle} J_{ij} s_i s_j$  (summed over all pairs of sites), where we assume that  $J_{ij} = J$  for  $i$  and  $j$  nearest neighbors, and 0 otherwise (and we have not included an external magnetic field). It can be shown (see Landau and Lifshitz [78]) that, for a lattice of  $N$  points, the partition function  $Z = \sum_{(s)} e^{-H(s)/kT}$  for a temperature  $T$  ( $k$  is the Boltzmann constant, and  $(s)$  denotes a whole configuration of spins) can be written as

$$Z = (1 - x^2)^{-N} 2^N \sum_r x^r g_r, \quad (3.3.5)$$

where  $x := \tanh(J/kT)$  and  $g_r$  is the number of multiloops with a total of  $r$  links

(where two loops may overlap at isolated points). This is similar to the total number of multiloops, although one cannot actually take the limit  $x \rightarrow 1$ , and the multiloops are not exactly like the ones we use in (3.3.2), because loops can intersect and their shape is not restricted by the condition that they be formed by a pair of causal paths. On the other hand, it is known (see Amit [73]) that, for temperatures near the instability value  $kT_0 = 2\gamma J$ , where  $\gamma$  is the number of nearest neighbors per point (4 in our case), the Ising model is equivalent to a local field theory: that of a scalar field with mass proportional to  $(T - T_0)/T_0$  (in fact, even the Ising model “action” in itself is a local expression in terms of the spins on the lattice).

The multiloop action is the only concrete proposal for a causal set action we have. There are however other classes of possible actions.

### Sums over Alexandrov sets

One way of ensuring that the action is local could be to build it, in a less “global” fashion than the above  $S_m$ , from contributions of Alexandrov sets in  $P$ —our usual tools for defining local but invariant concepts in lorentzian settings. Thus, we could define

$$S_a[P] := \sum_{p < q} f[A(p, q)], \quad (3.3.6)$$

where  $f$  is any (non-integer) function defined on the Alexandrov sets of  $P$ . In particular,  $f$  could have small values for large Alexandrov sets, and it can be arranged to satisfy (5).

Functions on Alexandrov sets are the elements of the so-called *incidence algebra* of the causal set, well-known in combinatorics (see, e.g., Stanley [65]). This set of functions inherits a linear space structure from the field in which the functions are valued ( $\mathbb{R}$  in this case), and has a multiplication defined by the

convolution

$$(fg)[A(p, q)] = \sum_{r \in A[p, q]} f[A(p, r)] g[A(r, q)]. \quad (3.3.7)$$

We will see some examples of elements of the incidence algebra in the next section.

### “Pixie” actions

Our last kind of possible action is perhaps the most promising. Let me illustrate it with an example. We have seen in §2.4 that there are two 2-pixies, duals (time-reversals) of each other, and three 3-pixies, of which one self-dual and two duals of each other, and that in some sense this trend has to continue in higher dimensions. Of the pixies that are not self-dual, indicate one kind by  $P^\wedge$  and the other by  $P^\vee$ . Then one definition of the action could be

$$S_P[P] := \ln(N_{P^\wedge}/N_{P^\vee}). \quad (3.3.8)$$

We don't know how this action grows with the total size of  $P$  (it might be necessary, e.g., to use “multipixies” instead of single pixies), but it satisfies condition (1) in the same sense as  $S_m$  does, and more importantly it satisfies (5). This proposal has some other nice features: it uses just pixies, and somehow the way these are put together to form the whole causal set, which, in the same intuitive sense as  $S_m$  captured curvature, captures the idea that small, locally flat regions—the definition of pixie came from causal sets embedded in Minkowski space—are put together to form a curved space; besides, the fact that pixies may be of small height, maybe just two- or three-layer causal sets, makes us think that an action defined in terms of pixies has more chances of making sense of a layer-by-layer evolution for causal sets, of the type outlined in §3.1.

As it stands, (3.3.8) is probably not the correct definition, but its properties indicate that we should look for actions based on small subsets of  $P$ , possibly

pixies. One way to get some insight into how pixies are distributed and related to each other in a large causal set is to generate by computer a large sprinkling of points in some background, e.g., Minkowski space, and look for them.

### 3.4 Computer calculations

Causal sets are in a way the ideal subject for computer calculations, being not only based on discrete sets of elements, but having as fundamental variables quantities which are themselves discrete, in fact even boolean: a 'yes' or 'no' for the existence of a causal relation between any two elements. Thus, it might seem that calculating values for the action of a causal set, e.g., with a computer, has the double advantage of involving no approximation inherent in the introduction of a lattice, and no roundoff error due to the precision of the machine. The observation is correct, but as regards physical applications it neglects one important factor: the size of the causal set. Furthermore, in many problems we do need continuous variables, like when we evaluate the effective curvature or some other kinematical quantity in a region of a causal set, or the action.

Nevertheless, let us outline some calculations which can in principle be done, once we choose some  $S[P]$ . These will be of two kinds: the first one, more "technical" is how to evaluate, say, the number of paths or loops between two elements, the number of multiloops in a causal set, or how to recognize pixies; the second deals with applications in which we assume we know how to calculate the action for a causal set.

#### Counting things in a causal set

The first question to solve when doing computer calculations with causal sets is how to store the causal sets. Suppose we have a *labelled* causal set  $P = \{p_i\}$  (the labelling is necessary here, and I will assume it is consistent with the partial order in the sense that  $p_i < p_j$  (but  $p_i \neq p_j$ )  $\Rightarrow i < j$ ). The most convenient way to store  $P$  for calculations (when storage size becomes a main concern, one could opt for some other approach) is that of an *incidence matrix* (or *zeta function*),

the element of the incidence algebra of  $P$  defined by

$$\zeta_{i,j} = \begin{cases} 1 & \text{if } p_i \prec p_j \\ 0 & \text{otherwise,} \end{cases} \quad (3.4.1)$$

an upper triangular, invertible matrix of 0's and 1's, which has, as it is easy to see, the following properties (see Stanley [65]):

- (1)  $(\zeta^2)_{ij}$  is the number of elements in  $A(p_i, p_j)$ ;
- (2)  $(\zeta^k)_{ij}$  is the number of multichains of length  $k$  between  $p_i$  and  $p_j$ , where a *multichain* is a chain with some elements possibly repeated;
- (3)  $(\zeta - 1)_{ij} = 1$  if  $p_i \prec p_j$  but  $p_i \neq p_j$ , and 0 otherwise;
- (4)  $(\zeta - 1)_{ij}^k$  is the number of chains of length  $k$  from  $p_i$  to  $p_j$ ;
- (5)  $(2 - \zeta)_{ij}^{-1}$  is the total number of chains from  $p_i$  to  $p_j$ .

Related to the incidence matrix is what we call "link matrix" (or *eta function*)  $\eta_{i,j}$ , whose definition is the same as that of  $\zeta_{i,j}$ , but with  $p_i \prec p_j$  replaced by  $p_i \prec p_j$ . The link matrix has similar properties to those of the incidence matrix, including:

- (1)  $(\eta^k)_{ij}$  is the number  $(N_k)_{ij}$  of  $k$ -link paths from  $p_i$  to  $p_j$ ;
- (2)  $(1 - \eta)_{ij}^{-1}$  is the total number  $(N_p)_{ij}$  of paths from  $p_i$  to  $p_j$ .

The matrices  $\zeta_{ij}$  and  $\eta_{ij}$  obviously contain the same amount of information, and there is a simple way to go from one to the other: e.g., if  $p_i \prec p_j$ , the Alexandrov set  $A(p_i, p_j)$  contains exactly two elements, so deleting from  $\zeta_{ij}$  all diagonal entries, and those whose corresponding entries of  $(\zeta^2)_{ij}$  are greater than 2, gives  $\eta_{ij}$ .

Some quantities associated with a given causal set are thus easy to calculate. For example, to calculate the number of paths between  $p_i$  and  $p_j$ , enter the causal set as a link matrix (obtained by reading them off some sprinkling of points in a background, assigning some desired relations, or generating re-

lations at random—in the last two cases, we must make sure that transitivity and non-circularity are satisfied), and find  $(N_p)_{ij} = (1 - \eta)_{ij}^{-1}$  either as  $\sum_{k=1}^{k_{\max}} (N_k)_{ij} = \sum_{k=1}^{k_{\max}} (\eta^k)_{ij}$ , where  $k_{\max}$  is the largest  $k$  for which  $\eta^k \neq 0$  (the height of  $A(p_i, p_j)$ ), or by directly inverting  $(1 - \eta)$  using some inversion routine. By the way, this method could possibly be used, with many sprinklings of points in  $n$ -dimensional Minkowski space, or by theoretical statistical arguments, to calculate the expected number of paths between two points, which we would like to know in connection with the definition of fractal dimension.

