Philosophy of Physics

Recovering General Relativity from a Planck Scale Discrete Theory of Quantum Gravity

RESEARCH

JEREMY BUTTERFIELD

FAY DOWKER

*Author affiliations can be found in the back matter of this article

LSE Press

ABSTRACT

An argument is presented that if a theory of quantum gravity is physically discrete at the Planck scale and the theory recovers General Relativity as an approximation, then, at the current stage of our knowledge, causal sets must arise within the theory, even if they are not its basis.

We lay out this argument in two claims. Roughly speaking, the first claim is that causal sets can recover continuum Lorentzian manifolds; and the second claim is that no other proposal for a set of discrete data that conforms to our sense of "fundamental discreteness at the Planck scale" is known to be able to recover continuum Lorentzian manifolds. To support this second claim, we show, in particular, that an apparent alternative discrete data set to causal sets, viz., a certain sort of combinatorial Lorentzian simplicial complex, cannot recover General Relativistic spacetimes in the appropriately unique way; for it cannot discriminate between Minkowski spacetime and a spacetime with a certain sort of gravitational wave burst.

CORRESPONDING AUTHOR:

Jeremy Butterfield

Trinity College, Cambridge, University of Cambridge, CB2 1TQ, United Kingdom

jb56@cam.ac.uk

TO CITE THIS ARTICLE:

Butterfield, Jeremy, and Fay Dowker. 2024. "Recovering General Relativity from a Planck Scale Discrete Theory of Quantum Gravity." *Philosophy of Physics* 2(1): 5, 1–31. DOI: https://doi.org/10.31389/pop.17

DOI: 10.31389/pop.17

1 INTRODUCTION

Many workers tackling the problem of quantum gravity believe that the differentiable manifold structure of spacetime in General Relativity (GR) breaks down at the Planck scale and will be replaced by something else in the eventually successful, deeper theory. Some go further and believe that this "something else" has a discrete character. The purpose of this paper is not to present directly the case for discreteness at the Planck scale. Instead, we will argue that, whatever might be the ultimate underlying degrees of freedom in a theory of quantum gravity: *if* that theory recovers General Relativity as an approximation in certain physical circumstances, *and if* it is discrete at the Planck scale, *then* at the current stage of our knowledge, the approach must produce a *causal set*, i.e., a locally finite partial order (Bombelli et al. 1987). For, we will argue, whatever might be the nature of the ultimate underlying degrees of freedom, a causal set is, so far as we currently know, the only entity that can do the job of being approximate-able by a Lorentzian geometry, while also being discrete at the Planck scale.

We should stress at the outset that our argument is of limited scope in two ways that correspond respectively to our two main claims (to be stated in Section 2). The first is about the adequacy of causal sets to "do the job," i.e., to be approximate-able by a Lorentzian geometry (cf. our claim 1). As we will discuss, a more precise statement of this adequacy is the *Hauptvermutung* (main conjecture) of causal set theory. So we stress here that we are by no means claiming to prove the *Hauptvermutung*. It will be enough for us that there is a good deal of evidence for it.

The second limitation of scope is about our claiming that causal sets are the only known way to "do the job." This will be made precise in Section 2, as our claim 2. But here, we stress that we will not claim to refute any other quantum gravity research programs which claim some "fundamental discreteness." This compatibility with our claim 2 will turn on the facts that (i) claim 2 focuses on, not theories or research programs, but what we will call the *discrete physical data sets* that a theory proposes to be approximated by a continuous spacetime; and (ii) what we mean by the phrase "fundamental discreteness," or by similar phrases such as "recover from discrete data a continuum spacetime." We stress that we will understand such words in a somewhat logically strong way. Namely, we take them to exclude both (i) a theory that "fills in" discrete data (e.g., a graph whose edges each have a length) with pieces of flat space or spacetime (like an architect's geodesic dome); and (ii) theories that are *defined* by taking a continuum limit of a discrete theory—in a similar way to how elementary path-integrals are defined. So, a research program that does take such a continuum limit (such as causal dynamical triangulations) falls outside the scope of our argument.

But we submit that our understanding of these words is a natural one. In any case, given this understanding, we support our claim 2 that causal sets "are the only known way," by showing that an apparently natural alternative—which *does* fall under our sense of the phrase "recover from discrete data a unique spacetime"—stumbles. For, this alternative cannot satisfy a uniqueness requirement: an accusation which we prove with an example of a gravitational wave burst.

We should also stress at the outset that, even within the limited scope just announced, we set aside what one might call the "dynamical aspects" of supporting our two claims: we consider only the "kinematical" aspects. This restriction will be clearer in Section 2, once the claims are stated. But the main point lies in a difference from the usual distinction between kinematics and dynamics.

DOI: 10.31389/pop.17

Usually, a theory's kinematics states what its fundamental degrees of freedom are. But the kinematics is silent about how these degrees of freedom change over time (especially as encoded in equations of motion)—that is the job of dynamics. But we instead mean a slightly different distinction: one which allocates still less to kinematics, i.e., to the part of the theory we will focus on. For we will take no view—we will not need to take a view—about what are the degrees of freedom of the putative theory of quantum gravity we are concerned with (which we will call "X"). We will only have to assume that: (1), somehow or other, X recovers an adequate set of spacetimes so as to justify the phrase "recovery of general relativity"; and (2) X does so by giving, for each spacetime it recovers, data that are discrete at the Planck scale. (In Section 2.1, (1) and (2) will be made precise as Assumptions 1 and 2. For the moment, it suffices to say that "recovering a spacetime from discrete data" will be understood in the somewhat logically strong way just explained.)

So for us, "considering only kinematics" means considering only the discrete data that X provides: nevermind what X's fundamental degrees of freedom are, and nevermind by what means (e.g., some kind of coarse-graining) X extracts that discrete data from them. That will be enough for our argument in favor of causal sets to go through. For as just explained, this paper's argument will be, in short, that given our Assumptions, X's discrete data will have to be, or provide, causal sets. For the apparently natural alternative that we will articulate stumbles on our example of the gravitational wave burst.[1]

So the plan is as follows. In Section 2, we state our main argument, in terms of our two claims, labeled "claim 1" and "claim 2." Then in Section 3, we develop details about causal sets, so as to justify claim 1. In Section 4, we support claim 2, by showing how a certain sort of simplicial complex—that is Lorentzian, yet fundamentally discrete in our required sense that it does not "fill in" discrete data with pieces of flat spacetime—cannot recover a GR spacetime. This will be shown with an example of a gravitational wave burst, and we argue that the conclusion is more general than the particular example. In Section 5, we conclude.

2 THE ARGUMENT

We begin by stating and discussing two assumptions that our argument will make (Section 2.1). Then we state our argument in terms of our two main claims (Section 2.2). Finally in Section 2.3, we review the central role of causal order in general relativity. This review forms a backdrop to our defense of the two claims in the succeeding Sections (claim 1 in Section 3, and claim 2 in Section 4).

2.1 TWO ASSUMPTIONS

Let us make two assumptions about a theory of quantum gravity, which we call "X."

Assumption 1: In certain physical situations and at large scales, X recovers General Relativity (GR) as an approximation.

Assumption 2: X is physically discrete at the Planck scale.

---

1 Nor will this paper's argument need dynamical assumptions *within* the causal set program. For, since our topic is recovery of a whole 4-dimensional spacetime, we can take each causal set as an entirety, independent of how it might come to be dynamically. But we note that in fact the causal set program: (i) has a classical dynamics for causal sets with many merits (Rideout and Sorkin 2000); and (ii) as to quantum dynamics, espouses a path integral approach though has yet to find a fully satisfactory dynamics. Some aspects of (i) and (ii) are discussed in a longer version of this paper (Butterfield and Dowker 2021).

Neither of these assumptions is precise; and at our current stage of knowledge, they cannot be made precise. But most workers will have an intuitive feeling for what they mean; and we will now develop them in three extended comments. The first two are about Assumption 1; Assumption 2 comes in to play in the third comment. Note that these comments introduce some jargon we will use. Comment 2 introduces *grounding state* for a special sort of state within the theory X. Comment 3 introduces *Discrete Physical Data (DPD)* for what a grounding state supplies as the material to which a spacetime in GR is an approximation.

Comment 1: Recovering GR:—

With Assumption 1, we commit to General Relativity as the theory of continuum spacetime and gravity that theory X recovers. In particular, Assumption 1 implies local Lorentz invariance in the recovered theory, where by "local Lorentz invariance," we mean the approximate symmetry enjoyed by General Relativity (sometimes called "local Lorentz symmetry"). In short, this is the equivalence principle: that for any required degree of accuracy, the physics in a sufficiently small laboratory over a sufficiently short time is indistinguishable from the physics in Minkowski spacetime. In concrete terms, this means that theory X is assumed not to give rise to any of the rival "modified gravity" theories that are characterized by having more than one spacetime metric, or aether fields or even more brutal violations of local Lorentz invariance such as background foliations of spacetime. (A useful recent review of such theories from the perspective of their causal properties is (Carballo-Rubio et al. 2020).)

Of course, this is not to say that Assumption 1 means that GR is exactly true. GR is assumed to be an approximation, and theory X will predict deviations from GR which, we hope, will be observable to us, large, late-time observers, in spite of the tininess of the Planck scale. But we intend with Assumption 1 to exclude those deviations being (manifestations of) violations of local Lorentz invariance.

These remarks mean that we intend Assumption 1 to encompass the view of General Relativity as arising as an effective local field theory. More precisely, it encompasses such a treatment when that treatment takes all the terms of its effective Lagrangian to be appropriate powers of derivatives of the metric in order to respect local Lorentz invariance (with the cosmological constant problem having been solved somehow). (This kind of effective Lagrangian for General Relativity is common in such treatments: for example, cf. equations (21) and (46) of (Donoghue 1997) and equation (3.1) of (Burgess 2004).)

We also intend Assumption 1 to include matter degrees of freedom, as well as the spacetime—manifold and Lorentzian metric—degrees of freedom. We call a vacuum or non-vacuum solution of the Einstein equations in four dimensions a "solution of GR" or "GR solution" for short. We will say "spacetime of GR" or "GR spacetime" when we want to talk solely about the component of a GR solution that is the 4-dimensional Lorentzian manifold.

To develop Assumption 1 further, we invoke three familiar examples of the recovery of a physical theory from a more fundamental one.

General Relativity recovers Newtonian gravity as an approximation in certain, contingent circumstances. There is much to say—and much has been said—about the details, both technical and philosophical, of GR's recovery of Newtonian gravity, see (Havas 1964; Malament 1986; Malament 2012; Trautman 1965) as a sample from this rich literature. An important part of what this "recovery" entails is that all the data needed to recover, approximately, a non-relativistic spacetime picture with a particular Newtonian

gravitational potential satisfying Poisson's equation are present within a particular spacetime in GR. It is this sufficiency of the physical data in the underlying theory for the recovery of the approximating theory that we will appeal to in applying Assumption 1 to our argument.

Another example we wish to invoke in developing Assumption 1 is fluid mechanics as an approximation to molecular dynamics in certain situations and at large, i.e., macroscopic scales. This is an example pertinent to the combination of Assumption 1 and Assumption 2 in that it involves a continuum approximation. Again, there is much to say—and much has been said—about the details, both technical and philosophical, of continuum approximations to more fundamental, atomic theories of matter, see (Cercignani 1987; Chapman and Cowling 1991) as a sample from this enormous literature. Here, we want to emphasize just one point: continuum *approximations* where the number N of fundamental constituents is large but *not* infinite—as opposed to strict continuum *limits*—exist and make sense as emergent from the more basic, more fundamental theories. (Here we understand "emergent" as in (Butterfield 2011a; Butterfield 2011b).)

Since X is a quantum theory, the understanding of Assumption 1 gleaned from the two examples above must be augmented by considering how ordinary quantum mechanics can, in certain, circumstances be approximated by classical mechanics. Once again, there is much to say—and much has been said—about the details, both technical and philosophical, of quantum mechanics' recovery of classical mechanics; see (Bell 2004; Landsman 2017; Saunders et al. 2010) for a sample from this enormous literature. The ongoing disagreements about the foundations of quantum mechanics mean that there is, here, rather *more* to say than on the two previous examples. Indeed, there is currently no consensus on whether or in what way we can understand classical mechanics as having, in fact, been recovered from quantum mechanics. We will cut through this ongoing scientific debate by taking Assumption 1 to cover both the possibility that the recovery of GR by theory X is based on a better understanding of quantum theory than the one we have now, and the possibility that the recovery of GR by theory X relies upon similar interpretational rules of thumb—such as Copenhagen-esque splits and/or anthropocentric reasoning—to those in use today.

