

Article

Causal Structure in Spin Foams

Eugenio Bianchi ^{1,2,*}  and Pierre Martin-Dussaud ³ 
¹ Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802, USA

² Department of Physics, The Pennsylvania State University, University Park, PA 16802, USA

³ Basic Research Community for Physics e.V., Mariannenstraße 89, 04315 Leipzig, Germany

* Correspondence: ebianchi@psu.edu

Abstract: The metric field of general relativity is almost fully determined by its causal structure. Yet, in spin foam models of quantum gravity, the role played by the causal structure is still largely unexplored. The goal of this paper is to clarify how causality is encoded in such models. The quest unveils the physical meaning of the orientation of the two-complex and its role as a dynamical variable. We propose a causal version of the EPRL spin foam model and discuss the role of the causal structure in the reconstruction of a semiclassical space–time geometry.

Keywords: loop quantum gravity; spin foams

1. Introduction

The information carried by a space–time metric is mainly of a causal nature. Indeed, Malament’s theorem states that the causal relations between the points of a 4D manifold fully determine the metric, up to a conformal factor given at each point [1–3]. In quantum models of space–time, the role of the metric is usually played by more fundamental objects, like spins and intertwiners in spin foam models [4–6] for the dynamics of loop quantum gravity [7–9]. There, it is less evident to see how causality enters the scene: how is it encoded? The question is of importance to understand more generally whether causality is a fundamental or an emergent property of space–time. Aspects of the causal structure of a quantum space–time are at the roots of the spin foam formalism [10–12] and have been studied in [13–33].

Our investigation proceeds as follows: In Section 2, we recall what is meant by causality over a Lorentzian manifold, and we show how it survives over a simplicial complex; in Section 3, we show how the causal structure can be represented on the dual skeleton; in Section 4, we show how the causal structure is determined by the dynamics of Regge calculus [34]; in Section 5, we elucidate the role of causality at the level of the path integral over geometries; in Section 6, we investigate the case of discrete BF theory; in Section 7, we propose a causal version of the Engle–Pereira–Rovelli–Livine (EPRL) spin foam model [35]; and finally, in Section 8, we discuss how the proposed causal EPRL model relates to previous proposals for implementing causality such as the Livine–Oriti vertex [16] and the Engle vertex [36,37].

2. Discrete Causal Structure

The geometry of a Lorentzian manifold is fully encoded in the metric g . The signature of g is either $(-, +, +, +)$ or $(+, -, -, -)$. It is generally held that this freedom is a pure convention, with no physical consequences. However, for the rest of our work, it is useful to let this choice be open and write the signature as $(\eta, -\eta, -\eta, -\eta)$ with $\eta \in \{-1, +1\}$.

The *causal structure* of g can be decomposed into two sub-notions that we call *bare causality*¹ and *time orientability* (see [3] for a standard reference).



Citation: Bianchi, E.; Martin-Dussaud, P. Causal Structure in Spin Foams. *Universe* **2024**, *10*, 181. <https://doi.org/10.3390/universe10040181>

Academic Editor: Parampreet Singh

Received: 18 March 2024

Revised: 10 April 2024

Accepted: 10 April 2024

Published: 14 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

2.1. Bare Causality

We call “bare causality” the property of each tangent space at any point of the manifold, and it is partitioned into three classes of tangent vectors: time-like, space-like and null. Formally, the classes are equivalence classes for the following relation:

$$u \sim v \iff \text{sign } g(u, u) = \text{sign } g(v, v), \quad (1)$$

where u and v are tangent vectors. The three equivalent classes are named

$$\text{sign } g(u, u) = \begin{cases} \eta & \text{time-like} \\ 0 & \text{null} \\ -\eta & \text{space-like.} \end{cases} \quad (2)$$

The definition of bare causality is local in the sense that it makes sense at each point of the manifold, but it can also be formulated as a global property. The “bare-causal structure” of a Lorentzian space–time consists in the possibility to say, given any two points, whether they are space-like, time-like, or null separated. To be clear with the definitions, two points are time-like separated if they can be joined by a smooth curve whose tangent vectors are time-like all along. It is important that the curve is smooth because otherwise one could turn around sharply and draw a time-like curve between any two points.

2.2. Time Orientability

On top of bare causality, one can define a notion of “time orientability”. At a local level, time orientability is the property of time-like vectors to be divided into two classes: past and future. Formally, the two classes are equivalence classes of time-like vectors for the following relation:

$$u \sim v \iff \text{sign } g(u, v) = \text{sign } g(v, v). \quad (3)$$

In the case of bare causality, the information contained in the metric g alone enables distinguishing one from another of the three classes of time, null and space without ambiguity. This is not the case for time orientability: the two classes defined by (3) are perfectly symmetric. Thus, the denomination “past” or “future” is arbitrary as long as no external arrow of time is imposed additionally. So we can pick a reference vector u_0 (the arrow of time), label the two classes by $\omega \in \{-1, +1\}$ and say that the time-like vector v is in the class ω if

$$\text{sign } g(u_0, v) = \omega \eta. \quad (4)$$

To fix the language, we declare that $\omega = +1$ is the future, so that u_0 is future-pointing.

As done previously, the local definition of time orientation can be turned into a global one by requiring continuity across the classes at different points. Then, space–time is said to be time orientable if it is possible to continuously define a division of time-like vectors in past and future classes. It is then possible to say that a point lies in the future of some other. As a consequence, one can define the *causal future* $I^+(p)$ and the *causal past* $I^-(p)$ of a point p . Again, the arrow of time, i.e., the labeling of “past” or “future,” is conventional, e.g., attached to a specific choice of reference vector u_0 , but it is not a geometric property of the metric.

Time orientability is conceptually different from bare causality. However, any Lorentzian metric locally defines light-cones with both a bare-causal and a time-orientable structure. So the conceptual difference is often overlooked, and the term “causality” is used indifferently to discuss either notions or both. Yet, it is important to have the distinction clear in mind, especially when moving to quantum models, because we expect the Lorentzian metric to make way for new objects, while the underlying physical notions may survive.

2.3. Discrete Bare Causality

At a discrete level, consider a Lorentzian 4-simplicial complex Δ , i.e., a set of Minkowskian 4-simplices nicely glued together [34]². It is again possible to define the notions of bare causality and time orientability, both at a local and at a global level.

Each 4-simplex σ comes with an embedding in Minkowski space–time. It is bounded by five tetrahedra, each with a unique normal 4-vector N of unit norm and directed outward. A priori, N can be time-like, space-like, or null. Two nearby 4-simplices share exactly one common tetrahedron T . The two 4-simplices are said to be time-like separated if N (computed with respect to any of the two 4-simplices) is time-like. A similar definition holds for space-like and null separation.

Unfortunately, this local notion of bare causality fails to extend straightforwardly to distant 4-simplices. Indeed, one could be tempted to say that two distant 4-simplices are time-like separated if there exists a sequence of time-like-separated nearby tetrahedra in-between. However, this definition fails because, for instance, two space-like separated nearby tetrahedra could be connected by a common time-like-separated nearby tetrahedron. In the previous continuous case, the smoothness of the time-like curve was preventing such a pathology, but this is not possible anymore in the discrete case. The difficulty can be circumvented by first introducing a local notion of time orientability.

2.4. Discrete Time Orientability

A tetrahedron is said to be space-like if it is embedded within a space-like hyperplane. In this case, its 4-normal N is time-like. Time orientability is the property that the space-like boundary tetrahedra of a 4-simplex can be divided into two classes by the following relation:

$$T_1 \sim T_2 \iff \text{sign}(N_1 \cdot N_2) = \text{sign}(N_1 \cdot N_1), \quad (5)$$

where the dot denotes the Minkowskian scalar product. A choice of time orientation indicates which class is called past or future (relative to the 4-simplex).

At a global level, we say that Δ is time orientable if there exists a consistent choice of time orientation for each 4-simplex, so that each space-like tetrahedron has an opposite time orientation relative to each of the two 4-simplices that bounds it: if a tetrahedron is in the future of a 4-simplex, it should be in the past of another.

Given two 4-simplices, σ_1 and σ_2 , sharing a space-like tetrahedron T , we say that σ_2 is in the future of σ_1 if T is in the future of σ_1 (hence in the past of σ_2). This definition allows us to define straightforwardly a notion of causal future and causal past of a 4-simplex: σ_2 is in the future of σ_1 if there exists a future-oriented chain of 4-simplices in-between. This definition encompasses the notion of time-like separation for distant 4-simplices that was initially looked for. Thus, both bare causality and time orientability are defined locally and globally in the discrete setting.

3. Causality on the Dual Skeleton

The previously defined discrete causal structure can be easily represented on Δ_1^* , the dual 1-skeleton of Δ . The 1-skeleton Δ_1^* is built from Δ by replacing each 4-simplex by a vertex, each tetrahedron by an edge and forgetting about triangles, segments and points of Δ .

Bare causality discriminates between space-like and time-like edges³, while time orientability provides an orientation to the time-like edges. Overall, causality is then represented by

1. An arrow from past to future on time-like edges,
2. No arrow on space-like edges.

In the following, we assume that all tetrahedra are space-like, which implies that all N are time-like. Dually, it means that all the edges of Δ_1^* carry an arrow. This simplifying assumption is made in many of the formulations of spin foams. It is important to note that this condition automatically implements some implicit assumptions about the fundamental

causal structure of space–time. Indeed, this assumption erases any local notion of bare causality: all nearby 4-simplices are time-like separated. Thus, at the most local level, the primacy is granted to time orientability. The notion of bare causality only emerges at a more global level as follows: given two distant vertices, if one is not in the past of the other, then they are said to be space-like separated.

3.1. Dual Causal Set

In mathematical terms, Δ_1^* is a 5-valent *simple oriented graph*⁴. Assuming it is also *acyclic*⁵, its *transitive closure* defines a *poset* (partially ordered set). The elements of the poset are the vertices of Δ_1^* , and the partial order $\leq$ comes from a unique extension of the set of arrows with the following properties:

1. Reflexivity: $v \leq v$ (by convention).
2. Anti-symmetry: $v_1 \leq v_2$ and $v_2 \leq v_1$ imply $v_1 = v_2$.
3. Transitivity: $v_1 \leq v_2$ and $v_2 \leq v_3$ imply $v_1 \leq v_3$.

In most reasonable cases, the poset of Δ_1^* is *locally finite*, meaning that for any pair of vertices (v_1, v_2) , the so-called *causal diamond* $\{v \mid v_1 \leq v \leq v_2\}$ is a finite set. Such a poset is a *causal set*, as defined originally in [39].

We have shown, without much surprise, that the discretization of a Lorentzian manifold naturally carries a causal set structure. Causal set theory takes the causal set structure as a starting point. Then, the question naturally poses itself as to whether or not it is possible to reconstruct Δ_1^* from its associated causal set only. Given a causal set, one can derive a notion of neighborhood by declaring that two vertices x and y , such that $x \leq y$, are next to each other if there is no $z \neq x, y$ such that $x \leq z \leq y$. In other words, the neighborhood relations are obtained by a *transitive reduction* of the causal set, i.e., a graph with the fewest possible arrows and the same “reachability relations” as the causal set. Interestingly, for a finite directed acyclic graph, such a transitive reduction is unique. However, *the transitive reduction of the transitive closure is not the identity*. Thus, it is not possible to recover Δ_1^* from the causal set by transitive reduction. In other words, the notions of neighborhood for causal set theory and for discrete Lorentzian geometry, as described over Δ_1^* , are not the same.