Suppose now that we want to use these matrices to calculate a less immediate quantity, the number of loops between two elements in  $P$ . This can be obtained as the sum over all  $k$  and  $l$  of the number of non-self-intersecting  $(k, l)$ -loops between  $p_i$  and  $p_j$ ,  $(N'_{k,l})_{ij}$ . Since the loops are non-intersecting,  $N'_{k,l}$  is not given by  $N_k N_l$  (or  $\frac{1}{2} N_k (N_k - 1)$  for  $k = l$ ), as in (2.5.17). (Actually, the number of  $(2, 2)$ - and  $(2, 3)$ -loops is correctly given by these formulae.) In terms of paths, the problem arises because some paths from  $p_i$  to  $p_j$  start off as distinct, but they merge before reaching  $p_j$ , and they should thus not be counted as distinct paths for loop calculation purposes, and/or they start off from  $p_i$  as one path and then branch off into several paths. Two or more paths merge at an element  $p_m \in A(p_i, p_j)$ , iff  $p_m$  can be reached in more than one way from  $p_i$ . Therefore, in principle, one should be able to solve the problem by changing to 1 the  $(i, m)$  entry in the appropriate matrix for the interval  $A(p_i, p_m)$  (which will be greater than 1). We have not yet found a way to translate this idea into a prescription that will correctly take care of all the possibilities.

Calculation of the number of multiloops in a causal set  $P$  is also made difficult by the requirement that the loops do not intersect each other. One way to simplify the problem is to impose the stronger condition that the Alexandrov sets defined

by the endpoints of the single loops be disjoint. This reduces the calculation to the sum over all maximal disjoint sets of Alexandrov sets in  $P$  of the product of the number of loops between the endpoints of each Alexandrov set.

### Monte Carlo calculations

Amplitudes for processes or expectation values of observables are defined in the path integral approach by summing over the contribution of all allowed paths. But in some situations one can get by with a sampling of the set of paths by a limited number of them. Using the Monte Carlo idea, one can randomly generate paths  $\gamma$  in the whole allowed set, calculate the quantity of interest (namely 1 if we are looking for an amplitude, or the value of the observable for the path otherwise), and weigh the result using the action, namely multiplying it by  $e^{-S_E[\gamma]}$ , where  $S_E[\gamma]$  is the euclidean action of  $\gamma$ .

In causal set theory, there are some conceptual difficulties with this procedure. First, the expectation values I mentioned are only expectation values in a mathematical sense: it is not clear if and how we can define physical expectation values of observables, meaning results we would get by averaging the outcome of repeated experiments. Second, euclideanizing the action, which is necessary in the Monte Carlo method to obtain thermalization, cannot have, in causal set theory, the meaning of performing an analytic continuation in imaginary time, to get a euclidean signature for the metric, since in a positive-definite metric there would be no causal order left, no structure to use in writing down a "euclideanized" action. We could try writing down, for  $S_E[P]$ , the same action as  $S[P]$ , and change however the  $i$  in front of it to  $-1$ . The mathematical meaning of this would be, if anything, that of an analytic continuation in imaginary coupling constant, but this does not tell us how to interpret physically the results. It is worth investigating whether the closed time version of the sum over histories



can give real values for these "expectation values" even in the lorentzian metric case.

If one wants nevertheless to do calculations with causal sets using the above method, the main difficulties in setting up a program to do so are the choice of a random sample of histories, and the calculation of the action. Once these are solved, one can calculate, e.g., the "expected dimension of space-time" as

$$\sum_P n_{eff}[P] e^{-S[P]}. \dagger$$

---

† This suggestion is due to N. Cabibbo.

### 3.5 Conclusions

What I have described in these two chapters is the proposal of a framework for a small scale structure of space-time; I have given the first results obtained in this program, and indicated the directions in which work should be done to establish the usefulness of the approach. Obviously, this is still a theory in the making, and most of the work still remains to be done, but a reasonably detailed qualitative picture has emerged, that I will now summarize.

We assume that space-time is, at a fundamental level, a causal set: a locally finite set of elements endowed with a partial order relation, that we interpret as going over to the causal structure of space-time at macroscopic scales, when we associate a manifold to the causal set. For suitably large causal sets, I have discussed how it may be possible to find such a manifold, which approximates well the causal set structure. In this manifold, either the original causal set or, more likely, some coarse-graining of it, appears faithfully embedded, with a density which depends on the amount of coarse-graining we had to perform, but which should in any case be not too different from Planck density. We conclude then that a causal set which reproduces the features of our universe must have a number of elements equal to the volume of the universe in Planck units, i.e.,  $\approx (10^{60})^4 = 10^{240}$ . The causal relation in this set is dynamically determined; dynamics however picks out not a single causal structure, but rather a whole class of them, which contribute a roughly equal amplitude in a sum-over-histories formulation of quantum mechanics. These different causal sets are also close to each other in the sense that the manifolds associated with them coincide at large scales, thus defining an effective classical space-time manifold, as our observations require for the visible universe (excluding however the very early universe and inside black holes). Since the action of a causal set goes over

approximately to the Einstein-Hilbert action, this effective space-time is expected to satisfy the Einstein equation, albeit a priori with the possible addition of a cosmological constant. We have seen furthermore that it is plausible that this effective cosmological constant will actually vanish, since the fact that space-time is a causal set forces us to impose a larger number of conditions for the classical limit than the equivalent of the field equations in the continuum case.

What does such a picture say about the issues that motivated us to look for a quantum theory of gravity in the first place, and what consequences does it have? As regards the gravitational aspects of the continuum approximation, one immediate consequence of this program is that all space-times “produced” by the theory are free, by definition, of closed timelike curves, and they are time-orientable. Also by definition, causal sets cannot run into singularities, and continuum space-times which develop singularities will correspond to causal sets with well-defined “evolution” everywhere. Near the singular region, these causal sets will admit faithful embeddings into manifolds that will look different from the classical solutions, or more likely will not admit any faithful embedding at all: this is where causal set dynamics will come into its own.<sup>†</sup> At the same time, causal sets have the necessary kinematic flexibility to represent topology changing processes, and, in general, a fluctuating geometry, with different topological properties at different scales.

A shift in scale in causal set theory is performed by a coarse-graining, and the change in topological properties it induces is accompanied by the emergence

---

<sup>†</sup> Discreteness is the main feature in causal set theory which makes it free of singularities. However, our formulation of dynamics, including the choice of time, reflected in the form of the Lagrange multiplier term added to the action when varying it, also represents a departure from the conventional formalism. In principle our time constraint can lead to quantum evolutions which even in the continuum case keep the wave functions for the geometry away from singular configurations. It is with this motivation that the continuum studies of quantum gravity with a time constraint, as I was mentioning in §3.1, have been undertaken.

of new structure, in the form of fields. The coupling of these fields to the metric at the continuum level will be determined by the form the fundamental causal set action takes when expressed in terms of them, and this coupling will be an important prediction of the theory, which can be compared with phenomenology. Just like for geometry, all calculations of amplitudes involving causal sets by their very nature will yield finite results (we restrict ourselves to finite causal sets), and we cannot encounter the ultraviolet divergences of quantum field theory. We can say that causal sets provide a physical basis for a momentum cutoff, although an understanding of the renormalization group properties of the various physical fields requires a more detailed analysis of the behavior of the action under coarse-graining.

Two remarks seem appropriate about the relationship of causal sets to continuum fields. First, I should point out that the prescription I outlined for recovering these fields may need to be generalized or complemented by other ideas, not based just on coarse-grainings. An example of what I mean is provided by spinor fields. These are in general hard to get just from geometry, but there are proposals for getting spin- $\frac{1}{2}$  from pure gravity on a manifold (Friedman and Sorkin [32]), and it might be that in causal set theory spinor fields will arise at a higher level than the manifold, and will encode, macroscopically, non-trivial topology that can be described in continuum terms at smaller scales. Second, a precise identification of coarse-graining fields with continuum fields will clarify the question of how to interpret physically the duality operation we defined on causal sets. Although at first it might seem that duality is most naturally associated with time-reversal  $T$ ,  $CPT$  could be a better guess, for it appears to be a more fundamental symmetry in continuum field theory, and, if, as we have been doing, we interpret the  $\star$  operation involved in the closed time approach to sums

over histories (see (3.1.3)) as duality for causal sets, then duality always leaves the basic amplitude in (3.1.4) invariant.

A success in the attempt to derive all observed fields from properties of causal sets, in the manner described or in any other way, would be a much more conceptually satisfactory reduction of geometry and particle phenomenology to causality than the one reached in the continuum theory, in addition to the practical gain represented by the natural notion of volume we have available, as already explained. Even with the results of §2.2 generalized to Kaluza-Klein space-times, and granting that one could recover from these the correct 4-dimensional phenomenology, the continuum theory does not provide a dynamical framework for causal spaces, or a constructive procedure for obtaining the manifold and metric.

#### Modifications or extensions

If the procedure I outlined for obtaining continuum fields other than the metric does not yield the correct phenomenology, one possible cure is to introduce fundamental “matter” fields at the fundamental level, together with the causal sets. This would be analogous to what is commonly done in Kaluza-Klein theory, where one partially gives up the original idea of recovering the observed phenomenology from a purely gravitational theory in higher dimensions and adds extra fields on the Kaluza-Klein manifold  $\mathcal{K}$ . Another possible modification of the theory, discussed in response to questions we have been asked several times, is to identify in causal sets structures which can be interpreted as string world-sheets, or to map the causal set elements to strings rather than to points of a manifold. The first suggestion is not easy to implement because the causal set structure does not single out natural 2-dimensional “timelike” subsets; furthermore, it would not really give us anything new, because the action would still be written ultimately in causal set terms, although possibly in string wording, and

most likely the continuum reconstruction theorems would still aim at identifying causal set elements with points, and discrete strings with continuum strings as a consequence. Regarding the second suggestion, see the remarks at the beginning of §2.3.

