Imperial College London

Department of Theoretical Physics

Causal Set Theory as a Discrete Model for Classical Spacetime

Author:
Omar F. Sosa Rodríguez

Supervisor:
Prof. Fay Dowker

Submitted in partial fulfilment of the requirements for the degree of Master of Science of Imperial College London, 2013-2014

Acknowledgements

I thank God for the great family I have been given: Carolina, Sergio and Abraxas. Your constant love and support, even from across the ocean, is what made this year possible for me. You are my main source of inspiration and motivation.

To my supervisor Fay Dowker. Thank you for the course on black holes which were simply the best lectures of the masters. Thank you for all the interesting and inspiring discussions, for supervising this work, and for being so patient while doing so. I look forward to meet you again.

Though I am grateful to all of my friends of the MSc, I am specially thankful to:

- Sonny. For rescuing me infinite times, for all the awesome days of adventures in London and for sharing the endless days of revision.
- Oscar. I would have not survived this master if you were not there. Thank you for all your advice and for making me part of your family.
- Mrs. Smith. Thank you for this last week specially. I know for a fact that you are an amazing friend. If you were there I would always feel more comfortable. *You really tied the room together.*

I love the three of you and I really hope we can meet again soon.

Last but not least. Montse, Liliana, Gerson and Esmeralda. If I have felt safe and confident in moving forward is because I know you have my back.

To Emma.
I'm sorry I wasn't there.

Contents

1 Introduction

Nowadays it is well known that General Relativity, which is our best theory to describe gravity, fails to fit a quantum description. Since we have managed to describe the three other known forces (electromagnetic, weak and strong) in a very successful quantum field theory scheme, it is our guess and hope that gravity exhibits a quantum behavior at small scales, although we have so far failed to prove such guess. Many attempts have been made to find a theory of quantum gravity (See the list of references at the end of the dissertation). At the moment theoretical physicist of all over the world are working hardly on the subject and yet there is no general agreement on which theory is the best candidate to succeed. String theory[28] and Loop Quantum Gravity[29, 30] are probably the most known examples. The reason for having so many different theories is that physicists must make a choice among several alternative approaches. For example, one must decide whether to work with the "observable and state" framework or the path integral approach of quantum mechanics; one must make a choice between dynamical or absolute topologies; a Lorentzian or a Riemannian metric, to name a few. The advantages and disadvantages of each of these approaches (and more) are reviewed in [26].

Another important choice to consider (very deeply studied by Rafael D. Sorkin as well) is whether to work with a discrete or a continuous model of spacetime. This is a very profound and delicate topic in a philosophical and physical sense. Even though most of the theories in which the community of theoretical physicists is working on regard spacetime as a continuum, there are many good reasons to believe that this is not the case. A very strong motivation is, for instance, the results obtained in the study of black hole thermodynamics[24, 25]. When studying the entropy of a black hole[1], one *deduces* that its value is given by

$$S_{BH} = \frac{A}{4}, \tag{1.1}$$

where A is the area of the black hole horizon measured in "Planck" units. This result is deduced using the analogy between the laws of thermodynamics and the laws of black hole mechanics. The above expression for the entropy of a black hole agrees with every other result derived in this context in such a perfect fashion that no physicist doubts that it is correct. However, there is no formal derivation for equation 1.1. Nevertheless, we can recall that in statistical mechanics it is shown that the entropy of a box of gas is given by the number of molecules of the gas contained in the box. Thus the entropy of a black hole, which is essentially the number of "Planck pixels" at the horizon, suggests that there is an underlying discrete behavior in the spacetime region that the horizon comprises. But we know that the laws of physics make no distinction between the horizon of a black hole and

[1]One must be careful, in fact, in saying whether the entropy is inside or outside the horizon of a black hole, but we will not address that matter here.

any arbitrary region of spacetime. This reasoning leads us to the conclusion that every region of spacetime (and not only the horizon of a black hole) should be fundamentally discrete.

A further motivation for a discrete model of spacetime has to do with the infinities that are encountered in current theories. For example, in Quantum Field Theory, one has to implement *renormalization* because without it the theory would yield infinite physical quantities[35]. In General Relativity we encounter events which we have to consider as excluded from the spacetime because we would have points with infinite curvature[10, 15]; these events are what we know as *singularities*. Note that these infinities are present in our theories even when we are not trying to implement a quantum description of gravity. Attempting to do so only makes the situation worse (See [35, 26]). We remark the fact that these "blow-ups" are evidenced when considering distances of the order of the Planck length[2]. The solution to these infinities may be the hypothesis that a finite spacetime volume has a finite number of degrees of freedom, i.e. that it is a discrete rather than a continuous entity, property that we would not be allowed to ignore if we consider the size of our "spacetime atoms" to be at Planck scale.

If the idea of a "quantum spacetime" is right, an important remark to keep in mind is that a quantized spacetime must be an hypothesis of the sought theory, since it could never be recovered by means of quantization of another entity in the continuum scenario [17]. This means that the discreteness we are looking for must be imposed *a priori*. Of course, there is more than one theory with a discrete spacetime hypothesis [18, 29, 31, 28, 32, 5, 33, 34].The question that remains is, which is the best choice? We know that some information must be encoded in the underlying discrete structure of spacetime. This information should be strong enough to recover the well-known continuum description at large scales. Causal Set Theory [22, 14]seems to fulfill these requirements and to avoid issues that other theories encounter such as Lorentz Invariance. Furthermore, the simplicity[3] of its assumptions makes it ideal to be a fundamental theory. In this work we will present the results that motivate our bias for Causal Set Theory.

Essentially, Causal Set Theory states that the most fundamental structure of spacetime is its causal order[14]. This causal order, which mathematically is no more than a partial order, is the relation induced in the spacetime by means of the Lorentzian metric. Physically, this relation tells us which are the events that can be *causally influenced* by a given (set of) event(s). Of course, in General Relativity we know what the causal relation is because the metric is specified. Thus one might be tempted to propose the metric as the most fundamental characteristic of a spacetime. However, a metric is a feature that

[2] 10^{-32}cm.

[3] This is, however, a matter of opinion.

we define in a continuous manifold and we do not know how to define a "metric" in a discrete set (see, however, [21]). A partial order, on the other hand, can be defined on any set regardless of its continuity or discreteness. Moreover, in the case of a continuous manifold a causal order, as explained in chapter 3, is enough to determine a topology, a differential structure and a conformal (Lorentzian) metric. The fact that we can derive these properties from the causal order provides an intuitive explanation of why we observe a Lorentzian manifold as a large scale approximation of a discrete set. This is the reason why we would like to regard *causality* as the fundamental property of spacetime defined independently of the metric.

Although the idea of a discrete space and time can be traced to be as old as the concept of spacetime itself, the first successful attempts to approach discreteness in a consistent manner were made by R. Sorkin until the end of the XX century. [22]. The great gap between the idea of discreteness and its first successful approach probably relies on the fact that the scientific community of the XX century focused heavily on the continuum description due to its clear success when describing every phenomenon encountered. Leibnitz and Riemann considered ideas such as the concept of a *discrete manifold*, but there was no place for these concepts in physics. However, the series of failed attempts in obtaining a quantum theory of gravity have pushed theoretical physicists to question the smoothness of the spacetime structure. On his paper, Sorkin explain the ideas of Einsntein and Riemann that led him to propose Causal Set Theory as an alternative approach for quantum gravity. David Finkelstein proposed similar ideas on his paper on "Causal Nets"[27], based on concepts originated before 1950. The Causal Set Theory idea was enhanced later in the 1970 when S. Hawking provided strong results ([10]) that deeply relate the causal structure of a strongly causal spacetime to its metric. A few years later Malament extended Hawking's results to a broader kind of spacetimes [7]. Presenting the ideas of Hawking and Malament is the main goal of this dissertation.

To present these remarkable results one must allude to several mathematics related to topology with which most part of physics students at MSc level are unfamiliarized [4]. For this reason the first chapter is dedicated to a short review of definitions and some fundamental results in topology. It should be clear that this chapter is not meant to be a complete nor a profound treatment of topology. We have chosen to include in this section those concepts which are central for the mentioned purpose. These are, for instance, the definition of a topological space, the definition of continuity from the topological point of view, and the Hausdorff separation axiom. We are assuming that the reader has knowledge of differential geometry. However, some concepts such as the definition of a differential structure are usually skipped (or its importance is not emphasized enough) in a standard differential geometry course for physicists. This is why we have decided to dedicate a short

[4] Like myself before starting this dissertation.

section at the end of chapter 1 to explain a few terms that we will need to use in the later sections. For a more in-depth revision we refer the reader to [1, 2].

The second chapter of this work is dedicated to study the causal properties of spacetime [4, 3, 15, 19]. We analyze these properties in the familiar continuum scenario, where spacetime is regarded as a manifold endowed with a Lorentzian metric. We are particularly interested in the chronology and horismos relations, time orientation, and both local and global behaviour of geodesics. The goal of this is to present some results that we expect to be able to recover from Causal Set Theory in the large scale approximation. How this approximation can be done is explained in [18, 14, 26], but note that we will not be dealing with this particular subject here. In this chapter we also review a shortened version of the "causal hierarchy" of spacetimes that is presented in [6]. The hierarchy is constructed by imposing topological conditions on the spacetime manifold. This classification is originally intended to rule out some situations we consider to be unphysical and which are not discarded by the mere postulates of General Relativity (see chapter 3 of [4] for an elegant presentation of these postulates). Thus, this section is also expected to present some of the properties that we would *not* like to recover from Causal Set Theory. Overall, this chapter is intended to give an insight on what are the best axioms to build Causal Set Theory.

Chapter 3 is the main part of this work. We start by introducing the definition of *Causal Spaces* [13] (not to confuse with a Causal *Set*). One may think of a Causal Space as a generalization of a spacetime manifold. Just as a Lorentzian manifold, a Causal Space is equipped with a family of relations that satisfy analogous properties to the causality, chronology and horismos[5]. The importance of presenting this concept is that it is an abstraction of both a Causal Set and a Lorentzian manifold. Studying causal spaces is the right choice since our final goal is to make a comparison between the discrete and the continuum approach. Followed by this, we present the definition of a Causal Set and we explain how some of the pathologies previously encountered with General Relativity are avoided if we give a Causal Set a physical interpretation. Finally, we present and prove what is currently know as the Malament-Hawking-King-McCarthy-Levichev theorem[20][6]. This theorem shows how an enormous amount of physical information is encoded in the causal order relation. However, even at this point, proving MHKML is a tedious task. This theorem has been constructed from a set of results developed over many years. A complete proof of this theorem has never been collected in a single reference although it is commonly taken for granted in the Causal Sets literature [20]. For this reason we have decided to include here the full derivation. This is our top reason to believe that Causal Set theory is the right candidate to provide a quantum description of spacetime.

[5]In fact, they are called the same.

[6]As we shall see, the Kroenheimer and Penrose should be cited instead of Levichev.

2 Mathematical preliminaries.

The purpose of this section is to provide a quick revision of some basic notions of topology that will be used throughout the whole dissertation. In order to make it as concise as possible we present the definitions and theorems that are particularly important for our goals. Thus, many results that are of great importance for the area of topology itself are not included here. Theorems are stated without proof in this section. For the proofs and a more complete treatment we refer the reader to [1] and references therein.

2.1 Topology review

Topology allows us to "measure", in a certain sense, the closeness between points of a given set X.

Definition 2.1. A *topology* on a set X is a family of subsets $\mathscr{T}$, called *open sets*, satisfying the following properties:

1) $\emptyset, X \in \mathscr{T}$.

2) If $A, B \in \mathscr{T}$, then $A \cap B \in \mathscr{T}$. In other words, the *finite* intersection of open sets is an open set.

3) If $\mathcal{A} \subseteq \mathscr{T}$, then $\bigcup \mathcal{A} \in \mathscr{T}$. This means that the *arbitrary* union of open sets is open.

The pair $(X, \mathscr{T})$ is called a *topological space.*

Given a set X, a topology can always be constructed for it. If the set X consists of more that a single element, there will be many ways in which we can endow X with a topology.

Definition 2.2. Given two topologies $\mathscr{T}_1$ and $\mathscr{T}_2$ on a set X, we say that $\mathscr{T}_1$ is *weaker* than $\mathscr{T}_2$ if every open of $\mathscr{T}_1$ is also an open of $\mathscr{T}_2$. $\mathscr{T}_2$ is said to be *stronger.* We write it $\mathscr{T}_1 \leq \mathscr{T}_2$.

Pairs of topologies on a given set are not always comparable.

Definition 2.3. A subset E of a topological space $(X, \mathscr{T})$ is called *closed* if $X - E \in \mathscr{T}$

If a set is closed this does not necessarily means that it is *not open.* For instance, $X, \emptyset$ are closed and open in any topology for X. It is not always necessary to deal with all open sets of a given topology, which is usually complicated. For this we have the following definition:

Definition 2.4. Let $(X, \mathscr{T})$ be a topological space. A subcollection $\mathcal{B}$ of $\mathscr{T}$ is called a *basis* for $\mathscr{T}$ if every open set can be obtained as a union of the open sets in $\mathcal{B}$.

The main quality of a basis is the next result.

Theorem 2.5. *A subcollection $\mathcal{B}$ of a topology $\mathscr{T}$ on X is a basis for $\mathscr{T}$ if and only if for each $O \in \mathscr{T} - \emptyset$ and each $x \in O$, there exists $B \in \mathcal{B}$ such that $x \in B \subseteq O$.*

The next notion is of great utility when we want to consider local properties of a topological space.

Definition 2.6. Let x be an element in the topological space $(X, \mathscr{T})$. A subset N of X is called a *neighborhood* of x if there exists an open set O such that $x \in O \subset N$. The collection of neighborhoods of x is called *neighborhood system of* x and is denoted by $\mathcal{N}(x)$.

Clearly any open set O is a neighborhood for each of its points.

Theorem 2.7. *A subset O of a topological space $(X, \mathscr{T})$ is open if and only if for each $x \in O$, there exists $N \in \mathcal{N}(x)$ such that $N \subseteq O$.*

A topological space *naturally* induces a topology on any of its subsets.

Theorem 2.8. *Given a topological space $(X, \mathscr{T})$ and $Y \subseteq X$, the collection*

$$\mathscr{T}\mid_Y := \{O \cap Y \mid O \in \mathscr{T}\}$$

is a topology on Y. We call $\mathscr{T}\mid_Y$ the relative topology on Y with respect to $(X, \mathscr{T})$.

2.2 Interior, Closure, and Boundary.

Our next objective is to establish the concepts of *closure*, *interior* and *boundary* of a subset of a topological space.

Definition 2.9. A point $x \in X$ is an *accumulation point* of $E \subseteq X$ if $(E \cap N) - \{x\} \neq \emptyset$ for every $N \in \mathcal{N}(x)$. That is, every neighborhood of x contains points of E which are different from x. The set of all accumulation points of E will be denoted by $\mathrm{Der}\,(E)$.