Similarly, the conformal factor, which is an important piece of information of the metric, can arise in several different ways. In causal set theory, it emerges by counting the number of vertices within a given causal diamond [39,40]. In discrete Lorentzian geometry, it can be given by the Lorentzian volume of the 4-simplices, which requires additional input, not deducible from the causal set alone. For instance, the additional information can be provided by coloring each vertex with a real number (the 4-simplex volume), by coloring each edge with the volume of the corresponding tetrahedron or by introducing faces and coloring them with the area of the corresponding triangles. The latter option is of course relevant for spin foam models as discussed also in [16,29,30,41].

To work algebraically with the causal structure, it will soon appear convenient to express the orientation of the edges as follows. Given a vertex v , we define the orientation of an edge⁶ $e \in v$ with respect to v as

$$\varepsilon_v(e) \stackrel{\text{def}}{=} \begin{cases} -1 & \text{if } e \text{ is incoming,} \\ 1 & \text{if } e \text{ is outgoing.} \end{cases} \quad (6)$$

This convention is similar to the earlier choice in Equation (4) to call $\omega = 1$ the future. We define a *causal structure on the 1-skeleton* as an assignment of an orientation $\varepsilon_v(e)$ to each pair (v, e) such that $e \in v$ under the constraint.

In Equation (6), a global convention was chosen to attribute a numerical value to ‘incoming’ and ‘outgoing’. This convention can be made local by introducing at each vertex v a variable $\mu_v \in \{-1, 1\}$ and defining instead

$$\varepsilon_v(e) \stackrel{\text{def}}{=} \begin{cases} -\mu_v & \text{if } e \text{ is incoming} \\ \mu_v & \text{if } e \text{ is outgoing.} \end{cases} \quad (7)$$

In this case, the constraint becomes

$$\mu_{v_1} \varepsilon_{v_1}(e) = -\mu_{v_2} \varepsilon_{v_2}(e), \quad (8)$$

which is more similar to what can be found in [16].

$$\varepsilon_{v_1}(e) = -\varepsilon_{v_2}(e), \quad (9)$$

where v_1 and v_2 are the two endpoints of e . The latter condition expresses the fact that an incoming edge, with respect to one vertex, is outgoing with respect to the other.

3.2. Causal Wedges

We have seen that the causality of Δ can be read on the edges of Δ_1^* . Now, we are going to show that it can also be read equivalently on the wedges of the dual 2-skeleton Δ_2^* . To proceed, let us go back to the 4-simplicial complex Δ . A pair $w = (t, \sigma)$ such that $t \in \sigma$ is called a *wedge*. Given a wedge w , there exists exactly two tetrahedra $T_1, T_2 \in \sigma$ that share t , to which are associated the normals N_1 and N_2 . The *dihedral angle* of w is defined as

$$\theta_w \stackrel{\text{def}}{=} \text{sign}(N_1 \cdot N_2) \cosh^{-1}(|N_1 \cdot N_2|). \quad (10)$$

This definition is a natural extension of the notion of the dihedral angle from Euclidean to Minkowskian geometry. Its absolute value $|\theta_w|$ depends only on the absolute value of the scalar product $|N_1 \cdot N_2|$ of the normals. On the other hand, its sign depends on the relative time orientation of the two normals

$$\text{sign}(\theta_w) = \begin{cases} +\eta & \text{if } N_1 \text{ and } N_2 \text{ are co-chronal,} \\ -\eta & \text{if } N_1 \text{ and } N_2 \text{ are anti-chronal.} \end{cases} \quad (11)$$

When the normals are co-chronal (resp. anti-chronal), the wedge is said to be *thick* (resp. *thin*). At the level of the wedges, causality shows up as follows: a thick wedge encloses a time-like region, while a thin wedge encloses a space-like region (Figure 1). The thin/thick distinction provides an orientation of the wedges. However, this orientation does not extend to triangles because several wedges of the same triangle may not have the same orientation.

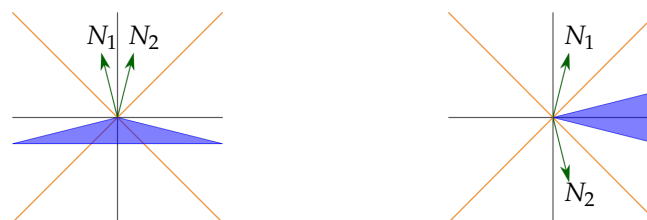


Figure 1. (Left): thick wedge, co-chronal normals. (Right): thin wedge, anti-chronal normals.

This notion translates easily on the dual complex. A pair of a face and a vertex (f, v) with $f \in v$ defines a (dual) wedge on the 2-skeleton Δ_2^* . There exists two unique edges e_1 and e_2 such that $e_1, e_2 \in f$ and $e_1, e_2 \in v$. The wedge is *thick* if e_1 and e_2 are both incoming

or both outgoing. It is *thin* otherwise. Algebraically, the wedge orientation can be defined as

$$\varepsilon_v(f) \stackrel{\text{def}}{=} \begin{cases} +\eta & \text{if thick,} \\ -\eta & \text{if thin.} \end{cases} \quad (12)$$

It is then easy to show that

$$\varepsilon_v(f) = \eta \varepsilon_v(e_1) \varepsilon_v(e_2). \quad (13)$$

As we have presented it, the wedge orientation is a byproduct of the edge orientation. However, is it possible to go the other way around and to compute the $\varepsilon_v(e)$ as a function of the $\varepsilon_v(f)$, i.e., to invert Equation (13)? The short answer is *no*, but not much information is actually missing to do this inversion.

Around the same vertex v , Equation (13) defines a system of 10 equations (one per face) with five unknowns (one per edge), so we may fear it to be over-constrained. However, the rank of the system (13) is only four. Indeed, given the orientation $\varepsilon_v(f)$ for any four wedges *that do not form a cycle*, one can deduce the orientation of the other six. A cycle is a sequence of faces that closes, bounding a 3D angle. Rather than a formal definition, this notion is best understood through few examples on the links of the vertex graph⁷ (Figure 2). Then, it is easy to show that the product of orientations along any cycle of wedges is

$$\prod_{f \in \text{cycle}} \varepsilon_v(f) = \eta^{\#\text{cycle}}, \quad (14)$$

where $\#\text{cycle}$ is the number of faces f in the cycle. Since any set of five wedges around v contains a cycle, then a four-wedge orientation is indeed sufficient to fix them all. Hence, the system (13) is actually under-constrained of exactly one dimension.

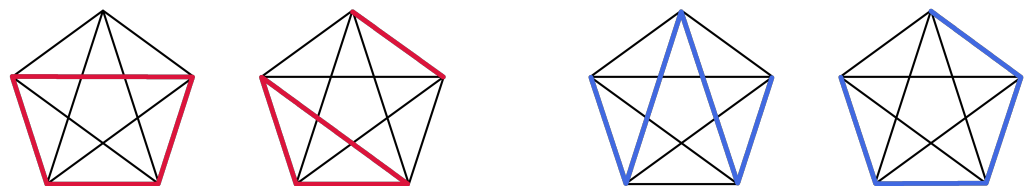


Figure 2. (Left) (Red): Two examples of sets of 4 wedges which form cycles. (Right) (Blues): Two examples of sets of 4 wedges which do *not* form cycles.

Given a vertex v , denote the surrounding edges by e_i , with $i \in \{1, \dots, 5\}$, and accordingly the surrounding faces by f_{ij} . To make the system invertible, let us add one independent equation by defining the orientation of the vertex v as follows:

$$\varepsilon_v \stackrel{\text{def}}{=} \prod_j \varepsilon_v(e_j). \quad (15)$$

Then, the set of Equation (13) augmented by (15), with unknowns $\varepsilon_v(e)$, is invertible, and one can show that

$$\varepsilon_v(e_i) = \varepsilon_v \prod_{k \neq i} \varepsilon_v(f_{ik}). \quad (16)$$

Thus, we see that we can recover the orientation of the edges from the orientation of the wedges, up to a vertex orientation.

Let us now consider a skeleton with many vertices. Can we deduce the orientation of the edges of the 1-skeleton from the orientation on the wedges of the 2-skeleton? From the previous analysis with a single vertex, we know that it will be possible to invert the system

of equations if one considers in addition one orientation ε_v per vertex. However, the set of ε_v is itself constrained because of the gluing condition (9), which now reads

$$\varepsilon_{v_1} \prod_{f|e \in f} \varepsilon_{v_1}(f) = -\varepsilon_{v_2} \prod_{f|e \in f} \varepsilon_{v_2}(f). \quad (17)$$

This constraint eliminates almost all the degrees of freedom introduced by ε_v , so that there remains finally only the freedom to fix the orientation of a single vertex in the whole skeleton. The orientation of all others can be deduced from it and the $\varepsilon_v(f)$. Indeed, assume that you have a 2-skeleton where all the $\varepsilon_v(f)$ have been fixed (satisfying the constraint along cycles). Then, if you fix only one ε_v , the orientation of the edges around v are fixed, and this “orientation fixing” will then propagate everywhere else, so the full set of ε_v will ultimately be fixed. Reversing the orientation of one ε_v , will reverse the entire skeleton, which corresponds to the time-reversal symmetry.

We have previously shown that the causal structure of discrete general relativity can be encoded in the dual 1-skeleton with oriented edges. Now, we have just seen that the *causal structure of a 2-skeleton* can be described as the assignation of $\varepsilon_v(f)$ to each wedge under the cycle constraint (14) and a global orientation ε , which can be regarded as a global arrow of time.

4. Lorentzian Regge Calculus

So far, we only focused on the kinematical aspects of causality. Now, we will see how causality shows up in the dynamics [34].

4.1. Lorentzian Regge Action

Following [16], we extend the formulae of [42] to the 4-dimensional case, the Lorentzian Regge action is a sum over the triangles t :

$$S_R \stackrel{\text{def}}{=} \sum_t A_t \delta_t, \quad (18)$$

with A_t the area of the triangle t and δ_t the deficit angle defined as a sum over the 4-simplices σ surrounding t :

$$\delta_t \stackrel{\text{def}}{=} \sum_{\sigma|t \in \sigma} \theta_{t\sigma}, \quad (19)$$

with the dihedral angle defined by equation (10). The order of the two sums can be exchanged:

$$S_R = \sum_{\sigma} \sum_{t|t \in \sigma} A_t \theta_{t\sigma}. \quad (20)$$

To derive the equations of motion by variational calculus, one should tell which are the independent variables of which S_R is a function of. In the original Regge calculus, it is shown that if the action is considered as a function of the lengths l_s of the segments s , the resulting equations of motions become Einstein equations in the continuous limit. However, this is not the only possible choice.

4.2. First-Order Regge Calculus

Barrett has proposed a formulation where the independent variables are both the lengths l_s and the angles $\theta_{t\sigma}$ [43]. This choice of variables mimics the Palatini formulation, which takes the metric and the torsion-less connection as primary fields of the Einstein–Hilbert action. Compared to the original Regge calculus, the introduction of $\theta_{t\sigma}$ extends the total number of variables. In order to recover the equations of motion, it is then necessary

to add constraints to the action, which is done with a Lagrange multiplier μ_σ per each 4-simplex σ . One obtains

$$S[l_s, \theta_{t\sigma}, \mu_\sigma] = \sum_\sigma \sum_{t|t \in \sigma} A_t(l_s) \theta_{t\sigma} + \sum_\sigma \mu_\sigma \det \gamma_\sigma \quad (21)$$

where γ_σ is the 5×5 matrix whose elements are the Minkowskian scalar products between the normals to the boundary tetrahedra:

$$[\gamma_\sigma]_{ij} = N_i \cdot N_j = \text{sign}(\theta_{ij}) \cosh \theta_{ij}, \quad (22)$$

where N_i is the (unit outward) normal to the i th boundary tetrahedron of σ and θ_{ij} is the dihedral angle between the tetrahedra i and j . The Lagrange multiplier imposes the constraint

$$\det \gamma_\sigma = 0 \quad (23)$$

which implements the closure of the normals, i.e.,

$$\sum_i V_i N_i = 0, \quad (24)$$

with V_i the volume of the i th tetrahedron.