For the time being, more suggestive possibilities seem to lie in quite different directions. One of these is given by the so-called Bergmann manifolds, a generalization of lorentzian manifolds to (certain) higher dimensions. A *Bergmann manifold* (see Finkelstein [75]) is a differentiable manifold  $\mathcal{B}_N$  of dimension  $n = N^2$ , with a complex vector space  $S(x)$  at each  $x \in \mathcal{B}_N$ , possessing a preferred skew-symmetric tensor with  $N$  indices,  $\epsilon_{AB\dots C}(x)$ , and a linear isomorphism  $\sigma_a^{AB'}(x)$  from hermitian forms in  $S(x)$  to cotangent vectors at  $x$ , called “spin vector” (or soldering form). Using this structure, we can define a “metric”

$$g_{ab\dots c} := \epsilon_{AB\dots C} \epsilon_{A'B'\dots C'} \sigma_a^{AA'} \sigma_b^{BB'} \dots \sigma_c^{CC'}, \quad (3.5.1)$$

which has  $N$  indices, and therefore makes  $\mathcal{B}_N$  into a Finsler space, i.e., the line element has the expression

$$ds^N = g_{\alpha\beta\dots\gamma} dx^\alpha dx^\beta \dots dx^\gamma. \quad (3.5.2)$$

For  $N = 2$ , a Bergmann manifold  $\mathcal{B}_2$  is the same as a 4-dimensional lorentzian manifold; but even for higher values of  $N$ , these manifolds have light cones that allow causal sets to be realized by suitable embeddings of their elements as points.

A true generalization of the theory would consist in allowing the relation defining the basic structure not to be antisymmetric. A set  $P$  with a reflexive and transitive relation  $<$  is called a *preposet* (or *quasi-ordered set*) (see Stanley [65]); the absence of the antisymmetry condition allows the existence of “closed paths”. Each preposet defines a unique poset by dividing by the equivalence

relation  $p \sim q$  iff  $p \prec q$  and  $q \prec p$ . There is no real motivation for considering such a generalization at this stage, but it has often been remarked that our observations do not imply that causality has to hold even at Planck scales, and there would be no fundamental reason for not accepting as valid space-times, ones with Planck-size closed timelike curves. If we did consider preposet theory as a serious possibility, we would then have to explain dynamically the absence of causality violation at macroscopic scales. Certainly, it is conceivable that manifold reconstruction theorems could be formulated and proved for preposet theory as well. For example, the 1-1 correspondence between finite posets and  $T_0$  finite topological spaces described in §2.7 becomes a 1-1 correspondence between finite preposets and finite topological spaces, the latter being  $T_0$  iff the preposet is a poset.

Finally, a useful approach to our causal set theory might consist in shifting the emphasis from the elements of a causal set to its past sets, or order ideals (defined in connection with finite topological spaces in §2.7). The set of order ideals of any causal set is also a poset, in which order is inclusion. But this is not any old poset, it is a *distributive lattice*: any two elements  $p$  and  $q$  have a unique minimal element to the future of both, called the *join*  $p \vee q$  (read: union), and a unique maximal element to the past of both, the *meet*  $p \wedge q$  (read: intersection), which in addition satisfy the distributivity properties  $p \vee (q \wedge r) = (p \vee q) \wedge (p \vee r)$  and  $p \wedge (q \vee r) = (p \wedge q) \vee (p \wedge r)$  (this is always true for any collection of subsets of any set, provided it is closed under union and intersection). Viceversa, from any distributive lattice we can define a unique (unlabelled) poset whose set of past sets is the given lattice. (This may seem strange, in view of the fact that, for a given number of elements, there are many more posets than distributive lattices, but one must realize that a given poset has many more order ideals than

elements—each element corresponds to a *principal* order ideal—and we saw how fast the number of posets increases with size.) Thus, a theory of distributive lattices is not a generalization of a theory of causal sets, but the former have many mathematical properties that can make them easier to work with. (This advantage is similar to that of working with Alexandrov sets and making use of the properties of the incidence algebra, that we have just started exploring in §3.4.)

But the similarity of the symbols for “join” and “meet” with those for “or” and “and” in logic is not a coincidence, and rewording causal set theory as a theory of distributive lattices makes it acquire a taste of propositional calculus.

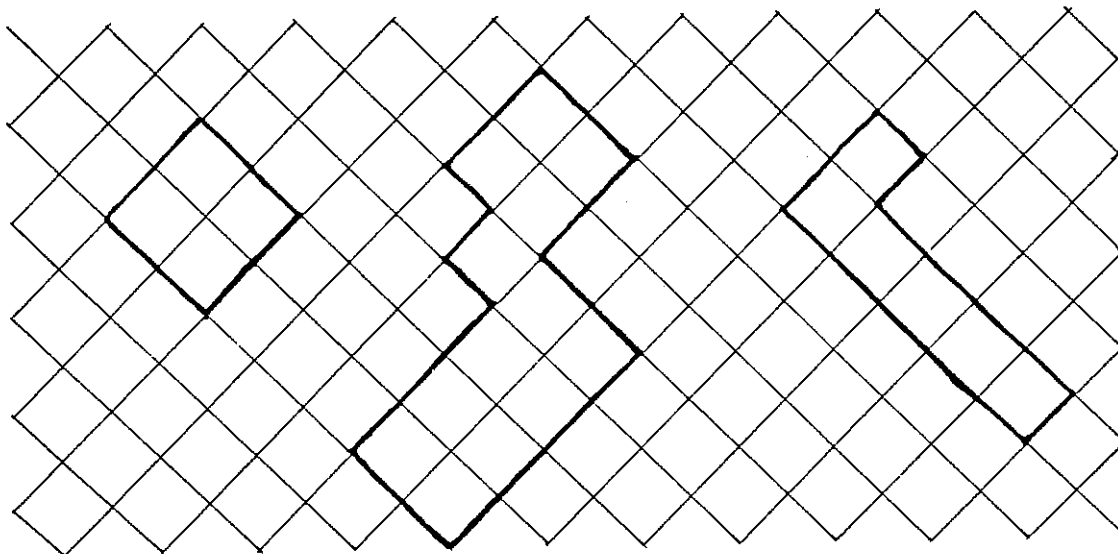


## APPENDICES

### A.1 Two-dimensional calculations with the “diamond lattice”

A case in which calculations with a large causal set can be done with relative ease (although it be considered not very representative of the generic causal set situation in some—perhaps even most—aspects), is that of a  $(1+1)$ -dimensional “diamond lattice” (a 2-dimensional square lattice tilted by  $45^\circ$ ), shown in figure A.1.1. This causal set is very unusual if one compares it with other causal sets embedded in Minkowski space: each point has only two links to its future and two to its past, which is a consequence of its noninvariance under Lorentz transformations (compare with figure 2.5.1 and the remark preceding it). The reason why we consider such a lattice is to get a feeling for the behavior of some causal set-related quantities, in an attempt to obtain a better understanding of the properties of various possible definitions for the action. In particular, we will study how the number of loops and multiloops increases with the volume of the causal set.

Consider the  $(1+1)$ -dimensional “diamond lattice” above, consisting of layers of points labelled from the bottom, the  $n$ -th layer containing  $n$  points, and the portion of lattice from the bottom to the  $n$ -th layer a total of  $N = \frac{1}{2}(n+1)(n+2)$  points. Due to the special symmetry of the lattice, a loop in it, as shown in figure A.1.1, can be equivalently thought of as a set of diamonds, piled in an appropriate way. Our first goal is to calculate the number  $L_n$  of different loops that can be drawn in an  $n$ -step diamond lattice.