One can show that a subset E of a topological space is closed if and only if $\mathrm{Der}\,(E) \subseteq E$. The next concept is of extreme importance for our purposes as it is for the area of topology in general.

Definition 2.10. The *closure* $\mathrm{Cl}(E)$ of a subset E of a topological space is given by $\mathrm{Cl}(E) = E \cup \mathrm{Der}\,(E)$.

The elements in the closure of a given set E are the only elements that we consider as "attached" to the set. As the name suggests, the closure of a set is itself a closed set. Furthermore, we have the following result:

Theorem 2.11. *Let $(X, \mathscr{T})$ be a topological space and let E, F be subsets of X. Then,*

1. *If* F *is a closed set and* $E \subset F$*, then* $\mathrm{Cl}(E) \subset F$*.*

2. E *is closed if and only if* $\mathrm{Cl}(E) = E$*.*

In other words, the closure of any set, $\mathrm{Cl}(E)$, is the "smallest" closed set containing E.

Definition 2.12. The interior of a subset E of a topological space, which we will denote by $\mathrm{Int}(E)$, is defined by

$$\mathrm{Int}(E) := \bigcup_{O \in \mathscr{T}} \{O : O \subseteq E\}.$$

Clearly $\mathrm{Int}(E)$ is an open set itself. This means that the interior of a a set E is the largest open set contained in E. In an analogous way to the result we have for closed sets, we have that O is an open set if and only if $O = \mathrm{Int}(O)$.
Finally, we have the definition of boundary, which plays a central role in our development.

Definition 2.13. A point $x \in X$ is a *boundary point* of E if any neighborhood N of x satisfies $N \cap E \neq \emptyset$ *and* $N \cap (X - E) \neq \emptyset$. The boundary of E, denoted ∂E, is the set of all boundary points of E.

There are many results relating the closure, interior, and boundary of a given set. Of our particular interest are the next relations.

Theorem 2.14. *For any subset* E *of a topological space* $(X, \mathscr{T})$ *the following statements hold:*

1. $\mathrm{Cl}(E) = E \cup \partial E$*.*

2. $\mathrm{Int}(E) = E - \partial E$*.*

3. $\mathrm{Int}(E) \cup \mathrm{Ext}(E) \cup \partial E = X$*.*

Where $\mathrm{Ext}(E) := \mathrm{Int}(X - E)$*.*

2.3 Continuity and Homeomorphisms.

Continuity is one of the most basics and important concepts of topology. Intuitively, we would like to regard a map between two sets as continuous if it takes "proximate" points to "proximate" points. The concept of proximity is provided by the topology of the sets which the map in consideration relates.

Definition 2.15. Let $f : X \longrightarrow Y$ be a function between the topological spaces $(X, \mathscr{T}_X)$ and $(Y, \mathscr{T}_Y)$. We say that f is continuous at $x_0 \in X$ if for any open subset $A \in \mathscr{T}_Y$ containing $f(x_0)$, there is an open subset $B \in \mathscr{T}_X$ that contains x_0 and such that $f[B] \subseteq A$. We say that f is continuous if it is continuous at every point of its domain.

It is convenient to keep in mind that the continuity of a function can be verified using only a basis for the topology. The continuity of a function can expressed in many different ways, here we present a few of the more useful ones.

Definition 2.16. If f is a function between the topological spaces $(X, \mathscr{T}_x)$ and $(Y, \mathscr{T}_Y)$, then the following statements are all equivalent.

1. f is continuous.
2. For any open O of Y, $f^{-1}[O]$ is open in X.
3. $f^{-1}[E]$ is closed in X for any closed E in Y.

Theorem 2.17. *If $f : X \longrightarrow Y$ is a continuous function and A is a subspace*[7] *of X, then $f\,|_A : A \longrightarrow Y$ is also continuous.*

In the case we are dealing with bijective functions, one finds that even if f is continuous it is not always the case that f^{-1} is continuous as well. Thus it is necessary to introduce the concept of homeomorphism.

Definition 2.18. A bijective function $h : X \longrightarrow Y$ between the topological spaces X and Y is called a *homeomorphism* if h *and* h^{-1} are continuous. If is possible to find such a function we say that X and Y are *homeomorphic* and write $X \cong Y$.

Homeomorphic spaces are regarded as equivalent in the topological sense. A necessary condition for two spaces to be homeomorphic is that they have the same cardinality. Furthermore, it is necessary that they have the same number of open subsets. Thus it might be possible that $(X, \mathscr{T}_1)$ and $(X, \mathscr{T}_2)$ are not homeomorphic, even when the set X in consideration is the same for both topological spaces.

The development of this work relies heavily on the convergence of sequences.

Definition 2.19. A *sequence* in X is a function from $\mathbb{N}$ to X. The sequence $n \mapsto x_n$ will be denoted by $\{x_n\}_{n\in\mathbb{N}}$.

Definition 2.20. A sequence $\{x_n\}_{n\in\mathbb{N}}$ *converges* in X to a point $x_0 \in X$ if for every neighborhood V of x_0 there exists $n(V) \in \mathbb{N}$ such that $x_m \in V$ for every $m \geq n(V)$. We write[8] $x_n \longrightarrow x_0$.

To determine the convergence of a sequence it is enough to consider a local basis in x_0, i.e. $x \longrightarrow x_0$ if, and only if, for every $B \in \mathcal{B}$ there exists $n(B) \in \mathbb{N}$ such that $x_m \in B$ for any $m \geq n(B)$, where $\mathcal{B}$ is a local basis for x_0.

For our purposes, the next two results will be of great importance.

[7] We will often make this abuse of notation and not write the topology of each space explicitly.

[8] Not to be confused with the horismos relation.

Theorem 2.21. *If E is a closed set and $\{x_n\}_{n\in\mathbb{N}}$ is a sequence in E that converges to x_0, then $x_0 \in E$.*

Theorem 2.22. *Let $f : X \longrightarrow Y$ be continuous. Then, if $\{x_n\}_{n\in\mathbb{N}}$ is a sequence in X that converges to x_0, we have that $f(x_n) \longrightarrow f(x_0)$.*

Hausdorff Spaces There are many different ways to demand a certain level of "separation" between the points in a topological space. These are called separation axioms. The study of each of these axioms individually is outside the scope of this work. However, of extreme importance for our future sections (and for the study of physics in general) is the *Hausdorff separation axiom*[9].

Definition 2.23. A topological space $(X, \mathscr{T})$ is a *Hausdorff space* if for any two *distinct* elements $x, y \in X$, there exist open sets U and V such that $x \in U$, $y \in V$ and $U \cap V = \emptyset$.

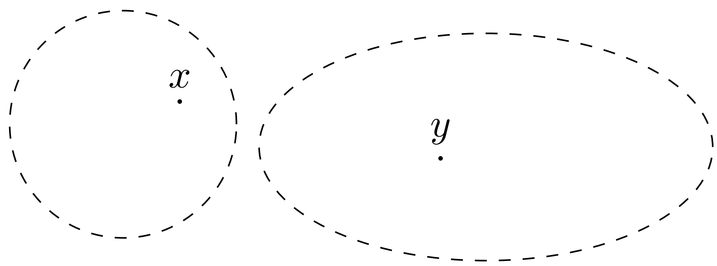

Figure 1: For any two points x and y we can find non-intersecting neighborhoods of these, here represented by the region enclosed in the dashed lines.

Two main properties of Hausdorff spaces are:

Theorem 2.24. *If $(X, \mathscr{T})$ is Hausdorff, then any convergent sequence converges to one point only.*

Theorem 2.25. *Any subspace of a Hausdorff space is Hausdorff (with the relative topology).*

2.4 Differentiable Structure

We now recall some concepts of differential geometry that we shall be using in the next section. Even though we have assumed that the reader has basic knowledge of differential geometry and General Relativity, the concept of a differentiable structure is very likely to be unfamiliar. The previous section on topology, allows us to approach differential geometry in a more formal way.

[9]Some times referred as axiom T_2.

Definition 2.26. A collection $\mathscr{C}$ of subsets of a set X is called a *cover* of X if $\bigcup \mathscr{C} = X$. In the particular case of X being a topological space and $\mathscr{C}$ being a cover of X whose elements are open sets, $\mathscr{C}$ is called an *open cover*.

Definition 2.27. A topological space M is called an n dimensional *manifold* if it has an open cover $\mathscr{U} = \{U_i\}_{i\in\Lambda}$ such that there is an homeomorphism $\varphi_i : U_i \to O \subset \mathbb{R}^n$ for each $i \in \Lambda$, where O is open in $\mathbb{R}^n$. Each pair (φ_i, U_i) is called a *chart* and the collection of all charts is called an *atlas*.

The previous definition just means that an n-dimensional manifold is *locally* homeomorphic to $\mathbb{R}^n$. Two charts (φ_i, U_i) and (φ_j, U_j) are said to be *r-compatible* if the transitions (or changes of coordinates)

$$\begin{aligned} \varphi_j \circ \varphi_i^{-1} : \varphi_i(U_{i\cap U_j}) &\longrightarrow \varphi_j(U_i \cap U_j) \\ \varphi_i \circ \varphi_j^{-1} : \varphi_j(U_{i\cap U_j}) &\longrightarrow \varphi_i(U_i \cap U_j) \end{aligned} \tag{2.1}$$

are of class C^r, i.e. r times differentiable and the derivatives are continuous. Note that r can be ∞. There is no ambiguity in what we mean here by differentiability because the transitions are maps between subsets of $\mathbb{R}^n$.

If every pair of charts of a given atlas $\mathcal{A}$ is r-compatible, we say that the $\mathcal{A}$ is of class C^r (any pair of charts with empty overlap is considered compatible). If that is the situation, then there is a unique maximal atlas $\mathbf{\Psi}$ which consist of all the charts that are r-compatible with the charts of $\mathcal{A}$.

Definition 2.28. A C^r *differential structure* on M is a maximal C^r atlas for M.

If M is a manifold and $\mathbf{\Psi}$ is a C^r differential structure, then the pair $(M, \mathbf{\Psi})$ is called a manifold of *class* C^r or simply a C^r manifold. What will be important later is that to determine the differential structure of a manifold it suffices to give an atlas contained in it [2].

Note that, even though a manifold is already a topological space, a C^r maximal atlas induces a natural topology on M. This is done by defining the open sets as the pre-images of the open sets in $\mathbb{R}^n$. The topology obtained is referred as the *manifold topology*, $\mathscr{T}^{man}$.

3 Causal properties of spacetime

The definitions and results presented in this section are taking from Penrose [3] and Hawking and Ellis [4].

A course on General Relativity (GR) is not an option in the formation of any physicist. General Relativity has now a hundred years old as a theory and thus the community of scientist have now become quite experienced in the subject. Physicists have dedicated so much effort and time to GR, that nowadays "basic GR" is presented in a very intuitive fashion, avoiding in this way many of the mathematical concepts that make it tedious and difficult to grasp[10]. Unfortunately, this represents a disadvantage for the purposes of this work. This is because questioning GR[11] (and, in fact, questioning any theory) demands first a deep and complete understanding of its foundations. A good place to start is to give a precise definition of *spacetime*. This definition is achieved relating our experience to what it could represent as a mathematical abstraction [15].

First of all, the fact that we can perceive *closeness* tells us that spacetime should have a topology. Furthermore, the fact that we can specify *where* and *when* events take place (coordinates), suggest that in fact spacetime should be a manifold. Because of the intuitive notion of distinct events, we would like to think that such a manifold is Hausdorff. This is a physically reasonable requirement and ensures the uniqueness of limits of convergent sequences. Now, since astronomical observations seem to suggest that there is no "edge" for the universe, we could physically associate this with the lack of a boundary in the sought manifold. The truly genius part in GR is the fact that gravity can be described by *curvature* of the manifold. The information about the curvature is encoded in the symmetric tensor g, called Lorentzian *metric*. Finally we take the following definition [4]:

Definition 3.1. A *spacetime* is the equivalent class of isometric pairs[12] (M, g), where M is a real, four-dimensional, connected, C^∞ Hausdorff manifold, and g is an at least C^2 Lorentzian metric globally defined on M. Points of the spacetime are called *events*.

When referring to a spacetime we will always work with a particular representative pair (M, g) of the isometry class. Even though we have fixed the dimension of spacetime in our definition, the results here presented are extendible to n-dimensional spacetimes, $n \geq 2$. The restrictions of it being a connected manifold is included for convenience, since in case of possible disconnectedness, only the connected component in which we live would be accessible to us. Note that we have not explicitly demanded $\partial M = 0$, this allows to keep our discussion more general.

[10] Consider, for example,[5].

[11] Which is what we are doing here.

[12] Two pairs, (M, g) and (M', g'), are isometric if there is a diffeomorphism $f : M \to M'$ such that $f_* g = g'$.

The Lorentzian metric allows us to classify the tangent vectors at each point of the manifold into essentially three categories.

Definition 3.2. [4] Let (M, g) be a spacetime, and let $x \in M$. A tangent vector $V \in T_xM$ is said to be *timelike*, *spacelike*, or *null* if $g(V, V)$ is respectively negative, positive or zero. A tangent vector is said to be *causal* if it is either timelike or null.

The set of null vectors at each tangent space defines the *null cone*, which separates the timelike vectors into two categories: *future-directed* and *past-directed*. When dealing with our spacetimes, we are implicitly assuming that a choice of past and future has been consistently made all over M. The existence of a Lorentzian metric globally defined on M guarantees that it is always possible to make such a choice (see theorem 2.4 in [6]).

Definition 3.3. [3] A *(smooth) path* is a continuous (resp. differentiable) map $\gamma : I \subset \mathbb{R} \to M$.

A smooth path will be called timelike if its tangent vector is timelike at each point. Null and causal paths are similarly defined. We shall use the term *curve* to refer to the image on M of a path. Hence, the equivalence class of paths equivalent under a smooth change of parameter will define the same curve. Following the convention adopted by Penrose [3] we shall consider the endpoints of any causal curve to be part of such curve.

Definition 3.4. An *affinely parametrized geodesic* is a path whose tangent vector V satisfies[13] $\nabla_V V = 0$ at all points of the curve (i.e. is parallel transported along the curve). Geodesics are classified into timelike, null, causal or spacelike according to the nature of its tangent vector.

In general, it is not true that there always exists a geodesic connecting two points of the manifold. Furthermore, if a geodesics exists between two points, this geodesic does not have to be unique. It will hence be convenient to make the following definition.

Definition 3.5. [3] An open set $\mathcal{N} \subset M$ is called *convex* if for any two points $p, q \in N$ there exists a unique geodesic $\gamma : [0, 1] \to \mathcal{N}$ entirely contained in $\mathcal{N}$ with $\gamma(0) = p$ and $\gamma(1) = q$.