4.3. Causal Structure from Dynamics

This choice of variables for the action makes it clear how causality plays its role in the dynamics. Indeed, as explained previously, the causal structure is encoded on the wedges as the sign of the dihedral angle between two neighbouring tetrahedra. This information can be directly obtained from the sign of the variable $\theta_{t\sigma}$. However, any configuration of signs of $\theta_{t\sigma}$ does not define an allowed causal structure, as it must also satisfy the cycle constraint (14). So, an assignment of $\theta_{t\sigma}$ defines an *orientation structure* but not necessarily a *causal structure*.

Moreover, it is also possible to obtain causal information from the lengths l_s . First of all, the sign of l_s tells whether the segment is space-like or time-like. Here, we are assuming that the segments are all space-like. Nevertheless, it is still possible to extract additional causal information. If some geometrical constraints are satisfied (generalisation of the triangular inequalities), then it is possible to reconstruct uniquely the geometry of a 4-simplex from the lengths of its segments. Then, there exists a formula expressing the dihedral angle as a function of the lengths l_s . This second derivation of the dihedral angle does not necessarily match with $\theta_{t\sigma}$ because $\theta_{t\sigma}$ and l_s are taken as independent variables in the first-order Regge calculus. So, the variables $\theta_{t\sigma}$ and l_s define independently two coexisting notions of causal structure on the wedges.

Of course, the two notions of causal structure must match when the equations of motion are imposed. In particular, the cycle constraint appears as a corollary of the equations of motion. Indeed, the constraint $\det \gamma_\sigma = 0$ implies the existence of a vector v such that $\sum_i \gamma_{ij} v_i = 0$. Then, the equation of motion obtained by varying θ_{ij} yields

$$A_t = \kappa_\sigma \mu_\sigma v_i v_j \sinh \theta_{ij} \quad (25)$$

for some $\kappa_\sigma \in \mathbb{R}$ (see [43]). Since $A_t > 0$,

$$\text{sign } \theta_{ij} = \text{sign}(\kappa_\sigma \mu_\sigma) \text{sign } v_i \text{sign } v_j \quad (26)$$

which implies the constraints (14). This shows that the equations of motion impose the structure of the wedges to be causal. This result should be stressed. In the standard metric formulation of general relativity, the causal structure is already well defined at the level of the kinematics because any metric defines a causal structure. In the first-order Regge calculus, we see that the causal structure does not necessarily exist at the kinematical level,

but the equations of motion impose the cycle constraint, which selects a surface of variables where every configuration has a well-defined causal structure.

5. Causal Path Integral

So far, the analysis has been purely classical, although discrete. When going to the quantum regime of gravity, it is reasonable to expect that the generic state of the metric field is a superposition of classical configurations. In particular, there might not be a definite causal structure. Several causal histories may interfere and thus generate in principle observable effects. A theory of quantum gravity should be able to predict these effects through the computation of transition amplitudes between different states of space.

Let us clarify the main ideas by proceeding heuristically, although a precise mathematical formulation may be more difficult to achieve. The standard procedure starts by foliating the space–time manifold into constant-time slices: $\mathcal{M} \cong \Sigma \times \mathbb{R}$. The classical states are 3-metrics h defined over Σ . At the quantum level, the 3-metric is an operator $\hat{h}$, with eigenstates $|h\rangle$ whose eigenvalues are the classical 3-metrics h . The sum-over-histories approach to quantum gravity [44–46] proposes to compute the transition amplitude between the state $|h_0\rangle$ at time t_0 and the state $|h_1\rangle$ at time t_1 as a path integral:

$$\langle h_1 | h_0 \rangle = \int [\mathcal{D}g] e^{\frac{i}{\hbar} S[g]}, \quad (27)$$

where $[\mathcal{D}g]$ is a measure on the set of 4-metrics g over $\Sigma \times [t_0, t_1]$, such that the restriction of g to the slice $t = t_0$ (resp. $t = t_1$) is h_0 (resp. h_1) and $S[g]$ is the Einstein–Hilbert action evaluated on such a metric.

5.1. General Boundary Formulation

The previous and standard formulation is not ideal because it relies upon a slicing of space–time into constant-time leaves, which may already fix too much structure for a general treatment of causality. Better suited for our purpose is the general boundary formulation developed by Oeckl in [47]. Consider a region of space–time M with boundary Σ . A 4-metric g on M induces a 3-metric h on Σ . The crux of a quantum theory of space–time is the computation of the *metric propagator*:

$$Z_M(h) = \int [\mathcal{D}g] e^{\frac{i}{\hbar} S[g]} \quad (28)$$

where the integral is carried over all the 4-metrics g bounded by h . The standard formulation (Equation (27)) is recovered when Σ is made of two disconnected components (past and future). In general, the boundary will have space-like, time-like and possibly null components. Here, we restrict attention to a finite space–time region with a boundary consisting of two space-like components. This lens-shaped space–time region is foliated by finite space-like leaves that meet at a fixed 2-dimensional corner.

5.2. Regge Path Integral

As a way towards the actual computation of the metric propagator, one can discretise the previous formula. Consider a 4-simplicial complex Δ . Its boundary is a 3-simplicial complex Σ . Working with the Regge action, the metric propagator is a function of the length of the segments of Σ , and it reads as follows:

$$\mathcal{A}_\Delta(l_\Sigma) = \int [dl_s] e^{\frac{i}{\hbar} S_R[l_s]}. \quad (29)$$

The integral is done over the lengths l_s of all the segments s in the bulk of Δ . Note that each integral could also be replaced by a sum with a cut-off in order to ensure a finite value to the propagator, but it does not seem useful in our quest for causality.

Using the first-order Regge calculus, the propagator reads:

$$\mathcal{A}_\Delta(l_\Sigma, \theta_\Sigma, \mu_\Sigma) = \int [dl_s][d\theta_{t\sigma}][d\mu_\sigma] \prod_\sigma e^{\frac{i}{\hbar}(\sum_{t \in \sigma} A_t \theta_{t\sigma} + \mu_\sigma \det \gamma_\sigma)}. \quad (30)$$

The integration over μ_σ can be formally carried over, which gives a δ -function that fixes the constraint:

$$\mathcal{A}_\Delta(l_\Sigma, \theta_\Sigma) = \int [dl_s][d\theta_{t\sigma}] \prod_\sigma \delta(\det \gamma_\sigma) e^{\frac{i}{\hbar} \sum_{t \in \sigma} A_t \theta_{t\sigma}}. \quad (31)$$

5.3. Causal Structure of the Boundary

A causal structure on Δ induces a causal structure on its boundary Σ . It indicates for each tetrahedron of the boundary whether it shall be regarded as future or past. It can be represented on the boundary of the dual 2-skeleton Δ_2^* , which is a 4-valent graph. The induced causal structure consists of assigning a sign to each node, depending on whether the edge attached to it is pointing inside or outside the bulk. Conventionally, we take the sign to be positive for an outward edge (future tetrahedron) and negative for an inward edge (past tetrahedron). An example is shown in Figure 3. It is important to notice that fixing a causal structure on the boundary does not in general impose a single causal history in the bulk: many different histories may share the same causal boundary.

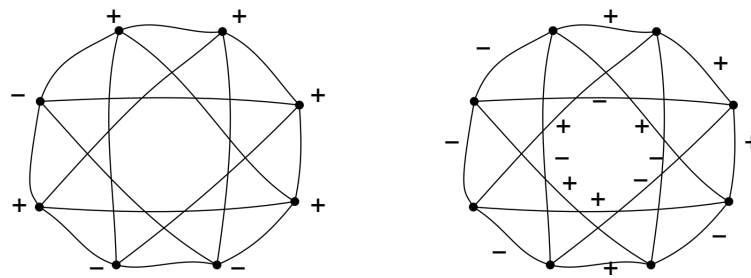


Figure 3. Example of causal structure on a boundary graph. (Left): encoded on the nodes. (Right): encoded on the links.

The causal information of the boundary can also be encoded on the links. To each link, one associates the product of the sign of the endpoints (Figure 3). For a 4-valent graph, there are twice as many links as there are nodes. But despite the double number of variables, this encoding is not injective but 2-to-1. Physically, by encoding causality on the links, we only provide information about bare causality, while encoding over the nodes also gives a time orientation.

A random assignment of signs to links only defines a causal structure on the boundary if it enables consistent assigning of signs to the nodes. This happens if, and only if, the signs on the links satisfy the constraint that their product around any loop of the graph is 1. The latter constraint is implied by the cycle condition (14) in the bulk. In fact, when two edges crossing the boundary share a common vertex, the sign of the wedge matches the sign of the link between the two corresponding nodes. More generally, the sign of the link is equal to the product of the signs of the wedges around the corresponding face in the bulk, which can be written as

$$\varepsilon_l = \prod_{v \in f} \varepsilon_v(f), \quad (32)$$

where f is the face that intersects the boundary along the link l .

5.4. Causal Amplitude

The metric propagator is a function of the boundary variables. In concrete situations, the bare-causal structure of the boundary may be fixed by an assignment of link orientations ε_l . In this case, the range of integration on the angles $\theta_{t\sigma}$ in Equation (31) must be restricted

for the wedges that belong to faces intersecting the boundary. This restriction consists of implementing the constraint (32). For instance, consider a link l bounding a face f (dual to t) with only one vertex v (dual to σ). If $\varepsilon_l = 1$ (resp. $\varepsilon_l = -1$), then the integration over $\theta_{t\sigma}$ shall be carried over $\mathbb{R}^+$ (resp. $\mathbb{R}^-$), instead of $\mathbb{R}$.

For the $\theta_{t\sigma}$ in the bulk, the integration is still done over all $\mathbb{R}$. We can rewrite the amplitude (31) as

$$\mathcal{A}_\Delta(l_\Sigma, r_\Sigma, \varepsilon_\Sigma) = \sum_{[\varepsilon_{t\sigma}]} \int [dl_s][dr_{t\sigma}] \prod_\sigma \delta(\det \gamma_\sigma) e^{\frac{i}{\hbar} \sum_{t \in \sigma} A_t \varepsilon_{t\sigma} r_{t\sigma}}, \quad (33)$$

where the sum is done over all possible orientations $\varepsilon_{t\sigma}$ of wedges in the bulk, compatible with the bare causality of the boundary ε_Σ , and the integration in r is performed over $\mathbb{R}^+$. Thus, the path integral is summing over both configurations that satisfy and do not satisfy the cycle condition (14).

To be more precise, at each vertex, the surrounding wedge orientations can satisfy the cycle condition for $\eta = 1$, $\eta = -1$ or none of them. The two first cases correspond to the possibility of locally defining a light cone. These light cones are only local because the value of the signature η may disagree from one 4-simplex to another. There is a global causal structure only when the signature is the same for all 4-simplices. Thus, the sum over orientations can be decomposed into several terms:

$$\sum_{[\varepsilon_{t\sigma}]} = \sum_{\substack{[\varepsilon_{t\sigma}] \\ \text{causal } \eta=1}} + \sum_{\substack{[\varepsilon_{t\sigma}] \\ \text{causal } \eta=-1}} + \sum_{\substack{[\eta_\sigma] \\ \text{signature} \\ \text{changes}}} \sum_{\substack{[\varepsilon_{t\sigma}] \\ \text{local} \\ \text{light-cones}}} + \sum_{\substack{[\varepsilon_{t\sigma}] \\ \text{spurious}}} \quad (34)$$

The latter terms gather “non-causal” or “spurious” configurations in the sense that there is at least one 4-simplex for which the set of wedge orientations $\varepsilon_{t\sigma}$ do not define a consistent local light cone. The presence of such non-causal histories contributing to the amplitude is the consequence of the peculiar choice of variables of the first-order Regge calculus using both the lengths l_s and the dihedral angles $\theta_{t\sigma}$. In the standard Regge calculus, only the lengths l_s are used; thus, the path integral is only summing over causal configurations.