Diamond lattice with loops

figure A.1.1

Define the following quantities:  $l_n$  is the number of loops starting at the minimal point and ending after  $n$  steps (at any point in the  $(n + 1)$ -th layer) and  $l_n^i$  is the number of configurations of diamonds which, after  $n$  steps, have  $i$  "peaks", so that  $l_n = l_n^1$ . We are interested in calculating  $l_n$ . The  $l_n^i$ 's are auxiliary quantities, which turn out to be useful because they satisfy a simple recursion relation: for any given configuration with  $i$  peaks after  $n$  steps, there are two ways it could have been obtained from one with  $i$  peaks after  $n - 1$  steps, one way it could have been obtained from one with  $i - 1$  peaks, and one way from a configuration with  $i + 1$  peaks, i.e.,

$$l_n^i = l_{n-1}^{i-1} + 2l_{n-1}^i + l_{n-1}^{i+1}. \quad (\text{A.1.1})$$

Thus, given the fact that, as can be easily checked,

$$l_n^i = 0 \quad \forall i > n - 1; \quad l_n^{n-1} = 1, \quad (\text{A.1.2})$$

we get, e.g.,  $l_2^1 = 1$ ,  $l_2^i = 0$  for  $i > 1$ , and this allows us to obtain all other  $l_n^i$ 's by iteration (notice in fact that we could set  $l_0^i = 0$ ,  $l_1^0 = 1$ , and  $l_n^0 = 0$  for  $n > 1$ , as

equivalent boundary conditions). We give here a table a few of the  $l_n^i$  obtained in this way:

$n; i$	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0
2	0	1	0	0	0	0	0	0
3	0	2	1	0	0	0	0	0
4	0	5	4	1	0	0	0	0
5	0	14	14	6	1	0	0	0
6	0	42	48	27	8	1	0	0
7	0	132	165	110	44	10	1	0

An alternative way of calculating these numbers is by noticing that  $l_{i+1}^i = l_n^{n-1} = 1$ ,  $l_{i+2}^i = 2i$ ,  $l_{i+3}^i = 2i^2 + 3i$ ,  $l_{i+4}^i = \frac{2}{3}(2i^3 + 9i^2 + 10i)$ , etc., and that in general

$$l_{i+k}^i = \sum_{l=0}^{i-1} (2l_{l+k}^{l+1} + l_{l+k}^{l+2}). \quad (\text{A.1.3})$$

The calculation of  $l_n$  still does not give us the number of loops contained in a given number of layers,  $L_n$ . To obtain this, we notice that  $L_n$  is the sum over all pairs of layers  $k < l$  of  $l_{l-k}$  times the number of points on the  $k$ -th layer where the loops can start:

$$\begin{aligned} L_n &= \sum_{p=0}^{n-2} \left[ (p+1) \sum_{q=2}^{n-p} l_q \right] \\ &= \sum_{p=0}^{n-2} (p+1) \lambda_{n-p}, \end{aligned} \quad (\text{A.1.4})$$

where  $\lambda_n := \sum_{q=1}^n l_q$  is the number of loops starting at the minimal point and ending after  $n$  steps or less. Notice that, if we define, similarly to  $l_n^i$ ,  $\lambda_n^i$  to be the number of diamond configurations which end after  $n$  steps or less with  $i$  peaks,

i.e.,

$$\lambda_n^i := \sum_{q=1}^n l_q^i, \tag{A.1.5}$$

so that  $\lambda_n = \lambda_n^1$ , then  $\lambda_n^i$  satisfies the same recursion relation as  $l_n^i$  does, but with different initial conditions, namely

$$\lambda_n^i = \lambda_{n-1}^{i-1} + 2\lambda_{n-1}^i + \lambda_{n-1}^{i+1}, \tag{A.1.6}$$

as can be checked using (A.1.1) and (A.1.5), with  $\lambda_0^i = 0$  and  $\lambda_n^0 = 1$  for  $n > 0$ .

Thus, we get the following values for  $\lambda_n^i$ :

$n; i$	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0
2	1	1	0	0	0	0	0	0
3	1	3	1	0	0	0	0	0
4	1	8	5	1	0	0	0	0
5	1	22	19	7	1	0	0	0
6	1	64	67	34	9	1	0	0
7	1	196	232	144	53	11	1	0

Values of  $L_n$ , calculated using (A.1.4), are given in the table below:

$n$	2	3	4	5	6	7	8
$L_n$	1	5	17	51	149	443	1362
$\ln(L_n)$	0	1.61	2.83	3.93	5.00	6.09	7.22

Based on these numbers, it seems that  $L_n$  grows exponentially with  $n$ , or with the total number of points in the lattice.

Let us try to estimate the asymptotic behavior of  $L_n$  for large  $n$  using a different approach. We will first replace the recursion relation (A.1.1), by a differential equation for a function  $l(n, y)$ , and we will try to obtain information

about the asymptotic behavior of  $l_n^i$  as  $n \rightarrow \infty$  from that of  $l(n, y)$  as  $n \rightarrow \infty$ , for fixed  $y$ . We rewrite (A.1.1) as

$$-3l_{n-1}^i - (l_{n-1}^i - l_n^i) = l_{n-1}^{i-1} - 2l_{n-1}^i + l_{n-1}^{i+1}, \quad (\text{A.1.7})$$

and replace it by

$$-3l(n-1, y) + \frac{\partial l}{\partial n}(n - \frac{1}{2}, y) = \frac{\partial^2 l}{\partial y^2}(n-1, y), \quad (\text{A.1.8})$$

or

$$-3l + \frac{\partial l}{\partial n} + \frac{1}{2} \frac{\partial^2 l}{\partial n^2} = \frac{\partial^2 l}{\partial y^2}. \quad (\text{A.1.9})$$

Solutions to this equation can be found by the method of separation of variables: if we set  $l_n^i = N(n)I(y)$ , then the usual procedure gives

$$-3 + \frac{1}{2N} \frac{\partial^2 N}{\partial n^2} + \frac{1}{N} \frac{\partial N}{\partial n} = \frac{1}{I} \frac{\partial^2 I}{\partial y^2}, \quad (\text{A.1.10})$$

or, introducing an arbitrary separation constant  $\alpha$ ,

$$\begin{aligned} N(n) &= A e^{\xi_+ n} + B e^{\xi_- n} \\ I(y) &= C e^{\sqrt{\alpha} y} + D e^{-\sqrt{\alpha} y}, \end{aligned} \quad (\text{A.1.11})$$

where  $A, B, C$  and  $D$  are arbitrary constants, and  $\xi_{\pm} := -1 \pm \sqrt{7 + 2\alpha}$ . Therefore, the general solution of the differential equation is

$$l(n, y) = \int_{-\infty}^{\infty} d\alpha \left( A(\alpha) e^{\xi_+ n} + B(\alpha) e^{\xi_- n} \right) \left( C(\alpha) e^{\sqrt{\alpha} y} + D(\alpha) e^{-\sqrt{\alpha} y} \right) \quad (\text{A.1.12})$$

where the arbitrary functions  $A, B, C$  and  $D$  depend on the imposed boundary conditions. The latter we would like to be equivalent to the ones specified above for  $l_n^i$ :  $l(n=0, y) = 0$  for  $y > 0$ , and  $l(n, y=0) = 0$ , unless  $n \approx 1$ .

What can we say now about the total number of *multiloops*,  $M_n$ , contained in a lattice of  $n$  steps? We can certainly put an upper bound on  $M_n$  by counting all possible combinations of used and unused diamonds: we thus get  $M_n < 2^N = 2^{(n+1)(n+2)/2}$ . The overcounting consists in the fact that this will include all multiloops in which loops are allowed to touch each other at the corners and have arbitrary shapes (i.e., they need not be formed by two timelike paths joined at their ends). In fact, the multiloops we count in this way are the ones used in the Ising model calculations mentioned in §3.3, and our upper bound on  $M_n$  is (up to a coefficient) the partition function (3.3.5) for  $x = 1$ .

## A.2 An application of simulated annealing

A method called *simulated annealing* has recently been developed for extremizing functions of many variables (for references and a review of this and other methods, see, e.g., Press, Flannery, Teukolsky and Vetterling [82], in particular §10.9). Usually, in the applications of simulated annealing, the variables can have only discrete values, but I will summarize the method without this restriction. Indicating the function to be minimized by  $f(q_1, q_2, \dots, q_N)$ , by a *configuration* of the system I will denote a set of values for  $(q_1, q_2, \dots, q_N) \equiv q$ , and  $f$  will be called an *energy function*, since it will play a role analogous to that of the potential energy of a physical system.

We start with an arbitrarily chosen configuration  $q_{(1)}$  for the system and calculate its energy  $f(q_{(1)})$ . Although some insight can go into the choice of the initial configuration, the latter could be very far from the one that minimizes  $f$ . We thus want to let the system “evolve” by generating a series of new configurations, trying to ensure that they get us closer and closer to the minimum. To check that they do is however a difficult problem, since  $f$  in general may have many local minima, and the fact that in one reconfiguration the value of  $f$  is lowered does not imply that this reconfiguration brings us closer to the minimum. To avoid getting stuck in local minima, we want to allow ourselves to occasionally make reconfigurations that increase the value of  $f$ , with a probability which depends on how close to the true minimum we think we are. This is achieved by imagining the system to have a temperature  $T$ , which superimposes a random thermal fluctuation to its evolution towards a minimum of the energy. We thus proceed as follows.

Choose a value for the temperature  $T$ . Given any configuration  $q_{(i)}$ , we

randomly generate a new one,  $\tilde{q}$ —exactly how we do this, i.e., what we mean by “randomly”, depends on the problem at hand, but we will see an example shortly—, and calculate the energy difference  $\Delta f = f(\tilde{q}) - f(q_{(i)})$ . If  $\Delta f < 0$  we accept the reconfiguration, and we set  $q_{(i+1)} = \tilde{q}$ , otherwise we do so with probability  $e^{-\Delta f/T}$ . Repeat this procedure for producing new configurations, either a fixed number of times, or until the value of  $f$  for many iterations changes so little between one configuration and the next, that we interpret the changes as being due to “thermal fluctuations”, which prevent the system from settling down to the minimum. Then lower  $T$  (normally by a fixed fraction), and again repeat the procedure. Lowering the temperature is what the annealing process consists in, and the success of the run may strongly depend on the way in which it is done. Ideally, when or before  $T = 0$ , the configuration of the system should be the true minimum of  $f$ .

