Aspects of Nonlocality: from Particles to Black Holes

by

Mehdi Saravani

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Physics

Waterloo, Ontario, Canada, 2016

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

This dissertation is partially the product of collaborative research and co-authored publications as follows:

- **Chapter 2** Mehdi Saravani and Siavash Aslanbeigi, "On the Causal Set-Continuum Correspondence", Class. Quant. Grav., 31(20):205013, 2014
- **Chapter 3** Siavash Aslanbeigi, Mehdi Saravani and Rafael D. Sorkin, "Generalized causal set d'Alembertians", Journal of High Energy Physics, 6:24, June 2014
- **Chapter 4** Mehdi Saravani and Siavash Aslanbeigi, "Dark Matter From Spacetime Non-locality", Phys. Rev., D92(10):103504, 2015
- **Chapter 5** Mehdi Saravani and Niayesh Afshordi, "Off-shell Dark Matter: *A Cosmological relic of Quantum Gravity*", arXiv:1604.02448
- **Chapter 6** Mehdi Saravani, Niayesh Afshordi and Robert Mann, "Dynamical Emergence of Universal Horizons during the formation of Black Holes", Phys. Rev., D89:084029, 2014
- **Chapter 7** Michael Meiers, Mehdi Saravani and Niayesh Afshordi, "Cosmic Censorship in Lorentz Violating Theories of Gravity", arXiv:1511.08969, 2015, Accepted by Phys. Rev. D

Abstract

This dissertation is a collection of works on different aspects of the following subjects: causal set–continuum correspondence, wave propagation on causal sets, nonlocal quantum field theory, potential connection between dark matter and quantum gravity and causal structure of black holes in Lorentz violating gravitational theories.

We present two results which concern certain aspects of the question: when is a causal set well approximated by a Lorentzian manifold? The first result is a theorem which shows that the number-volume correspondence, if required to hold even for arbitrarily small regions, is best realized via Poisson sprinkling. The second result concerns a family of lattices in 1+1 dimensional Minkowski space, known as Lorentzian lattices, which we show provide a much better number-volume correspondence than Poisson sprinkling for large volumes. We argue, however, that this feature should not persist in higher dimensions and conclude by conjecturing a form of the aforementioned theorem that holds under weaker assumptions, namely that Poisson sprinkling provides the best number-volume correspondence in 3+1 dimensions for spacetime regions with macroscopically large volumes.

We then study wave propagation on a background causal set. We introduce a family of generalized d'Alembertian operators in D-dimensional Minkowski $\mathbb{M}^D$ spacetimes which are manifestly Lorentz-invariant, retarded, and non-local. The prototypes of these operators arose in earlier work as averages of matrix operators meant to describe the propagation of a scalar field in a causal set. We generalize the original definitions to produce an infinite family of "Generalized Causet Box (GCB) operators" parametrized by certain coefficients $\{a, b_n\}$, and we derive the conditions on the latter needed for the usual d'Alembertian, $\Box$, to be recovered in the infrared limit. The continuum average of a GCB operator is an integral operator in $\mathbb{M}^D$, and it is these continuum operators that we mainly study. To that end, we compute their action on plane waves, or equivalently their Fourier transforms $g(p)$ [p being the momentum-vector]. For timelike p, $g(p)$ has an imaginary part whose sign depends on whether p is past or future-directed. For small p, $g(p)$ is necessarily proportional to $p \cdot p$, but for large p it becomes constant, raising the possibility of a genuinely Lorentzian perturbative regulator for quantum field theory in $\mathbb{M}^D$. We also address the question of whether or not the evolution defined by the GCB operators is stable, finding evidence that the original 4D causal set d'Alembertian is unstable, while its 2D counterpart is stable.

Following our earlier work on wave propagation on a causal set, we study the quantum theory of a massless scalar field whose evolution is given not by the the d'Alembertian, but by an operator $\widetilde{\Box}$ which is Lorentz invariant, reduces to $\Box$ at low energies, and defines an explicitly retarded evolution: $\widetilde{\Box}\phi(x)$ only depends on $\phi(y)$, with y is in the causal past

of x. This modification results in the existence of a continuum of massive particles, in addition to the usual massless ones, in the free theory. When interactions are introduced, these massive or off-shell quanta can be produced by the scattering of massless particles, but once produced, they no longer interact, which makes them a natural candidate for dark matter which we dub off-shell dark matter (OfDM). Finally, we generalize this idea to massive scalar fields too.

We then consider phenomenological predictions of OfDM model in the context of cosmology. OfDM particles rate of production is suppressed by the scale of nonlocality (e.g. Planck length). As a result, we show that OfDM is only produced in the first moments of big bang, and then effectively decouples (except through its gravitational interactions). We examine the observational predictions of this model: In the context of cosmic inflation, we show that this proposal relates the reheating temperature to the inflaton mass, which narrows down the uncertainty in the number of e-foldings of specific inflationary scenarios. We also demonstrate that OfDM is indeed cold, and discuss potentially observable signatures on small scale matter power spectrum.