These three examples of recovery lead one to expect that each of the "physical situations" referred to in Assumption 1, in which a solution of GR is obtained as an approximation, may involve not only: (i) a certain state in the physical state-space of the theory X,[2] but also: (ii) a choice of a range of values of certain parameters (or ratios of parameters) defining an approximation scheme—an obvious example being a choice of a relevant physical "observation" scale and some concomitant coarse-graining scheme. Besides, we expect that some physical states in X will correspond to GR spacetimes with singularities, and that such a state—if theory X is to fulfill our expectations of quantum gravity—treats the singularity physically and predictively although the continuum, Lorentzian geometric description breaks down close to the singularity. So a single state in theory X may include both a "physical situation" away from the singularity that has a continuum approximation and a "physical situation"—close to and at the singularity—that does not. Thus we expect a variety of components in the definition of the "physical situations" referred to in Assumption 1. All of that being said, for ease of writing, we will sometimes refer to these "physical situations" as "states."

---

2 This state, for all we now know, may well *not* be a conventional quantum state such as a vector in—or density operator on—a physical Hilbert space. It may, for example, be one *co-event* from a collection of physically allowed co-events in a path integral-based framework (Sorkin 2007), or some other concept.

Comment 2: Some, but not all, GR solutions:—

Assumption 1 does not require that *all* the solutions allowed by the postulates of (some precise formulation of) General Relativity be recoverable from X as approximations. For example, spacetimes with closed causal curves might not be recovered by X.

But crucially, Assumption 2 implies that a GR spacetime can only be recovered from X if the characteristic distance over which the GR spacetime varies appreciably is everywhere much greater than the Planck length. This condition for spacetimes to be recoverable from X will be part of important conditions which we will introduce later; namely, the *discrete-continuum correspondences*, for causal sets in Section 3 and for simplicial complexes in Section 4.1. So, we need to be more precise about its meaning. We say that the characteristic distance over which a GR spacetime, (M, g), *varies appreciably is everywhere much greater than the Planck length* if, for every point p in (M, g) there exists a normal neighborhood, U_p of p that:

- **(i)** is covered by local inertial coordinates centred at p,
- **(ii)** contains an Alexandrov interval—a causal diamond—that contains p and has large volume in Planck units; and
- **(iii)** is such that everywhere in U_p, all the components of the Riemann tensor in the local inertial coordinates are small in Planck units; so we bound the value of the curvature itself, not just its variation, throughout the neighborhood U_p. This is necessary so as not to have physical structure on smaller than Planckian scales—for example, in anti-de Sitter space, if the magnitude of the Ricci scalar were not bounded above then the radius of curvature could be smaller than the Planck scale.

In orde is Newton's gravitational constant r to claim that GR is recovered from X, however, a very substantial collection of GR solutions must be recovered from X. This collection certainly includes all solutions that are currently known to be phenomenologically important, such as large portions of Minkowski spacetime, solutions containing black holes, expanding cosmologies, and solutions containing gravitational waves. Here is a very concrete example: There must be a state in X from which it is possible to glean data to reconstruct, approximately, a GR solution including gravitational waves propagating from a binary black hole merger event to an observation event that coordinates with the LIGO data for GW150914 (Abbott et al. 2016). To put the point contrapositively: If there is no state in theory X that gives data to which a GR solution with gravitational waves is a continuum approximation, then GR is not recoverable from X.

Let us refer to the states in X that give rise to data from which solutions of GR can be recovered as **grounding states**: since they are the ground or basis of recovering GR solutions. (We avoid: (i) "semiclassical states," as having perhaps, too specific connotations; and (ii) "basic states," as connoting that the states are fundamental within X—whereas we want to allow that the grounding states might well be specified using criteria that are *not* fundamental in X.) One can conceive of the whole collection of grounding states in X as the *continuum regime* or *continuum sector* of X. Comment 3 will fill out the idea of a grounding state, by developing Assumption 2.

To conclude Comments 1 and 2: we, of course, admit that Assumption 1 is not *compulsory*. There is however, to date, no convincing observational evidence for any theory of gravity apart from GR. And so we expect widespread, though not unanimous, assent to Assumption 1.

DOI: 10.31389/pop.17

Comment 3: From discrete data to one spacetime:—

For our argument, we take Assumptions 1 and 2 as implying that:

> *a grounding state of X contains, or gives rise to some Discrete Physical Data (DPD),*[3] *to which a GR spacetime is an—essentially unique—approximation.*

This Comment spells out the meaning of this italicized statement, in three remarks. The first is about DPD, the second about essential uniqueness; and the third is about discreteness and about the Planck scale.

A: The grounding state and the DPD: In what follows, we will be mostly interested in the data to which the GR spacetime is an approximation, and will refer to it as *the DPD* in the grounding state. Agreed: If the grounding state recovers a non-vacuum GR solution, then that grounding state will produce, in addition to the DPD to which the GR spacetime is an approximation, other discrete data that encode the matter distribution of the GR solution. So, if we need to refer to the additional matter data, we will do so explicitly, in addition to the DPD.

We do not specify how the grounding state contains, or produces, or gives rise to the DPD. For example, this data could be eigenvalues of or expectation values of certain operators on a Hilbert space—or not. Whatever theory X is, if it recovers GR as an approximation, there must come a stage at which we have in our hands a set of DPD and there is a procedure, in principle, for obtaining a GR spacetime, approximately and essentially uniquely from that set of DPD. This stage is *after* the quantal nature of the grounding state—small scale quantum fluctuations, interference between histories, multitudes of branches or what have you—has been "dealt with" by coarse-graining, collapse of the wave-function, anthropocentric reasoning, or what have you. In the end, if the theory is to work for *this* purpose—recovering a GR spacetime as an approximation in some state—then the data it produces must be data one can store as bits in a classical computer memory.

B: Essential uniqueness: By "essentially uniquely" we mean that if two continuum spacetimes (M, g) and (M', g') are both recoverable from the DPD provided by a grounding state in X, then they must be approximately isometric on scales large compared to the Planck scale.

There are two concerns that one might have about this requirement. Broadly speaking, (i) is mathematical, while (ii) is physical.

(i): The mathematical concern is how to define "approximately isometric": i.e., how to define a scale-dependent metric on the space of continuum spacetimes (Lorentzian geometries) that would allow us to say when two such are indistinguishable above a certain scale. Indeed, as we mentioned in Section 1: the *Hauptvermutung* ("main conjecture") of causal set theory is that there is such a definition, for which two spacetimes recoverable from a single appropriately chosen causal set are approximately isometric above the Planck scale. We shall make this more precise when discussing causal sets in Section 3.2. For the moment, it suffices to make two points:

---

3 An alternative phrase that we might have adopted instead of "DPD" is Riemann's phrase *Discrete Manifold* in his *Habilitationschrift* (Riemann, 1868; see Riemann 2016 for an English translation) where "manifold" means simply "multitude." But, owing to the development of Riemannian geometry (which, of course, Riemann's lecture began), the term "manifold" has come to so strongly connote the continuum that "discrete manifold" can sound to modern ears like an oxymoron.

DOI: 10.31389/pop.17

(1) There are various proposed metrics, and some results about them, e.g., (Bombelli 2000); though lack of space will prevent this paper's reviewing the evidence for the *Hauptvermutung*.

(2) It is interesting to compare the *Hauptvermutung* with the second example under Comment 1 above, about how a fluid-mechanical description of a sample of, e.g., water, is an approximation to a molecular-dynamical one, accurate (i.e., describing "an essentially unique sample") at distances much larger than the molecular. For, there is not just the obvious analogy between the two cases: that in both cases, one in principle needs to make some single mathematically precise statement of the approximation, and prove it. There is also a contrast. Namely, in the fluids case, there is (so far as we know) no research program explicitly devoted to finding such a statement and proof. Rather, there is a collection of various results about such approximations (i.e., for various features of various fluids, and for various degrees of accuracy); cf. (Cercignani 1987; Chapman and Cowling 1991). Agreed: since each of the frameworks, the fluid-mechanical and the molecular-dynamical, has been very well confirmed in many ways, this collection is enough to reassure one's scientific conscience that there is no problem about the molecular-dynamical framework securing essentially unique descriptions of a fluid. On the other hand, in *our* case, i.e., general relativity as an approximation to causal set theory, only one of the two frameworks, viz. general relativity, is well confirmed. So one's scientific conscience naturally prompts the effort to give a precise statement and proof of essential uniqueness. (Though, it might, in due course, turn out that causal set theory has enough scientific success that a collection of results about approximation lessens the demand for the *Hauptvermutung* as a single theorem).

So much, for the moment, about the *Hauptvermutung*. Summing up: For our overall argument, we can and will rely on a heuristic understanding of the idea of spacetimes being "approximately isometric above some scale." For, we expect that the phrase makes intuitive sense to most workers on quantum gravity—as shown, for example, by how they understand Kaluza-Klein theory.

(ii): One might worry that orthodox quantum theory (on at least some views of it) licenses superpositions of macroscopically very different mass distributions: which, in the context of GR, would seem to lead to superpositions of spacetimes that are *not* approximately isometric. Indeed, maybe the grounding states in theory X are superpositions of this kind. Our answer to this has already been given in remark A. Namely, our assumption that theory X recovers GR includes the assumption that such superpositions, if they are in fact present, can be dealt with, and have been dealt with by the stage at which we consider the DPD to which a spacetime in GR is a good approximation.

One might also ask: "what about dualities?" In recent decades, fundamental physics has formulated—and in some cases proven—various dualities between quantum theories formulated on spacetimes. A duality is, in effect, a surprising isomorphism between the state-spaces and algebras of quantities of the two theories, like a giant symmetry between the theories (De Haro and Butterfield 2017; De Haro and Butterfield 2019). Indeed, in the most celebrated case, viz., gauge-gravity duality in string theory, the two spacetimes concerned even differ in their dimension, one being the boundary of the other, which certainly counts as surprising! So if we countenance dualities such as these, what becomes of our "essential uniqueness" requirement?

We answer this concern in the same way as the concern about quantum superpositions: They are important concerns, *but they are not our concerns, here*. In GR, the physical world is just

DOI: 10.31389/pop.17

one spacetime (and matter distribution). If theory X enjoys a duality or dualities, and if it recovers GR, then this singleness of spacetime must be recovered, somehow, in X. We step in and take the DPD in our hands *after* this singleness has been established, however it has been established: i.e., established by whatever argument disposes satisfactorily of all but one of the dual descriptions. (It could, for example, be a more-or-less anthropocentric argument.)

C: Planck scale discreteness For definiteness let us define[4] the Planck time, Planck length, and Planck 4-volume, respectively:

$$t_p := \sqrt{\frac{8\pi G\hbar}{c^5}}, \quad l_p := \sqrt{\frac{8\pi G\hbar}{c^3}} \quad \text{and} \quad V_p := t_p l_p^3 = (8\pi G\hbar)^2 c^{-7}, \tag{1}$$

where G is Newton's gravitational constant, c is the speed of light, and $\hbar$ is the (reduced) Planck constant. In GR, V_p, t_p, and l_p are not independent: Knowing one of the three quantities above is sufficient to fix the other two.

At our current stage of knowledge of quantum gravity based on work in any existing approach, we cannot say more than that, in our putative successful theory X, the discreteness time scale, say, will be *of order* t_p. However, Assumptions 1 and 2 imply that, in our theory X that successfully recovers GR, the actual value of the discreteness scale should be calculable. In other words, theory X should tell us, via the discrete-continuum correspondence, the ratio of the fundamental discreteness time scale t_f, say, and t_p:

$$t_f = z t_p, \quad l_f = z l_p, \quad V_f = z^4 V_p, \tag{2}$$

where z is a number of order 1 that theory X should determine. One way z might be determined is by calculating the entropy of a black hole in theory X in terms of the fundamental discreteness length: $S_{BH} = \alpha \frac{A_{BH}}{l_f^2}$. Equating this to the known value $S_{BH} = 2\pi \frac{A_{BH}}{l_p^2}$ would then give the value of $z = \sqrt{\frac{\alpha}{2\pi}}$.

The condition that theory X is physically discrete at the Planck scale has meaning only for the grounding states. For it is only these states (by definition) that provide DPD that can be approximated by Lorentzian geometries. So, for a grounding state—and only for a grounding state—the DPD are approximated by a GR spacetime so that the *discrete-continuum correspondence* between the set of DPD and the spacetime does justice to the concept of Planck scale discreteness.

We do not specify exactly how the DPD are discrete. One might want to demand of theory X that the geometry of a GR spacetime in a region of finite spacetime volume be recovered from a finite amount of combinatorial data—as in a causal set—but our argument will not hinge on it. The physical data provided by theory X in a grounding state may be completely digital and combinatorial, or they may include real numbers such as edge lengths on a graph.