If a convex neighborhood $\mathcal{N}$ has compact closure $\bar{\mathcal{N}}$ contained in a convex neighborhood, then we say $\mathcal{N}$ is *simple*. Given a point $p \in M$, it is always possible to find a simply convex neighborhood $\mathcal{N}_p$ such that $p \in \mathcal{N}_p$. An important property to keep in mind is that open convex sets form a basis for the manifold topology [7].

[13] Here ∇ denotes the torsion-free connection for which $\nabla g = 0$.

3.1 Elements of Causality theory

The main objects to study here are the relations that dictate whether or not there can be a *cause and effect* connection between events in the spacetime. Before analyzing any of the properties of these relations, we first need formally define what a *relation is*.

Definition 3.6. Let X be any set. A *relation* R on X is a subset of $X \times X$. We write aRb meaning $(a, b) \in R$. If Y is a subset of X, we write $aR_Y b$ if $(a, b) \in R \cap Y \times Y$, R_Y is called the *restriction* of R to Y.

Note that defining a relation does not demand any extra structure on the set it is defined.

Let (M, g) be a spacetime, and let x, y be two events in M. We define the natural causal relations as follows[4]:

Definition 3.7. Chronology, $\ll$: We say that x *chronologically precedes* y, written $x \ll y$ if there exists a future-oriented timelike curve with past end-point x and future end-point y.

Definition 3.8. Causality, $\prec$: We say that x *causaly precedes* y, written $x \prec y$ if there exists a future-oriented causal, *possibly degenerate*, curve with past end-point x and future end-point y. We regard null vectors as future-directed if they can be obtained as a limit of future-directed timelike vectors.

Since we have allowed the possibility of a degenerate curve in the definition of the causality relation, we immediately get that $x \prec x$ for all events in M. However, note that this is not true in general for the chronology. If it happens that there exists $x \in M$ such that $x \ll x$, this means that there exists a closed timelike curve in M (i.e. an everywhere timelike curve whose past and future end-points are identical). In an analogous way, the existence of two *different* events $x \neq y$ such that $x \prec y$ and $y \prec x$ signifies the existence of a closed non-degenerate causal curve.

From the definition of the causal relations we immediately get the following properties [3]:

$$a \ll b \quad \Longrightarrow \quad a \prec b \tag{3.1}$$

$$a \ll b \quad \text{and} \quad b \ll c \quad \Longrightarrow \quad a \ll c \tag{3.2}$$

$$a \prec b \quad \text{and} \quad b \prec c \quad \Longrightarrow \quad a \prec c. \tag{3.3}$$

Notice that the converse of equation 3.1 is not necessarily true. For our later convenience, we here introduce the horismos relation,

Definition 3.9. Horismos, $\rightarrow$: We say that x horismotically precedes y, written $x \rightarrow y$, if $x \prec y$ but *not* $x \ll y$.

Definition 3.10. Let $\mathcal{S}$ and $\mathcal{U}$ be subsets of a spacetime M. The *chronological future* $I^+[\mathcal{S},\mathcal{U}]$ *of* $\mathcal{S}$ *relative to* $\mathcal{U}$ is defined as

$$I^+[\mathcal{S},\mathcal{U}] = \{x \in \mathcal{U} \mid y \ll_{\mathcal{U}} x \quad \text{for some} \quad y \in \mathcal{S}\}, \tag{3.4}$$

where $y \ll_{\mathcal{U}} x$ means that there exists a future directed timelike curve from y to x *contained* in $\mathcal{U}$.

In other words, $I^+[\mathcal{S}](:= I^+[\mathcal{S}, M])$ are the points in M which can be reached from $\mathcal{S}$ by a future-directed timelike curve. Around any given point $p \in I^+[\mathcal{S}]$ we can always find a neighborhood of p which can also be reached by timelike curves, i.e. the sets $I^+[\mathcal{S}]$ are open. This property will be important for later developments.

A remark worth mentioning here is that in section 6.2 of "*The large scale structure of spacetime*"[4], Hawking and Ellis state that $I^+[\mathcal{S},\mathcal{U}]$ cannot contain the set $\mathcal{S}$. This is an error: even if the spacetime does not allow closed timelike curve since, simply consider $I^+[\gamma]$ with γ being an inextendible timelike curve.

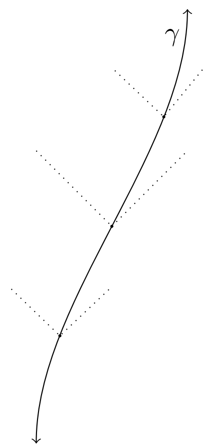

Figure 2: Each point of an inextendible timelike curve is contained in the chronological future of some other point in the curve, so $I^+[\gamma]$ contains γ. This is not true for an extendible curve because of its past end-point.

Just as the chronological future, we similarly define the *causal future* $J^+[\mathcal{S}]$,

$$J^+[\mathcal{S},\mathcal{U}] := \{x \in \mathcal{U} \mid y \prec x \quad \text{for some} \quad y \in \mathcal{S}\}, \tag{3.5}$$

and the *horismotic future* $E^+[\mathcal{S},\mathcal{U}]$:

$$E^+[\mathcal{S},\mathcal{U}] := J^+[\mathcal{S},\mathcal{U}] - I^+[\mathcal{S},\mathcal{U}]. \tag{3.6}$$

When no relative set is specified, we implicitly mean that the relative set is the whole manifold. The chronological, causal and horismotic pasts, denoted $I^-[\mathcal{S}]$, $J^-[\mathcal{S}]$ and $E^-[\mathcal{S}]$ respectively, are defined in the dual obvious way.

As opposed to the chronological future (and past), the sets $J^+[\mathcal{S},\mathcal{U}]$ are, in general, not necessarily open nor necessarily closed. Some useful properties to keep in mind are the following, they are proven in theorem 2.18 of [3]:

Theorem 3.11. *In any spacetime the next properties of between the causal relations hold.*

i) $a \ll b \quad \text{and} \quad b \prec c \implies a \ll c$

ii) $a \prec b \quad \text{and} \quad b \ll c \implies a \ll c$

iii) If α is a null geodesic from event a to b, and β is a null geodesic from b to c, then $\alpha \cup \beta$ is a null geodesic itself or $a \ll c$.

iv) If $a \to b$ then there exists a null geodesic from a to b.

We here see that the relations $\ll$ and $\prec$ are transitive. However, property (iii) tell us that the converse of property (iv) is not always true: one can have a null geodesic between two events A and B and yet have $A \ll B$.

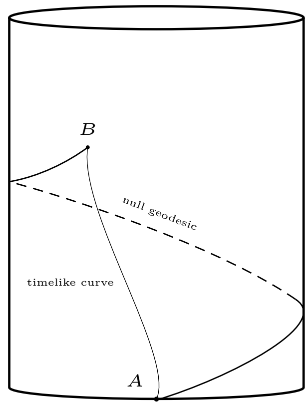

Figure 3: In this pathological cylindrical spacetime B lies on a null geodesic through A, but B can also be reached by a timelike curve from A. Thus A and B are not related by the horismos.

Also note that $a \to b \to c$ does not necessarily implies $a \to c$. This is illustrated in figure 4.

If we restrict ourselves to convex neighborhoods, the the following statements hold (lemma 14.2 in [8]):

Theorem 3.12. *Let (M, g) be a spacetime and let $\mathcal{S} \subset M$ be a convex neighborhood. Then,*

1. *If $q \in I^+[p,\mathcal{S}]$, the unique geodesic connecting p and q is timelike and future-pointing.*

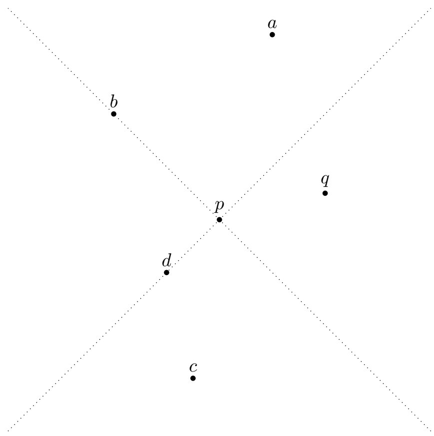

Figure 4: As we would do in Minkowski spacetime, the lightcone of p, $E^{\pm}[p] := E^{+}[p] \cup E^{-}[p]$, is represented by the dotted lines at 45 degrees. Some of the causal relations in the figure are: $p \ll a$, $p \to b$, $c \ll p$, $d \to p$. The events p and q are not related. Also, we do not have $d \to b$.

2. *$I^{+}[p, \mathcal{S}]$ is open in $\mathcal{S}$ (and in M).*

3. *$J^{+}[p, \mathcal{S}] = Cl_{\mathcal{S}}(I^{+}[p, \mathcal{S}])$*

We are now in position to present the next result , which is used by Malament [7] to extend the results of Hawking, King and McCarthy (HKM) in [9]. HKM is presented is the next chapter. The theorem and its proof are presented just as in lemma 6.2.1 of [4].

Theorem 3.13. *Let X be a spacetime and let $O \subseteq X$ be an open set. Denote by $\{\lambda_n\}_{n\in\mathbb{N}}$ an infinite sequence of non-spacelike curves in O which are future inextendible in O. If $p \in O$ is a limit point*[14] *of the sequence $\{\lambda_n\}_{n\in\mathbb{N}}$, then there exists a non-spacelike curve λ through p which is future inextendible in O and which is a limit curve*[15] *of the λ_n.*

Proof. Since every open set can be regarded as a spacetime on it own, we can simply consider $O = X$. Let U_1 be a convex normal neighborhood of p and let $B_a(q)$ be the *open ball*[16] of radius a about q. Take also an open ball of radius b about p and let $\{\lambda(1,0)_n\}_{n\in\mathbb{N}}$ be a subsequence of the part of the sequence $\{\lambda_n\}_{n\in\mathbb{N}}$ that lies inside U_1 and which converges to p. Now, $\partial B_b(p)$ is compact so it will contain limit points of the subsequence $\{\lambda(1,0)_n\}_{n\in\mathbb{N}}$. Any of the limit points y must be to the causal future (or past) of p relative to U_1, because otherwise we could find neighborhoods of y and p between which there would be no non-spacelike curve in U_1. Take $x_{11} \in J^{+}[p, U_1] \cap \partial B_b(p)$ to be one

[14]i.e. every neighborhood of p intersects an infinite number of the λ_n.

[15]For every point $p \in \lambda$ we can find a subsequence of $\{\lambda_n\}_{n\in\mathbb{N}}$ that converges to p.

[16]Open balls in X are defined by the inverse image of open balls in $\mathbb{R}^n$ of the coordinate functions of our atlas.

of such limit points. We can now consider a subsequence $\{\lambda(1,1)_n\}_{n\in\mathbb{N}}$ of $\{\lambda(1,0)_n\}_{n\in\mathbb{N}}$ converging to x_{11}.

This same idea can be repeated inductively, by dividing the interval $[0,b]$ into smaller sections on each step. Define $x_{ij} \in J^+[p,U_1] \cap \partial B_{i^{-1}jb}(p)$ as the limit point of the subsequence $\{\lambda(i-1,i-1)_n\}_{n\in\mathbb{N}}$ if $j=0$, or $\{\lambda(i,j-1)_n\}_{n\in\mathbb{N}}$ if $1 \le j \le i$. Of such subsequences, we can then take a subsequence $\{\lambda(i,j)_n\}_{n\in\mathbb{N}}$ which converges to x_{ij}. The closure of the union of all the limit points x_{ij}, $j \le i$, will be a non-spacelike curve λ from $p = x_{i0}$ to $x_{11} = x_{ii}$, because any pair of x_{ij}'s have non-spacelike separation.

Now take $\{\lambda'_n\}_{n\in\mathbb{N}}$ as a subsequence of $\{\lambda(m,m)_n\}_{n\in\mathbb{N}}$ which intersects each open ball $B_{\frac{b}{m}}(x_{mj})$, $0 \le j \le m$. This way, λ is by construction a limit curve of $\{\lambda_n\}$ from p to x_{11}. If we now take an convex normal neighborhood U_2 of x_{11} and repeat the construction this time for $\{\lambda'_n\}_{n\in\mathbb{N}}$ we can extend λ indefinitely. □

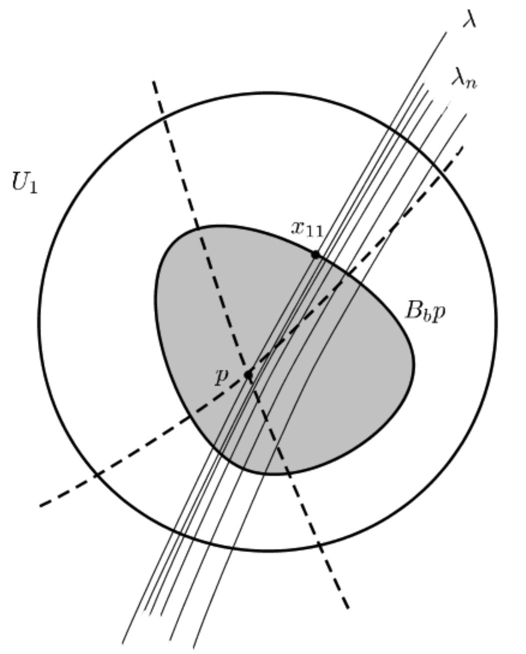

Figure 5: The figure represents the first step in the "inductive" proof of theorem 3.13.

To conclude this section I include a theorem of R. Penrose. The proof for this can be found in proposition 4.9 of [3].

Theorem 3.14. *Let N be a simple region, Q an open set contained in N, and $p \in Q$. Then there exists $u, v \in Q$ such that p is contained in the 'interval"*

$$\langle\langle u,v\rangle\rangle_N := \{q \in N \mid u \ll_N q \ll_N v\},$$

and $\langle\langle u,v\rangle\rangle_N$ is entirely contained in Q.

3.2 Causal hierarchy of spacetimes

In the present form, our definition of spacetime allows many "pathological" situations, such as the existence of closed timelike curves (CTC's) which lead to several contradictions [11, 4]. Hence, we would like to impose further conditions on our manifold in order to get rid of these anomalies so that we are left with physically "reasonable" behaviors. We will refer to these conditions as *causal restrictions*. In the next chapter we shall address the question of which causal restrictions arise naturally if spacetime is regarded as a *causal set*. The hierarchy presented here is a shortened version of the compilation made by [6], paying particular attention to those spacetime that we shall deal with later on.

Chronological and causal spacetimes. We will say that a space time is chronological if it admits no closed timelike curves. This sentence can be rephrased in terms of the chronology relation: A spacetime is chronological if the chronology relation $\ll$ is irreflexive, i.e. there is *no* $p \in X$ such that $p \ll p$. A main property of these kind of spacetimes is that they cannot be compact [6]. In an analogous way, a spacetime is causal if admits no closed causal curves. This can be rephrased in terms of the causality relation as: if $x \prec y$ *and* $y \prec x$, then $x = y$. Keep in mind however that in any spacetime it is true by definition that $x \prec x$.