Finally, we explore the validity of Cosmic Censorship conjecture in Lorentz violating theories of gravity. Is Cosmic Censorship special to General Relativity, or can it survive a violation of local Lorentz invariance? Studies have shown that singularities in Lorentz violating Einstein-Aether (or Hořava-Lifshitz) theories can lie behind a universal horizon in simple black hole spacetimes. Even infinitely fast signals cannot escape these universal horizons. We review this result and extend it, for an incompressible aether, to 3+1D dynamical or spinning spacetimes which possess inner Killing horizons, and show that a universal horizon always forms in between the outer and (would-be) inner horizons. This finding suggests a notion of Cosmic Censorship, given that geometry in these theories never evolves beyond the universal horizon (avoiding potentially singular inner Killing horizons). A surprising result is that there are 3 distinct possible stationary universal horizons for a spinning black hole, only one of which matches the dynamical spherical solution. This motivates dynamical studies of collapse in Einstein-Aether theories beyond spherical symmetry, which may reveal instabilities around the spherical solution.

Acknowledgements

The past five years have been one of the most memorable stages in my life. I owe this, first and foremost, to my advisor Niayesh Afshordi. His sharp intuition and broad knowledge of physics have taught me how to grow as a researcher. I would like to thank him for his constant support, endless patience and for his trust, to let me pursue research subjects that I found interesting.

I consider myself lucky that I have learned, worked and collaborated with Rafael Sorkin. His deep understanding of physics, his ability to explain the most complex subjects in a simple form and his careful way of reasoning have impressed me greatly. I would like to thank him for all he had taught me.

I would also like to thank my good friend and collaborator Siavash Aslanbeigi. Whenever we started working on a subject together, he always impressed me with his persistence, scrutiny and step by step study of the subject. His approach in doing research often made us overcome obstacles we faced in our work together.

I want to express my gratitude to Achim Kempf, for helping me in different stages of my PhD. I had the chance of learning from him and collaborating with him. I would also like to acknowledge my other collaborators Robert B. Mann, Yasaman K. Yazdi, Kurt Hinterbichler and Michael Meiers.

I am thankful to Heidar Moradi and Seyed Faroogh Moosavian to give me feedback on the first draft of this thesis. During my PhD I have discussed various matters of physics and mathematics with Latham Boyle, Fay Dowker, Ghazal Geshnizjani, Mansour Karami, Nosiphiwo Zwane, Chiamaka Okoli Nwosu, Elizabeth Gould, Trevor Rempel, Tim Koslowski, Dionigi Benincasa, Alessio Belenchia, Stefano Liberati, Vasudev Shyam, Jacob Barnett, Razieh Pourhasan, John Moffat, Peter Lunts, Joao Caetano, Andres Felipe Schlief Carvajal, Natacha Altamirano, Dalimil Mazac, Anton Van Niekerk, Daniel Carrasco Guariento and Mohammad Shalaby.

I am thankful to my good friend Farbod Kamiab, together we spent countless hours discussing matters of physics and life.

I am especially indebted to Perimeter Institute for Theoretical Physics for hosting me in the past five years and providing an outstanding research environment.

Finally, I would like to thank my dearest Pegah whose continuous support has been incredibly important to me.

To my parents:
Malihe and Reza

and my lovely Pegili ♥

Table of Contents

List of Figures

Chapter 1

Introduction

The past five years have been very fruitful for fundamental physics. Due to the measurement of Cosmic Microwave Background (CMB) anisotropies by Planck satellite, we now have a better understanding of the initial state and cosmological history of our universe [1, 2, 3]. The Planck data was a great improvement over the previous measurement of CMB done by Wilkinson Microwave Anisotropy Probe [4]. It is quite remarkable that with this measurement we can partially understand the state of our cosmos at a fraction of a second after the big bang and at the energy scales of just a few orders below the quantum gravity regime.

On the particle physics side, the Higgs boson, the last unverified ingredient of the Standard Model of particle physics, was detected at Large Hadron Collider (LHC) with $m_H = 126.0 \pm 0.4(stat) \pm 0.4(sys)$ GeV. This observation confirmed the 50 years old prediction of Higgs boson [5], the first ever scalar field detected in nature. It is interesting that the observed mass of the Higgs boson is in the range that makes the Standard Model vacuum meta-stable which opens up theoretical speculations for the origin of this near-criticality [6].

Recently, gravitational waves were detected for the first time by LIGO [7] which their existence was theorized 100 years ago by Albert Einstein [8]. With this detection, a new era of astrophysical observations is opened that could immensely impact our progress in fundamental physics.

These exciting discoveries are all compatible with the well-established theories of cosmology, particle physics and gravity. However, it is quite unfortunate that they provide us with no clue on how to proceed with resolving some longstanding problems/puzzles in fundamental physics: dark matter, dark energy and quantizing gravity. The Planck data is

well fitted by the so called ΛCDM cosmological model, and no deviation from a late time cosmological constant is detected. While the data rules out many inflationary models, it is still consistent with many others. Moreover, no tensor mode was detected which could have greatly narrowed down the viable theories of inflationary epoch.

While the Standard Model of particle physics was confirmed at LHC, no sign of any extension to the Standard Model, any supersymmetric particle or dark matter candidate is detected so far. A deviation would have meant additional particles in nature that could have pointed out new structures or symmetries in particle physics, but unfortunately the data is well-described by the Standard Model of particle physics.

The area of gravitational wave observation is also still too young to discover any deviation from General Relativity (GR). Hopefully, with increasing number of gravitational wave detections in future we can put stringent constraints on deviations from GR or possibly detect one.