But we do take Assumption 2 to demand that the DPD and its approximating spacetime together embody the following two heuristics:

(a) the DPD do not give information about geometry on scales smaller than the Planck scale; and

(b) for the DPD to recover the continuum physics in and of a finite region in a GR spacetime requires only a finite amount of information.

As an example of (a): A lattice whose spacing is many orders of magnitude smaller than the Planck length cannot arise as DPD in a grounding state of X. As an example of (b):— The

4 We use $8\pi G$—the cosmologist's convention—rather than just G in the definitions.

DOI: 10.31389/pop.17

geometrical version of a simplicial complex, i.e., as a piecewise flat continuum manifold—a union of pieces of flat Minkowski space—cannot arise as DPD in X. For the flat "filling" of the interiors of the simplices is continuum information. Thus, on our understanding of "discrete," a simplicial complex is only discrete if it is a *combinatorial* complex, perhaps decorated with edge lengths, triangle areas, spins, etc. This exclusion (announced in Section 1) will be important in Section 4's defense of claim 2.

The scale of the discreteness of theory X is presented in this paper as discreteness in spacetime. To conclude this Remark about what we mean by "Planck scale discreteness," we briefly consider what it would mean for this discreteness to be expressed in terms of energy, because some workers conceive of the "Planck scale" in the first instance as very large (in energy) and not, primarily, as very small (in spacetime). Discreteness at the Planck scale means that there is a physical cutoff in spacetime that one might want to translate directly into a high frequency cutoff at the Planck frequency, $\nu_p := t_p^{-1}$, and, since frequency is a frame-dependent concept, one might think that a high frequency cutoff must imply a preferred frame in which to express that cutoff and thereby violate local Lorentz invariance. The relationship between spacetime discreteness, a high frequency/energy cutoff, and local Lorentz invariance is more subtle than this suggests, however, and the phrase "frequency is a frame-dependent concept" is, in this context, a little too glib. Frequency *is* a frame dependent concept *in the sense* that a free wave packet in Minkowski spacetime whose frequency is peaked around $\nu < \nu_p$ in one frame is peaked at a different frequency, ν', in another frame. Depending on the boost factor, ν' can be greater than ν_p. The point is that such a packet really has no associated covariant energy, except the rest mass if it is a massive particle/field. Thinking about this in covariant, spacetime terms, one sees that the crucial question for compatibility with discreteness is whether the *spacetime support* of the packet corresponds to a large enough subset of the DPD to recover all the contours of the packet. This *will* be the case if there exists a frame in which the typical frequency of the packet is small compared to ν_p and if the discreteness of the DPD is Planck scale, and Lorentz invariant. For then, in every frame *including the one in which the packet happens to have low frequency*, there will be enough DPD corresponding to the spacetime support to carry the information about the contours of the packet (Sorkin 2009, 1–3).

Lorentz invariant discreteness, then, is compatible with free wave packets of any frequency in any frame in infinite Minkowski spacetime; and consideration of these packets does not really probe the concept of a high energy cutoff (except in the trivial sense that the rest mass of an elementary particle should be less than the Planck mass—for, otherwise, the frequency, $(2\pi)^{-1}\sqrt{m^2+|\vec{p}|^2}$, is Planckian or higher in *every* frame). It is in considering *interactions* that the concept of "high frequency cutoff" becomes meaningful because there *are* covariant, frame-independent measures of the energy of an interaction. For example, the center of mass energy of a collision of two particles is Lorentz invariant. Again, translating this into a covariant spacetime picture, the question becomes: Are there enough DPD corresponding to the interaction region—the intersection of the spacetime supports of the two wave packets, say—to support the physical details of the interaction? Now, if the centre of mass energy of a two particle collision is higher than the Planck energy and the spacetime volume of the interaction region is smaller than the Planck volume, there will be very few or even no corresponding DPD to support the interaction, effectively cutting off such high energy interactions.[5] In this way, a Lorentz invariant high energy cutoff arises from Lorentz invariant spacetime discreteness.

5 In the context of causal set theory, Sorkin has pointed out that this is a potential signature of discreteness: such packets propagating on a causal set background could pass through each other without interaction (Sorkin 2009).

DOI: 10.31389/pop.17

2.2 TWO CLAIMS

We now state our main argument in terms of two claims, "Claim 1" and "Claim 2." We will make explicit two limitations of our argument before turning to support the Claims: Claim 1 in Section 3, and Claim 2 in Section 4.

> **Claim 1:** *A causal set—a locally finite partial order—is a set of Discrete Physical Data in our sense (DPD) that, taken as being discrete on the Planck scale, can recover a GR spacetime as a continuum approximation.*
>
> **Claim 2:** *There is in the current literature no other proposal for a set of Discrete Physical Data (DPD) in our sense, discrete at the Planck scale, that can recover a GR spacetime as a continuum approximation.*

Thus, Claims 1 and 2, as clarified by Assumptions 1 and 2 in Section 2.1, express the argument we announced in Section 1. Namely, whatever the ultimate underlying degrees of freedom in a theory X of quantum gravity, *if* X is discrete at the Planck scale (cf. Comment 3 in Section 2.1), and recovers General Relativity as an approximation in certain circumstances (cf. Comments 1 and 2 in Section 2.1), *then* according to our current state of knowledge, X needs to produce a causal set.

But before we support the Claims in Sections 3 and 4, we should stress that our argument has two important limitations of scope: (as announced in Section 1).

A: We only consider kinematics: We stress that Claim 1 does not say that *every* causal set recovers a GR spacetime. Only some do, and which ones do will be spelled out in Section 3; along with the *Hauptvermutung*'*s* idea of recovering an essentially unique GR spacetime. For, the claim is made at what one might call a "kinematical level."[6] We submit that at the current stage of our knowledge, no theory of quantum gravity, neither causal set theory nor any other theory, explains dynamically why grounding states with DPD that can recover GR spacetimes occur. In the case of causal set theory, finding a quantum dynamics that will do this job is a major outstanding task. It is addressed with a path-integral, or sum-over-histories, in which the histories summed over are causal sets. For reasons of space, we cannot give details. But the longer version of this paper (Butterfield and Dowker 2021) does so, and argues that our Claims 1 and 2 will not be undermined by a quantum dynamics that takes the form of a sum-over-histories in which the histories are discrete: either causal sets (Claim 1) or some other discrete entity (Claim 2).

B: We are only concerned with theories that produce discrete physical data in our sense: Here as announced in Section 1, we understand "discrete data" in a logically strong sense—though, we claim, a natural one.

We, of course, admit that there are theories, and frameworks for theories, of quantum gravity, with many merits, that *use* discreteness, but in such a way as to *not* fall in the scope of our argument. This can happen for one or both of two reasons. (i): The first reason is that the theory uses discreteness either as a regulator to be taken away in a strict continuum limit, or more generally to define an approximation to a continuum theory. (ii): The second reason is that, even away from the so-called continuum limit, at a stage at which the regulator is still finite (non-zero), the spacetimes considered are actually continuous: for example, because they are *geometric* simplicial complexes—piecewise flat Lorentzian manifolds—of the kind we discussed (and set aside) in Remark C of Comment 3 in Section 2.1.

---

6 Recall from Section 1 that for us, the word "kinematics" is logically weak: the "kinematical level" need not encode the fundamental degrees of freedom.

Both these points (i) and (ii) are well illustrated by the construction of the path integral in elementary non-relativistic quantum mechanics. Thus recall that in the process of defining the path integral, we consider zigzag (i.e., piecewise straight) paths in spacetime with successively shorter time divisions, so that at each finite stage of the construction all the paths considered ("skeletonizations") are everywhere continuous but not differentiable at their corners. Taking an appropriate limit of infinitely many time divisions, we get the path integral, with all its new mathematical structures and physical ideas. (So this construction exemplifies the idea of a singular limit, in the—non-technical—sense of a limit in which some facts or structures, mathematical and/or physical, exist at the limit, but not before: as mentioned in (Butterfield 2011a; Butterfield 2011b).) Though one might casually say that the skeletonizations, being zigzags, "look discrete," that is obviously loose talk, since they are continuous paths. Moreover, these zigzag paths are just stages in the construction, and none of them (no matter how fine, i.e., how minuscule the time duration of their straight segments) give the physical interpretation of the path integral eventually constructed. Thus, our point is that, for all its novelty and richness—its glories!—the use of a form of discreteness in defining the path integral in quantum mechanics does not in itself produce discrete physics. It is wrong to say that the path integral in quantum mechanics is discrete.

Similarly, here: Theories such as Causal Dynamical Triangulations (CDT) (Ambjorn and Loll 1998; Ambjorn et al. 2000; Ambjorn et al. 2001a; Ambjorn et al. 2001b), and Lorentzian Quantum Regge Calculus (Williams 1986) are examples of quantum gravity approaches that use piecewise flat Lorentzian manifolds. CDT uses a path integral approach in which the piecewise flat Lorentzian manifolds play the role of the skeletonized trajectories in the quantum mechanics example, and the full CDT theory is defined in the limit in which the lengths of the edges of the simplices in the (Euclideanized) piecewise flat manifold are taken strictly to zero. Such use of piecewise flat Lorentzian manifolds in quantum gravity theories and frameworks is *not* targeted by our argument, in particular by Claim 2. Nor does our argument directly target the use of combinatorial Lorentzian simplicial complexes—*if* the discreteness scale is not physical, but is a regulator that is taken strictly to zero in the definition of the quantum gravity theory. In short: the coherence, and merits, of these two theories illustrate our argument's limited scope.

We also agree that there are several quantum gravity approaches—in addition to CDT and Quantum Regge Calculus, which we have mentioned—that have a discrete "flavor," including: Energetic Causal Sets (Cortês and Smolin 2014), Entropic Gravity (Verlinde 2011), Group Field Theory (Oriti 2009), Holography (Banks 2020), Loop Quantum Gravity (Rovelli 2008), Spin Foams (Rovelli and Vidotto 2014), Quantum Graphity (Konopka et al. 2008), Spacetime Code (Finkelstein 1969), Thermodynamic Gravity (Padmanabhan 2014), and the Wolfram model (Gorard 2020), among others. So, agreed: It would be a good project (though a large one—beyond this paper) to assess these "discretely-flavored" quantum gravity approaches as to whether they only employ discreteness à *la* (i) and/or (ii) above, or whether they aim to recover a GR spacetime, in Section 2.1's sense, by means of Planck scale discrete physical data in our sense—and thereby fall within the scope of our argument. Thus, our point is that any quantum gravity approach within our scope is in danger of stumbling in the way that Section 4 will show combinatorial Lorentzian simplicial complexes do.

2.3 LORENTZIAN GEOMETRY

Our arguments in this paper depend on the all-important discrete-continuum correspondence, i.e., the correspondence between a set of DPD and the continuum

spacetime that approximates it. In analyzing and assessing this correspondence, it will be important that, in a Lorentzian geometry, the relation between metrical and topological structures is different than in Euclidean geometry. So, in this Section, we briefly discuss this.

A Lorentzian manifold is not isotropic on small scales because Minkowski space itself, though flat, is not isotropic: It has preferred directions, the *null* directions. Moreover, the points along a null geodesic are totally ordered in the spacetime causal order:[7] a point a is to the past of point b on a null geodesic in Minkowski spacetime even though the distance between them is zero. This means that Lorentzian geometry, if it is not actually nonlocal, teeters on the very edge of being nonlocal. This verging-on-nonlocal character can be illustrated by considering all the points less than some geodesic distance from a point p in four-dimensional Minkowski spacetime. All the points along the future and past light cones from p are zero distance from p. And the locus of points that are one Planck unit of geodesic proper time, say, to the past or to the future of p, is a double-sheeted three-dimensional spacelike hyperboloid of infinite volume that asymptotes to the past and future light cones from p. Thus, the set of points that are physically, geometrically close to p is very nonlocal. Besides, the locus of points that are one Planck unit of spacelike geodesic distance from p is a three-dimensional single-sheeted hyperboloid that asymptotes to both the past and future light cones. So the analogue of a Euclidean geodesic ball around a point p is an *unbounded* four-dimensional region of infinite spacetime volume between these hyperboloids, a neighborhood of the past and future light cones from p: as shown in Figure 1.