Distinguishing spacetimes. This is our next step in our causal ladder. A spacetime is called *future* distinguishing if it satisfies any of the following equivalent properties:

1. $I^+[q] = I^+[p] \implies p = q$.
2. The function that associates each point in the spacetime to its chronological future, $p \mapsto I^+[p]$, is one to one.
3. Given any $p \in X$ and any neighborhood U containing p, there exists a neighborhood $V \subset U$ of p such that $J^+[p, V] = J^+[p] \cap V$
4. For any $p \in X$ and any neighborhood U containing p, there exists a neighborhood $V \subset U$, $p \in V$, which distinguishes p in U to the future.

We say that a spacetime is *past distinguishing* if it satisfies any of the properties in the above list, but replacing each item with the analogous past assertion. If a spacetime happens to be both future *and* past distinguishing we shall refer to it simply as *distinguishing*. The proof for the equivalence of the previous properties can be found in [6].

The main result that will concern us about distinguishing spacetimes is the following theorem:

Theorem 3.15. *Let (X_1, g_1), (X_2, g_2) be two spacetimes, with (X_1, g_1) distinguishing. If we can find a diffeomorphism $f : X_1 \longrightarrow X_2$ that preserves the causality relation $\prec$, then (X_2, g_2) is also distinguishing and $g_1 = f^* g_2$.*

Before continuing introducing out causal hierarchy, we present some results that will motivate some future definitions in section 3, when we need to introduce causal curves in our spacetime without knowing the metric with which our set is endowed.

Theorem 3.16. *A continuous curve $\gamma : I \longrightarrow X$ is future directed and timelike if, and only if, for all $t_0 \in I$ and all open convex sets O containing $\gamma(t_0)$, there exists an open subinterval $I' \subseteq I$ containing t_0 such that*

$$\begin{aligned} t \in I' \quad \textit{and} \quad t < t_0 &\implies \gamma(t) \ll \gamma(t_0) \\ t \in I' \quad \textit{and} \quad t_0 < t &\implies \gamma(t_0) \ll \gamma(t) \end{aligned} \tag{3.7}$$

If we replace $\ll$ by $\prec$ in equation (3.7) we obtain instead that γ is *future directed* and *causal*. Finally the following definition will be of great importance in what follows:

Theorem 3.17. *A continuous curve $\gamma : I \longrightarrow X$ is a future directed null geodesic if, and only if, for every open convex set O containing $\gamma(t)$, $t \in I$, there exists an open subinterval $I' \subseteq I$ containing t such that*

$$t_1, t_2 \in I' \quad \textit{and} \quad t_1 < t_2 \implies \gamma(t_1) \rightarrow \gamma(t_2) \tag{3.8}$$

Dual results to 3.16 and 3.17 hold for past directed curves.

In the case of distinguishing spacetimes we can make a further characterization of causal curves:

Theorem 3.18. *Let (X, g) be a distinguishing spacetime. A continuous curve $\gamma : I \subset \mathbb{R} \longrightarrow X$ is causal if and only if it is totally ordered by $\prec$.*

Proof. Since any connected subset of a causal curve is itself a causal curve, the implication to the right is trivial. Thus we focus on the proof of the implication to the left. Assume, thus, that either $t < t' \implies \gamma(t) \prec \gamma(t')$ or $t < t' \implies \gamma(t') \prec \gamma(t)$ for all t and t' in I. Let us see that if we have the first case, $t_1 < t_2 \implies \gamma(t_1) \prec \gamma(t_2)$, then we have $t_1 < t_3 \implies \gamma(t_1) \prec \gamma(t_3)$. Otherwise, since the spacetime is distinguishing, defined $L^+[p] := J^+[p] - \{p\}$, it is, $\mathrm{Cl}(L^+[p]) \cap L^-[p] = L^+[p] \cap \mathrm{Cl}(L^-[p]) = \emptyset$. Hence, if we choose $\gamma(t_1) = p$, there will be $\bar{t}$ contained in either $[t_2, t_3]$ or $[t_3, t_2]$ such that $r = \gamma(\bar{t}) \in \mathrm{Cl}(L^+[p]) \cap \mathrm{Cl}(L^-[p])$, which is a contradiction to the assumed fact that the spacetime is distinguishable (we must have either $r \in L^+[p]$ or $r \in L^-[p]$. Analogous reasoning shows that $t_3 < t_1 \implies \gamma(t_3) \prec \gamma(t_1)$.

Now, take $t_0 \in I$ and denote $p_0 = \gamma(t_0)$, and let U be a convex neighborhood of p_0 and $V \ni p_0$, $V \subset U$ a neighborhood which distinguishes p_0 in U. If $t_0 < t$ then $p_0 \prec \gamma(t)$ and, hence, $p_0 \prec_V \gamma(t)$, $p_0 \prec_U \gamma(t)$. □

Keep in mind that this previous result is a property of distinguishing spacetimes, but does not characterize them exactly. One may have that continuous curves totally ordered by the causality relation are either future or past directed causal curves in only future (or only past) distinguishing spaces.

Strongly causal spacetimes This is the kind of spacetime we will deal with the most throughout this dissertation. The proof of the following theorem can be found in lemma 3.21 of [6].

Theorem 3.19. *For any spacetime (X, g), the following two sentences are equivalent:*

1. *For any neighborhood $U \ni p$, there exists a neighborhood $V \subset U$, $p \in V$, such that V is causally convex in X (thus, also causally convex in U).*
2. *For any neighborhood $U \ni p$, there exists a neighborhood $V \subset U$, $p \in V$, such that any future (or past) directed causal curve $\gamma : I \longrightarrow X$ with endpoints at V is entirely contained in U.*

With this result as motivation, we define a spacetime (X, g) to be *strongly causal at p* if it satisfies any of the equivalent items in 3.19. If the properties hold at every point of the spacetime, we simply say that (X, g) is strongly causal. A further equivalent definition for strong causality will be given in the next chapter in terms of the topology only.

The following theorem about strongly causal spacetimes is important for the last part of this dissertation. The proof for this theorem can be found in proposition 2.16 of [3].

Theorem 3.20. *Let $p \in X$. Then strong causality fails at p if and only if there exists $q \prec p$, with $q \neq p$, such that: $x \ll p$ and $q \ll y$ together imply $x \ll y$ for all x, y.*

The time-reversed version also holds. Thus, an immediate consequence of 3.20 is that strong causality cannot fail only at one point of the spacetime. This is because if strong causality fails at p, and q is like in 3.20, then strong causality also fails at q.

These spacetimes presented so far are the ones we will be concerned with mainly when we introduce the notion of a causal set and how these kind of spaces can be used to describe spacetime. The order in which these spacetimes were defined was such that the strength of the causal restriction increases in each step. Furthermore, each "step" is a *proper* subspace of the previous one. Thus, for example, distinguishable spacetimes are always causal and chronological, but not necessarily strongly causal.

It is important to mention that each spacetime in the previous "causal ladder" exhibits certain properties that lead us to assume that these are not physically reasonable spaces. For example, a space that is chronological is not necessarily distinguishable, whereas our experience tells us that the past and the future of each event are two sets which are perfectly distinguishable from event to event. In a similar way, spacetimes which are not

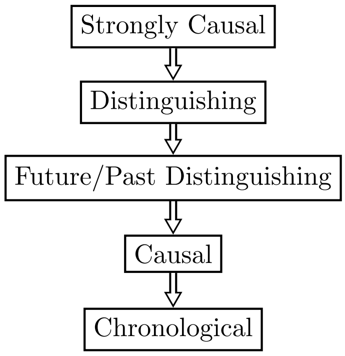

Figure 6: Shortened version of the causal ladder. Each “step” satisfies the causal properties of all the “steps” below it.

chronological lead to all the paradoxes that arise when discussing time-traveling. Yet, even though we have strong reasons to believe that these spaces are not physical, our best theory of gravity so far[17] does not eliminate these anomalies *a priori*. Furthermore, we have situations like the existence of closed causal curves in the Kerr Black Hole solution for Einstein equations. This solution arises from what we consider to be a very reasonable physical situation, and yet it leads to something we consider as non physical. Even strongly causal spacetimes are not physically ideal when considering quantum effects: quantum gravity is expected to take place in at least stably causal spacetimes [6]. So further restriction are needed to describe the most physically ideal scenario. Having these situations on mind, we would like to stress that when aiming to construct a new theory of space and time it is not necessarily a good idea to impose causality restrictions as a part of the postulates of such a theory. Instead, we would like to see what kind of restrictions we obtain from the given set of assumptions in our theory. This way, the causal properties can be used as a test for the theory. For example, in the causal set approach to spacetime that we will study in the next chapter, we see that a causal set spacetime cannot lead to a weaker than chronological spacetime. Thus, if we ever observe time-traveling for instance, we would know that our approach is incorrect.

[17] General relativity, without a doubt.

4 The causal approach to spacetime

As discussed in the main introduction, there are many different motivations to propose a discrete theory of spacetime. The causal set theory approach is the most successful of these discrete theories, one of the main reasons for this is that it preserves Lorentz invariance. However for the purposes of this chapter it is important that we consider *causal spaces* rather than *causal sets*. Precise definitions for these entities are given below, but the main difference between these two sets is that a causal *space* is not necessarily locally finite, as opposed to a casual set. The reason for choosing to deal with these kind of spaces is that the goal of this chapter is to compare the manifold approach of spacetime versus the causal set approach, and since a causal space allows both discreteness and continuity it is the most useful bridge between these two theories. The objective of the comparison is to give further motivation to the causal set theory of spacetime.

Lorentzian geometry is our main mathematical tool when aiming to describe spacetime. The causal structure $(X, \prec)$ studied in chapter 2 arises from the metric and the differential structure of our manifold spacetime (X, g). However, the set $(X, \prec)$ itself is rich in geometric information. These information should be treated very carefully as it may easily lead to confusions and misinterpretations.

We divide this chapter into two main sections. The first one is dedicated to introduce precisely the definition of a causal space. We have already dealt with some of the concepts presented in this section in an imprecise way. However, we write them here in a more appropriate way because the magnitude of the result presented in the second section demands a lot of unambiguity. Probably the most representative work on causal spaces is "*On the structure of Causal Spaces*" by Kroenheimer and Penrose [13], and for this reason most of this section is taken from their paper. The second section is dedicated to what is known as the *Malament-Hawking-King-McCarthy(MHKM) theorem*, which provides an important relation between $(X, \prec)$ and (X, g). These is a big result constructed from "smaller" theorems, but the complete proof of the theorem has never been gathered in a single reference, this is why we have decided to include it here. Each bit of the theorem is presented here and some very important observations are made.

4.1 Causal spaces

As we have seen, Lorentzian manifolds studied as models for spacetime allow us to define some causality relations between the events. When studying these relations in an axiomatic way we obtain what is called a causal space. However a causal space may lead to a wider class of sets and not only Lorentzian manifolds. For in what follows we drop the assumption that X is a manifold.

Definition 4.1. The quadruple $(X, \prec, \ll, \rightarrow)$ will be called a *causal space* if X is a set and $\prec, \ll$, and $\rightarrow$ are three relations on X satisfying the following conditions:

(I) $x \prec x$;

(II) if $x \prec y$ and $y \prec z$, then $x \prec z$, i.e. $\prec$ is transitive;

(III) if $x \prec y$ and $y \prec x$, then $x = y$;

(IV) No $x \ll x$;

(V) $x \ll y \implies x \prec y$;

(VI) If either $x \prec y$ *and* $y \ll z$, or $x \ll y$ *and* $y \prec z$, then $x \ll z$;

(VII) $x \rightarrow y$ if, and only if, $x \prec y$ *and* not $x \ll y$.

We refer to X as the *underlying set* and the relations $\prec, \ll, \rightarrow$ are called *causality, chronology* and *hormismos* respectively.

Note that no structure is assumed about the underlying set X, it could be a manifold spacetime as a particular example but this is not necessarily the case. Furthermore, although we have repeated the notation for the relations, note that in this context $\prec, \ll$ and $\rightarrow$ did not arise from the existence of a particular kind of path connecting two points in X, in this sense we are working in an axiomatic way. Nonetheless it is this previous situation what we are trying to resemble, hence the reason for repeating the notation.

Remark: The concept of a causal space is *not exactly* a more general type of spacetimes. That is, not every spacetime manifold constitutes a causal space. This is precisely because the chronology relation in spacetime manifolds may have closed loops, thus the property (IV) in 4.1 is not forced to hold. However, we do consider that causal spaces are a generalization of *physical* spacetimes.

In the case that we want to associate the relations on a causal space with the natural causal relations of a manifold spacetime, we note that such an association can only be made accurately if the spacetime in consideration is at least chronological, this is due to item (IV) in the definition of a causal space. Even though we mentioned it was a better option not to impose restrictions on the causal properties of the theory, such conditions are included here because it is known that, in any case, this result comes up in the causal set theory of spacetimes.

The chronological, causal and null futures (and pasts) of a given subset $A \subset X$ are defined in terms of the chronology, causality and horismos respectively in the same way as we did in section 2. Keep in mind however that this is a more abstract construction and that, although we are using the same terminology, we are not necessarily dealing with the physical situations we are used to consider when working with manifolds. Thus we denote

$$\begin{aligned} J^+[a] &= \{x \in X : a \prec x\} \\ I^+[a] &= \{x \in X : a \ll x\} \\ E^+[a] &= \{x \in X : a \to x\} \end{aligned} \tag{4.1}$$

Definition 4.2. A causal space is called *regular* if for distinct points $x_1, x_2, y_1, y_2 \in X$ and $x_i \to y_j$ for each i and j, then $x_1 \| x_2$ if and only if $y_1 \| y_2$. By $x \| y$ we mean that x and y are not related by any of the causal relations.

It is convenient to introduce here the *Alexandrov intervals*. These are defined as

$$\begin{aligned} [p, q] &:= J^+[p] \cap J^-[q] \\ \langle\langle p, q \rangle\rangle &:= I^+[p] \cap I^-[q]. \end{aligned} \tag{4.2}$$

If we now wish to study the local properties of a causal space we need to specify a topology. Given any set, we know that it is always possible to define a topology on it. However, of main importance is the so-called *Alexandrov topology*.

Definition 4.3. The *Alexandrov topology* $\mathscr{T}^*$ on a set X equipped with a non reflexive partial order $\ll$ (for instance, a causal space) is the weakest topology on X in which each $I^+[x]$ and $I^-[x]$ are open sets for any $x \in X$.

In this way, the open Alexandrov intervals $\langle\langle x, y \rangle\rangle$, with $x \neq y$, form a basis for the Alexandrov topology[18].