With the current status of our observations, how can we make progress and find solutions to the unresolved problems in fundamental physics? I believe there are two strategies to move forward. One is to follow up on the well-known fundamental physical theories, like local Quantum Field Theory (LQFT) and GR, and build new models within this context. Another approach is to work on alternative theories with different fundamental assumption(s) and look for new phenomenology, hoping it could result in a viable solution to these problems. This thesis is a collection of my efforts following the latter approach.

1.1 Locality and Lorentz Invariance

Quantum field theory and General Relativity, the two main pillars of modern physics, are established on two fundamental assumptions: Lorentz invariance and locality. From observational point of view, Lorentz invariance has been tested in many experiments and to a very high accuracy [9, 10] while the situation with locality is different. There is no direct experiment testing locality; it is verified indirectly by the confirmation of local theories.

In this thesis, we explore the consequences of abandoning each one of these assumptions. In summary, this thesis is an attempt to answer the following questions: *Why locality? And why Lorentz invariance?* From the phenomenological point of view, it is worth mentioning that in many approaches to quantum gravity one of these assumptions is relaxed and it would be necessary to study the phenomenological consequences of this modification at low energies to confirm/constrain each theory.

1.1.1 Lorentzian Nonlocality and Causal Set

Nonlocality can arise in many different contexts. One could think that it is impossible to probe scales smaller than the Planck length, since it requires Planckian energies which would end up making (Planckian) black holes. This in effect implies the existence of a Planck size nonlocality coming from gravitational effects. However, nonlocality does not have to be restricted to the Planckian regime. Almost any effective field theory shows a degree of nonlocality which could be much larger than the Planck length (e.g. [11]). Moreover, nonlocality could be not just an effective phenomena, but a direct consequence of the fundamental structure of spacetime, as it appears in Causal Set approach [12] to quantum gravity which I explain in what follows.

However, if Lorentz symmetry remains to be a fundamental symmetry of nature (as current experiments suggest), the nature of non-locality should be Lorentzian, i.e. non-locality in spacetime. As an example, let us consider Causal Set theory that postulates the fundamental structure of spacetime is discrete and Lorentzian. This implies physics cannot remain local beyond some scale (at least the discreteness scale).

But how could a discrete structure be Lorentzian? Obviously, this cannot be done if we require invariance under the full Lorentz group exactly. So, the best we can do is to introduce an approximate notion of Lorentz symmetry, for example invariance under a discrete subgroup of Lorentz transformations. Due to the stochastic nature of causal sets, the approximate Lorentz symmetry is defined statistically in the Causal Set approach. Then the following question arises which is central to the Causal Set program: which discrete structure keeps Lorentz invariance at best? Chapter 2 addresses different aspects of this question.

The Lorentzian nonlocality manifests itself concretely when one seeks to describe the wave propagation of a scalar field on a causal set by defining a discrete counterpart of the d'Alembertian operator. An infinite family of these operators, controlled by a nonlocality scale parameter, is defined in Chapter 3. Moreover, their continuum limit defines nonlocal (Lorentzian) evolution for scalar fields on a continuum spacetime.

Having a nonlocal Lorentzian evolution for scalar fields, a natural question then arises: what are the physical consequences of modifying local evolution of fields to a nonlocal one described by the above-mentioned operators? Chapter 4 concerns quantization of the nonlocal field theory and shows how this modification results in the appearance of a new set of excitations of the field. Once interaction is introduced, the cross-section of any scattering process which contains one or more of the new excitations in the "in" state is zero; a behaviour that makes the new modes a natural candidate for dark matter. Chapter 5 discusses phenomenological predictions of this new dark matter candidate.

As we discussed above, abandoning the assumption of locality which is inspired by quantum gravity approaches like Causal Set could result in interesting phenomenology and potentially a solution to the dark matter problem.

1.1.2 Lorentz Violation

While Lorentz symmetry is well tested in the matter sector, its tests in the gravitational sector are much weaker and many Lorentz violating theories of gravity are still consistent with observations. From theoretical point of view, main reason for studying alternative gravitational theories stems from quantizing gravity. GR, while being a very successful *classical* theory, has failed to cope with quantum mechanics. Therefore, one approach to quantum gravity has been to abandon diffeomorphism invariance in order to make the gravitational theory renormalizable (e.g. [13]).

Theories with broken Lorentz invariance exhibit a different causal structure compare to GR. Since the notion of black holes are tightly connected to the causal structure of each theory, black holes are primary objects to look for deviations from GR. In fact, in many examples of theories with broken Lorentz invariance, superluminal degrees of freedom appear (see [14, 15]) which potentially can escape the traditional Killing horizon of a black hole. A natural question then arises which is addressed in Chapters 6 and 7: can black holes (objects with a singularity protected by an event horizon) exist in these theories?

In the next Section, we will provide an overview of each chapter and how they are connected in a bigger scheme. The material in each chapter is also presented as self-contained as possible.

1.2 Outline

Here, we present a summary of each chapter of the thesis, its motivations and results. We first start this thesis by the following question in Chapter 2: when is a causal set well-approximated by a Lorentzian manifold?

From the viewpoint of Causal Set theory[1], the continuum spacetime of general relativity is only fundamental to the extent that it provides a good approximation to an underlying causal set. Therefore, criteria must be established to determine how well a Lorentzian geometry (M, g) approximates a causal set $(C, \prec)$. One natural criterion is to require the

[1]For a brief review on Causal Set theory see Chapters 2.2 and 2.3 of [16].

existence of an injective map $f : C \to M$ which preserves causal relations: $\forall\ x, y \in C$, $x \prec y$ if and only if $f(x) \in J^-(f(y))$, where $J^-(f(y))$ is the set of all points in M which lie in the causal past of $f(y)$.