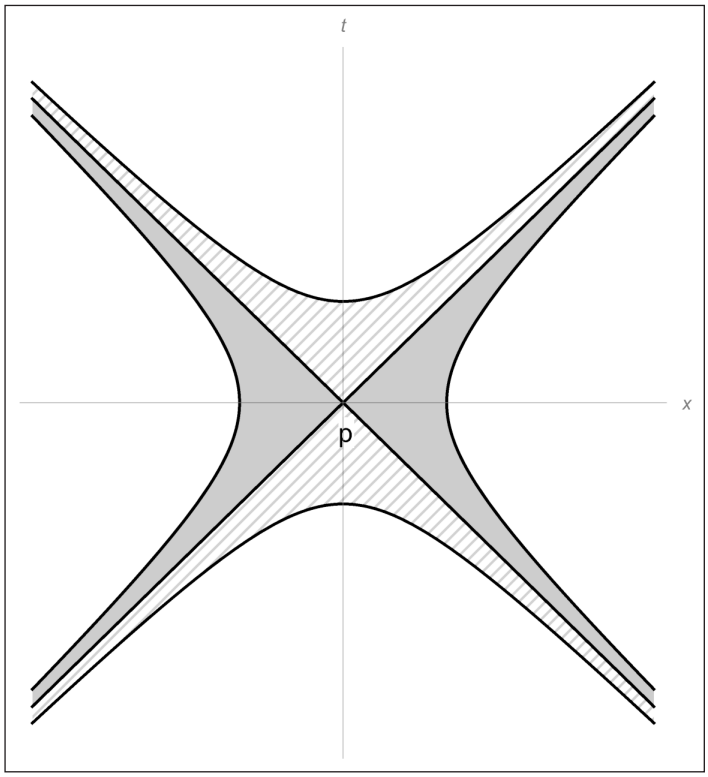


Figure 1 The figure shows the $t - x$ plane of four-dimensional Minkowski spacetime. The figure should be imagined as rotated about the t-axis into the suppressed y and z dimensions. The point p is at the origin and the light cones of p are shown. The striped shaded region is the locus of points in the causal past or causal future of p that are within one unit of geodesic proper time of p. The solid shaded region is the locus of points spacelike to p that are within one unit of geodesic proper spatial distance of p. Both shaded regions are unbounded and of infinite spacetime volume and their union is the (analogue of the) geodesic ball around p. The boundaries of this ball are hyperboloids.

7 This use of the word "causal" in GR does not imply "causation" (whatever that might mean!): to say x is to the causal past of y is merely to say that x is before y and y is after x.

DOI: 10.31389/pop.17

The nonlocal character of these Lorentzian "geodesic balls" immediately calls into question the relationship between the metric and the topology of Lorentzian geometries. In Euclidean space, the geometry and the topology are compatible in the sense that the set of open geodesic balls is a base for the manifold topology; and this is also the case for Riemannian manifolds generally. In Minkowski spacetime, the nonlocality of the geodesic "balls" means they cannot form a base for the topology. Remarkably, the compatibility between geometry and topology nevertheless holds for Minkowski spacetime: not via geodesic balls at points but, instead, via Alexandrov intervals—"causal diamonds"—between *pairs* of points. The Alexandrov interval between points x and y, where $x \in I^-(y)$ (i.e., there is a past directed timelike curve from y to x), is the intersection of the chronological future $I^+(x)$ of x (i.e., the set of points to which there is a future directed timelike curve from x) and the chronological past $I^-(y)$ of y. The Alexandrov topology is the smallest topology containing these Alexandrov intervals, and the set of Alexandrov intervals is a base for the Alexandrov topology. For Minkowski spacetime, the Alexandrov topology is equal to the manifold topology. This result continues to hold for other Lorentzian geometries, but—in contrast to the Riemannian case—not for all of them: The Alexandrov topology equals the manifold topology if and only if the Lorentzian geometry is strongly causal (Kronheimer and Penrose 1967, 487).[8] Note the difference between the Riemannian and Lorentzian cases. In the Riemannian case, the metric determines the topology locally and via the concept of geodesic distance. In the Lorentzian case the metric determines the topology bi-locally and via the concept of causal order.[9]

Remarkably, the causal order of a strongly causal Lorentzian manifold determines much more than just its topology. It also determines the differentiable structure in dimensions greater than 2, thanks to a 1966 theorem of Hawking (Hawking 2014; Hawking et al. 1976). With the tangent space in hand, the metric can then be determined in terms of one unknown fiducial component, which component can be traded in for an overall conformal factor (Hawking and Ellis 1973, 60–61). Thus, for a strongly causal spacetime of dimension greater than 2, the causal order determines the metric up to an overall conformal factor. This result was later extended by Malament to cover Lorentzian manifolds satisfying a strictly weaker causality condition (Malament 1977). This final result, the Kronheimer-Penrose-Hawking-Malament (KPHM) theorem, states that the causal order of a distinguishing[10] spacetime of dimension greater than 2 determines its topology, differentiable structure, and metric, up to a conformal factor. It is the strongest statement of how much of the structure of a Lorentzian manifold is determined by the causal order.[11] In short, the causal order captures almost the full geometry, and all that is missing is a local scale. The KPHM theorem can be encapsulated in the slogan: "Order + Volume = Lorentzian Geometry." There is no Riemannian analogue.

---

8 Strong causality is a weaker causality condition than global hyperbolicity but a stronger condition than being merely causal: See (Minguzzi and Sanchez 2008) for a review of the rungs of the "causal ladder."

9 Recall that the causal order is the totality of relations $x \in J^-(y)$ (which means there is a future directed causal curve from x to y). For any Lorentzian spacetime, the causal order determines the chronological order and thence the Alexandrov topology (Kronheimer and Penrose 1967, 485–487).

10 A past/future distinguishing spacetime is one in which any two distinct points have distinct chronological pasts/futures. A distinguishing spacetime is both past and future distinguishing.

11 Here, "the strongest statement" means that any weaker causality condition renders the theorem false.

DOI: 10.31389/pop.17

Causal order is not just at the heart of the mathematics of Lorentzian geometry in GR, it is also at the heart of the physics of GR. Carter-Penrose diagrams and other spacetime diagrams representing the causal structure of spacetimes abound in GR textbooks. The epitome of GR is a black hole, and the event horizon of a black hole is defined in causal terms: The boundary of the closure of the causal past of future null infinity. From this concept of event horizon flows much of the physics of black holes. For example, the Second Law of black hole mechanics, Hawking's Area Theorem, is a result about the event horizon (Hawking 1971); and assuming an event horizon forms, it can be proved that the Kerr family of solutions is very likely to be the physical description of the final state of any black hole in GR (either with the assumption the final state is axisymmetric (Carter 1971; Robinson 1975) or dropping the assumption of axisymmetry but assuming that the metric is analytic (Hawking and Ellis 1973, 323–347; Chruściel et al. 2012, 20–24), Theorem 3.3 of (Chruściel et al. 2012).

"Gravity is geometry" is a common slogan to sum up the lesson of GR. But a *Riemannian* curved spacetime would not give us the physics of GR. Perhaps we should always say "Gravity is Lorentzian geometry" to remind ourselves of what GR is really teaching us.

3 CLAIM 1: CAUSAL SETS RECOVER GR SPACETIMES

In this Section, we will support Claim 1 by referring to existing results in the literature on causal sets. Indeed, the central points we make are present already in the founding papers of causal set theory (Bombelli et al. 1987; Myrheim 1978; 't Hooft 1979), though more evidence for them has accrued since then.

So, the Claim we wish to defend is: Given any distinguishing GR spacetime, (M, g), whose geometry varies appreciably only on scales much larger than the Planck scale, there is a causal set C that is:

(i) well approximated by (M, g), and

(ii) such that the correspondence between C and (M, g) is discrete at the Planck-scale.

More precisely, there are many causal sets that are very likely to satisfy (i) and (ii). Here, of course, "very likely" signals, not a probability, but our having good theoretical reasons to believe this. This use of "very likely" echoes the distinction we made in Section 1 and Remark A in Section 2.2, between kinematical and dynamical aspects of causal set theory. That is, we take Claim 1 as a kinematical claim, about the approximation relation between a GR spacetime and a causal set. So, our argument does not depend on assumptions or results about the dynamics of causal sets.

As we will explain, there *are* issues to address in causal sets' satisfying (i) and (ii); issues that have been recognized since the inception of the causal set approach to quantum gravity. Describing these issues, and the progress made so far in addressing them, will be the main theme of this Section. We begin with a warning that the discrete-continuum correspondence—the approximation relation—is not as straightforward as one might think (Section 3.1). In Section 3.2, we give the all-important formulation of the discrete-continuum correspondence for causal sets (as mentioned in Section 1: often discussed under the label "*Hauptvermutung*," i.e., "main conjecture"), and briefly point to various sorts of evidence in its favor.

3.1 APPROXIMATING THE DISCRETE BY THE CONTINUOUS

3.1.1 A variety of analogies

In Comment 1 in Section 2.1, we gave fluid mechanics as an example of an emergent continuum from a discrete substructure. Bulk matter provides other examples, including metals and other crystals that are formed from regular arrays of discrete atoms. Everyday life provides yet more examples. A film reel of a sequence of finitely many still images is experienced by the watcher as continuous motion. A geodesic dome formed of many flat triangles appears, when viewed from far enough away, as a smooth surface.

These examples and experiences give rise to intuitions about emergent continua that may tempt us into thinking that it is straightforward to recover a GR spacetime from a set of DPD that is discrete at the Planck scale. "Straightforward" because such examples abound and because the Planck scale is so very very small—not only compared to us humans but also compared to the scale of physics we have so far probed in any particle accelerator—so that *of course* a discrete manifold can appear as a continuous spacetime to us.

We endorse the idea that a continuum spacetime can emerge from a discrete underpinning. But, we *caution* against the thoughts that it will be straightforward, and that we can take *any* of the concrete examples of emergent continua mentioned above as a guide to the relation between the DPD and the GR spacetime. For, we will argue, in this Section on causal sets and the next on simplicial complexes, that the approximate local Lorentz invariance of GR and our assumption that the DPD are discrete at the Planck scale mean that the discrete-continuum correspondence by which theory X recovers GR needs to be very carefully considered. In particular, we will see in Section 4 that a natural discrete-continuum correspondence for combinatorial simplicial complexes *fails*.

3.1.2 Agnosticism about the existence of structure "below" the Planck scale

Recall how, early in our mathematical education, points are introduced as being the limit of ever smaller regions of space or spacetime. The teacher or textbook urges on us that there is "surely" no lower bound to the diameter or volume of these regions, and we come to accept points as extensionless in the above sense. (Hence the label, in pedagogy about spacetime, "point-events.") Assumption 2, of Planck scale physical discreteness, is that what the teacher or textbook urged is wrong in quantum gravity. For, the Assumption denies the physical meaningfulness of any continuum topological or geometrical concepts—such as dimension, length, and volume—below the Planck scale in a GR spacetime.

But, this is not to say that there cannot be any physical discrete structure *in addition* to the DPD and to data that recover any matter degrees of freedom in the GR solution. One might be tempted to call such extra structure *sub-Planckian*; but that could be misleading since such structure really has no scale as such, because it does not contribute to the recovery of the approximating Lorentzian geometry (M, g). To illustrate this idea of extra structure, consider the case of causal sets. We are about to argue (in Section 3.2) that, using this paper's jargon of grounding states and DPD, a causal set can be the Planck scale DPD in a grounding state that recovers a GR spacetime. The elements of such a causal-set-as-DPD, C, are to be considered as *atoms of spacetime* in the sense of the original Greek word *atomos*, i.e., indivisible. That is: an element of C has no internal structure and cannot be analyzed ("divided") into parts. But that is not to say that there can be no physical data at

DOI: 10.31389/pop.17

all in addition to C: It is just that the extra data will not be data about the approximating Lorentzian geometry.

Now, if there *are* further data as well as the DPD in the grounding state, then who knows what they are, and who knows how the grounding state contains, or implies, or gives rise to[12] the DPD themselves. A certain amount of some sort of coarse-graining will have had to have been done to the full discrete data provided by the grounding state in order to arrive at the DPD alone. In the specific case of the causal-set-as-DPD, C, from which the GR spacetime (M, g) is recovered, one possibility is that the grounding state might provide a causal set C' that is not itself faithfully embeddable in a GR spacetime, but that, on decimation of C' by a random deletion of elements chosen with fixed probability, produces the DPD-set C that does faithfully embed in (M, g) at Planckian density.

3.2 THE DISCRETE-CONTINUUM CORRESPONDENCE FOR CAUSAL SETS

In Section 2.3, we reviewed key aspects of Lorentzian manifolds, the GR spacetimes that must be recovered by grounding states in theory X. Consider, now, a *particular* four-dimensional, distinguishing GR spacetime (M, g) that must be recovered in quantum gravity theory X. In this Section we will state the discrete-continuum correspondence—or approximation relation— between a causal set as Planck scale DPD[13] and (M, g) (Bombelli et al. 1987). But first, for completeness' sake, we give the formal definition of a causal set.