However, in the classical limit, when $\hbar \rightarrow 0$, the configurations that contribute the most are the stationary points of the action, which satisfy the classical equations of motion and thus, as seen previously, have a proper causal structure. So the causal configurations are selected in the classical limit. These configurations can be partitioned in two subsets depending on the choice of signature $\eta \in \{-1, 1\}$ for which the cycle condition (14) is satisfied.

In this framework, the existence of a consistent causal structure is an emerging feature of space–time. It originates from orientation degrees of freedom located on the wedges. At the quantum level, only a minority of configurations define proper light cones. The unveiling of this structure offers the possibility of defining alternative amplitudes by fixing some of the degrees of freedom.

For instance, one could keep in (33) only the terms $[\varepsilon_{t\sigma}]$ that satisfy the cycle condition (14). One could also further restrict the sum by imposing a choice of signature η . Eventually, one could keep only a single term that has a fixed causal structure.

The suggestion of considering such restrictions in the range of the path integral was initially suggested by Teitelboim [48] and first applied to spin foams by Livine and Oriti [16]. In the context of the standard formulation (equation (27)), h_0 is regarded as “past” and h_1 as “future”. So Teitelboim proposed to restrict the range of integration over the 4-metrics g for which the proper time from the first to the second slice is positive. Working in the ADM formalism, it amounts to restricting the range of integration of the lapse N to positive values only. The resulting “causal amplitude” is not anymore gauge-invariant, i.e., it is not a solution of the Hamiltonian constraint.

However, Teitelboim argued that such a restriction could be worth considering by drawing an analogy with the propagator of the free relativistic particle. Indeed, the ampli-

tude (33) can be seen as the analog of the Hadamard propagator, which is a symmetrized 2-point correlation function solving the Klein–Gordon equation. However, to compute the transition probability between two positions and times, one must use the Feynman propagator, which is the time-ordered Green function of the Klein–Gordon equation. Both propagators can be computed using the path integral method. It is then shown that they differ by a different range of integration. In the Feynman propagator, one only sums over these trajectories to evolve forward in time (time-ordered), which appears as a restriction to the range of the path integral giving the Hadamard function.

In a nutshell, the choice of including the causal terms or not depends on what we want to compute: a projector on the physical Hilbert space or an evolution operator.

6. BF Theory

As a first step toward spin foams, let us consider discrete BF theory. It is a topological theory, so we do not expect any causal structure to arise, but it makes use of an orientation structure that is worth looking at as a warm-up.

6.1. Discrete BF Theory

Following [49], the discretization of BF theory is done over a 2-complex $\mathcal{C}$. The variables are one group element $g_e \in G$ per edge $e \in \mathcal{C}$. Then, the amplitude is defined as

$$Z_{\mathcal{C}} \stackrel{\text{def}}{=} \int [\mathrm{d}g_e] \prod_f \delta(U_f(g_e)). \quad (35)$$

There is one integral per edge, $\mathrm{d}g_e$ is the Haar measure over G , δ is the Dirac δ -function over G and the product is carried over all the faces f of $\mathcal{C}$. Moreover, we define the circular product

$$U_f(g_e) \stackrel{\text{def}}{=} \prod_{e \in f}^{\circlearrowleft} g_e, \quad (36)$$

where the product is made over the edges e surrounding f . Although sometimes overlooked, we want to draw attention to the orientation structure that is required to define correctly the circular product. We need the following additional structure over $\mathcal{C}$:

1. a distinguished edge to each face that serves as a starting point in the product;
2. an orientation to each face that tells the order of the following edges.

Although this structure is required to define U_f , $\delta(U_f)$ does not actually depend on it, due to the invariance of the δ -function under inversion and cyclic permutation.

It is common to rewrite $Z_{\mathcal{C}}$ by splitting the δ -function into a sum over the irreducible representations (irreps) of G :

$$\delta(U) = \sum_{\rho} \dim \rho \operatorname{Tr} \rho(U) \quad (37)$$

Then, for each edge, the integral over the group element g can be rewritten as a sum over intertwiners, which formally reads

$$\int \mathrm{d}g \bigotimes_{f \in e}^{\circlearrowleft} \rho_f(g) = \sum_i u^*. \quad (38)$$

Again, the definition of the circular tensor product requires us to introduce an additional structure:

3. a distinguished face to each edge that serves as a starting point in the tensor product;
4. an orientation of the faces around each edge, which can be thought of as an arrow on the edge (with the right-hand convention to turn around for instance).

Of course, Z_C remains blind to this structure. The amplitude finally becomes:

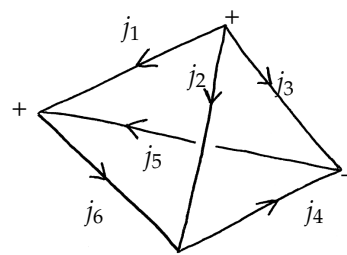
$$Z_C = \sum_{\rho} \sum_{\iota} \prod_f \dim \rho_f \prod_v A_v. \quad (39)$$

The sum in ρ (resp. ι) is made over all the possible labeling of the faces (resp. edges) by irreps (resp. intertwiners). The *vertex amplitude* A_v is a function of the irreps ρ_f and intertwiners ι_e attached to the faces and edges surrounding a vertex v .

The four orientation structures just introduced enter the computation of A_v . These structures lie on the edges and faces of the 2-complex. Motivated by our analysis in the previous sections, we want to investigate the idea that causality of spin foams could arise from the breaking of the invariance of Z_C with respect to the orientation of C .

6.2. Ponzano–Regge Model

To proceed concretely, let us consider the simple example of the *Ponzano–Regge model* [50], for which C is dual to a 3-dimensional simplicial complex Δ , and $G = SU(2)$. In this case, the irreps are labeled by spins $j \in \mathbb{N}/2$, and there is no sum over the intertwiners (because it is unique). The vertex amplitude can be nicely represented pictorially as a graph where each node stands for an adjacent edge and each link for a face in-between. The *vertex graph* takes typically the following form:



(40)

The arrows on the links are induced by the orientation of the faces, and the signs on the nodes are induced by the orientation of the edges (+ for incoming). Any combination of arrows and signs can be found, but the topology of the graph is the same for every vertex. The labels j on the links are inherited from the irreps on the faces.

The graphical calculus is defined by the following rules:

1. To each link l , associate a variable m_l that will be summed over;
2. The 3jm-Wigner symbol⁸ is associated with the following nodes:

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = \begin{array}{c} j_3 \quad j_2 \quad j_1 \\ \diagdown \quad \diagup \quad \diagup \\ \text{+} \end{array} = \begin{array}{c} j_1 \quad j_2 \quad j_3 \\ \diagdown \quad \diagup \quad \diagdown \\ \text{-} \end{array} \quad (41)$$

The sign on the node indicates the sense in which the attached links shall be read.

3. If an arrow is reversed, replace in the formula above m_l with $-m_l$ and multiply by $(-1)^{j_l - m_l}$, like

$$\begin{array}{c} j_3 \quad j_2 \quad j_1 \\ \diagdown \quad \diagup \quad \diagup \\ \text{+} \end{array} = (-1)^{j_3 - m_3} \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & -m_3 \end{pmatrix} \quad (42)$$

or

$$\begin{array}{c} j_1 \\ \swarrow \\ \downarrow j_2 \\ \searrow j_3 \end{array} = (-1)^{j_1-m_1} \begin{pmatrix} j_1 & j_2 & j_3 \\ -m_1 & m_2 & m_3 \end{pmatrix} \quad (43)$$

A positive (resp. negative) node with an incoming (resp. outgoing) link corresponds to a counter-alignment of the face and the edge.

4. Multiply all factors and sum over all m_l from $-j_l$ to j_l (integer steps).

As an example, graph (40) evaluates to

$$A_v = \sum_{m_i} (-1)^{j_4 - m_4 + j_1 - m_1} \begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \times \begin{pmatrix} j_4 & j_5 & j_3 \\ m_4 & -m_5 & m_3 \end{pmatrix} \begin{pmatrix} j_6 & j_2 & j_4 \\ m_6 & m_2 & -m_4 \end{pmatrix} \begin{pmatrix} j_6 & j_5 & j_1 \\ m_6 & -m_5 & -m_1 \end{pmatrix} \quad (44)$$

The power of graphical calculus is apparent when comparing this cumbersome formula to diagram (40). Up to a sign, the vertex amplitude equals the 6j-symbol:

$$A_v = \pm \begin{Bmatrix} \dot{j}_1 & \dot{j}_2 & \dot{j}_3 \\ \dot{j}_4 & \dot{j}_5 & \dot{j}_6 \end{Bmatrix}. \quad (45)$$

The sign $\pm$ is a function of the spins j_i , and it depends on the orientation of the links and nodes. It matters when several vertices are glued together.

The importance of the Ponzano–Regge model was historically revealed by its semi-classical limit, which makes it a good candidate for 3D Euclidean quantum gravity [50]. Indeed, the vertex amplitude admits a graphical representation as a tetrahedron depicted in (40). This shape initially expresses the invariance of the $6j$ -symbol under the action of the tetrahedral group. But it turns out that it also carries a deeper geometric meaning when the labels j_i are interpreted as the edge lengths of the tetrahedron. Denoting V the volume of this tetrahedron, one can prove the following behavior for the vertex amplitude $A_v(\lambda j_i)$ when $\lambda \rightarrow \infty$:

$$\left\{ \begin{array}{ccc} \lambda_{j_1} & \lambda_{j_2} & \lambda_{j_3} \\ \lambda_{j_4} & \lambda_{j_5} & \lambda_{j_6} \end{array} \right\} \sim \frac{1}{4\sqrt{3}\pi\lambda^3 V} (e^{iS} + e^{-iS}) \quad (46)$$

with the action

$$S \stackrel{\text{def}}{=} \sum_i \left(\lambda_{j_i} + \frac{1}{2} \right) \zeta_i + \frac{\pi}{4} \quad (47)$$

with ξ_i the exterior dihedral angle along the edge i [52].

Graphical calculus also clarifies how the orientation enters the computation of the vertex amplitude. The invariance of the total amplitude $Z_{\mathcal{C}}$ under changes of orientation of the edges and faces is checked in Appendix A. To be precise, the invariance is only true for faces and edges, which lie in the bulk of $\mathcal{C}$. In general, $\mathcal{C}$ is bounded by some 3-valent graph Γ over which an orientation is induced by $\mathcal{C}$. The total amplitude $Z_{\mathcal{C}}$ is sensitive to the orientation of Γ . The case is similar to the amplitude defined by Equation (33). Although topological in the bulk, BF theory is non-trivial on the boundary. The boundary orientation provides a prototype of boundary causal structure.

A change in boundary orientation affects the value of Z_C in a simple way:

- A flip of a link orientation brings a global factor $(-1)^{2j}$ if the two endpoints carry opposite signs, none otherwise.
- A flip of a node orientation brings an overall factor $(-1)^{j_1+j_2+j_3}$.

As for Feynman diagrams, this simple way of modifying the causal structure of the boundary can be understood as a *crossing symmetry*.