As far as the causal set-continuum correspondence is concerned, preserving causal relations is not enough, because causal relations are left invariant under conformal transformations. To break this degeneracy, information about scale is needed. Causal sets contain information about scale implicitly through counting of elements, since they are locally finite (i.e. discrete). In fact, discreteness allows one to count elements, which is thought to provide information about scale: a spacetime region with volume V should contain about ρV causal set elements, where ρ is a constant, thought to be set by the Planck scale, which represents the number density of points.

Then, one has to require a number–volume (N-V) correspondence: the number N_S of embedded points in a spacetime region $S \subset M$ should "reflect" its volume V_S:

$$N_S \approx \rho V_S = \rho \int_S \sqrt{-g(x)} d^D x. \tag{1.1}$$

Having the N-V formulation at hand, the key question becomes: what is the map that realizes the number–volume correspondence with the least noise?

The attitude in the causal set program is that this mapping is best done through *Poisson sprinkling*. In this approach, one first reverses direction by obtaining a causal set $C(M)$ from a given spacetime (M, g): randomly select points from M using the Poisson process at density ρ and endow the selected points with their causal relations.

But does Poisson sprinkling provide the best number-volume correspondence? We show two results in Chapter 2 in regard to this question. The first result is that the number–volume correspondence is best realized via Poisson sprinkling if required for any (even arbitrarily small) volume. Quite surprisingly, we also show that $1 + 1$-dimensional Lorentzian lattices provide a much better number–volume correspondence than Poisson sprinkling for large volumes. We present evidence, however, that this feature should not persist in $3 + 1$ dimensions and conjecture that the Poisson process should indeed provide the best number–volume correspondence for macroscopically large spacetime regions.

Having studied the causal set–continuum correspondence, we turn into dynamics of scalar fields on a background causal set in Chapter 3. If the fundamental structure of spacetime really is of the form of causal sets, physics cannot be expected to remain local at all scales. To appreciate why this should be, consider a covariant notion of closeness or locality in a causal set. The only available measure of closeness between two elements is some approximate notion of their Lorentzian distance. A small Lorentzian distance,

however, does not mean that the two elements are in a small neighbourhood of each other: any point may have many "close" neighbours near its past and future light cones. It is in this sense that the continuum notion of locality, as induced by the topology of spacetime as a manifold, is lost.

Dynamics of matter fields which propagate on the causal set would certainly be sensitive to this loss of locality. This feature manifests itself concretely when one seeks to describe the wave propagation of a scalar field on a causal set by defining a discrete counterpart of the d'Alembertian operator, $\Box$.

We shall denote a discrete causal set d'Alembertian in D dimensions by $B_\rho^{(D)}$, where ρ is a scale that controls the extent of the non-locality. For causal sets which are well-approximated by D-dimensional Minkowski space $\mathbb{M}^D$, averaging $B_\rho^{(D)}$ over all such causal sets leads to a manifestly non-local, *retarded* and Lorentzian continuum operator $\Box_\rho^{(D)}$ defined in $\mathbb{M}^D$. This operator is designed in a way that it reproduces the usual d'Alembertian in the limit of zero non-locality scale: $\Box_\rho^{(D)}\phi \to \Box\phi$ as $\rho \to \infty$.

We compute the spectrum of eigenvalues of this operator and show that for timelike momenta it contains also an imaginary part (unlike d'Alembertian operator), which changes sign under interchange of past with future. This property has a significant consequence when we discuss the quantization of the nonlocal field theories in Chapter 4. The UV behaviour of the spectrum also differs from that of $\Box$ in a way which suggests a genuinely Lorentzian, perturbative regulator for quantum field theory.

At the next step in Chapter 4, we proceed to quantize this nonlocal field theory. Let's consider modifying the evolution of a massless scalar field ϕ coupled to a source $J(x)$ via $\Box \to \widetilde{\Box}_\rho$[2]:

$$\Box\phi(x) = J(x) \to \widetilde{\Box}_\rho\phi(x) = J(x). \tag{1.2}$$

The first obstacle to quantizing this theory comes from the fact that Equation (1.2) cannot be derived from an action. One might propose to substitute $\Box$ with $\widetilde{\Box}_\rho$ in the action of a massless scalar field:

$$S[\phi] = \int d^4x \left(\frac{1}{2}\phi(x)\widetilde{\Box}_\rho\phi(x) - J(x)\phi(x)\right). \tag{1.3}$$

However, variation of $S[\phi]$ with respect to ϕ gives

$$\frac{1}{2}(\widetilde{\Box}_\rho + \widetilde{\Box}_\rho^T)\phi(x) = J(x), \tag{1.4}$$

[2] $\widetilde{\Box}_\rho$ is a generic retarded and nonlolcal operator which reduces to $\Box$ at low energies, i.e. $\rho \to \infty$. $\Box_\rho^{(D)}$ is an example in this class of operators that arises from the evolution on an underlying causal set.

where $\widetilde{\Box}_\rho^T$ is defined in Fourier space via complex conjugation of the spectrum of $\widetilde{\Box}_\rho$. Since eigenvalues of $\widetilde{\Box}_\rho$ are generally complex, Equation (1.4) is different from the desired Equation (1.2), which further means that we cannot use (1.3) in Feynman path integral to quantize this theory. The quantization scheme which we believe is the most suited in this case is the Schwinger-Keldysh (or double path integral) formalism, since it naturally incorporates a retarded operator.