Definition: A *causal set* is a set C with a relation $\prec$, called "precedes," on C that satisfies the following conditions:

(1) if $x \prec y$ and $y \prec z$ then $x \prec z \; \forall x, y, z \in C$ (transitivity);

(2) $x \not\prec x \; \forall x \in C$ (acyclicity);

(3) for any pair of elements x and z of C such that $x \prec z$, the set $\{y \,|\, x \prec y \prec z\}$ is finite (local finiteness).

Of these axioms, (1) and (2) together say that $(C, \prec)$ is a partially ordered set, or poset for short. A poset is sometimes called simply an "order." The third axiom of local finiteness expresses the discreteness of the causal set. Note that the continuum spacetime causal order is reflexive: $p \in J^-(p), \forall p \in M$. So it is the relation "precedes or equals" in the causal set, with notation $\preceq$, that coordinates with the continuum spacetime causal order in the discrete-continuum correspondence below. An alternative convention—"the reflexive convention"—in the literature on partial orders is to give the definition in terms of $\preceq$.

Phrases that can be taken as synonymous with causal set are "discrete order," "locally finite partial order" and "transitive directed acyclic graph."

Now let *the discrete-continuum correspondence for causal sets*, which we will call **DCC-C**, be as follows:

---

12 Here, the phrase "contains, or implies, or gives rise to" deliberately echoes the italicized statement at the start of Comment 3, Section 2.1.

13 When we say "the set of DPD is a causal set," we mean, strictly speaking, an order-isomorphism equivalence class of causal sets. However, we adopt the standard practice whereby we think of and talk about the set of DPD as a single representative causal set rather an equivalence class of causal sets; and we then compensate for this by making sure that the mathematical identity of, and any labels on, the elements of the causal set are ignored in the causal set's role as DPD, so that only the *order relation*, $\prec$, between the elements and the *number* of the elements have meaning as physical data.

DOI: 10.31389/pop.17

A causal set, $(C, \prec)$, recovers the GR Spacetime (M, g) if there exists a *Planck scale faithful embedding* (Bombelli et al. 1987): that is, an injective map $\phi : C \hookrightarrow M$ satisfying the following conditions.

(i) (Planck-scale uniform): The number of causal set elements embedded in any sufficiently large, physically nice region of M is approximately equal to the spacetime volume of the region in fundamental volume units $V_f = z^4 V_p$ (see Remark C of Comment 3 in Section 2.1). "Physically nice" means that the region contains large, approximately flat Alexandrov intervals and it has no Planck scale features such as very wiggly boundaries.

(ii) (Order-preserving): Elements x and y of C are ordered, $x \prec y$, if and only $\phi(x) \in J^-(\phi(y))$.

(iii) The characteristic distance over which the continuous geometry (M, g) varies appreciably is everywhere much greater than the Planck length/time. (As mentioned in Comment 2 in Section 2.1, this is a condition on any (M, g) recovered by a Planck scale discrete theory X. We state it here as a reminder, and because it is in the original definition of faithful embedding for causal sets (Bombelli et al., 1987.)

Thus the DCC-C claims that when a faithful embedding of $(C, \prec)$ into (M, g) exists, all the geometric information of (M, g) on scales large compared to the Planck scale is encoded in $(C, \prec)$. Clearly, the concept of faithful embedding derives directly from Section 2.3's Kronheimer-Penrose-Hawking-Malament theorem. If the macroscopic causal order of (M, g) is recovered from the microscopic order of $(C, \prec)$ and the volume measure of the continuum, i.e., (M, g)'s one remaining "degree of freedom," is recovered from the counting measure on C, then, the theorem suggests, that is sufficient to recover the geometry of (M, g).

The faithful embedding is the all-important concept of the discrete-continuum correspondence for causal sets (DCC-C): without it, the claim that causal sets "do the job" cannot be assessed. We emphasize that this correspondence via faithful embedding means that the causal set is not merely discrete but clearly, meaningfully, discrete *at the Planck scale*. This condition on DPD—that the discreteness has a physical scale, which scale is the Planck scale—is meaningless at the level of the DPD *in itself*. It only acquires meaning in the context of the discrete-continuum correspondence, as emphasized in Remark C of Comment 3 in Section 2.1.

Note also that this DCC-C implies there is a whole class of causal sets *each* of which recovers (M, g)—just as there are many microscopically distinct molecular states of a gas each of which recovers the same continuum fluid state.

In Comment 3 in Section 2.1 (Remark B), we noted that, in order for theory X to recover GR, the set of DPD must recover an *essentially unique* large-scale continuum spacetime. In the context of causal sets, we require that if $(C, \prec)$ faithfully embeds in two GR spacetimes (M_1, g_1) and (M_2, g_2) then (M_1, g_1) and (M_2, g_2) must be approximately isometric in the following way. Let $\phi_1 : C \hookrightarrow M_1$ and $\phi_2 : C \hookrightarrow M_2$ be the two faithful embeddings. Then there exists a C-preserving diffeomorphism $f : M_1 \mapsto M_2$ that is an approximate isometry, where C-preserving means that $f \circ \phi_1 = \phi_2$ (Bombelli et al. 1987).

This fuller statement of the *Hauptvermutung* says, essentially, that the DCC-C as stated above *works*: all the topological, differentiable and metrical structure of (M, g) at scales large compared to the Planck scale is indeed encoded in any $(C, \prec)$ that is faithfully embeddable in (M, g). To state it another way, if the *Hauptvermutung* fails then so does the DCC-C.

DOI: 10.31389/pop.17

For reasons of space, we cannot go into details of the accumulated evidence for the *Hauptvermutung*: these are contained in the longer version of this paper (Butterfield and Dowker 2021) and also discussed in (Butterfield 2024). Here, it must suffice to say that:

(i) One main approach to providing evidence for the Hauptvermutung is to use a *Poisson sprinkling*: a random selection of points of a given Lorentzian manifold (M, g) that satisfies (iii) of DCC-C, at some density. These points are then endowed with the order induced by the causal order of (M, g), so as to produce a random causal set that is faithfully embedded in (M, g);

(ii) Many studies using this approach have shown that a lot of information about the topology and geometry of (M, g), e.g., about its dimension, causal structure, and curvature, can be extracted from *just* the discrete data provided by such a random, faithfully embedded causal set.

4 CLAIM 2: NO OTHER PROPOSAL IS KNOWN TO SUCCEED

In this Section, we turn to supporting Claim 2 from Section 2.2: that causal sets are the only proposal, known so far, for DPD that (i) recovers GR spacetimes (cf. Comments 1 and 2 in Section 2.1) and (ii) is fundamentally discrete at Planck scale in our sense (i.e., as announced in Section 1's second limitation of scope, and explained in Comments 1 and 2 of Section 2.1 and in Remark B of Section 2.2).

We make the claim on the basis of an argument that an alternative, namely combinatorial simplicial complexes which include lattices, that might seem to be counterexamples to our claim, do *not* do the job. As discussed in Remark B of Section 2.2, we concede that these may not be the only alternatives; but they are natural, and it would be a large project to assess the other quantum gravity programs to see whether they use other alternatives.

We present this argument in Sections 4.1 and 4.2. In short, we will argue that (a natural formulation of) the *Hauptvermutung* for a combinatorial Lorentzian simplicial complex *fails*. That is, a natural discrete-continuum correspondence for such a complex fails to recover an essentially unique GR spacetime.

Before giving the argument, we first clarify what we mean by a combinatorial simplicial complex. As already mentioned in Remark C of Comment 3 in Section 2.1 and Remark B in Section 2.2: there are two conceptions of a simplicial complex, which we call the *geometric simplicial complex* and the *combinatorial simplicial complex*. The former is (in dimension d) a union of k-simplices, for $k = 0, 1, 2, \ldots d$, which are pieces of k-dimensional flat space (either Euclidean space if this is a Riemannian geometric simplicial complex, or Minkowski spacetime if this is a Lorentzian geometric simplicial complex) satisfying certain conditions. If this geometric simplicial complex satisfies certain further conditions, it will be a piecewise flat (PF) d-dimensional manifold. On the other hand, a combinatorial simplicial complex is discrete: It is a collection of vertices, pairs-of-vertices, triples-of-vertices *etc*. satisfying a set of conditions, but without any flat space "in-filling." So, for us, *geometrical* simplicial complexes do not satisfy our requirement of being discrete data sets.

Since this is an important point, let us emphasize it. Suppose one did allow geometrical simplicial complexes as DPD. Consider the particular case in which the Lorentzian geometry to be recovered is 4-dimensional Minkowski spacetime itself, and the DPD is a *geometric* simplicial complex that is some particular triangulation of Minkowski spacetime. The DPD in this case, then, simply *is* Minkowski spacetime. But Minkowski

DOI: 10.31389/pop.17

spacetime is *not* discrete, let alone discrete at the Planck scale. More generally: A piece of a continuum flat spacetime is not the absence of data; it is substantial, it is a continuum with all its topological, differentiable, metrical, and causal structure. It contains physical data of the sort we are excluding with our Assumption 2 (Section 2.1). However, a combinatorial simplicial complex—with or without edge lengths, triangle areas, or other decorations—*is* discrete. So, when we refer to a simplicial complex, we will mean a combinatorial simplicial complex (unless specified otherwise).

Thus, we stress again (cf. Remark B in Section 2.2) that our argument is no obstacle to a *geometric* Lorentzian simplicial complex being considered, and used to approximate a continuum Lorentzian geometry.[14] Indeed, the triangulation of Minkowski spacetime just considered is simply a case in which the approximation of the continuum spacetime is *perfect*.

4.1 THE COMBINATORIAL LORENTZIAN SIMPLICIAL COMPLEX DOES NOT DO THE JOB

Consider a four-dimensional Euclidean geometric simplicial complex, $\overline{S}$, that is also a piecewise flat Riemannian manifold. Let S be the *combinatorial* simplicial complex corresponding to $\overline{S}$. Let each edge (1-simplex) of S be decorated by a *Lorentzian edge-length*, where a Lorentzian edge-length is a timelike, null (zero) or spacelike length, and also a future pointing direction in the case of timelike or null edges. Denote these decorations collectively by dec. We call (S, dec) a *Combinatorial Lorentzian Regge Complex* (CLRC).[15]

Now let the *discrete-continuum correspondence*—which we will call: **DCC-CLRC**—be as follows. (Note the contrast with Section 3.2's discrete-continuum correspondence for causal sets, DCC-C.)

A CLRC, (S, dec), recovers the GR Spacetime (M, g) if the following conditions hold.

(i) There is a piecewise flat manifold $\overline{S}$ that is a triangulation of M such that S is the combinatorial simplicial complex corresponding to $\overline{S}$.

(ii) There is a homeomorphism $f : \overline{S} \to M$ such that the f-image of each edge of $\overline{S}$ is a—spacelike, timelike or null—geodesic in (M, g). f induces an embedding of the vertices of S in (M, g).

(iii) The spacetime length and future pointing direction (if appropriate) of the f-image of each edge of $\overline{S}$ in (M, g) equals the Lorentzian edge-length of the corresponding edge of (S, dec) up to some tolerance.

(iv) The edge-lengths are no greater that a few Planck units.

(v) The characteristic distance over which the continuous geometry (M, g) varies appreciably is everywhere much greater than the Planck length/time.

14 Note the direction of the approximation relation here.

15 We call this a Regge Complex because, in Regge calculus, the edges of a geometric simplicial complex are decorated by edge-lengths and those edge-lengths determine the geometry of each simplex of the complex if the interior metric is flat, as is assumed in Regge calculus (Regge 1961; Sorkin 1974). In Regge calculus there are consistency conditions on the Lorentzian edge-lengths of a simplex if the interior flat metric is to have a Lorentzian signature (Sorkin 1974). We do not impose signature conditions on the Lorentzian edge-lengths, *dec*, at this stage of defining a CLRC. Signature conditions will be (implicitly) imposed within the upcoming proposal for a discrete-continuum correspondence between a CLRC and a GR spacetime.

DOI: 10.31389/pop.17

We admit that this definition, DCC-CLRC, is not the only possible definition of a Planck scale discrete-continuum correspondence for CLRCs. But, it is reasonable, and it builds on one's intuitions about geodesic domes and similar examples, as discussed in Section 3.1.1. It is also intuitive in the sense that the metric in GR is, more often than not, given by a line element, $ds^2 = g_{\mu\nu}(x)dx^\mu dx^\nu$ and the geodesic edge-length data corresponds directly to continuum metric information on short line segments.