6.3. Causal Ponzano–Regge Model

Now, we want to go further and suggest a way to break the orientation invariance in the bulk of the Ponzano–Regge model. A proposal of this kind can be found in [22], starting from a formulation of causality in terms of the flux variables of the discretized BF theory. Here, we adopt a different strategy, using the spin representation, motivated by the semi-classical limit. It exemplifies the strategy that will be adopted in the next section for the EPRL model [35].

Consider $\mathcal{H}_j$, the spin- j irreducible representation of $SU(2)$, with the canonical basis $|jm\rangle$. The coherent states are defined as

$$|j, z\rangle \stackrel{\text{def}}{=} u(z)|j, -j\rangle \quad (48)$$

with $u : \mathbb{C}^2 \rightarrow SU(2)$ a (well-chosen) surjective map. Then, the intertwiner can be written as the state

$$|\iota\rangle = \frac{1}{N} \int dg \, g \cdot \bigotimes_{i=1}^3 |j_i, z_i\rangle \quad (49)$$

with $N = N(j_i, z_i)$ such that $\langle \iota | \iota \rangle = 1$. Up to a global phase, the vertex amplitude is

$$A_v = \frac{1}{N} \int_{SU(2)} [dg_n] \prod_l K_l(g_{t_l}, g_{s_l}). \quad (50)$$

t_l and s_l are, respectively, the source and the target of l . The wedge amplitude is defined as

$$K(g_t, g_s) \stackrel{\text{def}}{=} \langle j, z_t | g_t^{-1} g_s | j, z_s \rangle = \langle z_t | g_t^{-1} g_s | z_s \rangle^{2j}. \quad (51)$$

For the purpose of studying the semi-classical limit, it is convenient to write the wedge amplitude as

$$K(g_t, g_s) = e^{2j(\log r + i\theta)} \quad (52)$$

with $r \in \mathbb{R}^+$ and $\theta \in (-\pi, \pi]$.

Now, to introduce a notion of causality, we can force some orientation structure to appear artificially: this is done here by writing the identity

$$K(g_t, g_s) = \sum_{\varepsilon \in \{1, -1\}} \Theta(\varepsilon\theta) e^{2j(\log r + i\theta)}, \quad (53)$$

where Θ is the step function and we are summing over the signs ε . We then define the causal wedge amplitude as

$$K^\varepsilon(g_t, g_s) \stackrel{\text{def}}{=} \Theta(\varepsilon\theta) e^{2j(\log r + i\theta)}. \quad (54)$$

for some choice of ε , which can be understood as a choice of wedge orientation. Given one orientation ε_l per wedge l , the causal vertex amplitude A_v^ε is defined by replacing the wedge amplitude K_l by its causal alternative $K_l^{\varepsilon_l}$ in Equation (50). The BF vertex amplitude is recovered as

$$A_v = \sum_{[\varepsilon_l]} A_v^{\varepsilon_l}, \quad (55)$$

where the sum is made over all possible sign assignments to the wedges. There is a total of 2^6 such configurations. This sum introduces a partition of the range of integration of (50) into as many sectors. In the semi-classical limit, only two sectors survive, as it appears in (46). Since the exterior dihedral angles of any tetrahedron are always such that $\sin \tilde{\xi} \leq 0$,

the two sectors are when ε_I are either all positive or all negative. Starting with the causal vertex amplitude with all ε_I negative thus leads to the asymptotic limit

$$A_v^- \sim \frac{1}{4\sqrt{3\pi V}} e^{iS}. \quad (56)$$

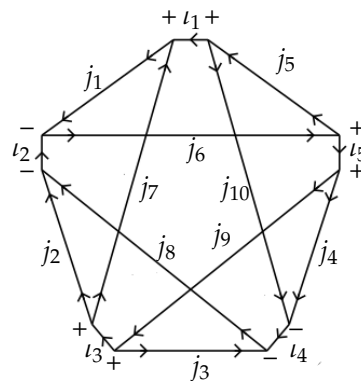
This provides a toy model for the appearance of causality, which we will apply to the EPRL model.

6.4. $\{15j\}$ BF Theory

Before moving to the EPRL model, let us look at BF theory in 4 dimensions. The main difference with respect to the $3d$ case comes from the fact that Δ is a 4-dimensional simplicial complex; thus, the amplitude Z_C includes a sum over the intertwiners. The graphical representation of the intertwiners requires the introduction of an additional structure:

5. at each edge, the surrounding faces are partitioned into two sets of two (there exist three such partitions);
6. these two sets are ordered (e.g., called left and right).

Then, the vertex amplitude is represented by a pentagram like



(57)

$\iota \in \mathbb{N}/2$ labels the intertwiners. It is surrounded by two positive (resp. negative) nodes, when the edge is incoming (resp. outgoing). When the nodes are positive (resp. negative), the arrow goes from the right set to the left set (resp. the other way around). One can check with the rules above that the overall amplitude Z_C is insensitive to the additional structure introduced. The same procedure as before can be used to define the causal vertex amplitude A_v^ε .

To sum up, discrete BF theory is defined over a 2-complex $\mathcal{C}$ with several auxiliary orientation structures. The value of the partition function Z_C is sensitive to the boundary values of these structures but not in the bulk. However, we have made a proposal to select a causal structure in the bulk as well.

7. EPRL Model and Its Causal Structure

General relativity can be formulated in a language close to the one of BF theory. The essential difference is that GR has local degrees of freedom, which mathematically arise from the implementation of constraints on the BF variables. Spin foam models build upon this insight and consist of a weak implementation of these constraints in the discrete BF theory.

The appearance of the local degrees of freedom shows up in the breaking of the topological invariance of BF theory. The question then arises if these constraints induce also a causal structure on the 2-complex. The analysis of the previous sections suggests that such a causal structure can arise as an orientation structure on the edges or the wedges.

Imposing constraints on these orientations then breaks the bulk orientation invariance of BF theory discussed above. In this section, we propose such a construction for the EPRL model.

7.1. Lorentzian EPRL Model

The EPRL model [35], as formulated in [53], is defined over a 2-complex $\mathcal{C}$ by the following partition function:

$$Z_{\mathcal{C}} = \int_{SU(2)} [dh_w] \prod_f \delta(U_f) \prod_v A_v(h_w). \quad (58)$$

The integral is made over the variables $h_w \in SU(2)$ associated with each wedge w in the bulk of $\mathcal{C}$. $Z_{\mathcal{C}}$ is a function of the variables $h_l \in SU(2)$ associated with each link of the boundary graph Γ . Similarly to BF theory, the precise definition requires an additional structure on $\mathcal{C}$:

1. a starting wedge per each face;
2. an orientation per each face;
3. an orientation to each wedge, i.e., each wedge w has a source edge s_w and a target edge t_w ;
4. a distinguished edge E_v per each vertex v .

Then, U_f is defined as the circular product

$$U_f(h_w) \stackrel{\text{def}}{=} \prod_{w \in f}^{\circlearrowleft} h_w \quad (59)$$

that starts with the starting wedge of f , circulates in the sense given by the orientation of f , and each h_w is inverted when the orientation of the wedge does not match with the orientation of the face. Besides, the vertex amplitude is

$$A_v(h_w) = \int_{SL_2(\mathbb{C})} [dg_e] \delta(g_{E_v}) \prod_{w \in v} K(h_w, g_{s_w} g_{t_w}^{-1}) \quad (60)$$

where there is one integration over $SL_2(\mathbb{C})$ per each edge surrounding v . The $\delta(g_{E_v})$ is only here to make the integral finite, but the value of A_v is actually independent of the choice of E_v . The wedge amplitude K is a function over $SU(2) \times SL_2(\mathbb{C})$ given by

$$K(h, g) = \sum_j (2j+1)^2 \int_{SU(2)} dk \overline{\chi^j(hk)} \chi^{\gamma j}(kg), \quad (61)$$

with χ^j and $\chi^{p,k}$, respectively, the characters of $SU(2)$ and $SL_2(\mathbb{C})$ and $\gamma \in \mathbb{R}$ the Barbero–Immirzi parameter.

The success of the EPRL model lies in its semi-classical limit, which was studied in [32,54] for a 2-complex $\mathcal{C}$ made of a single vertex dual to a Lorentzian 4-simplex σ . In this case, the partition function reduces to a single vertex amplitude $A_v(h_w)$, function of one $SU(2)$ element h_w per each of the 10 links of the boundary graph. The boundary of σ is made of 5 tetrahedra, which can be described by the areas j and the normals $\vec{n}$ to their faces. Then, the kinematics of loop quantum gravity prescribes a construction of a coherent state $\Psi_{j,\vec{n}}(h_w)$ that is “peaked” on the boundary geometry of σ . In the limit of large areas, $\lambda \rightarrow \infty$,

$$\begin{aligned} \langle \Psi_{\lambda j, \vec{n}} | A_v \rangle &\stackrel{\text{def}}{=} \int_{SU(2)} [dh_w] \Psi_{j, \vec{n}}^*(h_w) A_v(h_w) \\ &\sim \frac{1}{\lambda^{12}} \left(N_{\sigma} e^{i\lambda S_R} + N_{P\sigma} e^{-i\lambda S_R} \right) \end{aligned} \quad (62)$$

with $N_\sigma, N_{P\sigma} \in \mathbb{R}$ and S_R is the Lorentzian Regge action of σ . This result takes a form similar to (46) for the semiclassical limit of the Ponzano–Regge model.

The EPRL model is blind to the bulk orientation structure that has been introduced to define it. Indeed, changing the starting wedge or the orientation of a face f changes the value of U_f but not of $\delta(U_f)$. Moreover, reversing the orientation of a wedge w changes both U_f and A_v . In U_f , it replaces h_w with h_w^{-1} . In A_v , it interchanges s_w and t_w , and so $K(h_w, g_{s_w} g_{t_w}^{-1})$ becomes $K(h_w, g_{t_w} g_{s_w}^{-1})$, which is easily shown to be equal to $K(h_w^{-1}, g_{s_w} g_{t_w}^{-1})$. A change in variables $h_w \rightarrow h_w^{-1}$ within the integral finally proves that the value of Z_C remains unchanged.

The model is nevertheless sensitive to the orientation structure on the boundary. This structure consists only of the orientation of the links on the boundary over which the variables h_l live. In the light of our preceding analysis, we expect this structure to carry causal information.