Upon quantizing the free massless scalar field $\phi(x)$, we find *off-shell modes* in the mode expansion of the quantized field operator $\widehat{\phi}(x)$. These are modes which do not satisfy any dispersion relation, unlike in usual local quantum field theory where every Fourier mode in the field expansion with four-momentum p is an on-shell quanta, i.e. it satisfies $p \cdot p = 0$. We show that the off-shell modes can exist in "in" and "out" states of scatterings [3] and the cross-section of any scattering process which contains one or more off-shell particle(s) in the "in" state is zero. That is to say, *on-shell quanta can scatter and produce off-shell particles, but once produced, off-shell particles no longer interact*. This behaviour makes these off-shell particles a natural candidate for dark matter. The phenomenological story, which we discuss in Chapter 5, would be that dark matter particles were produced in the early universe as off-shell modes of quantum fields which we shall refer to as off-shell dark matter (OfDM). This feature of the theory can be traced back to the fact that $\widetilde{\Box}_\rho$ defines an explicitly retarded evolution. We demonstrate that the root of this behaviour comes from the fact that off-shell modes are a continuum set of excitations and they have an (infinitely) larger phase space compared to on-shells.

Chapter 5 concerns the phenomenological predictions of OfDM model in the context of cosmology. In this chapter, we address different aspects of the following questions: can OfDM model explain the observed energy density of cosmological dark matter for a reasonable value of parameters (e.g. nonlocality scale)? Does OfDM behave as a cold dark matter (CDM)? What are the observational signatures that could distinguish OfDM from other dark matter candidates, for example WIMPs?

It is quite interesting that all the observational evidences for CDM is through its gravitational interactions. This property has a natural explanation in our model. As we mentioned earlier, OfDM particles cannot be detected through scattering experiments. This means that if the majority of the observed cosmological dark matter is made of OfDM particles, we would not be able to detect the dark matter particles directly. However, they can be produced in scattering experiments and this is one way to indirectly confirm their existence by detecting missing energy in scatterings.

[3] These modes are different from virtual particles which exist as intermediate states in Feynman diagrams.

We discuss predictions of OfDM model in the context of cosmology and show that it is intertwined with the physics of inflation and reheating. In a very simple reheating scenario, this model effectively fixes the reheating temperature of the universe. By constraining the reheating temperature, we can narrow the predictions of spectral index and tensor-to-scalar ratio for a given inflationary potential, by fixing the number of e-foldings. Furthermore, since OfDM particles are not interacting with other particles (and themselves), they do not thermalize. This behaviour results into much shallower suppression of matter power spectrum on small scales compared to a Gaussian cutoff of thermal dark matter candidates. We show that this feature in principle could be another way to test this model via the observations probing matter power spectrum in sub-pc scales.

We end our journey on nonlocal Lorentzian field theory at this point and mention a few unanswered questions which could be the subject of future research:

1. In Chapter 5 we assumed that off-shell modes of a nonlocal field gravitate like ordinary (on-shell) matter. This assumption has to be verified.

2. We presented the quantization of a nonlocal scalar field in Chapter 4. But how about spinor or gauge fields? This is specially important in the case of gauge fields. Can we define a nonlocal version of gauge transformation to keep gauge invariance? or a Planck suppressed violation of gauge invariance is expected? What are the physical consequences of breaking gauge invariance at high energies?

3. We show in Chapter 4 that off-shell modes of a nonlocal field cannot be detected in scattering experiments. But how about other types of experiments? Is there a way of observing off-shell modes in laboratory directly?

Now we turn our attention to Lorentz violation in alternative gravitational theories in Chapters 6 and 7. There are various motivations to consider the existence of a fundamental preferred frame which range from pure phenomenology to attempts to solve the non-renormalizability of quantum gravity and the cosmological constant problem(s). In many explicit constructions, such as Einstein-Aether [15] or Gravitational Aether theories [17], K-essence [18], Cuscuton theory [19], or (non-projectable) Hořava-Lifshitz gravity [13, 14], the low energy theory contains a fluid (which defines a preferred frame) with superluminal or incompressible excitations.

The existence of superluminal excitations points out that a different causal structure exists in these theories compared to GR, even when the back-reaction of these excitations on the geometry is negligible. This property is especially of significance in the black hole

solutions, since superluminal signals can potentially escape the traditional Killing horizon of a black hole and make the classical theory unpredictable.

We lay out necessary foundations in Chapter 6 to study spherical gravitational collapse of neutral matter in the presence of a superluminal fluid in details. Although the resulting black hole solutions are close to the Schwarzschild spacetime, they possess a new feature: they contain a trapped surface forbidding the escape of any signal, no matter how fast its propagation speed is. This new type of horizon has been called a "Universal" horizon, as it is universal to all signals with arbitrary speed.

We extend this result to charged and spinning black holes which possess inner Killing horizons in Chapter 7, and show that a universal horizon always forms between the outer and (would-be) inner horizons. This finding suggests a notion of Cosmic Censorship even in Lorentz violating theories of gravity. A surprising result is that there are 3 distinct possible stationary universal horizons for a spinning black hole, only one of them matches the dynamical spherical solution. This motivates dynamical studies of collapse beyond spherical symmetry, which may reveal instabilities around the spherical solution.