Our problem was that of trying to embed a given causal set  $P$  of size  $N$  in a Minkowski space of a given dimensionality  $n$ . It can be cast into the form of a minimization problem if we give a function  $f$  which we interpret, for each set of positions of  $N$  points in  $n$ -dimensional Minkowski space, as the “badness” of the agreement between the causal relations induced on these points and those defining  $P$ . Thus, the energy is the sum over all pairs of points of a quantity which is zero if the points are correctly related, and otherwise depends on the distance they have to be displaced by to become correctly related. The good configurations are not a set of zero measure in the set of all configurations, and they are distinguished by the vanishing of the energy. This has the advantage that a good configuration is immediately recognized when it is reached, and the program does not need to keep running, e.g., until  $T = 0$ , as in the general case.



### The program<sup>†</sup>

The points in  $n$ -dimensional Minkowski space from which we want to obtain a realization of  $P = \{p_i\}$  are represented as spheres, the intersection of a  $t = \text{const}$  hyperplane with the past light cones of the points (see lemma 1), with radius  $R_i$  and center at  $x_i^a$ . Thus, in a causal embedding of  $P$ ,  $p_i < p_j$  iff  $-(R_j - R_i)^2 + \sum_{a=1}^{n-1} (x_j^a - x_i^a)^2 < 0$  and  $R_i < R_j$ . The solution we have given to the problem of the choice of energy and reconfiguration algorithm can be best understood if we start from the latter. Once the program decides whether a particular sphere is to be moved in a reconfiguration (this happens with a probability depending on how correctly it is related to the others, which actually makes ours a "biased annealing", not really thermal), the position of the sphere is changed from the current one by a normally distributed displacement in a uniformly random direction in space, and the radius is changed to a new value, normally distributed around the current one. The displacement of the sphere looks therefore like a random displacement in  $n$ -dimensional Euclidean space, and the contribution to the energy due to a pair of points has been chosen, accordingly, to be proportional to the minimum euclidean distance one of them has to move in this space in order to be correctly related to the other. More details can be found by looking at the listing of the program, below.

A more satisfactory choice of displacements and energy would use in a more direct way the lorentzian nature of the Minkowski metric, and an improved version of the program will be written, taking this fact into account. The present version of the program finds correct embeddings of simple causal sets, like the embedding of the six-element crown  $P_6^c = P_{(3)}^{DM}$  in 3 dimensions, but does not converge to a correct embedding for more complicated causal sets.

---

<sup>†</sup> Written in collaboration with D. Meyer.

## Listing of the source code (written in Pascal)

```

program cones(input,output);

const
  maxN = 20; maxdim = 10;
type
  pmatrix = array[1..2,1..100] of real;
var
  name: packed array[1..12] of char;
  infile, outfile: text;
  i, j, zij, dim, N, count, Ndata, gliy: integer;
  seed: integer:= 1959; gliset, totalcount: integer:= 0;
  glir: array[1..97] of integer;
  change: array[1..maxN] of boolean;
  z: array[1..maxN,1..maxN] of boolean;
  Xnew, Xold: array[1..maxN,1..maxdim] of real:= zero;
  Rnew, Rold: array[1..maxN] of real:= zero;
  E: array[1..100] of real:= zero;
  Enew, Eold: array[1..maxN,1..maxN] of real:= zero;
  deltaE, Rave, R, Rmin, T, Eave, Evar, spheat, glgset: real;
  data: pmatrix;

#include 'rand.pas'

procedure energy;
const
  roottwo = sqrt(2);
var
  Rij, Xij, s2: real;
  i, j, k: integer;
begin
  deltaE:= 0;
  for i:=1 to N-1 do
    for j:=i+1 to N do
      if change[i] or change[j] then begin
        Rij:= Rnew[i]-Rnew[j];
        Xij:= 0;
        for k:=1 to dim do Xij:=Xij+(Xnew[i,k]-Xnew[j,k])**2;
        s2:= -(Rij**2)+Xij;
        Xij:= sqrt(Xij);
        if z[i,j] then
          if s2>0 then Enew[i,j]:=(Xij+Rij)/(roottwo*Rave)
          else
            if Rij>0 then Enew[i,j]:=sqrt(s2+2*(Rij**2))/Rave
            else Enew[i,j]:= 0
        else
          if s2>0 then Enew[i,j]:= 0
          else Enew[i,j]:= (abs(Rij)-Xij)/(roottwo*Rave);
        deltaE:= deltaE+Enew[i,j]-Eold[i,j]
      end
    end
  end; { energy }

```

```

procedure reconfigure;
var
  i, k: integer;
  Efraction, norm: real;
  displacement: array[1..maxdim+1] of real;
begin
  Rave:= R;
  for i:=1 to N do begin
    Efraction:= 2*Eold[i,i]/E[1];
    if ran2(seed)<Efraction then change[i]:= true;
    repeat
      repeat
        norm:= 0;
        for k:=1 to dim+1 do begin
          displacement[k]:= 2*ran2(seed)-1;
          norm:= norm+displacement[k]**2
        end
      until norm<1;
      norm:= Eold[i,i]*R*gasdev(seed)/sqrt(norm);
      for k:=1 to dim do
        Xnew[i,k]:= Xold[i,k]+displacement[k]*norm;
        Rnew[i]:= Rold[i]+displacement[dim+1]*norm
      until Rnew[i]>0;
      Rave:= Rave+(Rnew[i]-Rold[i])/N
    end
  end; { reconfigure }

procedure update;
var
  i, j, k: integer;
begin
  for i:=100 downto 2 do E[i]:= E[i-1];
  E[1]:= 0;
  Rmin:= Rnew[1];
  for i:=1 to N do begin
    if Rnew[i]<Rmin then Rmin:= Rnew[i];
    if change[i] then begin
      for k:=1 to dim do Xold[i,k]:= Xnew[i,k];
      Rold[i]:= Rnew[i]
    end;
    for j:=i+1 to N do
      if change[i] or change[j] then Eold[i,j]:= Enew[i,j];
    Eold[i,i]:= 0;
    for j:=1 to i-1 do Eold[i,i]:= Eold[i,i]+Eold[j,i]/2;
    for j:=i+1 to N do Eold[i,i]:= Eold[i,i]+Eold[i,j]/2;
    E[1]:= E[1]+Eold[i,i]
  end;
  for i:=1 to N do begin
    for k:=1 to dim do Xold[i,k]:= Xold[i,k]/Rmin;
    Rold[i]:= Rold[i]/Rmin
  end;
  Rave:= Rave/Rmin;
  R:= Rave
end; { update }

```

```

procedure statistics;
var
  i: integer;
  E2: real;
begin
  Eave:= 0;
  E2:= 0;
  for i:=1 to 100 do begin
    Eave:= Eave+E[i];
    E2:= E2+E[i]**2
  end;
  Eave:= Eave/100;
  E2:= E2/100;
  Evar:= sqrt(E2-(Eave**2));
  if totalcount=0 then T:= 100;
  if T>0 then spheat:= (Evar/T)**2;
  totalcount:= totalcount+100
end; { statistics }

procedure startup;
begin
  write('Input data file: ');
  readln(name);
  open(infile,name,old);
  reset(infile);
  open(outfile,'cone.out');
  rewrite(outfile);
  readln(infile,N);
  for i:=1 to N-1 do
    for j:=i+1 to N do begin
      read(infile,zij);
      if zij=0 then z[i,j]:= false
      else
        if zij=1 then z[i,j]:= true
        else writeln('wrong data in input file')
      end;
    write('how many space dimensions? ');
    readln(dim);
    for i:=1 to N do begin
      change[i]:= true;
      for j:=1 to dim do Xnew[i,j]:= 0;
      Rnew[i]:= i+1
    end;
  Rave:= (N+4)/2;
  energy;
  update;
  writeln('Energy of initial configuration: ',E[1]:9);
  writeln(outfile,
    'Energy of initial configuration: ',E[1]:9)
end; { startup }

```

```

procedure warmup;
begin
  count:= 0;
  while (count<100) and (E[1]>0) do begin
    reconfigure;
    energy;
    update;
    count:= count+1;
    writeln(count)
  end;
  statistics;
  writeln('
    Count   Temp.   E[1].   Eave.   Evar.   spheat');
  writeln('
    warmup -----',E[1]:8,Eave:8,Evar:8,spheat:8);
  writeln(outfile,'
    Count   Temp.   E[1].   Eave.   Evar.   spheat');
  writeln(outfile,'
    warmup -----',E[1]:8,Eave:8,Evar:8,spheat:8);
  write('waiting for <cr> ... ');
  readln
end; { warmup }

procedure decide;
begin
  if deltaE<0 then begin
    update;
    count:= count+1
  end
  else
    if T>0 then
      if ran2(seed)<4*exp(-deltaE/T) then begin
        update;
        count:= count+1
      end
end; { decide }

procedure anneal;
var
  continue: char;
begin
  Ndata:= 0;
  continue:= 'y';
  while continue='y' do begin
    count:= 0;
    while (count<100) and (E[1]>0) do begin
      reconfigure;
      energy;
      decide
    end;
    statistics;
    writeln('
      totalcount:5,T:8,E[1]:8,Eave:8.Evar:8,spheat:8);
    writeln(outfile,'
      totalcount:5,T:8,E[1]:8,Eave:8,Evar:8,spheat:8);
  end;
end;