7.2. Causal Structures in the EPRL Model

Analogously to our previous toy model for BF theory, one can break the orientation invariance in the bulk as follows. First, one performs the integral over $k \in SU(2)$ in (61), which yields

$$K(h, g) = \sum_{j, m, n} (2j+1) \overline{D_{mn}^j(h)} D_{jmn}^{\gamma j, j}(g). \quad (63)$$

One can switch from the magnetic basis $|jm\rangle$ to the overcomplete coherent state basis $|j, z\rangle$ by introducing a resolution of the identity

$$\mathbb{1}_j = \frac{2j+1}{\pi} \int_{\Gamma} \frac{\Omega(z)}{\|z\|^4} |j, z\rangle \langle j, z| \quad (64)$$

with the measure

$$\Omega(z) \stackrel{\text{def}}{=} \frac{i}{2} (z_0 dz_1 - z_1 dz_0) \wedge (z_0^* dz_1^* - z_1^* dz_0^*). \quad (65)$$

Here, Γ is the image of a path $\mathbb{CP}^1 \rightarrow \mathbb{C}^2$ that crosses each vector line of $\mathbb{C}^2$ once and only once⁹. One gets

$$K(h, g) = \sum_j \frac{(2j+1)^3}{\pi^2} \int_{\Gamma' \times \Gamma''} \frac{\Omega(z') \Omega(z'')}{\|z'\|^4 \|z''\|^4} \langle j, z' | h^\dagger | j, z'' \rangle \langle \gamma j, j, j, z'' | D^{\gamma j, j}(g) | \gamma j, j, j, z' \rangle. \quad (66)$$

Following [32,54], the $SL_2(\mathbb{C})$ matrix element appearing above can be expressed in terms of an auxiliary $\mathbb{CP}^1$ as

$$\langle \gamma j, j, j, z'' | D^{\gamma j, j}(g) | \gamma j, j, j, z' \rangle = \frac{2j+1}{\pi} \int_{\Gamma} \frac{\Omega(\zeta)}{\|\zeta\|^4} \mathcal{A} e^{iS_\gamma}, \quad (67)$$

where

$$\mathcal{A} = \frac{\langle \zeta | z''^* \rangle^{2j} \langle z'^* | g^T \zeta \rangle^{2j} e^{2ij(\arg z'_1 - \arg z'_1)}}{\|z'\|^{2j} \|z''\|^{2j} \|\zeta\|^{2j-2} \|g^T \zeta\|^{2j+2}} \quad (68)$$

and the wedge action

$$S_\gamma = \gamma j \log \frac{\|g^T \zeta\|^2}{\|\zeta\|^2}. \quad (69)$$

In the semi-classical, the wedge action becomes proportional to the wedge dihedral angle. This suggests defining the causal wedge amplitude as

$$K_\epsilon(h, g) = \sum_j \frac{(2j+1)^4}{\pi^3} \int \frac{\Omega(\zeta) \Omega(z') \Omega(z'')}{\|\zeta\|^4 \|z'\|^4 \|z''\|^4} \langle z' | h^\dagger | z'' \rangle^{2j} \Theta(\epsilon S_\gamma) \mathcal{A} e^{iS_\gamma}. \quad (70)$$

The causal vertex amplitude A_v^ε is then defined by replacing K with K_ε in Equation (60), for some choice of ε on each wedge. This defines a causal EPRL model.

The epithet ‘causal’ can be further motivated by showing indeed that the extra variable $\varepsilon \in \{1, -1\}$ encodes a causal structure on the wedges, as described previously.

First, the usual vertex amplitude is recovered by summing over all possible sign assignments to wedges:

$$A_v = \sum_{[\varepsilon_w]} A_v^\varepsilon. \quad (71)$$

There are 2^{10} terms in the sum. If one interprets ε as wedge orientations, then a configuration only properly defines a causal structure if the cycle condition (14) is fulfilled. So, one can properly call A_v^ε a “causal vertex amplitude” when the configuration $[\varepsilon_w]$ satisfies (14), for either choice of signature $\eta \in \{-1, 1\}$.

Secondly, assuming that the $[\varepsilon_w]$ indeed defines a proper causal structure, the word “causal” for A_v^ε is only deserved if, in the asymptotic limit, ε really captures the causal orientation of the boundary state. More precisely, given a Lorentzian 4-simplex, it determines a set of $\tilde{\varepsilon}$ for each wedge, such that $\tilde{\varepsilon}_w = \text{sign } \theta_w$, where θ_w is the dihedral angle of the wedge w . Then, its boundary coherent state $\Psi_{j,\vec{n}}$, when contracted with A_v^ε , should have the expected classical limit, i.e.,

$$\langle \Psi_{\lambda j, \vec{n}} | A_v^\varepsilon \rangle \sim \frac{1}{\lambda^{12}} N_\sigma e^{i\lambda S_R} \quad (72)$$

One can check that this is the case. Indeed, $\langle \Psi_{\lambda j, \vec{n}} | A_v \rangle$ is an integral over the variables g, z, z' . The sum (71) introduces a partition of the range of integration in a number of sectors $[\varepsilon_w]$ characterized by $\varepsilon_w S_\gamma(x_w) > 0$, where x_w stands for all the variables g, z, z' on the wedge w . In the asymptotic limit, the two terms of (62) arise from two stationary points that are related by a parity transformation. The asymptotic analysis of [54] shows that one of them, denoted σ , is such that $\tilde{\varepsilon}_w S_\gamma(x_w^\sigma) > 0$. Then, the other, denoted $P\sigma$, satisfies $S_\gamma(x_w^{P\sigma}) = -S_\gamma(x_w^\sigma)$, so that $P\sigma$ and σ are in opposite sectors. By construction, $\langle \Psi_{\lambda j, \vec{n}} | A_v^\varepsilon \rangle$ selects only the sector of σ .

Our analysis suggests interpreting the two saddle points of the semi-classical limit as related by a switch in signature convention η . This interpretation differs from the one of [54] where the two configurations were presented as arising from two parity-related 4-simplices. It also differs from the interpretation of [36,37] as two sectors of the Plebanski formulation intertwined with dynamical orientations. All these interpretations are related, but it is not completely clear in which precise sense.

The asymptotic behavior (72) is the only criterion that constrains the definition of the causal wedge amplitude. So, the dichotomy operated by $\Theta(\varepsilon S_\gamma)$ is to some extent arbitrary, and other functions could work as well. Another choice of K_ε would define a different quantum theory with the same classical limit. Our choice appears to us as the simplest one. Its technical properties will be discussed in a second article.

To summarize, the EPRL amplitude can be understood as the sum of three contributions:

$$A_v = \sum_{\substack{[\varepsilon_w] \\ \text{causal } \eta=1}} A_v^\varepsilon + \sum_{\substack{[\varepsilon_w] \\ \text{causal } \eta=-1}} A_v^\varepsilon + \sum_{\substack{[\varepsilon_w] \\ \text{spurious}}} A_v^\varepsilon \quad (73)$$

The two first terms correspond to configurations of ε_w that satisfy the causal constraint (14), respectively, for the choice of signature $\eta = 1$ and $\eta = -1$. The last term contains all other configurations of ε_w . That both signatures enter in the amplitude is not surprising because both choices are allowed at the classical level. The two terms of the asymptotics in (62) can then be understood as arising from two different choices of signature. This again is to be expected because the Einstein–Hilbert action is such that

$$S_{EH}[-g_{\mu\nu}] = -S_{EH}[g_{\mu\nu}], \quad (74)$$

so that reverting the signature sends S to $-S$.

Barrett et al. [55] have also shown that the amplitude is not exponentially suppressed when the boundary is compatible with a Euclidean geometry in the bulk. In such a case, the group element g at the critical point belongs to $SU(2)$ so that the wedge action (69) vanishes. Thus, the Euclidean configurations cannot be identified with a configuration of wedge orientations using our prescription.

When considering a full spin foam, we obtain the following decomposition

$$Z_C = \sum_{\substack{[\varepsilon_w] \\ \text{causal } \eta=1}} Z_C^{\varepsilon_w} + \sum_{\substack{[\varepsilon_w] \\ \text{causal } \eta=-1}} Z_C^{\varepsilon_w} + \sum_{\substack{[\eta_v] \\ \text{signature} \\ \text{changes}}} \sum_{\substack{[\varepsilon_w] \\ \text{local} \\ \text{light-cones}}} Z_C^{\varepsilon_w} + \sum_{\substack{[\varepsilon_w] \\ \text{spurious}}} Z_C^{\varepsilon_w}. \quad (75)$$

It then becomes an option to only consider some terms in the sum and thus partially or completely fix the causal structure in the bulk. The discussion of Section 5 still holds. The choice to consider the Feynman-like propagator or Hadamard-like function then depends on what is meant to be computed: a causal propagator or a projector on the physical Hilbert space. This conclusion answers some questions that were raised previously in the context of loop quantum cosmology [56,57].

The causal EPRL model is not a new model but rather an interpretation of different components of the standard vertex amplitude in terms of causal structures. This interpretation is motivated *a priori* by the understanding that discrete causal structures can be encoded on wedges and *a posteriori* by the asymptotics of the causal vertex.

8. Relation to Earlier Proposals

We discuss how the proposal of the previous section relates to earlier results.

8.1. Livine–Oriti Barrett–Crane Causal Model

Our approach is closely related to an earlier proposal on the implementation of causality by Livine and Oriti [16]. In the context of the Barrett–Crane model, the wedge amplitude is

$$K^p(x_1, x_2) = \frac{2 \sin(\beta(x_1, x_2) p/2)}{p \sinh \beta(x_1, x_2)}, \quad (76)$$

where x_1 and x_2 can be understood as the normals to two boundary tetrahedra and $\beta(x_1, x_2)$ is the Lorentzian angle in-between (see appendix B for a complete definition of the symbols). The Livine–Oriti proposal consists of expanding the sine as

$$K^p(x_1, x_2) = \frac{1}{p \sinh \beta(x_1, x_2)} \sum_{\varepsilon=\pm 1} \varepsilon e^{i\varepsilon \beta(x_1, x_2) p/2}, \quad (77)$$

interpreting ε as an orientation on the wedges and selecting one of the two sectors only. The implementation of causality discussed in this paper can be understood as a direct generalization of the Livine–Oriti proposal to the EPRL model. The non-trivial step introduced here is in the identification of how to introduce the splitting in (70).

8.2. Divergence and Spikes

The EPRL model and the Ponzano–Regge model both suffer from infrared divergences. In the latter case, the simplest example is provided by a triangulation Δ made of four tetrahedra subdividing a bigger tetrahedron. The dual 2-complex has 4 vertices and 10 faces. The four interior faces enclose a *bubble*. The partition function is given by

$$Z_C = \sum_{j_i} \left(\prod_f (2j_f + 1) \right) A_{v_1} A_{v_2} A_{v_3} A_{v_4}. \quad (78)$$

The sum is made over the spins j_i attached to the interior faces. The spins of the other faces are not summed over because they are fixed by the boundary conditions. The range of the sum for each interior face is a priori restricted by its neighboring faces according to the Clebsch–Gordan conditions. But the existence of a bubble implies that the sums are actually unbounded. This would not be a problem if the vertex amplitude A_v was decreasing fast enough for large spins, but such is not the case.

In the asymptotic limit, A_v consists of two conjugate terms like

$$A_v \sim A_v^+ + A_v^-, \quad (79)$$

as shown in Equation (46). The sign $\pm$ can be seen as an orientation of the tetrahedron dual to v . Thus,

$$A_{v_1} A_{v_2} A_{v_3} A_{v_4} \sim \sum_{\epsilon_i} A_{v_1}^{\epsilon_1} A_{v_2}^{\epsilon_2} A_{v_3}^{\epsilon_3} A_{v_4}^{\epsilon_4}, \quad (80)$$

where the sum is carried over all possible sign assignments to the four vertices. In [58], it is suggested that only some of the terms in this sum, dubbed “spikes”, contribute to the divergence. The spikes are specific configurations among those for which the light cones are locally well-defined but with signature switches between the vertices. In other words, a wise selection of vertex orientation can cure the model from divergences. It is suggested that a similar behavior could cure the EPRL model as well.

Our proposed causal EPRL model (70) satisfies the requirements identified in [58]. However, there are two main differences with respect to their proposal: (i) We fix the orientation at the level of wedges, which is a finer scale than that of tetrahedra. (ii) The orientation is fixed in the definition of the amplitude, while theirs only holds in the asymptotic limit.

8.3. Engle’s Proper Vertex

The EPRL model has been criticized on the basis of its asymptotics, featuring the so-called *cosine problem* [59–61]. The presence of two critical points is expected to cause problems when several vertices are considered [62]. Engle has argued that the origin of this phenomenon is the fact that the EPRL model is built from discrete BF theory by imposing the simplicity constraint, but the latter is not strong enough as it admits three sectors out of the five of the Plebanski formulation [36,37]. The proposal of a *proper vertex* [37] includes further constraints that restrict the model to the Einstein–Hilbert sector only. This is done by introducing in each wedge amplitude a spectral projector that concretely acts as a step function Θ . As a result, only one of the two terms in (62) is selected.