Chapter 2

On the Causal Set-Continuum Correspondence

2.1 Background

From the viewpoint of causal set theory, the continuum spacetime of general relativity is only fundamental to the extent that it provides a good approximation to an underlying causal set [12, 20, 21, 22, 23]. Once a full dynamical theory of causal sets is available, it is necessary to judge whether or not the result of evolution looks anything like the universe we observe at low energies. Therefore, criteria must be established to determine how well a Lorentzian geometry (M, g) approximates a causal set $(C, \prec)$. [1] One natural criterion is to require the existence of an injective map $f : C \to M$ which preserves causal relations: $\forall\ x, y \in C$, $x \prec y$ if and only if $f(x) \in J^-(f(y))$, where $J^-(f(y))$ is the set of all points in M which lie in the causal past of $f(y)$. We would then say that C is *embeddable* in M. Of course, it is not very likely for a causal set which has emerged out of the dynamics to be *exactly* embeddable in *any* spacetime. Close to the discreteness scale, for instance, one would expect the causal set to be fairly chaotic. Therefore, a certain degree of *coarse graining* must be done before embedding is possible. It might also be necessary to introduce some notion of approximate embedding, because matching *all* causal relations exactly (and there would be a lot of them) seems too stringent a requirement. Once these issues are settled and embedding is possible, one last piece of information is required: scale. This is

[1] A causal set (causet) is a set C endowed with a binary relation $\prec$ such that for all $x, y, z \in C$ the following axioms are satisfied: (1) transitivity: $x \prec y\ \&\ y \prec z \Rightarrow x \prec z$, (2) irreflexivity: $x \not\prec x$, (3): local finiteness: $|\{y \in C | x \prec y \prec z\}| < \infty$.

because preserving causal relations cannot distinguish between spacetimes whose metrics are conformally related. Causal sets contain information about scale implicitly through counting of elements, because they are locally finite (i.e. discrete). To make use of this property, one also requires a *number–volume (N-V) correspondence*: the number N_S of embedded points in any spacetime region $S \subset M$ should "reflect" its volume V_S:

$$N_S \approx \rho V_S = \rho \int_S \sqrt{-g(x)} d^D x, \tag{2.1}$$

where ρ is a constant, thought to be set by the Planck scale, which represents the number density of points. Of course, this correspondence cannot be *exactly* true, the most obvious reason being that ρV_S is not always an integer. Also, for any embedding, there would always be infinitely many empty regions meandering through the embedded points. These issues can be addressed by first settling on the types of "test regions" S, and then requiring the correspondence in a statistical sense. To do so, let us first note that the causal set-continuum correspondence is only physically meaningful on scales much larger than the discreteness scale. Therefore, S should be a region whose spacetime volume is much larger than that set by the discreteness scale. The shape of S can be picked to disallow regions that meander through the embedded points but have large volumes. A natural choice, given that spacetime is Lorentzian, is the causal interval $I(x, y)$: given any two timelike points $x \prec y \in M$, $I(x, y)$ is the collection of all points in the causal future of x and the causal past of y. Having decided on the types of test regions, the number–volume correspondence can be formulated as follows: pick at random M causal intervals $S_1, S_2, \ldots, S_M$ with the same volume $V \gg \rho^{-1}$, and let $N_1, N_2, \ldots, N_M$ be the number of embedded elements in these regions, respectively. We then require that as $M \to \infty$:

$$\langle N \rangle = \rho V, \qquad \frac{\delta N}{\langle N \rangle} = \frac{\sqrt{\langle (N - \langle N \rangle)^2 \rangle}}{\langle N \rangle} \ll 1. \tag{2.2}$$

Having the N-V formulation at hand, [2] the key question becomes: what is the map that realizes the number–volume correspondence with the least noise?

The attitude in the causal set program is that this mapping is best done through *Poisson sprinkling*. In this approach, one first reverses direction by obtaining a causal set $C(M)$ from a given spacetime (M, g): randomly select points from M using the Poisson process

[2] It may seem more natural to require instead $|N_S - \rho V_S| \ll \rho V_S$ for all test regions S. This requirement, however, is a bit too stringent. Even if there is only one region which violates this condition, the N-V correspondence would be rendered unsatisfied. Requiring (2.2) ensures that *almost all* regions have volumes representative of the number of embedded points in them.

at density ρ and endow the selected points with their causal relations. The probability of selecting n points from a region with volume V is [3]

$$P(n) = \frac{(\rho V)^n e^{-\rho V}}{n!}. \tag{2.3}$$

Both the expectation value and variance of the number of selected points in a region with volume V is equal to ρV:

$$\langle N \rangle_{Pois} = \rho V, \qquad \frac{\delta N_{Pois}}{\langle N \rangle_{Pois}} = \frac{1}{\sqrt{\rho V}}. \tag{2.4}$$

The causal set-continuum correspondence is then judged as follows: *a Lorenztian manifold (M, g) is well-approximated by a causal set C if and only if C could have arisen from a sprinkling of (M, g) with "high probability"*. This definition is consistent with the N-V requirement formulated above: if C is embeddable as a "large enough" sprinkling of (M, g), (2.2) would be satisfied because of the ergodic nature of the Poisson process. The "high probability" requirement is necessary to ensure that a large enough sprinkling is indeed obtained. Ultimately, one needs to decide how high "high probability" is. A practical meaning could be that observables (such as dimension, proper time, etc) are not too wildly far from their mean [23]. It is interesting to note that *any* embeddable C has a finite probability of being realized through a Poisson sprinkling. This formulation of the causal set-continuum correspondence can be used for any point process (i.e. not just Poisson) which satisfies the N-V requirement on average.