We now show that the *Hauptvermutung* for DCC-CLRC *fails*—by providing a counterexample. More specifically, we will exhibit a CLRC that, according to DCC-CLRC, "recovers" *both* Minkowski spacetime *and* a spacetime with a gravitational wave burst (and, in fact, many spacetimes with different gravitational wave bursts). In view of our requirement of essential uniqueness (Remark B in Comment 3 of Section 2.1), this "double win" is a failure.

Let $\overline{S}$ be the triangulation of 4-d Euclidean space, $\mathbb{R}^4$, used in (Roček and Williams 1981).[16] The vertices of the triangulation are the vertices of the integer lattice. The unit 4-d hypercube at the origin with the 16 vertices $\{(0,0,0,0),(0,0,0,1),(0,0,1,0),(0,0,1,1),\ldots(1,1,1,1)\}$ is triangulated into 24 4-simplices. Each of these 24 4-simplices is the convex hull of a monotonic path on the lattice from $(0,0,0,0)$ to $(1,1,1,1)$. There are 15 edges or "lattice vectors" that point from $(0,0,0,0)$ to each of the other vertices of the hypercube and these lattice vectors are labeled 1–15 by reading their vector components as binary numbers, so that the lattice vector $(1,0,0,0)$ is labeled 8, for example. This fundamental triangulated unit hypercube is translated all over the integer lattice—one such hypercube at each lattice vertex—to give $\overline{S}$. We take the combinatorial complex, S, corresponding to $\overline{S}$, and label its edges with Lorentzian length labels given in the following table in units of the Planck length, l_p, and Planck time, t_p:

TIMELIKE (1)	NULL (0)	SPACELIKE (1)	SPACELIKE ($\sqrt{2}$)	SPACELIKE ($\sqrt{3}$)
8: (1,0,0,0)	9: (1,0,0,1)	1: (0,0,0,1)	3: (0,0,1,1)	7: (0,1,1,1)
	10: (1,0,1,0)	2: (0,0,1,0)	5: (0,1,0,1)	
	12: (1,1,0,0)	4: (0,1,0,0)	6: (0,1,1,0)	
		11: (1,1,0,1)	15: (1,1,1,1)	
		13: (1,1,0,1)		
		14: (1,1,1,0)		

Also, the timelike and null edges are future-directed in the direction of increasing first coordinate. This, then, is our CLRC, (S, dec). It corresponds to the geometric Lorentzian simplicial complex in (Williams, 1986) which is the triangulation of Minkowski spacetime corresponding to $\overline{S}$ and the Lorentzian edge-lengths dec are the Minkowski spacetime proper lengths of the geodesics which are straight lines in the canonical inertial coordinate system defined by the fact that $M = \mathbb{R}^4$.

According to the DCC-CLRC, (S, dec) recovers Minkowski spacetime, (M_{mink}, g_{mink}). This is straightforward but for completeness, let us explicitly check that the conditions (i) to (v) of DCC-CLRC are satisfied. (i) By definition of S, the required $\overline{S}$ is the triangulation of $\mathbb{R}^4$ described above. (ii) The manifold for Minkowski spacetime is $M_{mink} = \mathbb{R}^4$ and its

16 Note that Rocek and Williams use the word "edge" to mean edges of the underlying lattice only and use "face diagonals," "body diagonals," and "hyperbody diagonal" to refer to the other 1-simplices. But we use "edge" to refer to any 1-simplex. Note also that Rocek and Williams are working in Riemannian signature, so that the edge-lengths they quote are Euclidean edge-lengths.

canonical coordinate system is inertial. The triangulating map $f : \overline{S} \to M_{mink}$ is the identity from $\mathbb{R}^4$ to $\mathbb{R}^4$ and the edges of S are embedded, via this map, as straight coordinate lines between vertices. (iii) These straight coordinate lines are geodesics in Minkowski spacetime and have the spacetime proper lengths matching the Lorentzian edge-lengths *dec* given in the table. Finally, (iv) and (v) are satisfied.

Now, this simplicial complex (S, dec) violates Lorentz invariance—just as its Riemannian version violates Euclidean symmetry and this may be enough reason for some to reject it as possible DPD in a quantum gravity theory. But we have a much more serious charge: (S, dec) does not recover Minkowski spacetime *at all* because there are many other GR spacetimes that, according to the discrete-continuum correspondence, DCC-CLRC, are "recovered" by (S, dec).

To construct an example of such a spacetime, we will use the widely known fact that the distribution of the embedded vertices of S induced by f is not actually uniform in Minkowski space. The distribution only *seems* uniform in the frame preferred by the complex in which the images of the vertices under f form the integer lattice. When boosted, in the z-direction say, this distribution is revealed to be very far from uniform: see for example Figure 1(a) of (Saravani and Aslanbeigi 2014). Indeed the boosted distribution of f-image-vertices has large *voids*: large, physically nice, regions of spacetime in which there are no f-image-vertices. By "large, physically nice regions," we mean regions of spacetime that contain approximately flat Alexandrov intervals of spacetime volume large on the Planck scale. Thus arises the possibility that there is curvature or physics—*e.g.*, a gravitational wave burst—in the voids, that does not register (get encoded) at all in the edge-lengths that decorate the CLRC. This will be the basis of our counterexample.

In Figure 2 is shown the $x = y = 0$ plane of the integer lattice in Minkowski space—with inertial coordinates $\{t, x, y, z\}$—with the x and y dimensions suppressed. Let $u := \frac{1}{\sqrt{2}}(t - z)$ and $v := \frac{1}{\sqrt{2}}(t + z)$ be null coordinates in the z direction. The shaded region in Figure 1 is

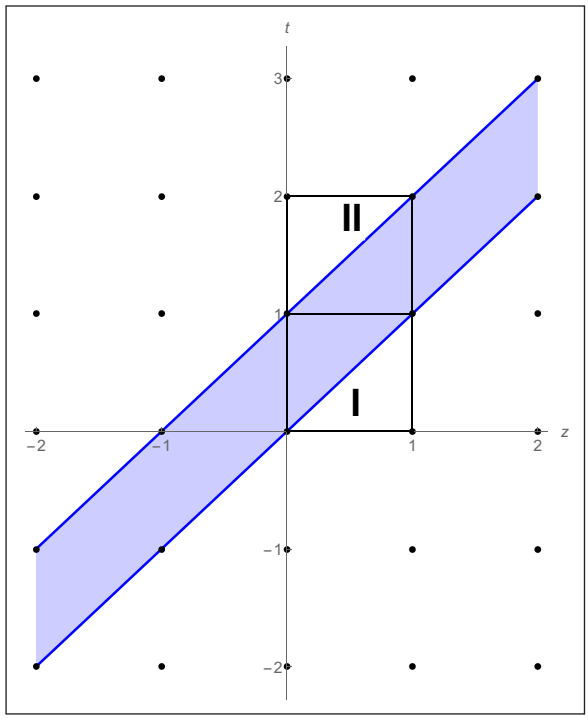


Figure 2 The diagram represents both Minkowski spacetime and the spacetime (3) with a plane gravitational wave burst. The integer lattice of image-vertices in the plane $x = y = 0$ is shown and most edges are not drawn. The shaded region is $0 < t - z < 1$ and is empty of vertices. In the gravitational wave spacetime the shaded region is the support of the wave burst. The labels I and II refer to two hypercubes at $(0, 0, 0, 0)$ and at $(1, 0, 0, 0)$, respectively which intersect the support of the wave burst. All other coordinate-hypercubes that intersect the support of the wave burst are isometric either to I or to II.

DOI: 10.31389/pop.17

the open region $0 < u < \frac{1}{\sqrt{2}}$, which is crossed by image-edges of the complex but is empty of image-vertices: It is a void. If the lattice is boosted in the z-direction, the distribution in the u direction stretches by some factor b and by factor b^{-1} in the v direction. When the γ factor of the boost is very large then $b \approx 2\gamma$. Choosing $\gamma = 10^{44}$, say, the height of the void in the time coordinate becomes of the order of one second and contains Alexandrov intervals of height one second: The void is physically nice and large.

We now exhibit a GR spacetime which is also "recovered" by (S, dec) according to DCC-CLRC. Let (M, g) be a perturbation of Minkowski spacetime that is a plane-fronted, transverse traceless gravitational wave burst in the z-direction:

$$ds^2 = -dt^2 + dz^2 + (1 + \epsilon h(u))dx^2 + (1 - \epsilon h(u))dy^2, \tag{3}$$

where the support of $h(u)$ is $0 < u < \frac{1}{\sqrt{2}}$, and $u := \frac{1}{\sqrt{2}}(t - z)$ as before. The spacetime is flat for $u \leq 0$ and $u \geq \frac{1}{\sqrt{2}}$. The manifold M is $\mathbb{R}^4$: the point $(t, x, y, z) \in \mathbb{R}^4$ is the point in (M, g) with coordinates (t, x, y, z), in the obvious way.

We now check that the conditions (i) to (v) of DCC-CLRC hold. (i) By definition of S, the required $\overline{S}$ is the triangulation of $\mathbb{R}^4$, exactly as in the case of Minkowski spacetime. (ii) Let $f : \overline{S} \to M$ be a homeomorphism that is a perturbation of the identity map chosen, such that:

- **(a)** f is the identity outside the support of $h(u)$ so that f maps the vertices of $\overline{S}$ to the points with integer coordinates in (M, g) and the support of $h(u)$ is therefore empty of image-vertices.
- **(b)** f maps the edges of $\overline{S}$ to the geodesics between the corresponding embedded vertices in (M, g).

Such a homeomorphism exists because (3) is a perturbation of Minkowski spacetime and there is a unique geodesic between two vertices of one coordinate-hypercube. (iii) Most of the geodesics are between vertices that lie outside the support of h where the metric is flat, and so they are straight lines in the coordinates $\{t, x, y, z\}$ of (3) and have the same length as in Minkowski spacetime. For the geodesic edges that cross the support of h, it can be shown that they have the same length as in Minkowski spacetime to first order in ϵ, provided that we choose h to satisfy the constraint:

$$\int_0^{\frac{1}{\sqrt{2}}} du h(u) = 0. \tag{4}$$

The details of the calculation are given in the Appendix of the longer version of this paper (Butterfield and Dowker 2021). Furthermore, condition (iv) holds. The only condition that might give one pause is (v). We deal with the doubt that (v) holds as a response to Objection 1, below.

There are many choices of wave packet contour $h(u)$ that satisfy the criterion (4); and for each choice, there is a GR spacetime that is "recovered" by (S, dec) according to DCC-CLRC. Thus, the requirement of essential uniqueness (Remark B of Comment 3 in Section 2.1) fails. And so we conclude that DCC-CLRC does not work as a statement of the recovery of a GR spacetime from a CLRC.

We now consider four possible objections to this conclusion.

Objection 1: The spacetime (3) does not satisfy condition (v) of the DCC-CLRC because it varies on the scale of the discreteness; and so it should be dismissed out of hand.

DOI: 10.31389/pop.17

Reply: Were this a Riemannian geometry, this objection would be valid. However, the gravitational wave spacetime is Lorentzian: and in a boosted coordinate system, the support of the wave burst is revealed to be a physically nice, large region of spacetime in which the contour of the burst varies slowly on the discreteness scale.

For definiteness, consider the coordinate transformation, $\{t, x, y, z\} \rightarrow \{t', x, y, z'\}$, which is a boost in the z direction with gamma factor $\gamma \gg 1$. Then the null coordinates transform as $\{u, v\} \rightarrow \{u', v'\}$, where $u' = bu$ and $v' = b^{-1}v$, where $b \approx 2\gamma$. Then the metric (3) in the new coordinates is

$$ds^2 = -dt'^2 + dz'^2 + (1 + \epsilon H(u'))dx^2 + (1 - \epsilon H(u'))dy^2, \tag{5}$$

where $H(u') = h(u)$ and the support of $H(u')$ is $0 < u' < \frac{1}{\sqrt{2}}b$. This is the same GR spacetime, expressed in different coordinates. The derivatives of H with respect to t' and z' are small, the curvature components are small on Planckian scales and the support is a physically nice region containing approximately flat Alexandrov intervals of proper height of order $\Delta t' = bt_{pl}$. For large enough γ factor, the height $\Delta t'$ can be as large as one likes. The concrete definition given in Comment 2 of Section 2.1 is satisfied because $\{t', x, y, z'\}$ is approximately an inertial coordinate system everywhere in the support of the burst, and in these coordinates the Riemann tensor is given approximately by second u'-derivatives of $\epsilon H(u')$. Therefore the spacetime (3) satisfies condition (v) of the DCC-CLRC.