Our proposed causal vertex (70) shares a similar feature in that it amounts to the introduction of a step function Θ on each wedge¹⁰. Yet, let us remark that the motivations are different. While Engle’s proper vertex is motivated by the restriction to the Einstein–Hilbert sectors, and therefore to address the cosine problem, we are motivated by our analysis of the causal structure. It would be interesting to investigate thoroughly the relation between the sectors of Plebanski, the signature convention and the orientation of space–time, following the analysis of Immirzi in [30]. Our analysis suggests that the appearance of the cosine in the asymptotics is a feature of the projector for the Hamiltonian constraint. Instead, when the spin foam model is used as a causal propagator (as in Teitelboim’s approach [16,48]), one selects only one class of causal configurations and finds only the contributions of the form $e^{i\lambda S_R}$ in the asymptotics.

9. Conclusions

The metric field of general relativity is almost fully determined by its causal structure. In this paper, we have investigated the role played by the causal structure in spin foam quantum gravity. We can summarise our main points as follows:

- The notion of causality in general relativity encompasses two related but conceptually different notions: bare causality and time orientability.

- There is a natural way to translate these notions to a simplicial complex.
- The causal structure can be implemented on the dual 1-skeleton (edges). It can be seen as the combination of a causal set with a neighborhood relation.
- It can also be encoded on the dual 2-skeleton (wedges) with a degeneracy of 2 that corresponds to a global time-reversal symmetry.
- Starting from the set of all possible wedge orientations, the Lorentzian Regge action determines equations of motion whose solutions fix a proper causal structure.
- The metric propagator can be written as a sum over all possible wedge orientations. By fixing the causal structure from the beginning, one defines a causal metric propagator, similar to the Feynman propagator.
- Discrete BF theory naturally carries an orientation structure on the edges and faces, although it is blind to it in the bulk.
- Discrete BF theory is sensitive to the orientation on the boundary. There are simple rules of crossing symmetry to go from one orientation to another.
- There is a simple way to break the orientation invariance in the bulk, which provides a toy model to study causality in spin foam models (70).
- The EPRL amplitude can be regarded as a sum over all possible configurations of wedge orientations ε_w , which provide additional dynamical variables encoding the causal structure. Only a subset of it corresponds to properly causal configurations.
- The causal EPRL vertex shares common traits with the Livine–Oriti causal version of the Barrett–Crane model and with Engle’s proper vertex.
- Whether one should use the causal or the full EPRL amplitude depends on what one wants to compute: a projector on the physical Hilbert space or a causal propagator.

Author Contributions: Investigation, E.B. and P.M.-D; writing—original draft preparation, E.B. and P.M.-D; writing—review and editing, E.B. and P.M.-D. All authors have read and agreed to the published version of the manuscript.

Funding: This work was made possible through the support of the ID# 62312 grant from the John Templeton Foundation, as part of the project “The Quantum Information Structure of Spacetime” (QISS). The opinions expressed in this work are those of the author(s) and do not necessarily reflect the views of the John Templeton Foundation. E.B. acknowledges support from the National Science Foundation, Grant No. PHY-2207851.

Data Availability Statement: Data are contained within the article.

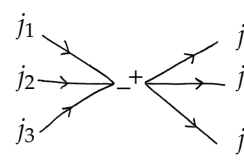
Acknowledgments: The authors thank Abhay Ashtekar, Pietro Donà, Francesco Gozzini, Alejandro Perez and Simone Speziale for insights and discussions during the course of this work. P.M.-D. thanks Alexandra Elbakyan for her help in accessing the scientific literature.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Orientation Invariance of the Ponzano–Regge Model

The Ponzano–Regge model is defined over a 2-complex $\mathcal{C}$. To write the partition function like (39), it is necessary to introduce an orientation structure on the edges and faces. However, the value of $Z_{\mathcal{C}}$ is independent of this structure in the bulk. This can be checked as follows.

Consider an edge. Its orientation plays a role in the amplitudes of the two vertices that terminate the edge. Graphically, the contribution of the edge is



$$(A1)$$

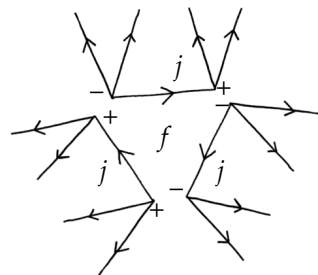
The face orientation, represented by the arrows, is actually irrelevant to the property that we are proving. What matters is that an edge spawns two nodes, one positive and one

negative. Reversing the orientation amounts to switching the signs of both nodes. With the above rules, it is easy to check that

$$\begin{array}{c} j_1 \\ j_2 \\ j_3 \end{array} \rightarrow + \begin{array}{c} j_1 \\ j_2 \\ j_3 \end{array} = \begin{array}{c} j_1 \\ j_2 \\ j_3 \end{array} \rightarrow - \begin{array}{c} j_1 \\ j_2 \\ j_3 \end{array} \quad (\text{A2})$$

which proves the invariance under a change in edge orientation.

Now, consider a face. Its orientation plays a role in the amplitudes of all the vertices around the face. For instance, in the case of a face labeled by j and surrounded by three vertices, its orientation appears as an arrow on the j -link of three vertex amplitudes, like



(A3)

Reversing the orientation of the face amounts to reversing the arrow on each of the links around. Now, using the above rules, it is easy to show that

1. Reversing the arrow of a link in-between two positive or two negative nodes amounts to multiplying by $(-1)^{2j}$.
2. Reversing the arrow of a link in-between a positive and a negative link does not change the vertex amplitude.

Overall, one gets a factor $(-1)^{2j}$ for each link surrounded by nodes of the same sign. To conclude, we need to prove the lemma that the number of such links around a face is always even. Indeed, choose a face and call V the number of vertices around and D (resp. S) the number of links around that have endpoint nodes with different (resp. the same) signs. We have $S + D = V$. Then, notice that there are in total as many positive as negative nodes, so that the product of the signs of the nodes around a face is $(-1)^V$. The same quantity can be computed differently as $(-1)^D$. This shows that V and D have the same parity, which implies that S is even and proves the lemma. As a conclusion, Z_C is indeed invariant under a change in face orientation.

Appendix B. Causal Barrett–Crane Model

In this appendix, we review the Livine–Oriti causal version of the Lorentzian Barrett–Crane spin foam [16].

Appendix B.1. Lorentzian Barrett–Crane

The Lorentzian Barrett–Crane model was introduced in [12]. The 2-complex $\mathcal{C}$ is dual to a 4-dimensional simplicial complex Δ , and each face f is labeled by a positive number $p_f \in \mathbb{R}_+$. As formulated in [16]¹¹, the partition function reads

$$Z_C = \int_{\mathbb{R}_+} [dp_f] \prod_f p_f^2 \prod_e A_e \prod_v A_v. \quad (\text{A4})$$

The integral is carried over the labels p_f for each face. The edge amplitude is given by

$$A_e = \int_{(H^+)^2} dx_1 dx_2 \prod_{f \in e} K^{p_f}(x_1, x_2). \quad (\text{A5})$$

The integration is carried over variables $x_1, x_2 \in H^+$ of the upper hyperboloid and the so-called kernel

$$K^p(x_1, x_2) = \frac{2 \sin(\beta(x_1, x_2) p/2)}{p \sinh \beta(x_1, x_2)} \quad (\text{A6})$$

with $\beta(x_1, x_2)$ the Lorentzian angle between x_1 and x_2 :

$$\beta(x_1, x_2) \stackrel{\text{def}}{=} \cosh^{-1}(x_1 \cdot x_2) \geq 0. \quad (\text{A7})$$

Surrounding a vertex v , the edges are labeled with an index $i \in \{1, \dots, 5\}$, and the faces are consistently labeled with a couple of such indices. Then, the vertex amplitude is computed as

$$A_v = \int_{(\mathcal{H}^+)^5} [dx_e] \delta(x_5 - 1) \prod_{(ij)} K^{p_{ij}}(x_i, x_j). \quad (\text{A8})$$

It is striking to realize that the formulae are expressed without any reference to the orientation of $\mathcal{C}$. In this sense, the Lorentzian Barrett–Crane model is completely blind to the causal structure to the extent that it does not even take care of the causal structure on the boundary.

Appendix B.2. Livine–Oriti Causal Model

The latter fact was recognised as unsatisfactory by Livine and Oriti in [16]. Rewriting the kernel as

$$K^p(x_1, x_2) = \frac{1}{p \sinh \beta(x_1, x_2)} \sum_{\varepsilon = \pm 1} \varepsilon e^{i\varepsilon \beta(x_1, x_2) p/2} \quad (\text{A9})$$

we see that the Lorentzian Barrett–Crane gives the amplitude:

$$Z(\Delta) = \sum_{\varepsilon_f} \int_{(\mathcal{H}^+)^5} [dx_e] \delta(x_5 - 1) \left(\prod_{f \in v} \frac{\varepsilon_f}{p_f \sinh \beta_f} \right) e^{i \sum_f \varepsilon_f \beta_f p_f/2}. \quad (\text{A10})$$

where the sum is carried over all possible sign assignments to the faces f around the vertex v . The p_f are related to the area of the boundary triangle by

$$p_f = \lambda A_f. \quad (\text{A11})$$

In the semi-classical limit, when $\lambda \rightarrow \infty$, the sum over ε_f and the integral over x_e in (A10) are dominated by only two terms for which the variables describe the two only possible 4-simplices, which have p_f as areas up to degenerate contributions.

The coexistence of two terms in the asymptotic limit can be interpreted as a time-reversal symmetry of the model. So Livine and Oriti proposed to change the vertex amplitude of the Barrett–Crane model by truncating the sum over ε_f to keep only the term that matches the causal structure of the dual 4-simplicial complex. This defines a causal amplitude with only one term in the asymptotic limit. When Δ consists of many 4-simplices, the same truncation leads to the selection of a single causal structure on Δ .

Notes

- ¹ The denomination is ours. Surprisingly, it seems that this notion does not carry a specific name in the literature. It is usually simply called “causality”, but here we need a specific name to be accurate.
- ² We leave the rigorous definition of “nicely glued together” unspecified here as it does not affect directly our investigation. See [32] for details and its relation to twisted geometries [38].

- 3 We deliberately ignore the null case, which does not seem to shed much light on our investigation.
- 4 A *directed graph* is given by a set of vertices and a set of ordered pairs of vertices (arrows). It is considered *simple* if there are no arrows from a vertex to itself. It is considered *oriented* if there is at most one arrow between any two vertices.
- 5 A directed graph is *acyclic* if it has no directed cycles, which means, in causal language, no closed time-like curves. If we do assume the presence of directed cycles, then the construction of the poset is still possible but subtler because the anti-symmetry implies the contraction of such cycles, so that more combinatorial information is lost.
- 6 We denote indifferently $e \in v$ or $v \in e$ when the vertex v is an endpoint of the edge e .
- 7 Given a vertex, the vertex graph associates a node to each edge and a link to each wedge in-between. In graph theory, the word “edge” is usually used instead of “link”. But, we stick to a common convention in loop quantum gravity (see [7]) where “edge” is reserved to the bulk of 2-complexes and “link” is used for the boundary.
- 8 We refer to [51] for an introduction to the mathematical material used in this section.
- 9 More abstractly, it can be regarded as a section of the Hopf bundle.
- 10 Note that, a priori, it is not clear that the restrictions introduced in the two proposals match away from the semi-classical limit. In fact, Engle’s proper vertex introduces a step-function Θ on each wedge, which depends on data on the full 4-simplex, while the step function Θ in (70) is local on the wedge but includes the wedge orientations ε_w as additional dynamical variables.
- 11 Note that in [16], the expressions are directly on the triangulation instead of the dual picture.