Poisson sprinkling has many desirable features. It has been proven that even its realizations do not select a preferred frame in Minkowski space [24]. If this mapping really does provide the best causal set-continuum dictionary, it is intriguing that Lorentz invariance should follow as a biproduct. Also, Poisson sprinkling works in *any* curved background. Even the extra requirement of the shape of test regions as causal intervals is not necessary in this context. On the way to proving that the causal set structure is in principle rich enough to give rise to a smooth Lorentzian manifold, Poisson sprinkling has played a central role. But is it unique?

This chapter contains two results which (we hope) shed some light on certain aspects of this question. The first result is that the number–volume correspondence, if required to hold even for arbitrarily small regions, is best realized via Poisson sprinkling. The

[3] The Poisson process can be obtained by dividing spacetime into small regions of volume dV so that (i) in each infinitesimal region one point can be selected at most, and (ii) this selection happens with the probability ρdV *independent* of outside regions. Then, the probability of selecting n points in a volume V is $P(n) = \binom{V/dV}{n}(\rho dV)^n(1-\rho dV)^{V/dV-n}$, which converges to (2) in the limit $dV \to 0$.

second result concerns a family of lattices in $1+1$-dimensional Minkowski space, known as Lorentzian lattices, which we show provide a better number–volume correspondence than Poisson sprinkling for large volumes[4]. We argue, however, that this feature should not persist in higher dimensions and that it is special to $1+1$-dimensional Lorentzian lattices. We conclude by conjecturing that Poisson sprinkling provides the best number–volume correspondence in $3+1$ dimensions for spacetime regions with macroscopically large volumes.

2.2 Nothing beats Poisson for Planckian volumes

In this Section we prove that the number–volume correspondence is best realized via Poisson sprinkling for arbitrarily small volumes. We set $\rho = 1$ in the statement and proof of the theorem.

Theorem 1. *Let ξ be a point process whose realizations are points of a smooth Lorentzian manifold (M, g). Let N_S be the random variable which counts the number of points in a causal interval $S \subset M$: it takes on a value $n \in \{0, 1, 2, \dots\}$ with probability $P_S(n)$. Assume also that ξ realizes the number–volume correspondence on average $\forall\ S$: $\langle N_S \rangle = \sum_{n=0}^{\infty} n P_S(n) = V_S$, where V_S is the spacetime volume of S. Then*

$$\langle (N_S - V_S)^2 \rangle \leq \alpha V_S \quad \textit{where} \qquad 0 \leq \alpha < 1. \tag{2.5}$$

cannot be satisfied for all S. Note that the Poisson process satisfies (2.5) for $\alpha = 1$.

Proof. It is shown in Appendix A.1 that the variance of any random variable N_S which takes on a value $n \in \{0, 1, 2, \dots\}$ with probability $P_S(n)$, and whose mean is $V_S > 0$, must satisfy the inequality

$$\langle (N_S - V_S)^2 \rangle \geq (V_S - n_*)(n_* + 1 - V_S), \tag{2.6}$$

where n_* is the largest integer which is smaller than or equal to V_S. To see why this should be true, consider choosing $P_S(n)$ to obtain the least possible variance for N_S. Intuitively, this can be done by letting $P_S(n) = 0\ \forall\ n \neq n_*, n_* + 1$. Requiring $\langle N_S \rangle = V_S$ and $\sum_{n=0}^{\infty} P_S(n) = 1$ then implies $P_S(n_*) = n_* + 1 - V_S$ and $P_S(n_* + 1) = V_S - n_*$, which leads

[4] The existence of Lorentzian lattices in $1+1$-dimensional Minkowski space, and that they might be a contender for the Poisson process, was suggested by Aron Wall to Rafael Sorkin, who then mentioned it to us.

to the variance $(V_S - n_*)(n_* + 1 - V_S)$. The formal proof of this result is given in Appendix A.1.

Let us now proceed to prove the theorem by contradiction. Assume there exists $0 \leq \alpha < 1$ such that $\langle (N_S - V_S)^2 \rangle \leq \alpha V_S$ for all S. It then follows from (2.6) that