Although the causal ladder presented in the previous section was defined for manifold spacetimes, we can note that each classification is made in terms of the causal relations. Thus, we adopt this same causal ladder for causal spaces according to the character of its chronology relation. For example, every causal space is by definition a chronological space.

When defining a causal space, we note that the horismos relation is determined by the causality and the chronology. However, we can always construct a causal space starting from any of the causal relations. These constructions will be different (in general) depending on which relation one uses to construct the other two. Of our particular interest is the construction made from the causality, because it is this relation the one that leads the structure of a causal *set* when dealing with the discrete case. To see more about the construction of causal spaces starting with the horismos or the chronology we refer the reader to [13, 6].

Definition 4.4. Let X be a set equipped with a reflexive partial order $\prec$. We define

[18] We assume that all of our causal spaces are *full* in the sense of [13].

- The $\mathfrak{C}$-horismos, $\to^{\mathfrak{C}}$. We say that y is to the $\mathfrak{C}$-horismos of x, written $x \to^{\mathfrak{C}} y$, if and only if, $x \prec y$ and $\prec$ linearly orders $[u, v]$ whenever $[u, v]$ is a *proper* subset of $[x, y]$.
- The $\mathfrak{C}$-chronology, $\ll^{\mathfrak{C}}$. We say that y is to the $\mathfrak{C}$-chronological future of x, written $x \ll^{\mathfrak{C}} y$, if and only if, $x \prec y$ and not $x \to^{\mathfrak{C}} y$.

By construction, the relations $\prec, \ll^{\mathfrak{C}}$ and $\to^{\mathfrak{C}}$ will satisfy properties (I) to (VII) in definition 4.1. Hence the quadruplet $(X, \prec, \ll^{\mathfrak{C}}, \to^{\mathfrak{C}})$ is a causal space.

Now, if we are given a set X and a reflexive partial order $\prec$ just as in the previous definition, it is true that the chronology and the horismos relation that would complete the pair $(X, \prec)$ to get a causal space are not unique. For this reason it is *not always* true that

$$(X, \prec, \ll, \to) = (X, \prec, \ll^{\mathfrak{C}}, \to^{\mathfrak{C}}). \tag{4.3}$$

The causal spaces for which equation 4.3 holds, are called $\mathfrak{C}$-spaces. If the space is regular, a sufficient condition for $(X, \prec, \ll, \to)$ to be a $\mathfrak{C}$-space is that $x \to^{\mathfrak{C}} y \implies x \to y$ (The implication in the reverse order is deduced from the regularity of the space). Important point worth mentioning here (proved in the next section) is that a manifold spacetime regarded a causal space is a $\mathfrak{C}$-space.

4.2 The axioms of causal set theory

The reason for studying causal spaces arises from the idea that the physics that we observe are merely a result of the fact that *cause* always precedes the *effect*. Of course, this thought is very general and Causal Set theory is just one way of formalizing the statement. It is not the only way in which the hypothesis can be approached, but it is definitely the most successful one. Following the line of [14], we shall depart from such assumption and refer to it as the *Causal Metric Hypothesis:*

" *The properties of the physical universe are manifestations of causal structure*".

Though we expect to recover *something like* General Relativity as some sort of large scale approximation, the philosophical building block of Causal Set theory is already very different from what we are used to have. General Relativity could be summarized with the following line of thoughts: *There is spacetime and there is matter; matter tells spacetime how to curve, spacetime tells matter how to move.* On the other hand, the Causal Metric Hypothesis follows the complete opposite direction: "*Things happen; spacetime and matter are ways of describing them"* [14]. General Relativity is a theory that describes the relation between events, whereas Causal Set theory (and any other theory based on the Causal Metric Hypothesis) states that these relations *are* what determines everything else. In this work, however, we do not aim to see how to indeed obtain *everything else* from the causal relations. Our study is entirely focused on seeing how some of the results of (classical) General Relativity can be related to those of Causal Set theory. A discussion on the quantum implications demands way more effort and time, see for example [14, 16]. Thus, we may restrict the Causal Metric Hypothesis a bit further:

"*The properties of classical spacetime are manifestation of causal structure*".

Moreover, we believe that the underlying causal structure is that of a causal set. Thus, we set the hypothesis to a more specific one and which is more useful for our purposes here:

"*The properties of classical spacetime arise from the structure of a causal set*",

we shall call it *Classical Causal Metric Hypothesis*[19].

Before discussing any physical implications, we must give the precise set of axioms that define Causal Set theory. There are two ways in which we can present this axiomatic construction. One is known as the *irreflexive formulation*, and the other one is referred as *partial order* formulation. It is very important to mention that these two approaches are not entirely equivalent. The categories[20] that each formulation defines are different. However, they are *object-wise* equivalent and this nuance has not led to different conclusions about the physics that the theory aims to describe. Thus, an in-depth discussion of both systems of axioms is pointless for the objective of this work. Due to its resemblance with the causality relation in GR, we have chosen to present here only the partial order formulation.

The partial order formulation of Causal Set theory is based on interval finite partial orders. A *partial order* $\prec$ on a set P is a reflexive, antisymmetric, transitive, binary relation on P. By this we mean that the binary relation is such that:

- For every $x \in P$ we have that $x \prec x$ (reflexivity),
- If $x \prec y$ and $y \prec x$, then $x = y$ (antisymmetry),
- If $x \prec y$ and $y \prec z$, then $x \prec z$ (transitivity).

If, in addition, the partial order is such that the intervals

$$[x, z] := \{y \in P \mid x \prec y \prec z\}$$

have finite cardinality, then we say that the partial order is *interval finite.*

We now introduce the formal definition.

[19]This one is a slightly stronger hypothesis than the one introduced by Dribus [14] with the same name, but we shall not worry about this subtle difference.

[20]The collection objects and morphisms

Definition 4.5. A *causal set* is a pair $(C, \prec)$ where C is a countable[21] set and $\prec$ is an interval finite partial order in C, physically interpreted according to the following axioms:

(C1) The events in classical spacetime can be modeled by the set C, and $\prec$ represents the causal relations between pairs of events.

(C2) The volume of a spacetime region modeled by a subset $B \subset C$, is equal to the cardinality of B in fundamental units, up to Poisson-type fluctuations.

In recent years, the symbol that is used to represent the partial order has been replaced with $\preceq$. This has the advantage that, due to its similarity with the *less or equal than* relation for real numbers[22], it emphasizes the reflexive property of the relation. However, using $\prec$ makes it easier for the reader that wishes to compare with the references mentioned in this work, since this is the most common terminology. The point to keep in mind here is that $x \prec x$ is a true statement.

When referring to causal sets, we will follow the tradition ([14, 18, 20, 22]) and abuse of the notation by making statements such as "*...the causal set C...*", i.e. we will regard the relation $\prec$ as implicitly defined (unless otherwise stated). Any subset of a causal set $B \subset C$ can be regarded as causal set itself by considering the pair $(B, \prec_B)$. Here $\prec_B$ is the restriction of $\prec$ to the set B. In other words, for $x, y \in B$ we say that $x \prec_B y$ if and only if $x \prec y$.

Now recall the causal hierarchy presented in section 2. This "causal ladder" was intended to underline the topological restrictions that have to be imposed *a posteriori* in order to get rid of unphysical situations. Conversely to that scenario, the definition of a causal set immediately places us *at least* on the step of causal spacetimes. The antisymmetry *and* transitivity of $\prec$ imply that C cannot have cycles (except for the trivial ones $x \prec x$). That is, the relations

$$
\begin{aligned}
x &\prec x_1 \\
x_1 &\prec x_2 \\
&\vdots \\
x_{n-1} &\prec x_n \\
x_n &\prec x
\end{aligned}
\tag{4.4}
$$

cannot hold unless $x_i = x$ for all $i = 1, ..., n$. It is important to remark that acyclicity is not a consequence of the antisymmetry property only. It is only when combining antisymmetry with transitivity that the result follows. This fact becomes important when studying other types of relations, such as the horsimos $\rightarrow$, which is not transitive. A more profound

[21] Uncountable sets lead to unreasonable physical results [39].

[22] $\leq$.

discussion of these mathematical properties has to do with the study of Acyclic Directed Sets [14, 37].

Due to the axiom 1, a cycle in the above sense, would correspond physically to the existence a Closed Causal Curve (CCC) in spacetime. Thus, Causal Set theory eliminates all possible spacetimes that encounter this situation. This means that Causal Set theory will not lead us to deduce, for example, the Kerr black hole solution of Einstein equations. Is this a problem or an advantage? The antisymmetry and transitivity of $\prec$ are deliberately included in order to obtain this result. The reason for this is that the existence of CCC's leads to discussions on time-traveling, which in turn lead to several paradoxes. Hence, the scientific community has agreed on the idea that time-traveling is physically impossible phenomenon (with a few sceptic members [12]). If this is indeed the case, an immediate consequence is that GR is not a complete theory not even on classical grounds[23]. This is further motivation for considering alternative theories. On the other hand, we are not sure if acyclicity is the right way to avoid these problem. Certainly, we know that the Kerr black hole solution arises from applying GR to a very physical situation: the collapse of a rotating star. We are not yet in position to say whether acyclicity is the right choice to make.

Now, a lot of the physics of spacetime that we know today rely on derivations obtained from concepts defined in terms of the Lorentzian metric g. Note that a causal set is not endowed with a metric, thus we have plenty of physics information lost. This is where axiom (C2) comes in. However, its physical significance will become evident until the end of the next section. MHKM theorem shows us how a causal order *may* be used to *almost* determine a Lorentzian metric in the spacetime. The reason for not being able to completely determine the metric is because the causal order is not enough to provide a volume measure. Hence a conformal factor is left undetermined. However, if a measure $\mu : \mathcal{P}(C) \to \mathbb{R}^+$ is specified, then interpreting it as the volume of spacetime regions would completely determine the missing conformal factor. The condition "up to Poisson-type fluctuations" is needed to preserve Lorentz invariance (but this topic is outside the scope this dissertation). This is what R. Sorkin means by

"*Order plus number equals geometry*"[24].

4.3 Physical information encoded by the causal order.

We are now in position to present the main focus of this dissertation: the MHKM theorem. This result arises when a set of other results is gathered. These are the theorems due to

[23]Of course, this is not the only classical situation in which GR breaks down. See for example, [4], where unstable solutions to Einstein equations are discussed.

[24]http://www.einstein-online.info/spotlights/causal_sets

Malament, Hawking, King, McCarthy and Penrose. ([7, 9, 3]). We collect each of those theorems in this section as well as their proofs.

In its present form, the theorem reads as follow [20]:

Theorem 4.6. *If a causal bijection f exists between two n-dimensional spacetimes, (X_1, g_1) and (X_2, g_2), which are both future and past distinguishing, then these space times are conformally isometric when $n > 2$.*

This result is a very powerful motivation for the causal set description of spacetime. The reason for this is that the theorem means, in other words, that the natural causal structure $(X, \prec)$ of an n-dimensional distinguishing spacetime is enough to determine its conformal geometry as well as its topology. As we present how this theorem arises, we give some important remarks.

In the first place, note that the statement of the theorem is between *spacetimes* in the classical GR sense. This means that we are working with Lorentzian manifolds and the metrics are known. For this reason we must be very careful when we state that $(X, \prec)$ is enough to know $(X, [g])$. Even when this is not needed for every result in this section, the pair $(X, \prec)$ is not assumed to be a causal set but a manifold endowed with a partial order. This is done because concepts such as metric or differential structure are in general not possible to express solely in topological terms. In other words, throughout this section we shall be comparing these two sets: $(X, \prec)$ and $(X, [g])$, where X is a Lorentzian, real, C^∞ Hausdorff manifold of dimension $n > 2$, $\prec$ is the natural causality on X, and $[g]$ is the equivalence class of metrics on X which are conformal to g. Note that in the second case we assume the metric is given whereas in the first one we assume that we only have knowledge of the natural causality.

4.3.1 Causality determines the chronology and the horismos.

Peter B. Kroenheimer and Roger Penrose, 1966 [13][25]

Note that, of the three causal relations that the metric gives place to, we have given a special place to the causality. Assuming the causality relation is known, we would like to see if it is possible to recover from it the chronology and horismos relation. This is indeed the case, as is shown by the following results.

Lemma 4.7. *The causal horismos, $\to^{\mathfrak{C}}$, coincides with the natural horismos $\to$.*

Proof. ($\Leftarrow$) Let $x \to y$. By definition we then have that $x \prec y$. Assume there exists a proper subset $[u, v]$ of $[x, y]$ which is not linearly ordered by $\prec$. In such case, this means that we can find $p, q \in [u, v]$ such that $p \parallel q$. But since $p, q \in [u, v] \subset [x, y]$ we must have

[25] The proof presented here, however, was obtained from [6]

$x \to p$, $u \to p$, $x \to q$ and $u \to q$ (any other possibility such as $x \ll p$ will imply, by proposition 3.11, that $x \ll y$, which is a contradiction to our initial assumption). Since the natural horismos $\to$ is a regular horismos[13] these relations imply that $x = u$. In an analogous way, we conclude $y = v$. Thus $[u, v]$ cannot be a *proper* subset of $[x, y]$ nonlinearly ordered by $\prec$. This proves that $x \to y \implies x \to^{\mathfrak{C}} y$.

($\Rightarrow$) If $x \to^{\mathfrak{C}} y$, by definition we have that $x \prec y$. Then either $x \ll y$ or $x \to y$. But if $x \ll y$ we then could take p, q, with $x \prec p \prec q \prec y$ such that $I^+[p] \cap I^-[q] \neq \emptyset$, and thus the set $J^+[p] \cap J^-[q]$ would not be linearly ordered by $\prec$, which is a contradiction to the initial assumption. Thus $x \to y$ is the only possibility. $\square$

Since the causal and natural chronologies, $\ll$ and $\ll^{\mathfrak{C}}$, have equivalent definitions in terms of the causal and natural horismos respectively we conclude that these chronologies are equal. We have hence proved the following theorem [6]:

Theorem 4.8. *The causal space constructed from the natural causal relations,* $(X, \prec, \ll, \to)$*, coincides with* $(X, \prec, \ll^{\mathfrak{C}}, \to^{\mathfrak{C}})$.

A more general approach can be found in [13]. Although we will eventually need to restrict ourselves to strongly causal spacetimes, note that for the previous result no assumptions have been made about the hierarchical position of (X, g) in the causal ladder. Keep in mind, however, that when working with the natural causality relations on (X, g) we are implicitly assuming that the set X possesses a given manifold structure; we are not assuming that $(X, \prec)$ is any *arbitrary* set endowed with a partial order. Yet, should we eventually find out that spacetime fails to have a continuous description at small scales, the previous equality between relations tells us that the causal structure with which we are familiarized can in principle be obtained from a more fundamental structure.