Objection 2: Although spacetime (3) satisfies condition (v) of the DCC-CLRC, it is unrealistic: the wave is a linearized wave packet on a Minkowski spacetime background and is infinite in extent. It is not included in the collection of GR spacetimes that must be recovered by theory X discussed in Comment 2 in section 2.1. A realistic gravitational wave burst in our observable universe would not be a counterexample to the CLRC *Hauptvermutung*.

Reply: In a more realistic spacetime containing a gravitational wave burst, the finite spacetime support of the packet, the cosmological setting, and the discreteness of the underlying physical data together indeed places limits on the Lorentz boosts that are relevant approximate physical symmetries. If the Hubble scale of the observable universe today sets the scale for the largest region of approximately flat space that theory X needs to recover, then the void region in (3) in unprimed coordinates has finite extent: $\Delta u = 1/\sqrt{2} \sim 1$, $\Delta v \lesssim 10^{60}$ in Planck units. Under a boost in the z-direction, $\Delta u \Delta v$ is invariant and so $\Delta u' \Delta v' \lesssim 10^{60}$ in the primed coordinates of (5). Gravitational waves known to be empirically relevant today do not satisfy this bound. For example, the highest frequency measurable by LIGO is $\sim 10^4\, Hz$, which is a timescale $\Delta u' \sim 10^{40}$ in Planck units. This restricts $\Delta v' \lesssim 10^{20}$ in Planck units, which is unrealistic.

If theory X only need recover GR spacetimes that are known to be empirically relevant today, therefore, this would indeed disqualify our particular counterexample to the CLRC Hauptvermutung (it would also severely restrict what we understand by GR). Our response is twofold.

(i) Empirical relevance to our actual observations today is a strong restriction on the spacetimes that quantum gravity should be able to recover, and to our knowledge no one has made such a proposal. Even if we take that restriction seriously but relax it to require that theory X recover all GR spacetimes that might be empirically relevant in the future, then our counterexample comes back into play, since, in the future, observational physics may require consideration of superhorizon scales much larger than today's Hubble scale and/or gravitational waves of very high frequency; for example, from small primordial black holes (Domènech et al. 2021).

DOI: 10.31389/pop.17

(ii) Whatever view one takes on what class of GR spacetimes theory X should recover, it is the case that the CLRC—thought of as recovering a piece of Minkowski space, whose relevance is not in doubt—has void regions that are very large compared to the Planck scale. There is non-gravitational physics—the Standard Model, for example—in which the relevant scales of phenomena satisfy $\Delta u' \Delta v' \lesssim 10^{60}$. The CLRC will not be able to recover that non-gravitational physics in the voids uniquely. In other words, extending the CLRC Hauptvermutung to *GR solutions including matter*, where the decorations *dec* on the edges and/or vertices of the CLRC now include matter degrees of freedom, we can construct two continuum solutions, both in Minkowski spacetime but with different matter distributions in the voids that both accord with the same decorated CLRC. For example, the spacetime support of a wave packet for a single ultra-high energy gamma ray photon with peak frequency $f \sim 10^{28} Hz \sim 10^{-16}$ Planck units that travels a distance $\sim 10^{11} m \sim 10^{44}$ Planck units from a gamma ray burst source would fit into a void region of the integer lattice CLRC.

Objection 3: The DPD could be a *set* of combinatorial simplicial complexes, such that for each local Lorentz frame there is a member of the set that is uniform in that frame.

Reply: The discrete-continuum correspondence for such a set of data would involve an embedding of the vertices of all the simplicial complexes into the recovered spacetime. With one complex for each frame, that would result in a density for the vertices' embedding much higher than Planckian, if not infinite. Such data would contradict our assumption that the discreteness is Planckian in scale.

Objection 4: "Here is a discrete-continuum correspondence for Combinatorial Complexes that will work: Fill the simplices of the CLRC with patches of flat spacetime so as to form the corresponding Geometric Regge Complex (GRC). Smooth the corners of this piecewise flat Lorentzian manifold to form the differentiable manifold (M_{CLRC}, g_{CLRC}). We then declare a GR spacetime (M, g) to be a good approximation to the CLRC if (M, g) is approximately isometric to (M_{CLRC}, g_{CLRC})."

Reply: This natural-seeming proposal for the discrete-continuum correspondence for CLRCs is based on the assumption that filling in the simplices with patches of flat spacetime is the essentially unique way to produce a spacetime that (i) agrees with the data of a CLRC and (ii) has curvature that only varies appreciably on scales much larger than the Planck scale. In other words, it assumes that if one were to fill in the simplices of a CLRC with non-flat Lorentzian geometry, then the resulting manifold (smoothed if necessary) would necessarily be unphysical because it would have curvature that varies on sub-Planckian scales.

Now, this assumption may be valid for combinatorial *Riemannian* simplicial complexes (and we believe it is). But, it is false for CLRCs: filling in with a non-flat metric does *not* necessarily give a spacetime with curvature that varies on sub-Planckian scale. Indeed, our gravitational wave example shows exactly that: The spacetime (3) is the result of filling in the simplices of (S, dec) with non-flat geometry; and yet, (3) does *not* have curvature that varies on smaller than Planckian scales, as shown by (5).

4.2 AVOIDING VOIDS?

The failure of the CLRC Hauptvermutung means that the discrete-continuum correspondence DCC-CLRC does not hold good: The continuum is *not* recovered by a CLRC that satisfies the conditions of DCC-CLRC. The underlying reason for the failure of DCC-CLRC in the example above is the presence of a large, physically nice void region

in the putative continuum, for which the only discrete data are the lengths of the geodesic edges that cross the void; but that information is not nearly rich enough to encode the Lorentzian geometry in the void. (Recall the Reply to Objection 1 in Section 4.1.) Moreover, there are voids *everywhere*, not just the one highlighted by the gravitational wave burst example. We conclude: *the CLRC*, (S, dec), *of our counterexample is not approximated by a Lorentzian manifold at all.*

One might attempt to avoid this conclusion by eliminating voids; that is, by amending the discrete-continuum correspondence for CLRCs so as to require the number-volume correspondence for the embedding of the vertices of the CLRC in the manifold, i.e., condition (i) in DCC-C of Section 3.2 (Planck-scale uniformity). But there are two problems with this suggestion.

Firstly, if the vertices and the timelike and null edges of the simplicial complex form a directed acyclic graph, then its transitive completion will *be* a causal set. Then, if Planck-scale uniformity holds for the embedded vertices of the complex, then either this causal set will be faithfully embedded in the GR spacetime or the order relation of the causal set defined by the directions on the complex's edges will not be consistent with the spacetime causal order of the embedded vertices in the GR spacetime. The latter would be an inconsistency between the DPD and the continuum and would be hard to work around. The former case implies that for a CLRC to provide a genuine alternative set of DPD to a causal set whilst incorporating Planck-scale uniformity of embedding of vertices, the CLRC would have to have mostly or all spacelike edges. Even in that case, one could argue that from the information about the spacelike edge-lengths a unique causal order on the vertices of the CLRC should be deducible and so again with the imposition of Planck-scale uniformity on the embedded vertices of the complex, the complex would be effectively providing a faithfully embeddable causal set.

A second, stronger argument against the possibility of eliminating voids by adding a requirement of Planck-scale uniformity is that a triangulation of a Lorentzian spacetime cannot be constructed from a properly uniform distribution of embedded vertices, without introducing a frame. We note in particular that the constructions of Voronoi graphs and Delaunay triangulations from a set of embedded vertices, which are so natural in Riemannian geometry (e.g., Itzykson 1984), do not work for Lorentzian manifolds, and a supplementary frame must be introduced and used—as, for example, in (Lee 1984).

More generally, the basic obstruction to a Lorentz-invariant simplicial complex is the contradiction between (i) the local nature of a simplicial complex with its concept, inherent in its structure, of nearest-neighbor vertices joined by 1-simplices and (ii) the non-locality of the Lorentzian geometry. For example, in a uniform, Poisson distribution of embedded vertices, a given vertex will have a nonlocal set of hugely, if not infinitely, many *physically nearest neighbors*, hugging its past and future lightcones—as discussed in Section 2.3—which a simplicial complex cannot accommodate.

Finally, we note that the counterexample and all the above arguments apply also to Lorentzian low valence graphs,[17] including the special cases of Lorentzian "lattices" (graphs-with-symmetry). For consider: a CLRC contains within itself a Lorentzian low valence graph that is its 1-skeleton, the set of its vertices and edges. So we can take the decorated 1-skeleton of the CLRC (C, dec) of our counterexample constructed from the integer lattice, and use it as a counterexample to any proposed DCC for Lorentzian low valence graphs.

17 Causal sets are graphs—transitive directed acyclic graphs—and the condition of low valency is to make sure these graphs are genuine alternative DPD sets to causal sets.

5 CONCLUSION

The unity of physics demands that before giving up on so much scientific progress already made, we should take seriously the aim of recovering General Relativity, with its Lorentzian geometry and local Lorentz invariance, from quantum gravity. We have argued that in our current state of knowledge, a causal set is the only kind of entity that is discrete at the Planck scale in our sense, and can adequately encode the geometric information in a Lorentzian spacetime at macroscopic scales.

Further, if one accepts our justification for the two claims (given in Sections 3 and 4), then it is tempting to conjecture that no challenger to causal sets will arise in the future. More precisely, it is tempting to conjecture that a causal set is the unique Planck-scale DPD-set that can recover a GR spacetime as a continuum approximation. If that is indeed so, the stronger is the motivation for an approach to quantum gravity that is based fundamentally on causal sets.

ACKNOWLEDGEMENTS

We thank two anonymous referees; and David Meyer, David Rideout, Rafael Sorkin, Sumati Surya, and Yasaman Yazdi for helpful discussions. We also thank seminar audiences at Banff, Milan, Nijmegen, Perimeter Institute and Vienna.

FUNDING INFORMATION

FD and JB acknowledge the support of the Leverhulme/Royal Society interdisciplinary APEX grant APX/R1/180098. FD is supported in part by Perimeter Institute for Theoretical Physics. Research at Perimeter Institute is supported by the Government of Canada through Industry Canada and by the Province of Ontario through the Ministry of Economic Development and Innovation. FD is supported in part by STFC grant ST/P000762/1.

COMPETING INTERESTS

FD has no competing interests to declare. JB is a member of the Editorial Board of this journal.

AUTHOR AFFILIATIONS

Jeremy Butterfield orcid.org/0000-0002-0215-5802
Trinity College, Cambridge, University of Cambridge, CB2 1TQ, United Kingdom

Fay Dowker orcid.org/0000-0002-6652-1058
Blackett Laboratory, Imperial College, Prince Consort Road, London, SW7 2AZ, United Kingdom; Perimeter Institute, 31 Caroline Street North, Waterloo ON, N2L 2Y5, Canada

REFERENCES

Abbott, B. P., Richard Abbott, Thomas D. Abbott, et al. 2016. "Observation of Gravitational Waves from a Binary Black Hole Merger." *Phys. Rev. Lett*. 116(Feb): 061102. DOI: https://doi.org/10.1103/PhysRevLett.116.061102

Ambjorn, Jan, and **Renate Loll.** 1998. "Non-Perturbative Lorentzian Quantum Gravity, Causality and Topology Change." *Nucl. Phys.* B536: 407–434. DOI: https://doi.org/10.1016/S0550-3213(98)00692-0

Ambjorn, Jan, Jerzy Jurkiewicz, and **Renate Loll.** 2000. "A Non-Perturbative Lorentzian Path Integral for Gravity." *Phys. Rev. Lett.* 85: 924–927. DOI: https://doi.org/10.1103/PhysRevLett.85.924

Ambjorn, Jan, Jerzy Jurkiewicz, and **Renate Loll.** 2001a. "Dynamically Triangulating Lorentzian Quantum Gravity." *Nucl. Phys.* B610: 347–382. DOI: https://doi.org/10.1016/S0550-3213(01)00297-8

Ambjorn, Jan, Jerzy Jurkiewicz, and **Renate Loll.** 2001b. "Non-Perturbative 3d Lorentzian Quantum Gravity." *Phys. Rev.* D64: 044011. DOI: https://doi.org/10.1103/PhysRevD.64.044011

Banks, Thomas. 2020. "Holographic Space-Time and Quantum Information." *Frontiers in Physics* 8: 111. DOI: https://doi.org/10.3389/fphy.2020.00111

Bell, John Stewart. 2004. *Speakable and Unspeakable in Quantum Mechanics*. Second ed. Cambridge: Cambridge Univeristy Press. DOI: https://doi.org/10.1017/CBO9780511815676