```

```

Ndata:= Ndata+1;
data[1,Ndata]:= T;
data[2,Ndata]:= Eave;
write('y/t/n: ');
readln(continue);
if continue='t' then begin
  write('new T: ');
  readln(T);
  continue:= 'y'
end;
if continue='y' then T:= T*0.9;
if Ndata=35 then continue:= 'n'
end
end; { anneal }

{ procedure prepareplot(Ndata: integer; data: pmatrix);
fortran; }

procedure writeout(var where: text);
var
  i, j, k: integer;
begin
  writeln(where);
  writeln(where,'Incidence matrix:');
  writeln(where);
  for i:=1 to N do begin
    for j:=1 to N do write(where,' ',z[i,j]:1);
    writeln(where)
  end;
  writeln(where);
  writeln(where,'dimension: ',dim:2);
  writeln(where);
  for i:=1 to N do begin
    write(where,'i=',i:2,', X=');
    for k:=1 to dim do write(where,' ',Xold[i,k]:8);
    writeln(where,' ', R=',Rold[i]:8)
  end;
  writeln(where);
  writeln(where,'Energy matrix:');
  writeln(where);
  for i:=1 to N do begin
    for j:=1 to N do write(where,' ',Eold[i,j]:8);
    writeln(where)
  end;
  writeln(where);
  writeln(where,'Total energy: ',E[1])
end; { writeout }

begin { main program }
  startup;
  warmup;
  anneal;
  { prepareplot(Ndata,data); }
  writeout(output);
  writeout(outfile)
end.

```

**Random number generators (from reference [82]).**

```

FUNCTION ran2(VAR idum: integer): real;
(* Programs using RAN2 must declare the following variables
VAR
  gliy: integer;
  glir: ARRAY [1..97] OF integer;
in the main program. *)
CONST
  m=714025; ia=1366; ic=150889;
  rm=1.400512e-6; (* 1.0/m *)
VAR
  j: integer;
BEGIN
  IF (idum<0) THEN BEGIN
    idum:=(ic-idum) MOD m;
    FOR j:=1 to 97 DO BEGIN
      idum:=(ia*idum+ic) MOD m;
      glir[j]:=idum END;
    idum:=(ia*idum+ic) MOD m;
    gliy:=idum END;
    j:=1+(97*gliy) DIV m;
    IF ((j>97) OR (j<1)) THEN BEGIN
      writeln('pause in routine RAN2'); readln
    END;
    gliy:=glir[j]; ran2:=gliy*rm;
    idum:=(ia*idum+ic) MOD m;
    glir[j]:=idum
  END;

FUNCTION gasdev(VAR idum: integer): real;
(* Programs using GASDEV must declare the variables
VAR
  gliset: integer; glgset: real;
in the main routine and must initialize gliset to
  gliset:=0; *)
VAR
  fac,r,v1,v2: real;
BEGIN
  IF (gliset=0) THEN BEGIN
    REPEAT
      v1:=2.0*ran2(idum)-1.0; v2:=2.0*ran2(idum)-1.0;
      r:=sqr(v1)+sqr(v2);
    UNTIL (r<1.0);
    fac:=sqrt(-2.0*ln(r)/r); glgset:=v1*fac;
    gasdev:=v2*fac;
    gliset:=1 END
  ELSE BEGIN
    gasdev:=glgset; gliset:=0 END
  END;

```





### Other possibilities

An alternative to the simulated annealing method for minimizing functions of many variables is the so-called ALOPEX algorithm (see Pandya, Unnikrishnan and Harth [81]). This method also uses the noise produced by an effective temperature to overcome the local minima problem, but one does not first generate a configuration and then decide whether to accept it or not. Rather, at each reconfiguration, each variable is changed in a direction which has a higher probability of being in the same direction as the one of previous change, if this led to a decrease in  $f$ , and of being in the opposite direction otherwise, the exact probability depending on the temperature. In a simple case, each parameter can change from its value  $q_{i(N)}$  in the  $N$ -th configuration to

$$q_{i(N+1)} = q_{i(N)} + \delta_i(N) = q_{i(N)} \pm \delta \quad (\text{A.2.1})$$

in the  $(N+1)$ -th, for some constant  $\delta$ . The probability that  $\delta_i(N) = +\delta$  is

$$P_i^+(N) = \frac{1}{1 + e^{-\Delta_i(N)/T}}, \quad (\text{A.2.2})$$

where  $\Delta_i(N) := [f(q_{i(N)}) - f(q_{i(N-1)})][q_{i(N)} - q_{i(N-1)}]$ , and, clearly,  $P_i^-(N) = 1 - P_i^+(N)$ .

Finally, a third minimization method which is useful sometimes is that of imagining the system to be subject to a force, which makes it move toward the equilibrium configuration (and subject also to an implicit additional damping, since we don't want to set it into an oscillatory motion).

### A.3 Sprinkling points in a given background

We have seen various instances, in previous chapters, of situations in which something could be learned about properties of causal sets faithfully embeddable in some manifold with metric, from looking at the corresponding properties of a large set of points sprinkled in this manifold with uniform density: statistical properties from which fractal dimension is defined (§2.6), from which scalar curvature can be estimated (§2.7), or from which definitions of the action can be evaluated (§3.3). Here, I will briefly describe the algorithms suitable for producing with a computer those random sprinklings, for simple cases of space-times.

As a starting point, I assume we have available a generator of “random” numbers uniformly distributed between 0 and 1, like the function RAN2 used in the program listed in appendix 2, and that we want to generate randomly distributed points in a region  $S$  of a space-time  $(M, g_{ab})$ .

The general procedure will be this. Suppose we can choose a set of coordinates  $x^\mu$  such that the volume element can be factored into the form

$$d^n V \equiv \sqrt{-g} d^n x = \prod_{\mu=0}^{n-1} f_{(\mu)}(x^\mu) dx^\mu, \quad (\text{A.3.1})$$

where  $f_{(\mu)}(x^\mu)$  is a function of one coordinate  $x^\mu$  only. Suppose in addition that the region  $S$  of interest can be specified by giving separate ranges  $x^\mu \in [a^{(\mu)}, b^{(\mu)}]$  for all the coordinates (think of the  $a$ 's and  $b$ 's as constants for now, although we will need a slightly more general case in a while). These metrics are not of the most general kind, but they are the ones I will restrict my attention to, while the condition on  $S$  is just for simplicity of discussion: in many cases it will be satisfied; if not, we just need to enclose  $S$  in a bigger region which does satisfy it, sprinkle points uniformly there, and ignore the ones that fall outside  $S$ .

Then a uniform sprinkling in  $S$  is obtained by generating, for each  $\mu$ , the coordinate  $x^\mu$  of each point according to the probability distribution  $f_{(\mu)}(x^\mu)$ .

So now the problem is reduced to that of how to generate a variable  $x \in [a, b]$  with a known distribution  $f$ . This problem has several solutions (see, e.g., Press, Flannery, Teukolsky and Vetterling [82], esp. §7.2), the most general one, and sometimes a quite inefficient one, being the Monte Carlo method. But let me assume here that we are lucky and that the function  $f$  is integrable, and its primitive  $\int_a^x f dx' =: F(x)$  is analytically invertible. Then, we just have to generate a (uniformly) random number  $\xi$  between 0 and  $F(b)$ , calculate  $x = F^{-1}(\xi)$ , and this  $x$  will have probability distribution  $f$ .

We can now see some applications, in which  $S$  will be an Alexandrov neighborhood  $A(x, y)$ :

### 2-dimensional Minkowski space

This first example is trivial, since for any Alexandrov neighborhood in 2-dimensional Minkowski space there are null coordinates  $u$  and  $v$  such that the metric and the Alexandrov neighborhood are expressed as

$$-ds^2 = \frac{1}{2} du dv, \quad \text{and} \quad u \in [0, l], \quad v \in [0, l]. \quad (\text{A.3.2})$$

This obviously satisfies our assumptions, but all we need to do is to generate both  $u$  and  $v$  uniformly from 0 to  $l$ . It is identical to generating points in a rectangular cartesian box (which, by the way, is how figure (2.5.1) was generated).