4.3.2 Causality determines the topology of the spacetime.

Roger Penrose, 1970 [3]

We have to be careful when presenting this result. At this point we assume that (X, g) is strongly causal. We state that the causality relation carries the information that determines the topology of spacetime because of the next result, which is proposition 4.24 in [3].

Theorem 4.9. *The following three conditions on a spacetime X are equivalent:*

i) (X, g) *is strongly causal;*

ii) $\mathscr{T}^* = \mathscr{T}^{man}$*;*

iii) $(X, \mathscr{T}^*)$ *is Hausdorff.*

Proof. $(i) \implies (ii)$ First, by item 2 in proposition 3.12, the *Alexandrov intervals* $I^+[p] \cap I^-[q]$ are open in M since they are the intersection of two open sets. Thus we immediately have $\mathscr{T}^* \leq \mathscr{T}^{man}$, this is always true regardless of the causality restrictions on the spacetime. Take now $P \in \mathscr{T}^{man}$ and $p \in P$. Let N be a simple region of p contained in P. Because of strong causality assumption, we can find a causally convex open set Q such that $p \in Q \subset N$. By proposition 3.14 we have $u, v \in Q$ such that $p \in \langle u, v\rangle_N \subset Q$. If it happens that $\langle u, v\rangle_N \neq \langle u, v\rangle$ this signifies the existence of a causal curve from u to v which is not entirely contained in N, hence not entirely contained in Q. This is a contradiction to the causal convexity of Q. Thus, $\langle u, v\rangle \subset Q \subset P$. Which means $\mathscr{T}^{man} \leq \mathscr{T}^*$. Hence, $\mathscr{T}^* = \mathscr{T}^{man}$.

$(ii) \implies (iii)$. This is immediate since the manifold topology of any spacetime is Hausdorff by definition.

$(iii) \implies (i)$. Suppose X is not strongly causal at some p. Then by lemma 3.20 there exists $q \prec p$ such that: $x \ll p$ and $q \ll y$ together imply $x \ll y$ for any x, y. Let $p \in \langle x, u\rangle$ and $q \in \langle v, w\rangle$. We have $q \prec p \ll u$, so $q \in I^-[u]$. We can choose y such that $q \ll y$, $y \in I^-[u]$ and $y \in \langle v, w\rangle$. But since $x \ll y$ we also have that $y \in \langle x, u\rangle$. Thus $\langle x, u\rangle \cap \langle v, w\rangle \neq \emptyset$. This means that every Alexandrov neighborhood of p intersects every Alexandrov neighborhood of q, which is a contradiction to the assumption (iii). Thus, X must be strongly causal. □

The importance of this result relies on the implication that, in a strongly causal spacetime, one can determine its topology by knowing the causal relations. We summarize this result in the next theorem.

Theorem 4.10. *If* (X, g) *is a strongly causal spacetime, then* $(X, \mathscr{T}^*)$ *coincides with* $(X, \mathscr{T}^{man})$ *as topological spaces.*

This has a remarkable implication. Consider, for instance, that all information of our spacetime is lost except for the natural causality relation and the events. We would like to know if the remaining bits correspond to what once was a strongly causal spacetime. The usual way to do this is to see if the Alexadrov topology and the manifold topology coincide. But in this situation this is not possible to do: if we have lost the information about the coordinate functions we do not have a manifold topology. However, if we know the causality, we can define from it the $\mathfrak{C}$-chornology. From it, we can construct the Alexandrov topology taking the intervals

$$\langle\langle x, y\rangle\rangle := \{z \in X \mid x \ll^{\mathfrak{C}} z \ll^{\mathfrak{C}} y\}$$

as a basis. Then we ask if it happens that the Alexandrov topology is Hausdorff. If that is the case, we would know that the spacetime that we had (before the unfortunate lost of information) was strongly causal.

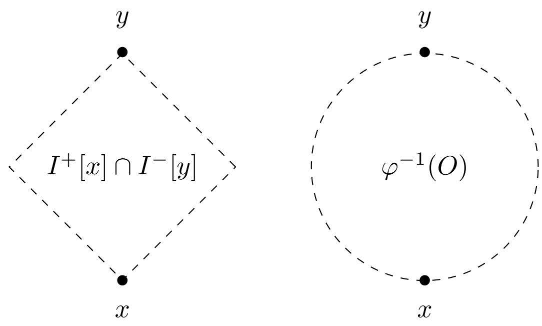

Figure 7: To the left: an Alexandrov open set, basis for $\mathscr{T}^*$. To the right: An open set in the manifold topology $\mathscr{T}^{man}$, formed by the pre-image of open sets in $\mathbb{R}$ of the coordinate functions.

Combining the results 3.16 and 3.17 of the previous chapter with the two previous theorems, we can take these results to *define* timelike and causal curves.

Definition 4.11. A continuous curve $\gamma : I \longrightarrow X$ is *future directed* and *timelike* if, and only if, for all $t_0 \in I$ and all open convex sets O containing $\gamma(t_0)$, there exists an open subinterval $I' \subseteq I$ containing t_0 such that

$$\begin{aligned} t \in I' \quad \text{and} \quad t < t_0 &\implies \gamma(t) \ll^{\mathfrak{C}} \gamma(t_0) \\ t \in I' \quad \text{and} \quad t_0 < t &\implies \gamma(t_0) \ll^{\mathfrak{C}} \gamma(t) \end{aligned} \tag{4.5}$$

Even though we have seen that the $\mathfrak{C}$-chronology coincides with the natural chronology, the notation $\ll^{\mathfrak{C}}$ has been used in place of $\ll$ to remind ourselves that our development has been built entirely from the causality relation. If we replace $\ll^{\mathfrak{C}}$ by $\prec$ in equation (4.5) we say instead that γ is *future directed* and *causal*. In the same way, we now take the result on null geodesics as a *definition*.

Definition 4.12. A continuous curve $\gamma : I \longrightarrow X$ is a *future directed null geodesic* if, and only if, for every open convex set O containing $\gamma(t)$, $t \in I$, there exists an open subinterval $I' \subseteq I$ containing t such that

$$t_1, t_2 \in I' \quad \text{and} \quad t_1 < t_2 \implies \gamma(t_1) \to^{\mathfrak{C}} \gamma(t_2) \tag{4.6}$$

Dual definitions to 4.11 and 4.12 can be given for past directed curves.

4.3.3 Causality encodes the differential structure.

Stephen Hawking, Andrew King and P. J. McCarthy, 1976 [9]

So far we have seen how the causality is enough to identify the topology of our spacetime. Furthermore, because the natural causal relations coincide with the $\mathfrak{C}$-relations, we can use these to define the class of timelike and spacelike curves, as well as null geodesics, just as we did in section 2. Because we know the topology, we then know which maps are continuous. We see here that this can identify the differential structure of X.

Theorem 4.13. *A homeomorphism $f : X \longrightarrow X$ that takes null geodesics to null geodesics is a diffeomorphism.*

Proof. Let $\mathcal{U}$ be a convex normal neighborhood and let $\gamma_i : F_i \longrightarrow \mathcal{U}$, $i = 1, ..., 4$, be four null geodesic paths satisfying the following properties:

1. For each $t_1 \in F_1$, there is a unique null geodesic λ in $\mathcal{U}$ from $\gamma_1(t_1)$ to the null geodesic γ_2.
2. For each $t_3 \in F_3$, there is a unique point $q \in \lambda$ for which q and γ_{t_3} lie on the same null geodesic in $\mathcal{U}$.
3. For each point $q \in \lambda$ there is a unique $t_4 \in F_4$, such that q and $\gamma_4(t_4)$ lie on the same null geodesic in $\mathcal{U}$.
4. If we define $\psi(t_1, t_3) := t_4$, with the parameters t_1, t_3 and t_4 as in (1)-(3), then ψ is a C^∞ map and, furthermore, $\partial\psi/\partial t_1 \neq 0 \neq \partial\psi/\partial t_3$.

These properties can always be satisfied in Minkowski spacetime. Hence, we can achieve (1)-(4)[26] by choosing $\mathcal{U}$ small enough.

Since f preserves null geodesics, $f(\gamma_i)$ $(i = 1, ..., 4)$, will be four null geodesics in $f(\mathcal{U})$. Thus one can find four paths $\tilde{\gamma}_i : \tilde{F}_i \longrightarrow f(\mathcal{U})$ with the same images as the $f(\gamma_i)$ but with the possibility of being parametrized in a different way. The matter to consider here is whether or not we can map between this parameters in a smooth way.

Define $h_i : F_i \longrightarrow \tilde{F}_i$ by

$$h_i := \tilde{\gamma}_i^{-1} \circ f \circ \gamma_i. \quad (i = 1, ..., 4). \tag{4.7}$$

Defined in this way h_i will be four continuous monotonic maps. Thus, by Lebesgue's theorem[27], h_i is differentiable almost everywhere. Define $\tilde{\psi} : \tilde{F}_1 \times \tilde{F}_3 \longrightarrow \tilde{F}_4$ in an analogous way as ψ. Then,

$$h_4(\psi(t_1, t_3)) = \tilde{\psi}(h_1(t_1), h_3(t_3)). \tag{4.8}$$

[26] Property (4) can only be achieved if the dimension of X is ≥ 3.

[27] This states that if $F : \mathbb{R} \longrightarrow \mathbb{R}$ is monotonic, i.e strictly increasing or decreasing, then F is almost everywhere differentiable, with respect to the Lebesgue measure [38].

Differentiating with respect to t_3 on the previous equation we get

$$h_4'(\psi(t_1,t_3))\frac{\partial\psi}{\partial t_3} = \frac{\partial\tilde{\psi}}{\partial\tilde{t}_3}h_3'(t_3). \tag{4.9}$$

Since h_3 is almost everywhere differentiable, property (4) implies that h_4' exists and is continuous, any anomaly of h_4 can be avoided appealing to its double argument, i.e. variating t_1 and t_3 in a convenient way. As above, we can show that each h_i is C^1 by choosing different combinations of the null geodesics. If we differentiate the previous equation again, this time with respect to t_1 we get

$$h_4''(\psi)\frac{\partial\psi}{\partial t_1}\frac{\partial\psi}{\partial t_3} + h_4'(\psi)\frac{\partial^2\psi}{\partial t_1\partial t_3} = \frac{\partial^2\tilde{\psi}}{\partial\tilde{t}_1\partial\tilde{t}_3}h_1'h_3'. \tag{4.10}$$

The same previous argument implies that h_4 is C^2. This can be repeated indefinitely, hence showing that h_4 is C^∞. And by taking different combinations of the geodesics we conclude that h_i is C^∞ for $i = 1,...,4$. Thus f is a map between C^∞ parameters on null geodesics.

Finally, let $\gamma_i : F_i \longrightarrow \mathcal{U}$, $i = 1,...,n$, be n null geodesics, and let $W \subset \mathcal{U}$ be a neighborhood. The map $\phi : W \longrightarrow O \subset \mathbb{R}^n$, with O open, defined by $\phi_i(q) := \gamma_i^{-1}(E^{\pm}[q,\mathcal{U}])$ can be made C^∞-compatible by choosing W small enough. Thus the pairs[28] (W,ϕ) form a C^∞ atlas for X, and this atlas is preserved by f if we choose the null geodesics γ_i as described in (1)-(4). Hence f is in this way a C^∞ diffeomorphism. □

This is the core result and one of the main motivations for the causal set approach to spacetime. There are, however, some important points that I believe should be clarified.

First of all, the theorem itself is an statement between spacetimes for which the metric is known. The reason why this remark is important is because *convex normal neighborhoods* are used for the proof of this theorem. As explained in chapter 2, these are a very special kind of neighborhoods since the behavior of *geodesics* is very special within them. This is a problem because the causality relation, as far as I concern, is not sufficient to characterize all kinds of geodesics but the null ones. Thus it is not clear how to construct convex normal neighborhoods using the causality relation nor if another kind of neighborhood (possibly Alexandrov neighborhoods) can be used to proof this theorem.

The second remark that I would like to note is that there are two references in which the proof of this theorem can be found. These are the Hawking's Adams prize essay [10] and the paper by Hawking, King and McCarthy [9]. The authors in [9] state that the version of the theorem presented in their paper is an improved version of the first proof constructed by Hawking. In any case, to claim that the causal relation encodes the differential structure the assumption of strong causality is made. The proof that I've chosen to write here is the "improved" version. However, I have made a slight modification:

[28] $\phi(q) = (\phi_1(q),...,\phi_n(q))$.

the coordinate function $\phi(q)$ is originally defined as $\Gamma(q) = \gamma_i^{-1}(I^+[q, \mathcal{U}])$. I have changed the definition of the coordinate function because I believe what I write here is what was originally meant, because the original function does not maps to a point in $\mathbb{R}$ but to an interval, thus it fails to be an appropriate function to use as coordinate. However, even if using $\phi(q) = \gamma_i^{-1}(E^{\pm}[q, \mathcal{U}])$, it not clear that this functions fulfill all the requirements to describe an atlas. The authors claim that this is obvious by "comparison with Minkowski's space", but I have not found this comparison helpful to show that the pairs (W, Γ) form indeed an atlas.

4.3.4 Causality implies the conformal metric

.

Stephen Hawking and George Ellis, 1973. [4][29]

In order to obtain this result, we need to assume that the statement that gives the title of the previous section holds. Although we suspect that the statement is true, we believe it would be convenient to construct a proof of the theorem 4.13 without using normal neighborhoods. Assuming thus, that we can get the differential structure from the causality relation, we then can define the tangent space at each point of X. Furthermore, we can classify the vectors at each tangent space as timelike, spacelike or null, according to the kind of curve for which they are tangent. Of particular importance are the null vectors. The proof of the following theorem is found in [4] p. 61.

Theorem 4.14. *The set of null vectors at $p \in X$, determines the metric g up to a conformal factor.*

Proof. Take $W, Z \in T_pX$ to be any two non-null vectors, i.e. two vector that are tangent to either a spacelike or a timelike curve. Since the metric g is a 2-tensor we have, by definition, that it must be bilinear. Furthermore, we know it must also be a symmetric mapping. Thus, we have that

$$g(W, Z) = \frac{1}{2}\left[g(W + Z, W + Z) - g(W, W) - g(Z, Z)\right]. \tag{4.11}$$

Now consider $V \in T_pX$ to be a timelike vector and $U \in T_pX$ a spacelike one. Then by looking for the *null combination*

$$0 = g(V + \lambda U, V + \lambda U) = g(V, V) + 2\lambda g(V, U) + \lambda^2 g(U, U), \tag{4.12}$$

we find that there are two possible choices of λ in order to satisfy the above equation (because $g(V, V) < 0$ and $g(U, U) > 0$). Since we have complete knowledge of the null

[29]The title of the section is not exactly what they proved, but it follows when combining it with the previous results.

cone, we can in principle determine the two roots for 4.12. Moreover, if λ_1 and λ_2 are such roots, we note that

$$\lambda_1 \lambda_2 = \frac{g(V,V)}{g(U,U)}. \tag{4.13}$$

Hence, by knowing the null cone we know the ratio between a spacelike and a timelike vector. Therefore, we can determine every term in the RHS of 4.12 up to a factor. Hence we can determine $g(W,Z)$ up to a factor.