References

1. Malament, D.B. The Class of Continuous Timelike Curves Determines the Topology of Spacetime. *J. Math. Phys.* **1977**, *18*, 1399–1404. [\[CrossRef\]](#)
2. Hawking, S.W.; King, A.R.; McCarthy, P.J. A New Topology for Curved Space-Time Which Incorporates the Causal, Differential, and Conformal Structures. *J. Math. Phys.* **1976**, *17*, 174–181. [\[CrossRef\]](#)
3. Hawking, S.W.; Ellis, G.F.R. *The Large Scale Structure of Space-Time*; Cambridge University Press: Cambridge, UK, 1973.
4. Perez, A. The Spin Foam Approach to Quantum Gravity. *Living Rev. Rel.* **2013**, *16*, 3. [\[CrossRef\]](#)
5. Bianchi, E. Spinfoam Gravity. In *Loop Quantum Gravity: The First 30 Years*; Ashtekar, A., Pullin, J., Eds.; WSP: Montreal, QC, Canada, 2017; pp. 97–124. [\[CrossRef\]](#)
6. Livine, E.R. Spinfoam Models for Quantum Gravity: Overview. *arXiv* **2024**, arXiv:2403.09364.
7. Rovelli, C.; Vidotto, F. *Covariant Loop Quantum Gravity*; Cambridge University Press: Cambridge, UK, 2014.
8. Thiemann, T. *Modern Canonical Quantum General Relativity*; Cambridge Monographs on Mathematical Physics; Cambridge University Press: Cambridge, UK, 2007. [\[CrossRef\]](#)
9. Ashtekar, A.; Bianchi, E. A short review of loop quantum gravity. *Rept. Prog. Phys.* **2021**, *84*, 042001. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Reisenberger, M.P.; Rovelli, C. ‘Sum over surfaces’ form of loop quantum gravity. *Phys. Rev. D* **1997**, *56*, 3490–3508.
11. Markopoulou, F.; Smolin, L. Causal evolution of spin networks. *Nucl. Phys. B* **1997**, *508*, 409–430. [\[CrossRef\]](#)
12. Barrett, J.W.; Crane, L. A Lorentzian signature model for quantum general relativity. *Class. Quant. Grav.* **2000**, *17*, 3101–3118. [\[CrossRef\]](#)
13. Markopoulou, F.; Smolin, L. Quantum geometry with intrinsic local causality. *Phys. Rev. D* **1998**, *58*, 084032.
14. Markopoulou, F. Quantum causal histories. *Class. Quant. Grav.* **2000**, *17*, 2059–2072.
15. Gupta, S. Causality in spin foam models. *Phys. Rev. D* **2000**, *61*, 064014. [\[CrossRef\]](#)
16. Livine, E.R.; Oriti, D. Implementing causality in the spin foam quantum geometry. *Nucl. Phys. B* **2003**, *663*, 231–279. [\[CrossRef\]](#)
17. Pfeiffer, H. On the causal Barrett-Crane model: Measure, coupling constant, Wick rotation, symmetries and observables. *Phys. Rev. D* **2003**, *67*, 064022. [\[CrossRef\]](#)
18. Hawkins, E.; Markopoulou, F.; Sahlmann, H. Evolution in quantum causal histories. *Class. Quant. Grav.* **2003**, *20*, 3839. [\[CrossRef\]](#)
19. Oriti, D. The Feynman propagator for spin foam quantum gravity. *Phys. Rev. Lett.* **2005**, *94*, 111301. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Freidel, L.; Livine, E.R. Ponzano-Regge model revisited III: Feynman diagrams and effective field theory. *Class. Quant. Grav.* **2006**, *23*, 2021–2062. [\[CrossRef\]](#)
21. Oriti, D. Generalised group field theories and quantum gravity transition amplitudes. *Phys. Rev. D* **2006**, *73*, 061502.
22. Oriti, D.; Tlas, T. Causality and matter propagation in 3-D spin foam quantum gravity. *Phys. Rev. D* **2006**, *74*, 104021.
23. Livine, E.R.; Terno, D.R. Quantum causal histories in the light of quantum information. *Phys. Rev. D* **2007**, *75*, 084001.
24. Bianchi, E.; Ding, Y. Lorentzian spinfoam propagator. *Phys. Rev. D* **2012**, *86*, 104040.
25. Rovelli, C.; Wilson-Ewing, E. Discrete Symmetries in Covariant LQG. *Phys. Rev. D* **2012**, *86*, 064002. [\[CrossRef\]](#)
26. Bianchi, E.; Hellmann, F. The Construction of Spin Foam Vertex Amplitudes. *SIGMA* **2013**, *9*, 008. [\[CrossRef\]](#)
27. Oriti, D. Group field theory as the 2nd quantization of Loop Quantum Gravity. *Class. Quant. Grav.* **2016**, *33*, 085005.
28. Immirzi, G. A note on the spinor construction of Spin Foam amplitudes. *Class. Quant. Grav.* **2014**, *31*, 095016.
29. Cort  s, M.; Smolin, L. Spin foam models as energetic causal sets. *Phys. Rev. D* **2016**, *93*, 084039. [\[CrossRef\]](#)
30. Immirzi, G. Causal spin foams. *arXiv* **2016**, arXiv:1610.04462.
31. Finocchiaro, M.; Oriti, D. Spin foam models and the Duflo map. *Class. Quant. Grav.* **2020**, *37*, 015010. [\[CrossRef\]](#)
32. Dona, P.; Speziale, S. Asymptotics of lowest unitary $SL(2, \mathbb{C})$ invariants on graphs. *Phys. Rev. D* **2020**, *102*, 086016.

33. Jercher, A.F.; Oriti, D.; Pithis, A.G.A. Complete Barrett-Crane model and its causal structure. *Phys. Rev. D* **2022**, *106*, 066019. [[CrossRef](#)]
34. Regge, T. General Relativity Without Coordinates. *Nuovo Cim.* **1961**, *19*, 558–571. [[CrossRef](#)]
35. Engle, J.; Livine, E.; Pereira, R.; Rovelli, C. LQG vertex with finite Immirzi parameter. *Nucl. Phys. B* **2008**, *799*, 136–149. [[CrossRef](#)]
36. Engle, J. Proposed proper Engle-Pereira-Rovelli-Livine vertex amplitude. *Phys. Rev. D* **2013**, *87*, 084048. [[CrossRef](#)]
37. Engle, J.; Zipfel, A. Lorentzian proper vertex amplitude: Classical analysis and quantum derivation. *Phys. Rev. D* **2016**, *94*, 064024. [[CrossRef](#)]
38. Freidel, L.; Speziale, S. Twisted geometries: A geometric parametrisation of SU(2) phase space. *Phys. Rev. D* **2010**, *82*, 084040. [[CrossRef](#)]
39. Bombelli, L.; Lee, J.; Meyer, D.; Sorkin, R. Space-Time as a Causal Set. *Phys. Rev. Lett.* **1987**, *59*, 521–524. [[CrossRef](#)]
40. Surya, S. The causal set approach to quantum gravity. *Living Rev. Rel.* **2019**, *22*, 5. [[CrossRef](#)]
41. Wieland, W.M. A new action for simplicial gravity in four dimensions. *Class. Quant. Grav.* **2015**, *32*, 015016.
42. Barrett, J.W.; Foxon, T.J. Semiclassical limits of simplicial quantum gravity. *Class. Quant. Grav.* **1994**, *11*, 543–556.
43. Barrett, J.W. First order Regge calculus. *Class. Quant. Grav.* **1994**, *11*, 2723–2730. [[CrossRef](#)]
44. Misner, C.W. Feynman quantization of general relativity. *Rev. Mod. Phys.* **1957**, *29*, 497–509. [[CrossRef](#)]
45. Wheeler, J.A. Geometrodynamics and the issue of final state. In *Proceedings of the Les Houches Summer School of Theoretical Physics: Relativity, Groups and Topology*; DeWitt, C., DeWitt, B., Eds.; Gordon and Breach Science Publishers: New York, NY, USA, 1964; pp. 317–522.
46. Hawking, S.W. Quantum Gravity and Path Integrals. *Phys. Rev. D* **1978**, *18*, 1747–1753. [[CrossRef](#)]
47. Oeckl, R. General boundary quantum field theory: Foundations and probability interpretation. *Adv. Theor. Math. Phys.* **2008**, *12*, 319–352. [[CrossRef](#)]
48. Teitelboim, C. Quantum Mechanics of the Gravitational Field. *Phys. Rev. D* **1982**, *25*, 3159. [[CrossRef](#)]
49. Baez, J.C. An Introduction to Spin Foam Models of BF Theory and Quantum Gravity. *Lect. Notes Phys.* **2000**, *543*, 25–93. [[CrossRef](#)]
50. Ponzano, G.; Regge, T. Semiclassical Limit of Racah Coefficients. In *Proceedings of the Spectroscopy and Group Theoretical Methods in Physics*; Bloch, F., Ed.; North-Holland Publ. Co.: Amsterdam, The Netherlands, 1968.
51. Martin-Dussaud, P. A Primer of Group Theory for Loop Quantum Gravity and Spin-foams. *Gen. Rel. Grav.* **2019**, *51*, 110. [[CrossRef](#)]
52. Roberts, J. Classical 6j-symbols and the tetrahedron. *Geom. Topol.* **1999**, *3*, 21–66. [[CrossRef](#)]
53. Rovelli, C. Zakopane lectures on loop gravity. *PoS* **2011**, *QGQS2011*, 003. [[CrossRef](#)]
54. Barrett, J.W.; Dowdall, R.J.; Fairbairn, W.J.; Hellmann, F.; Pereira, R. Lorentzian spin foam amplitudes: Graphical calculus and asymptotics. *Class. Quant. Grav.* **2010**, *27*, 165009. [[CrossRef](#)]
55. Barrett, J.W.; Dowdall, R.J.; Fairbairn, W.J.; Gomes, H.; Hellmann, F. Asymptotic analysis of the EPRL four-simplex amplitude. *J. Math. Phys.* **2009**, *50*, 112504. [[CrossRef](#)]
56. Ashtekar, A.; Campiglia, M.; Henderson, A. Casting Loop Quantum Cosmology in the Spin Foam Paradigm. *Class. Quant. Grav.* **2010**, *27*, 135020. [[CrossRef](#)]
57. Henderson, A.; Rovelli, C.; Vidotto, F.; Wilson-Ewing, E. Local spinfoam expansion in loop quantum cosmology. *Class. Quant. Grav.* **2011**, *28*, 025003. [[CrossRef](#)]
58. Christodoulou, M.; Langvik, M.; Riello, A.; Roken, C.; Rovelli, C. Divergences and Orientation in Spinfoams. *Class. Quant. Grav.* **2013**, *30*, 055009. [[CrossRef](#)]
59. Halliwell, J.J.; Hartle, J.B. Wave functions constructed from an invariant sum over histories satisfy constraints. *Phys. Rev. D* **1991**, *43*, 1170–1194. [[CrossRef](#)] [[PubMed](#)]
60. Engle, J. A spin-foam vertex amplitude with the correct semiclassical limit. *Phys. Lett. B* **2013**, *724*, 333–337. [[CrossRef](#)]
61. Vojinović, M. Cosine problem in EPRL/FK spinfoam model. *Gen. Rel. Grav.* **2014**, *46*, 1616.
62. Bianchi, E.; Satz, A. Semiclassical regime of Regge calculus and spin foams. *Nucl. Phys. B* **2009**, *808*, 546–568.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.