Bombelli, Luca. 2000. "Statistical Lorentzian Geometry and the Closeness of Lorentzian Manifolds." *Journal of Mathematical Physics* 41(10): 6944. DOI: https://doi.org/10.1063/1.1288494

Bombelli, Luca, Joohan Lee, David Meyer, and **Rafael D. Sorkin.** 1987. "Space-Time as a Causal Set." *Phys. Rev. Lett.* 59(Aug): 521–524. DOI: https://doi.org/10.1103/PhysRevLett.59.521

Burgess, Cliff P. 2004. "Quantum Gravity in Everyday Life: General Relativity as an Effective Field Theory." *Living Reviews in Relativity*. DOI: https://doi.org/10.12942/lrr-2004-5

Butterfield, Jeremy. 2011a. "Emergence, Reduction and Supervenience: A Varied Landscape." *Foundations of Physics* 41: 920–959. DOI: https://doi.org/10.1007/s10701-011-9549-0

Butterfield, Jeremy. 2011b. "Less is Different: Emergence and Reduction Reconciled." *Foundations of Physics* 41: 1065–1135. DOI: https://doi.org/10.1007/s10701-010-9516-1

Butterfield, Jeremy. 2024. "En Route to Reduction: Lorentzian Manifolds and Causal Sets." Forthcoming in *Time and Timelessness in Fundamental Physics and Cosmology*, edited by S. De Bianchi, M. Forgione, and L. Marongiu. Cham: Springer. arxiv: 2401.15474

Butterfield, Jeremy, and **Fay Dowker.** 2021. "Recovering General Relativity from a Planck Scale Discrete Theory of Quantum Gravity." DOI: https://doi.org/10.48550/arXiv.2106.01297

Carballo-Rubio, Raúl, Francesco Di Filippo, Stefano Liberati, and **Matt Visser.** 2020. "Causal Hierarchy in Modified Gravity." *Journal of High Energy Physics* 2020(12). DOI: https://doi.org/10.1007/JHEP12(2020)055

Carter, Brandon. 1971. "Axisymmetric Black Hole Has Only Two Degrees of Freedom." *Phys. Rev. Lett.* 26(Feb): 331–333. DOI: https://doi.org/10.1103/PhysRevLett.26.331

Cercignani, Carlo. 1987. *The Boltzmann Equation and Its Applications*. Applied Mathematical Sciences. New York: Springer. DOI: https://doi.org/10.1007/978-1-4612-1039-9

Chapman, Sydney, and **T. G. Cowling.** 1991. *The Mathematical Theory of Non-uniform Gases: An Account of the Kinetic Theory of Viscosity, Thermal Conduction and Diffusion of Gases*. Third ed. Cambridge University Press.

Chruściel, Piotr T., João Lopes Costa, and **Markus Heusler.** 2012. "Stationary Black Holes: Uniqueness and Beyond." *Living Reviews in Relativity* 15. DOI: https://doi.org/10.12942/lrr-2012-7

Cortês, Marina, and **Lee Smolin.** 2014. "Quantum Energetic Causal Sets." *Physical Review D* 90(4). DOI: https://doi.org/10.1103/PhysRevD.90.044035

De Haro, Sebastian, and **Jeremy Butterfield.** 2017. "A Schema for Duality, Illustrated by Bosonization." In *Foundations of Mathematics and Physics One Century After Hilbert*, edited by Joseph Kouneiher. Springer. DOI: https://doi.org/10.1007/978-3-319-64813-2_12

De Haro, Sebastian, and **Jeremy Butterfield.** 2019. "On Symmetry and Duality." *Synthese, Symmetries and Asymmetries in Physics*, 1–41. DOI: https://doi.org/10.1007/s11229-019-02258-x

Domènech, Guillem, Volodymyr Takhistov, and **Misao Sasaki.** 2021. Exploring Evaporating Primordial Black Holes with Gravitational Waves. *Physics Letters B* 823(dec): 136722. DOI: https://doi.org/10.1016/j.physletb.2021.136722

Donoghue, John F. 1997. "Introduction to the Effective Field Theory Description of Gravity." In *Proceedings of the Advanced School on Effective Theories*, edited by Fernando Cornet, and María José Herrero; Almunecar, Granada, Spain, 26 June–1 July 1995. World Scientific.

Finkelstein, David. 1969. "Space-Time Code." *Phys. Rev.* 184(5): 1261–1271. DOI: https://doi.org/10.1103/PhysRev.184.1261

Gorard, Jonathan. 2020. "Some Relativistic and Gravitational Properties of the Wolfram Model." *Complex Systems* 29(2): 599–654. DOI: https://doi.org/10.25088/ComplexSystems.29.2.599

Havas, Peter. 1964. "Four-Dimension Formulations of Newtonian Mechanics and their Relation to the Special and the General Theory of Relativity." *Rev. Mod. Phys.* 36: 938–965. DOI: https://doi.org/10.1103/RevModPhys.36.938

Hawking, Stephen. 1971. "Gravitational Radiation from Colliding Black Holes." *Phys. Rev. Lett.* 26(May): 1344–1346. DOI: https://doi.org/10.1103/PhysRevLett.26.1344

Hawking, Stephen, A. R. King, and **P. J. McCarthy.** 1976. "A New Topology for Curved Space-Time which Incorporates the Causal, Differential, and Conformal Structures." *J. Math. Phys.* 17: 174–181. DOI: https://doi.org/10.1063/1.522874

Hawking, Stephen. 2014. "Singularities and the Geometry of Spacetime." *EPJ H* 39: 413–513. DOI: https://doi.org/10.1140/epjh/e2014-50013-6

Hawking, Stephen W., and **George F. R. Ellis.** 1973. *The Large Scale Structure of Space-Time*. Cambridge, UK: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511524646

Itzykson, Claude. 1984. "Fields on a Random Lattice." In *Progress in Gauge Field Theory*, edited by G. 't Hooft, A. Jaffe, G. Lehmann, P. K. Mitter, and I. M. Singer, 337–371. New York: Plenum. DOI: https://doi.org/10.1007/978-1-4757-0280-4_11

Konopka, Tomasz, Fotini Markopoulou, and **Simone Severini.** 2008. "Quantum Graphity: A Model of Emergent Locality." *Physical Review D* 77(10). DOI: https://doi.org/10.1103/PhysRevD.77.104029

Kronheimer, E. H., and **Roger Penrose.** 1967. "On the Structure of Causal Spaces." *Mathematical Proceedings of the Cambridge Philosophical Society* 63(4): 481–501. DOI: https://doi.org/10.1017/S030500410004144X

Landsman, Klaas. 2017. *Foundations of Quantum Theory*. New York: Springer. DOI: https://doi.org/10.1007/978-3-319-51777-3

Lee, T. D. 1984. "Discrete Mechanics." In *How Far Are We from the Gauge Forces? The Subnuclear Series Vol. 21*, edited by A. Zichichi, 15–91. New York: Plenum. DOI: https://doi.org/10.1007/978-1-4684-5086-6_2

Malament, David B. 1977. "The Class of Continuous Timelike Curves Determines the Topology of Spacetime." *J. Math. Phys.* 18: 1399–1404. DOI: https://doi.org/10.1063/1.523436

Malament, David B. 1986. "Newtonian Gravity, Limits, and the Geometry of Space." In *From Quarks to Quasars*, edited by R. G. Colodny, 181–201. Pittsburgh: University of Pittsburgh Press. DOI: https://doi.org/10.2307/jj.5973221.7

Malament, David B. 2012. *Chapter 4. Newtonian gravitation theory*. Chicago: University of Chicago Press.

Minguzzi, Enrico, and **Miguel Sanchez.** 2008. "The Causal Hierarchy of Spacetimes." In *Recent Developments in Pseudo-Riemannian Geometry*, edited by H. Baum, and D. Alekseevsky, 299–358. Zurich: Eur. Math. Soc. Publ. House. DOI: https://doi.org/10.4171/051-1/9

Myrheim, Jan. 1978. *Statistical Geometry, CERN preprint TH-2538*. https://cds.cern.ch/record/293594

Oriti, Daniele. 2009. "The group field theory approach to quantum gravity." In *Approaches to Quantum Gravity: Towards a New Understanding of Space, Time and Matter*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511575549

Padmanabhan, T. 2014. "Gravity and Spacetime: An Emergent Perspective." In *Springer Handbook of Spacetime, Springer Handbooks*, edited by A. Ashtekar, and V. Petkov. Berlin, Heidelberg: Springer. DOI: https://doi.org/10.1007/978-3-642-41992-8_12

Regge, Tullio. 1961. "General Relativity Without Coordinates." *Nuovo Cim.* 19: 558–571. DOI: https://doi.org/10.1007/BF02733251

Rideout, David, and **Rafael Sorkin.** 2000. "A Classical Sequential Growth Dynamics for Causal Sets." *Phys. Rev.* D61: 024002. DOI: https://doi.org/10.1103/PhysRevD.61.024002

Riemann, Bernhard. 1868. *Über die Hypothesen, welche der Geometrie zu Grunde liegen*. Riemann's Habiliationsschrift, Göttingen, 1954. DOI: https://doi.org/10.1002/phbl.19540100702

Riemann, Bernhard. 2016. "On the Hypotheses Which Lie at the Bases of Geometry." In *On the Hypotheses Which Lie at the Bases of Geometry*, edited with an introduction by Jürgen Jost. Basel: Birkhäuser. DOI: https://doi.org/10.1007/978-3-319-26042-6

Robinson, D. C. 1975. "Uniqueness of the Kerr Black Hole." *Phys. Rev. Lett.* 34(Apr): 905–906. DOI: https://doi.org/10.1103/PhysRevLett.34.905

Rovelli, Carlo. 2008. "Loop Quantum Gravity." *Living Reviews in Relativity* 11. DOI: https://doi.org/10.12942/lrr-1998-1

Rovelli, Carlo, and **Francesca Vidotto.** 2014. *Covariant Loop Quantum Gravity: An Elementary Introduction to Quantum Gravity and Spinfoam Theory*. Cambridge University Press. DOI: https://doi.org/10.1017/CBO9781107706910

Roček, Martin, and **Ruth M. Williams.** 1981. "Quantum Regge Calculus." *Physics Letters B* 104(1): 31–37. DOI: https://doi.org/10.1016/0370-2693(81)90848-0

Saravani, Mehdi, and **Siavash Aslanbeigi.** 2014. "On the Causal Set-Continuum Correspondence." *Class. Quant. Grav.* 31(20): 205013. DOI: https://doi.org/10.1088/0264-9381/31/20/205013

Saunders, Simon, Jon Barrett, Adrian Kent, and **David Wallace** (eds.). 2010. *Many Worlds? Everett, Quantum Theory and Reality*. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199560561.001.0001

Sorkin, Rafael D. 1974. *The Development of Simplectic Methods for the Metrical and Electromagnetic Fields*. Ph.D. thesis, California Institute of Technology.

Sorkin, Rafael D. 2007. "Quantum Dynamics without the Wave Function." *J. Phys.* A40: 3207–3222. DOI: https://doi.org/10.1088/1751-8113/40/12/S20

Sorkin, Rafael D. 2009. "Light, Links and Causal Sets." *J. Phys. Conf. Ser.* 174: 012018. DOI: https://doi.org/10.1088/1742-6596/174/1/012018

't Hooft, Gerard. 1979. "Quantum Gravity: a Fundamental Problem and Some Radical Ideas." In *Recent Developments in Gravitation (Proceedings of the 1978 Cargese Summer Institute)*, edited by M. Levy, and S. Deser. New York: Plenum. DOI: https://doi.org/10.1007/978-1-4613-2955-8_8

Trautman, Andrezj . 1965. "Foundations and Current Problems of General Relativity." In *Lectures on General Relativity*, edited by S. Deser, and K. Ford. Prentice-Hall.

Verlinde, Erik. 2011. "On the Origin of Gravity and the Laws of Newton." *Journal of High Energy Physics* 2011(4). DOI: https://doi.org/10.1007/JHEP04(2011)029

Williams, Ruth M. 1986. "Quantum Regge Calculus in the Lorentzian Domain and its Hamiltonian Formulation." *Class. Quantum Grav.* 3: 853–869. DOI: https://doi.org/10.1088/0264-9381/3/5/015

TO CITE THIS ARTICLE:
Butterfield, Jeremy, and Fay Dowker. 2024. "Recovering General Relativity from a Planck Scale Discrete Theory of Quantum Gravity." *Philosophy of Physics* 2(1): 5, 1–31. DOI: https://doi.org/10.31389/pop.17

Submitted: 06 December 2022
Accepted: 21 March 2024
Published: 10 May 2024

COPYRIGHT:
© 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.

Philosophy of Physics is a peer-reviewed open access journal published by LSE Press.