□

Conclusion

If we were to construct physics from scratch, knowing that

all things are made of atoms

would save us many years of working in the wrong direction [40]. A persistent difficulty in building a physical theory is that we usually do not have enough mathematical experience to describe it in a formal abstract way. For instance, when studying fundamental GR the main mathematical tool to do this is Lorentzian geometry. However, the mathematicians community has an obvious preference for the Riemannian one. This situation is not even as unfortunate as the case of Newton, when he had to develop calculus by himself. Developing a mathematical theory and a physical one at the same time requires an immense amount of effort. After many centuries we finally know that *things* are indeed made of atoms. But, is *everything* made of "atoms"? Causal Set theory is based on the assumption that even spacetime is fundamentally discrete. Although it results complicated to imagine a discrete spacetime[30], it is a reasonable idea if we think that our measurement devices are too big to detect the discontinuity, in the same sense that things appear to be continuous to our naked eye.

The special part of Causal Set theory is that it regards causal order as the rule that determines the behaviour of this discreteness. Mathematically, the study of Causal Sets corresponds to study *directed* graphs. Again, mathematicians have preferred to focus on *undirected* graphs. This makes our task of constructing a physical theory more difficult and imposes many limitations. Nevertheless, we have managed to show that causal order is indeed a very strong relation. From it, we can determine the topology, the differential structure, and a conformal metric. Although we did not explicitly show it, the *full* metric can then be recovered once a volume element has been specified[14]. This is why it is important to include axiom (C2) if we want to accurately describe the physics of GR from Causal Set theory. However, keep in mind that GR is only part of all the physics that we know today. If Causal Set theory is in fact the way of describing fundamental physics, then we still have a long way to go. For instance, we need to be able to explain QFT. Plus, aiming to be a theory of quantum gravity, Causal Set theory must be able to explain what happens at the final stage of the "evaporation" of a black-hole[41]. Related to this topic is the question of whether Causal Set theory provides a proof for the Cosmic Censorship conjecture[42]. Furthermore, all of these problems does not even involve the phenomenological side of the theory.

[30] One is usually tempted to think of these spacetime "atoms" as embedded in another set. This is not the case we are aiming to describe with Causal Sets.

Yet, even at this early stage of its development, Causal Set theory has already remarkable implications. For example, in section 4.2 we saw how the properties of the relation $\prec$ make it impossible to obtain cycles. This is in agreement with the fact the we have not observed time traveling or any equivalent situation. Note, however, that one might consider a similar theory in which the order $\prec$ does allow cycles. Then, even in that case, one could not model the situation of a CCC with such cycles. If there are cycles allowed, the points that constitute the cycle are all equivalent in the following sense: Take $(C, \prec)$, and $x, y, z \in C$ such that $x \prec y \prec z \prec x$.

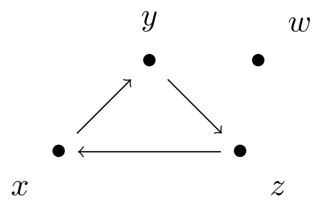

Figure 8: A cycle in a directed transitive set. The point w outside the cycle has no information to distinguish between points of the cycle.

Then, any other point outside the cycle does not make distinction between the points in it. This is because, if for example $w \in C$ and $x \prec w$, then $y \prec w$ and $z \prec w$. Similarly, if $v \| x$, then $v \| y$ and $v \| z$. In this sense, the points in the cycle are equivalent. Any topology we can construct from the *transitive* order relation $\prec$ does not resemble the situation of a closed causal curve:

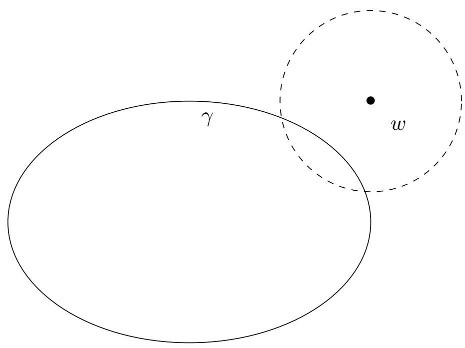

Figure 9: A closed causal curve or CCC γ. Points outside the loop can distinguish different points in the CCC by means of the manifold topology.

A point w outside the CCC has different degree of "closeness" with each point of the curve: some open sets can contain a and *only part* of the CCC. Thus even if cycles are allowed, the physical situation these would describe is not that of a CCC.

Note that when presenting the theorems that lead us to conclude that the causal order $\prec$ determine the conformal metric, we always took for granted that the underlying set was a manifold. If this assumption is relaxed, the situation gets complicated. The particularly complicated step is how to construct a differential structure. As we have mentioned before, a topology is not always enough to obtain a differential structure. Causal Set theory would be even more convincing if we could recover the metric *and* the manifold from the pair $(C, \prec)$. The answer to this problem might be in the study of *discrete manifolds*.

Even in the case of a continuous manifold, we pointed out that there are some steps that must be clarified. The definition of the coordinate function should be revised if we want to give the exact expression for the atlas. However, comparing with the construction made by Zeeman for Minkowski spacetime [43], we suspect that it can be proved that constructing an atlas in the same fashion is always possible as long as one consider sufficiently small (Alexandrov) neighborhoods. Another point to remark here is that we can only recover a C^∞ differential structure if the metric of the manifold is also C^∞. This important point is explicitly stated in Hawking's essay [10] but is not mentioned in the "improved" paper. Another point to remark is that on Malament's extension to FPD spacetimes (appendix), the manifold topology is no explicitly recovered and thus it is not clear to which $\mathbb{R}^n$ the manifold is homeomorphic. This subtlety is treated by O. Parrikar [20].

To conclude, we would like to remark that this central result for Causal Set theory should be attributed (mainly) to Malament, Hawking, Penrose and Kroenheimer.

A Malament's extension of HKM to distinguishing space-times.

On 1976, D. Malament published his work "The class of continuous timelike curves determines the topology of space-time"[7]. This work extends the results previously mentioned in the previous sections to distinguishing space-times. Malament does this by showing that strong causality is a *more than sufficient* conditions to recover the space-time topology. In fact, Malament shows that the weaker conditions in which we can obtain the topology are distinguishing space-times. By knowing the topology we then know the causal homeomorphisms and the last two sections then follow as a corollary under weaker conditions. Malament's theorem reads

Theorem A.1. *Let X and $\tilde{X}$ be space-times. If $f : X \longrightarrow \tilde{X}$ is a one-to-one mapping that preserves future directed timelike curves, then it is a homeomorphism with respect to the manifold topology.*

This previous is used to obtained the conclusion mentioned when combined with the following lemma:

Lemma A.2. *If f homeomorphism (with respect to the manifold topology) that preserves future directed timelike curves, then f also preserves null geodesics. Hence, because of theorem 4.13, it is a conformal diffeomorphism.*

Malament proves his theorem by characterizing continuity in terms of the convergence of sequences. He develops some properties that must be satisfied by the set of points in which the function in consideration is discontinuous. The properties derived lead to the conclusion that such a set must be empty. His proof is lengthy because it requires proving first a series of lemmas. Unfortunately, we do not know any shorter versions of his method. Thus, the proof presented here is exactly the one employed by Malament.

In what follows, f will always denote a bijection between the space-times X and $\tilde{X}$ that preserves (the same as f^{-1}) continuous future directed timelike curves, this type of bijections are usually called *causal bijections*. $\mathcal{D}$ and $\tilde{\mathcal{D}}$ will denote the subsets in which f and f^{-1} are respectively discontinuous. We shall prove at the end of these lemmas that $\mathcal{D}$ and $\tilde{\mathcal{D}}$ are in fact empty. Note that it is not obvious that f is continuous in the first place, since f is not assumed to preserve all but a restricted class of continuous curves (the future directed and timelike ones).

Lemma A.3. *If O is an open set in X, and $\tilde{O}$ is an open convex set in $\tilde{X}$, and $f[O] \subseteq \tilde{O}$, then $O \subseteq X - \mathcal{D}$.*

Proof. Let $\tilde{U}$ denote any open set in $\tilde{X}$ containing $f(p)$, $p \in X$. Since $\tilde{O}$ is convex, as a space-time it must be strongly causal. Thus, the Alexandrov topology on $\tilde{O}$ is equal to its relative manifold topology. Hence, $\tilde{U} \cap \tilde{O}$ is open in the Alexadrov topology on $\tilde{O}$. But, since f is assumed to preserve continuous timelike curves, then the restriction $f|_O \colon O \longrightarrow \tilde{O}$ is clearly continuous with respect to the Alexandrov topologies on O and $\tilde{O}$. So $f^{-1}[\tilde{U} \cap \tilde{O}]$ must be open in the Alexandrov topology on O, and hence it is also open in the manifold topology. □

Lemma A.4. *For any $p \in X$ there exists an open set O containing p such that $I^{\pm}[p, O] \subseteq X - \mathcal{D}$.*

Proof. Let $\tilde{O}$ be an open convex set containing $f(p)$. Assume that for every open set containing p it happens that $f\left[I^{+}[p, O]\right] \nsubseteq \tilde{O}$. Then for any open set O_1 containing p there must be a point $p_1 \in O_1$ such that $p_1 \in I^{+}[p, O_1]$ but $f(p_1) \notin \tilde{O}$. Take an open set $O_2 \subseteq I^{-}[p, O_1]$ containing p [31]. There must exist $p_2 \in O_2$ with $p_2 \in I^{+}[p, O_2] \subseteq I^{+}[p, O_1]$ but $f(p_2) \notin \tilde{O}$. This way we can generate a sequence of nested open sets $\{O_i\}$ all containing p, and a sequence of points $\{p_i\}$ such that $p_i \in O_i$, $p_{i+1} \ll^{\mathfrak{C}}_{O_i} p_i$ and $p \ll^{\mathfrak{C}}_{O_i} p_i$, but $f(p_i) \notin \tilde{O}$. If we choose this sequence of open sets such that $\bigcap O_i = p$, we can then join each p_{i+1} with p_i by means of a continuous future directed timelike curve γ_i entirely contained in O_i. Hence, the $\bigcup \gamma_i$ will constitute a continuous future directed timelike curve from p to p_1. Then, by construction, no initial segment of $f \circ \gamma$ can intersect $\tilde{O}$, which is a contradiction to the fact that $f \circ \gamma$ must be a continuous timelike curve through $f(p) \in \tilde{O}$.

Therefore, there must exist some open set O_a containing p such that $f\left[I^{+}[p, O_a]\right] \subseteq \tilde{O}$. In an analogous way we can find an open set $O_b \ni p$ such that $f\left[I^{-}[p, O_b]\right] \subseteq \tilde{O}$. Take $O = O_a \cap O_b$. Then $f\left[I^{\pm}[p, O]\right] \subset \tilde{O}$, and thus lemma A.3 implies that $I^{\pm}[p, O] \subseteq X - \mathcal{D}$. This means that f is continuous over local futures and pasts. □

Lemma A.5. *Both f and f^{-1} preserve continuous causal curves.*

Proof. Let $\gamma : I \longrightarrow X$ be a continuous future directed causal curve with $\gamma(t_0) = p$ for some $t_0 \in I$. Let $\tilde{O} \ni f(p)$ be an open convex set. As shown in lemma A.4, we can find an open set $O \ni p$ such that $f\left[I^{\pm}[p, O]\right] \subseteq \tilde{O}$. Such an open set can be chosen convex. Take a subinterval $\bar{I}$ of I containing t_0 so that

$$\begin{aligned} t \in \bar{I} \quad \text{and} \quad t < t_0 &\implies \gamma(t) \prec p \\ t \in \bar{I} \quad \text{and} \quad t_0 < t &\implies p \prec \gamma(t). \end{aligned} \tag{A.1}$$

If it happens that $\gamma(t) \prec p$, the properties in 3.1 imply that every timelike curve through $\gamma(t)$ must have a nonempty intersection with $I^{-}[p, O]$. Hence every continuous timelike curve through $(f \circ \gamma)(t)$ must intersect $I^{-}[f(p), \tilde{O}]$. Thus, $(f \circ \gamma)(t) \in Cl[I^{-}[f(p), \tilde{O}]] = J^{-}[f(p), \tilde{O}]$, where the last equality follows from the assumed convexity of $\tilde{O}$. This means that:

$$t \in \bar{I} \quad \text{and} \quad t < t_0 \implies (f \circ \gamma)(t) \prec_{\tilde{O}} f(p). \tag{A.2}$$

[31] This is always possible because the sets $I^{\pm}[p, O]$ are always open.

We can make the same argument by using the chronological futures instead, and hence get

$$t \in \bar{I} \quad \text{and} \quad t_0 < t \implies f(p) \prec_{\tilde{O}} (f \circ \gamma)(t) \tag{A.3}$$

. This two equations mean, by our definition 3.16, that $f \circ \gamma$ is a future directed causal curve. The symmetric argument follows for f^{-1}. □

Lemma A.6. *i) $\mathcal{D}$ and $\tilde{\mathcal{D}}$ are closed in the respective manifold topologies.*

ii) $p \in \mathcal{D} \iff f(p) \in \tilde{\mathcal{D}}$.

iii) If $p \in \mathcal{D}$, there exists a continuous inextendible future directed causal curve through p entirely contained in $\mathcal{D}$

Proof. (i). Assume f is continuous at p, and let $\tilde{O} \ni f(p)$ be an open convex set. By continuity of f we can find an open set $O \in X$ such that $p \in O \subseteq f^{-1}[\tilde{O}]$. Then, by lemma A.3, $O \subseteq X - \mathcal{D}$. Thus, $X - \mathcal{D}$ is an open set. Similarly for $\tilde{X} - \tilde{\mathcal{D}}$.

(ii) and (iii). Take $p \in \mathcal{D}$. Then as in the proof of lemma A.4, we can find a sequence $\{p_i\}$ converging to p, and an open convex $\tilde{O} \subseteq \tilde{X}$ that contains $f(p)$ but does not contain any of the $f(p_i)$. Take two sequences $\{r_i\}$ and $\{s_i\}$ which converge to p from the chronological past and future respectively such that for each I there is (locally) a future directed continuous timelike curve γ_i through p_i with r_i and s_i as respective past and future endpoints. p is the only accumulation point for the curves γ_i.