### $n$ -dimensional Minkowski space

Here generating points in a cartesian box is just as trivial as in 2 dimensions, but for  $S = A(x, y)$  we have to use the procedure outlined above. Thinking of  $A(x, y)$  as two base-to-base cones with base an  $(n-1)$ -ball of radius  $l/2$  and

height  $l/2$ , we choose as coordinates  $t$ ,  $r$  and the angles on an  $(n-2)$ -sphere  $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(n-2)}$ , in terms of which the metric and  $A(x, y)$  are given by

$$\begin{aligned} ds^2 &= -dt^2 + r^2 d\Omega_{n-2}^2 \\ &= -dt^2 + dr^2 + r^2 \left\{ (d\theta^{(1)})^2 + \sin^2 \theta^{(1)} \left[ (d\theta^{(2)})^2 + \sin^2 \theta^{(2)} (\dots) \right] \right\} \\ t &\in [-l/2, +l/2]; \quad r \in [0, l/2 - |t|]; \\ \theta^{(i)} &\in [0, \pi] \quad \text{for } i = 1, \dots, n-3; \quad \theta^{(n-2)} \in [0, 2\pi], \end{aligned} \quad (\text{A.3.3})$$

where  $d\Omega_n^2$  is the line element on the unit  $n$ -sphere  $S^n$ . Then,

$$\sqrt{-g} = r^{n-2} \sin^{n-3} \theta^{(1)} \sin^{n-4} \theta^{(2)} \dots \sin \theta^{(n-3)}, \quad (\text{A.3.4})$$

and it is straightforward to find an algorithm for generating each coordinate with the corresponding probability distribution. In the case of Minkowski space, it is much simpler to use cartesian coordinates and generate points in a larger rectangular box, and later throw away all points that fall outside the Alexandrov neighborhood, although for high dimensionalities this might require a greater amount of computer time. Therefore, I will not give explicitly the form of the  $F_{(\mu)}$ 's for spherical coordinates.

### de Sitter space-time

In  $(3+1)$ -dimensional de Sitter space-time, the space we can use when we need to do calculations in the presence of a constant positive curvature, the metric in the proper time gauge is given by (see, e.g., Hawking and Ellis [48])

$$ds^2 = -dr^2 + \frac{3}{\Lambda} \cosh^2 \left( \sqrt{\frac{\Lambda}{3}} r \right) d\Omega_3^2. \quad (\text{A.3.5})$$

Since  $d\Omega_3^2 = d\chi^2 + \sin^2 \chi d\theta^2 + \sin^2 \chi \sin^2 \theta d\phi^2$ ,

$$\sqrt{-g} = \left( \frac{3}{\Lambda} \right)^{3/2} \cosh^3 \left( \sqrt{\frac{\Lambda}{3}} r \right) \sin^2 \chi \sin \theta, \quad (\text{A.3.6})$$

and, to generate points randomly, we use the fact that

$$\begin{aligned} \int \sin x \, dx &= -\cos x & \int \sin^2 x \, dx &= \frac{x}{2} - \frac{1}{4} \sin 2x \\ \int \cosh^3 x \, dx &= \sinh x + \frac{1}{3} \sinh^3 x. \end{aligned} \quad (\text{A.3.7})$$

Alternatively, we could use the metric in the form

$$ds^2 = -dt^2 + e^{2t\sqrt{\Lambda/3}}(d\hat{x}^2 + d\hat{y}^2 + d\hat{z}^2), \quad (\text{A.3.8})$$

depending on which coordinates are more convenient for defining an Alexandrov neighborhood, or we could use the fact that de Sitter space is the hyperboloid  $-v^2 + w^2 + x^2 + y^2 + z^2 = 3/\Lambda$  in  $\mathbb{R}^5$ , to sprinkle points randomly in a thickened hyperboloid (between  $\Lambda$  and  $\Lambda + \Delta\Lambda$ ) in this space, and then project the points onto the desired hyperboloid.

### Other space-times

I will just mention two other kinds of metric in which we could in principle sprinkle points uniformly with the method described. In  $(3 + 1)$ -dimensional anti-de Sitter space-time (the hyperboloid  $-u^2 - v^2 + x^2 + y^2 + z^2 = 1$  in  $\mathbb{R}^5$ ), our model for a constant negative curvature space, the metric can be given in the forms

$$\begin{aligned} ds^2 &= -dt^2 + \cos^2 t [d\chi^2 + \sin^2 \chi (d\theta^2 + \sin^2 \theta d\phi^2)] \\ &= \cosh^2 r \, dt'^2 + dr^2 + \sinh^2 r (d\theta^2 + \sin^2 \theta d\phi^2). \end{aligned} \quad (\text{A.3.9})$$

Finally, a class of space-times for which the method applies, and in which physically it might be of some interest to sprinkle points is that of Robertson-Walker space-times.

## References

## On fundamental structures for space-time

- [1] Y. Ahmavaara, "The structure of space and the formalism of relativistic quantum theory", *J. Math. Phys.* **6**(1965)87-93.
- [2] V. Ambarzumian and D. Iwanenko, "Zur Frage nach Vermeidung der unendlichen Selbstrückwirkung des Elektrons", *Z. f. Physik* **64**(1930)563-7.
- [3] L. Bombelli, J. Lee, D. Meyer and R.D. Sorkin, "Space-time as a causal set", *Phys. Rev. Lett.* **59**(1987)521-4.
- [4] H.R. Coish, "Elementary particles in a finite world-geometry", *Phys. Rev.* **114**(1959)383; Coish refers to, and in fact his work is based on, earlier papers by Jarnefelt and Kustaanheimo.
- [5] H.S.M. Coxeter and J. Whitrow, "World-structure and non-euclidean honeycombs", *Proc. Roy. Soc. (London)* **A201**(1950)417-37.
- [6] D. Finkelstein,  
 "Space-time code", *Phys. Rev.* **184**(1969)1261-71;  
 "Space-time code. II", *Phys. Rev.* **D5**(1972)320-8;  
 "Space-time code. III", *Phys. Rev.* **D5**(1972)2922-37;  
 "Space-time code. IV", *Phys. Rev.* **D9**(1974)2219-31;  
 with G. Frye and L. Susskind,  
 "Space-time code. V", *Phys. Rev.* **D9**(1974)2231-6;  
 "Space-time code. VI", unpublished.
- [7] R. Geroch, "Einstein algebras", *Comm. Math. Phys.* **26**(1972)271-5.
- [8] B. Hasslacher and M.J. Perry, "Spin networks are simplicial quantum gravity", *Phys. Lett.* **B103**(1981)21-4.
- [9] H. Hata, K. Itoh, T. Kugo, H. Kunitomo and K. Ogawa, "Pregeometrical string field theory: creation of space-time and motion", *Phys. Lett.* **B175**(1986)138-144.
- [10] E.G. Hellund and K. Tanaka, "Quantized space-time", *Phys. Rev.* **94**(1954)192-5.
- [11] E.L. Hill, "Relativistic theory of discrete momentum space and discrete space-time", *Phys. Rev.* **100**(1955)1780-3.
- [12] G.T. Horowitz, J. Lykken, R. Rohm and A. Strominger, "Purely cubic action for string field theory", *Phys. Rev. Lett.* **57**(1986)283-6.
- [13] G.T. Horowitz and D.M. Witt, "Toward a string field theory independent of space-time topology", Santa Barbara preprint UCSBTH-87-36.
- [14] M. Lehto, H.B. Nielsen and M. Ninomiya, "Pregeometric quantum lattice: a general discussion", *Nucl. Phys.* **B272**(1986)213-27.

- [15] M. Lorente, "Space-time groups for the lattice", *Int. J. Theor. Phys.* **25** (1986)55-65.
- [16] J. Myrheim, "Statistical geometry", CERN preprint TH-2538 (1978).
- [17] R. Penrose, "Structure of space-time", in C. de Witt-Morette and J.A. Wheeler, eds., *Battelle rencontres 1967*, Benjamin 1968.
- [18] R. Penrose,  
 "An analysis of the structure of space-time", Adams prize essay 1966 (see, in particular, the appendix);  
 "Angular momentum: an approach to combinatorial space-time", in E.T. Bastin, ed., "Quantum theory and beyond", Cambridge 1970 (this volume contains also other papers of related interest);  
 "On the nature of quantum geometry", in J. Klauder, ed., *Magic without magic: John Archibald Wheeler*, Freeman 1972;  
 "Theory of quantized directions", unpublished notes.
- [19] G.F.B. Riemann, "Über die Hypothesen, welche der Geometrie zu Grunde liegen", inaugural lecture at Göttingen, 10 june 1854, *Gött. Abh.* **13**(1867) 133; reprinted in H. Weyl, ed., Springer-Verlag 1919; part of the relevant passage is quoted in Weyl [55], p. 97.
- [20] A. Schild,  
 "Discrete space-time and integral Lorentz transformations", *Phys. Rev.* **73** (1948)414-5;  
 same title, more detailed version, *Canad. J. Math.* **1**(1949)29-47.
- [21] E. Schrödinger, *Science and humanism*, Cambridge U. Press 1956.
- [22] L. Silberstein, *Discrete space-time*, U. of Toronto studies, Physics series, 1936.
- [23] H.S. Snyder,  
 "Quantized space-time", *Phys. Rev.* **71**(1947)38-41;  
 "The electromagnetic field in quantized space-time", *Phys. Rev.* **72**(1947) 68-71.
- [24] R.D. Sorkin,  
 "Posets as lattice topologies", in B. Bertotti, F. de Felice and A. Pascolini, eds., *General relativity and gravitation, Vol. 1*, Consiglio Nazionale delle Ricerche, Roma 1983;  
 paper in preparation.
- [25] R.D. Sorkin, unpublished.
- [26] L. Szabo,  
 "Quantum causal structures", *J. Math. Phys.* **27**(1986)2709-10;  
 "Simple example of quantum causal structure", *Int. J. Theor. Phys.* **26** (1987)833-43.