Since f preserves timelike curves, $\{f(r_i)\}$ and $\{f(s_i)\}$ must converge to $f(p)$. Furthermore, we can take the sequences $\{r_i\}$ and $\{s_i\}$ in such manner that all the $f \circ \gamma_i$ begin and end in $\tilde{O}$. But each $f \circ \gamma_i$ must leave $\tilde{O}$ at some point because $f(p_i) \notin \tilde{O}$. By 5 there must exists a future directed inextensible continuous causal curve Δ through $f(p)$ for which every point of it is an accumulation point of the $f \circ \gamma_i$. It follows that $\Delta - \{f(p)\} \subseteq \tilde{\mathcal{D}}$ because the only accumulation point of the γ_i is p. But we have shown that $\tilde{\mathcal{D}}$ is a closed set, thus $\Delta \subseteq \tilde{\mathcal{D}}$. Finally we have $p \in \mathcal{D} \implies f(p) \in \tilde{\mathcal{D}}$. The converse implication follows with a symmetric argument, using $f(p)$ and f^{-1}. In the proof of this converse implication, (iii) follows with 5. □

Lemma A.7. *If $\mathcal{D} \neq \emptyset$, there exists an open convex set O with $\mathcal{D} \cap O \neq \emptyset$ such that:*

i) $\mathcal{D}$ is achronal in O.

ii) A unique null geodesic Γ_p with $\Gamma_p \cap O \subseteq \mathcal{D}$ passes through each point $p \in \mathcal{D} \cap O$.

iii) Given any continuous null geodesic Γ intersecting $\mathcal{D} \cap O$, either $\Gamma \cap O \subseteq \mathcal{D}$ or $\Gamma \cap O \cap \mathcal{D}$ contains exactly one point.

Proof. (i). Assume the opposite, i.e. $\mathcal{D} \neq \emptyset$ and there is no open convex set O intersecting $\mathcal{D}$ in which $\mathcal{D}$ is achronal. Let O_1 be an open set with $O_1 \cap \mathcal{D} \neq \emptyset$. Since, by our assumption D is not achronal in O_1 we can find two points $r_1, s_1 \in O_1 \cap \mathcal{D}$ which are chronologically related, say $r_1 \ll_{O_1} s_1$. Take an open convex set O_2 such that $r_1 \in O_2 \subseteq I^-[s_1, O_1]$. Repeating the previous argument for O_2, we can find $r_2, s_2 \in O_2 \cap \mathcal{D}$ that are chronologically related, $r_2 \ll_{O_2} s_2$. Now he have $s_2 \ll_{O_1} s_1$. In this way we can generate a sequence $\{s_i\}$ in $O_1 \cap \mathcal{D}$ with $s_{i+1} \ll_{O_1} s_i$ for all i. Denote by s the accumulation point of such a sequence. By lemma A.4 we can find an open set $O \ni s$ such that $I^+[s, O] \subseteq M - \mathcal{D}$. But this is a contradicts the fact that $O_1 \cap \mathcal{D} \neq \emptyset$.

(ii) and (iii). By finding an achronal set, these propositions follow from lemma A.6. □

We are now in position to prove A.1:

Proof. Suppose f is not continuous, i.e. $\mathcal{D} \neq \emptyset$. By lemma A.7 we can find an open convex set O intersecting $\mathcal{D}$ such that $\mathcal{D}$ is achronal in O. Hence the statements (ii) and (iii) in such lemma will hold as well. Let $p \in \mathcal{D} \cap O$ and denote by Γ_p the unique continuous null geodesic that passes through p such that $\Gamma_p \cap O \subseteq \mathcal{D}$. Keep in mind that by lemma A.4 $I^{\pm}[\Gamma_p \cap O, O] \subseteq X - \mathcal{D}$. Since f is discontinuous at p there must exist a sequence $\{p_i\}$ converging to p, and an open convex set $\tilde{O} \in \tilde{X}$ containing $f(p)$ but such that $f(p_i) \notin \tilde{O}$ for all i.

Let Ω be a sufficiently short[32] segment of *any* future directed continuous null geodesic passing through p, distinct from Γ_p. We can find a sequence of continuous null geodesic segments Ω_i within O such that Ω_i passes through p_i and every open set intersecting Ω also intersects all the Ω_i. Choose the sequence $\{\Omega_i\}$ such that it has no convergence points off Ω. Eventually all the Ω_i intersect $I^{\pm}[\Gamma_p \cap O, O]$, and consequently all of them intersect $X - \mathcal{D}$. Part (iii) in lemma A.7 then implies that $\Omega_i \cap \mathcal{D}$ is either empty or a single point. We will assume that the intersection point $\Omega_i \cap \mathcal{D}$ (if it exists) is causally preceded by p_i for all i. Denote by Ω_i^- the segment of Ω_i that is to the past of p_i (with p_i included). We can then find a subsequence of continuous null geodesic segments $\{\Omega_i^-\}$ in O such that the following properties hold:

1. $\{\Omega_i^-\}$ converges to the segment of Ω that is to the past of p, Ω^-, and the sequence has no convergence points off Ω^-.

2. For each i, $\Omega_i^- \cap \mathcal{D}$ is either p_i or empty.

By item 2 and lemma A.5, it follows that each $f \circ \Omega_i^-$ is a continuous null geodesic segment in $\tilde{X}$. From $\Omega^- - \{p\} \subseteq X - \mathcal{D}$ and item 1, we conclude that $\{f \circ \Omega_i^-\}$ converges to $f \circ \Omega^-$. However, each $f \circ \Omega_i^-$ must leave $\tilde{O}$ at some point before $f(p_i)$ because $f(p_i) \notin \tilde{O}$ for all

[32] i.e. such that $f \circ \Omega \subset \tilde{O}$.

i. Let $f(q)$ be any point of the null geodesic extension $f \circ \Omega^-$. Now, $f(q)$ can't belong to $\tilde{X} - \tilde{\mathcal{D}}$. If it did, then, since $f(q)$ is a convergence point of $\{f \circ \Omega_i^-\}$, q must be a convergence point of $\{\Omega_i^-$. But $q \notin \Omega^-$, so this is not possible. Hence $f(q) \in \tilde{\mathcal{D}}$.

If in our construction we drop the requirement that the Ω_i intersect $\mathcal{D}$ to the future of p_i, we then get the following conclusion. If we denote by Ω^- and Ω^+ the corresponding past and future segments of Ω with respect to p, then either the future null geodesic extension of $f \circ \Omega^-$ or the past extension of $f \circ \Omega^+$ is a future directed continuous causal curve segment through $f(p)$ contained in $\tilde{\mathcal{D}}$. This holds for all future directed continuous null geodesic segments since Ω was chosen arbitrarily. Thus it follows that there exist distinct[33] future directed continuous causal curves through $f(p)$ contained in $\tilde{\mathcal{D}}$. Hence their preimages must be distinct future directed continuous causal curves through p contained in $\mathcal{D}$. This contradicts the initial assumption that $\mathcal{D}$ is achronal in O. Therefore, $\mathcal{D} = \emptyset$. By item (ii) in lemma A.6 $\tilde{\mathcal{D}} = \emptyset$ as well. □

Finally, we will prove A.2. As opposed to A.1, this lemma has a short proof. We only need to see if it is possible to express the null geodesics in terms of the continuous timelike curves.

Proof. First note the following: Let $p, q \in U$ with U open. We have that $q \in \partial I^+[p, U]$ if and only if for all future directed continuous timelike curves $\sigma : (0,1) \longrightarrow U$, if $\sigma(t_0) = q$ for some $t_0 \in (0,1)$, then there exist t_1, t_2 where $0 < t_1 < t_0 < t_2 < 1$, such that $\sigma(t_1) \notin I^+[p, U]$, but $\sigma(t_2) \in I^+[p, U]$.

Now, if $\gamma : I \longrightarrow M$ is a continuous curve, then γ is a future directed null geodesic if and only if for all $t_0 \in I$ and all open sets O containing $\gamma(t_0)$, there exist an open set $U \subseteq O$ containing $\gamma(t_0)$ such that for all $t_1, t_2 \in I$ with $t_1 < t_2$, if $\gamma(t_1), \gamma(t_2) \in U$ then $\gamma(t_2) \in \partial I^+[\gamma(t_1), U]$. □

Another important point covered in Malament's work is the fact that *only* in the case of Distinguishing space-times a causal isomorphism is in fact a conformal diffeomorphism. That is, if we lower the condition of "distinguishing" one step to only future/past distinguishing, then a causal isomorphism is not necessarily a homeomorphism of the manifold topology.

[33]Because f is bijective.

References

[1] Casarrubias Segura, F. and Tamariz Mascarúa, Á. (2012) *Elementos de topología general.* Sociedad Matemática Mexicana, México.

[2] Hirsch, M. W. (1997) *Differential topology.* (Graduate text in mathematics; 33), Springer-Verlag.

[3] Penrose, R. (1970) *Techniques of Differential Topology in Relativity.* Society for Industrial and Applied Mathematics, Philadelphia.

[4] Hawking, S. w. and Ellis, G. F. R. (1973) *The Large Scale Structure of Space-time.* Cambridge University Press.

[5] Misner, C. W.; Thorne, K. S.; and Wheeler, J. A. (1973) *Gravitation.* San Francisco, CA: W. H. Freeman.

[6] Minguzzi, E. and Sánchez, M. (2008) The causal hierarchy of spacetimes. Available from: arXiv:gr-qc/0609119

[7] Malament, D. B. (1977) The class of continuous timelike curves determines the topology of space-time. *Journal of Mathematical Physics*, 18 (7), 1399-1404.

[8] Beem, J.K., Ehrlich, P. E. and Easley, K.L (1996) *Global Lorentzian Geometry*, 2nd edition, New York, Marcel Dekker Inc.

[9] Hawking, S. W., King, A. R. and McCarthy, P. J. (1976) A new topology for curved space-time which incorporates the causal, differential, and conformal structures. *Journal of Mathematical Physics*,

[10] Hawking, S. W. (unpublished) Singularities and the geometry of space-time. Adam's prize essay.

[11] Hawking, S. W. (1992) Chronology protection conjecture. *Phys. Rev. D*, 46 (2), 603-611.

[12] Lloyd, S. et. al. (2011) The quantum mechanics of time travel through post-selected teleportation *Phys. Rev. D*, 84, 025007

[13] Kroenheimer, E. H and Penrose, R. (1966) On the structure of causal spaces. *Mathematical Proceedings of the Cambridge Philosophical Society*, 63, pp. 481-501.

[14] Dribus, B. F. (2013) On the axioms of causal set theory. Available from: arXiv:1311.2148 [gr-qc]

[15] P. S. Joshi. (2007) *Gravitational Collapse and Spacetime Singularities.* Cambridge Monographs on Mathematical Physics, New York.

17, No. 2, 174-181.

[16] t'Hooft, G. (1984) On the quantum structure of a black hole. *Nuclear Physics*, B256, 727-745.

[17] Henson, J. (2006) The causal set approach to quantum gravity. Available from: arXiv:gr-qc/0601121v2

[18] Sorkin, R. D. (2002) Causal Sets: Discrete Gravity. Available from: arXiv:gr-qc/0309009

[19] Carter, B. (1971) Causal Structure in Space-Time. *Journal of General Relativity and Gravitation*, 1 (4), 349-391.

[20] Parrikar, O. and Surya, S. (2011) Causal Topology in Future and Past Distinguishing Spacetimes. Available from: arXiv:1102.0936 [gr-qc]

[21] Benincasa, D. M. T. and Dowker, F. (2011) The Scalar Curvature of a Causal Set. Available from: arXiv:1001.2725 [gr-qc]

[22] Sorkin, R.D. (1990) First steps with causal sets, Capri 1990: *Proceedings of the Ninth Italian Conference on General Relativity and Gravitational Physics.* Singapore: World Scientific, pp. 68-90.

[23] Levichev, A. V. (1987) Methods of investigation of the causal structure of Lorentz manifolds. Translated from *Sibirskii Matematicheskii Zhurnal*, 31 (3), 39-54.

[24] Bekenstein, J. D. (1975) Statistical black-hole thermodynamics. *Phys. Rev. D*, 12 (10), 3077-3085.

[25] Bombelli, L., Koul, R. K., Lee, J. and Sorkin, R. D. Quantum source of entropy for blackholes. *Phys. Rev. D*, 34 (2), 373-383.

[26] Sorkin, R. D. (1993) Forks in the road, on the way to quantum gravity. Available from: arXiv:gr-qc/9706002.

[27] Finkelstein, D. (1969) Space-time Code. *Phys. Rev.* 184 (5),1261-1271

[28] Aharony, O., Gusber, S. S., Maldacena, J. Ooguri, H. and Oz, Y. (1999) Large N Field Theories, String Theory and Gravity. Available from: arXiv:hep-th/9905111

[29] Doná, P. and Speziale, S. (2010) Introductory lectures to loop quantum gravity. Available from: arxiv.org/abs/1007.0402v2

[30] Gaul, M. and Rovelli, C. (1999) Loop quantum gravity and the meaning of diffeomorphism Invariance. *Lect. Notes Phys.* 541 (2000) 277-324.

[31] Regge, T. (1961) General relativity without coordinates, *Nuovo Cim.* 19, 558.

[32] Frhlich, J. (1992) Regge Calculus and Discretized Gravitational Functional Integrals. *Non-Perturbative Quantum Field Theory: Mathematical Aspects and Applications, Selected Papers*. Singapore: World Scientific, pp. 523-545.

[33] Pérez, A. (2012) The Spin Foam Approach to Quantum Gravity. Available from: arXiv:1205.2019

[34] Ambjørn, J.; Görclich, A.; Jurkiewicz, J. and Loll R (2013) Quantum Gravity via Causal Dynamical Triangulations. Available from: arXiv:1302.2173

[35] Weinberg, S. (1979) Ultraviolet divergences in quantum theories of gravitation, in *General Relativity: Einstein century survey*, eds. Hawking, S. W. and Israel, W. Cambridge University Press, Cambridge, UK, 790-831.

[36] Bardeen, J. M.; Carter, B. and Hawking, W. (1973) The four laws of black hole mechanics. *Commun. math Phys*, 31, 161-170.

[37] Trudeau, R. J. (2003) *Introduction to Graph Theory*. Dover Books on Mathematics.

[38] Cohn, D. L. (1980) *Measure Theory*, USA, Boston. Birkhäuser.

[39] Rideout, D. P. and Sorkin, R. D. (1999) A Classical Sequential Growth Dynamics for Causal Sets. *Phys.Rev.D*61:024002,2000.

[40] Feynmann, R.; Leighton, R. B. and Sands, M. (1962) *The Feynmann Lectures on Physics*, Vol. 1. United States, Addison-Wesley

[41] Lloyd, S. and Preskill, J. (2014) Unitarity of black hole evaporation in final-state projection models. *Journal of High Energy Physics* Available online at: http://link.springer.com/article/10.1007%2FJHEP08%282014%29126

[42] Penrose, R. (1999) The question of cosmic censorship. *J. Astrophys. Astr.*, 20, 233-248.

[43] Zeeman, E. C. (1964) Causality implies the Lorentz group. *J. Math. Phys.*, 5, 490-493.