- [27] G. 't Hooft, "Quantum gravity: a fundamental problem and some radical ideas", in M. Levy and S. Deser, eds., *Recent developments in gravitation. Cargèse 1978*, Plenum 1979.
- [28] Y. Takeno, "Fluctuation of space-time and elementary particles", *Prog. Theor. Phys.* **38**(1967)1185-6.
- [29] P. Yodzis, "An algebraic approach to classical space-time", *Proc. Roy. Irish Acad.* **A75**(1975)37-47.

### On quantum gravity, quantum geometry, topology change

- [30] A. Ashtekar,  
 "New variables for classical and quantum gravity", *Phys. Rev. Lett.* **57**  
 (1986)2244-7;  
 "New hamiltonian formulation of general relativity", *Phys. Rev.* **D36**(1987)  
 1587-602.
- [31] A. Borde, "Topology change in general relativity", in preparation.
- [32] J.L. Friedman and R.D. Sorkin,  
 "Spin- $\frac{1}{2}$  from gravity", *Phys. Rev. Lett.* **44**(1980)1100-3;  
 "Half-integral spin from quantum gravity", *Gen. Rel. Grav.* **14**(1982)615-  
 20.
- [33] R. Geroch, "Topology in general relativity", *J. Math. Phys.* **8**(1967)782.
- [34] S.W. Hawking, "Space-time foam", *Nucl. Phys.* **B144**(1978)349-62.
- [35] T. Jacobson and L. Smolin, "Non-perturbative quantum gravity", Santa  
 Barbara preprint.
- [36] R.D. Sorkin,  
 "Topology change and monopole creation", *Phys. Rev.* **D33**(1986)978-82;  
 "Non-time orientable lorentzian cobordism allows for pair creation", *Int. J.*  
*Theor. Phys.* **25**(1986)877-81.
- [37] R.D. Sorkin, "Introduction to topological geons", in P.G. Bergmann and  
 V. de Sabbata, eds., *Topological properties and global properties of space-*  
*time*, Plenum Press 1986.
- [38] R.D. Sorkin, to appear in M. Dresden and A. Rosenblum, eds., *History of*  
*modern gauge theories* (Proc. of the 19-26 july 1987 conference, Logan,  
 Utah), Plenum Press.
- [39] F.J. Tipler, "Singularities and causality violation", *Ann. Phys. (N.Y.)* **108**  
 (1977)1-36.
- [40] J.A. Wheeler, "Geometrodynamics and the issue of the final state", in  
*Relativity, groups and topology. Les Houches 1963*, Gordon & Breach 1964.
- [41] P. Yodzis,  
 "Lorentz cobordism", *Comm. Math. Phys.* **26**(1972)39-52;  
 "Lorentz cobordism. II", *Gen. Rel. Grav.* **4**(1973)299-307.



## General relativity and structures on manifolds

- [42] A.D. Alexandrov, "A contribution to chronogeometry", *Can. J. Math.* **19**(1967)1119-28; see also references therein to previous work.
- [43] B. Carter, "Causal structure in space-time", *Gen. Rel. Grav.* **1**(1971)349-91.
- [44] J. Ehlers, F.A.E. Pirani and A. Schild, "The geometry of free fall and light propagation", in L. O'Raiheartaigh, ed., *General relativity. Papers in honour of J.L. Synge*, Clarendon 1972, pp. 63-84.
- [45] R. Geroch, "General relativity in the large", *Gen. Rel. Grav.* **2**(1971)61-74.
- [46] R. Göbel, "Zeeman topologies on space-times of general relativity theory", *Comm. Math. Phys.* **46**(1976)289-307.
- [47] S.W. Hawking, "Stable and generic properties in general relativity", *Gen. Rel. Grav.* **1**(1971)393-400.
- [48] S.W. Hawking and G.F.R. Ellis, *The large scale structure of space-time*, Cambridge 1973.
- [49] S.W. Hawking, A.R. King and P.J. McCarthy, "A new topology for curved space-time which incorporates the causal, differential, and conformal structures", *J. Math. Phys.* **17**(1976)174-81.
- [50] E.H. Kronheimer and R. Penrose, "On the structure of causal spaces", *Proc. Cambr. Phil. Soc.* **63**(1967)481-501.
- [51] D. Malament, "The class of continuous timelike curves determines the topology of space-time", *J. Math. Phys.* **18**(1977)1399-404.
- [52] H. Mehlberg, "Essai sur la théorie causale du temps. I. La théorie causale du temps chez ses principaux représentants", *Stud. Philos. (Lemberg)* **1** (1935)119-260;  
 "Essai sur la théorie causale du temps. II. Durée et causalité", *Stud. Philos. (Lemberg)* **2**(1937)111-231;  
 translated in V.1 of *Time, causality and the quantum theory*, Reidel.
- [53] C.W. Misner, K.S. Thorne and J.A. Wheeler, *Gravitation*, Freeman 1973.
- [54] A.A. Robb,  
*A theory of time and space*, Cambridge 1914;  
*Geometry of time and space*, Cambridge 1936 (a revised version of the first reference).
- [55] H. Weyl, *Space-time-matter*, Dover reprint 1962.
- [56] N.J.M. Woodhouse, "The differentiable and causal structures of space-time", *J. Math. Phys.* **14**(1973)495-501.
- [57] E.C. Zeeman,  
 "Causality implies the Lorentz group", *J. Math. Phys.* **5**(1964)490-3;  
 "The topology of Minkowski space", *Topology* **6**(1967)161-70.

## Mathematical references on posets

- [58] G. Behrendt, "Multiposets and the complexity of posets", *Ars Combinatoria* **23**(1987)69-74.
- [59] L. Comtet, *Advanced combinatorics*, Reidel 1974.
- [60] B. Dushnik and E.W. Miller, "Partially ordered sets", *Am. J. Math.* **63** (1941)600-10.
- [61] P.C. Fishburn, *Interval orders and interval graphs. A study of partially ordered sets*, Wiley 1985.
- [62] D. Kelly and W.T. Trotter, "Dimension theory for ordered sets", in Rival [64], pp. 171-211.
- [63] D.J. Kleitman and B.L. Rothschild, "Asymptotic enumeration of partial orders on a finite set", *Trans. Am. Math. Soc.* **205**(1975)205-20.
- [64] I. Rival, ed., *Ordered sets* (NATO ASI series C83), Reidel 1982.
- [65] R.P. Stanley, *Enumerative combinatorics*, V. 1, Wadsworth & Brooks/Cole 1986 (the exercises in chapter 3, the one on posets, contain many interesting results).

## More philosophical references

- [66] J. Earman, "Are spatial and temporal congruences conventional?", *Gen. Rel. Grav.* **1**(1970)143-57.
- [67] A. Grünbaum,  
*Philosophical problems of space and time*, 2nd ed., Reidel 1973;  
"Absolute and relational theories of space and time", in J. Earman, C. Glymour and J. Stachel, eds., *Foundations of space-time theories*, University of Minnesota Press: Minneapolis 1977, pp. 303-74.
- [68] H. Reichenbach, *Philosophie der Raum-Zeit-Lehre*, W. de Gruyter: Berlin 1928; english translation: *The philosophy of space and time*, Dover 1957.
- [69] R.U. Sexl, "Universal conventionalism and space-time", *Gen. Rel. Grav.* **1**(1970)159-80.
- [70] articles by M. Taketani and S. Sakata, in *Suppl. Progr. Theor. Phys.* **50**(1971).
- [71] R. Torretti, *Relativity and geometry*, Pergamon 1983.
- [72] J.A. Wheeler, preface to: J.C. Graves, *The conceptual foundations of contemporary relativity theory*, MIT Press 1971.

## Other references

- [73] D.J. Amit, *Field theory, the renormalization group and critical phenomena*, McGraw-Hill 1978, pp. 18-28.
- [74] L. Castell, M. Drieschner and C.F. von Weizsäcker, "Quantum theory and the structure of space-time", Hanser, Munich 1975, 1977, 1979.
- [75] D. Finkelstein, "Finite physics", to be published in *The universal Turing machine*.
- [76] C. Itzykson, "Fields on a random lattice", in G. 't Hooft et al., eds., *Progress in gauge field theory*, Plenum 1984, pp. 337-71.
- [77] J.M. Jauch, "Foundations of quantum mechanics", Addison-Wesley 1968.
- [78] L. Landau and E.M. Lifshitz, *Statistical physics*, part 1, 3rd ed., Pergamon 1980, pp. 498-506.
- [79] T.D. Lee, "Discrete mechanics", in A. Zichichi, ed., *How far are we from the gauge forces*, Plenum 1985.
- [80] A.R. Marlow, "Empirical topology: topology from partially ordered sets", *Int. J. Theor. Phys.* 19(1980)515-9.
- [81] A.S. Pandya, K.P. Unnikrishnan and E. Harth, "ALOPEX: a stochastic parallel optimization algorithm", in preparation.
- [82] W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, *Numerical recipes*, Cambridge 1986.
- [83] L.A. Santalo, *Integral geometry and geometrical probability*, Addison-Wesley 1976.