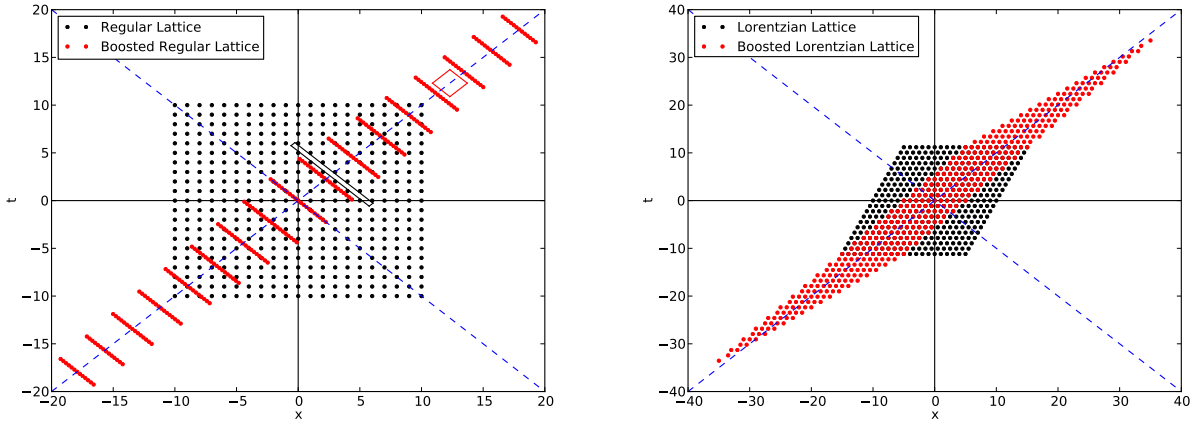
$$(V_S - n_*)(n_* + 1 - V_S) \leq \alpha V_S \qquad \forall \qquad S. \tag{2.7}$$

This, however, is clearly false because any region S with $V_S < 1 - \alpha$ violates this condition. □

The proof of this theorem rests heavily on regions with Planckian volumes. For instance, had we required the condition (2.5) for regions with $V_S > 1$, the proof would not have gone through. As we mentioned previously though, the causal set-continuum correspondence is only physically meaningful on scales much larger than the discreteness scale. In order to show that nothing *really* beats Poisson, our result would have to be generalized to the case of larger volumes. We have, however, found a counter example to this conjecture in the case of $1 + 1$-dimensional Minkowski space. As we shall see in the next Section, 2D Lorentzian lattices realize the number–volume correspondence much better than Poisson sprinkling for large volumes.

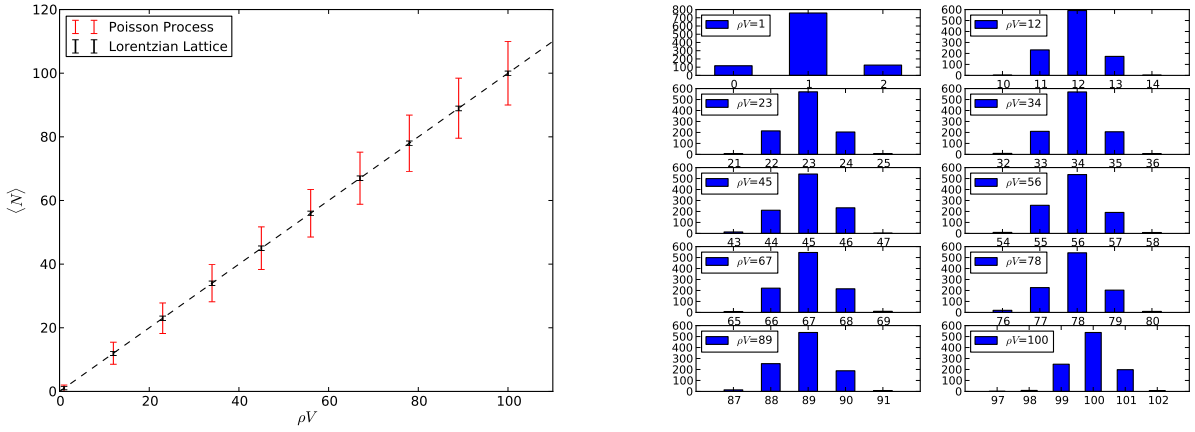
2.3 2D Lorentzian Lattices

Why is a *random*, as opposed to regular, embedding of points thought to provide the best number–volume correspondence? Consider, for instance, a causal set which is embeddable as a regular lattice in $1 + 1$-dimensional Minkowski space. Our intuition from Euclidean geometry dictates that such a lattice should at least match, if not beat, a random sprinkling in uniformity. Why not, then, use a regular lattice as opposed to Poisson sprinkling? Figure 2.1a shows what goes wrong in Lorentzian signature. Although the lattice is regular in one inertial frame, it is highly irregular for a boosted observer. Therefore, there are many empty regions with large volumes, which leads to a poor realization of the number–volume correspondence. Are there any regular lattices in $1 + 1$ that do not have this problem? As it turns out, the answer is yes: Lorentzian lattices. These are lattices which are invariant under a discrete subgroup of the Lorentz group. Such a lattice is shown in Figure 2.1b: it goes to itself under the action of a discrete set of boosts. We have classified all 2D Lorentzian lattices in Appendix A.2. In the case of the integer lattice shown in Figure 2.1a, the more it is boosted, the more irregular it becomes. A Lorentzian lattice, however,



(a) (b)

Figure 2.1: (a) The black dots show a lattice on the integers. The red dots are an active boost of this lattice by velocity $v = \tanh(1.5)$. The red diamond is a causal interval in the boosted frame which contains no points. The black diamond is the same causal interval as seen in the original frame. (b) The black dots show a Lorentzian lattice generated by the timelike vector $\xi_{(0)} = (\sqrt{5}/2, 1/2)$, and the spacelike vector $\xi_{(1)} = (0, 1)$. The red dots are boosts of the Lorentzian lattice by $v = \sqrt{5}/3$, showing that this particular boost takes the lattice to itself.



(a) (b)

Figure 2.2: The number–volume correspondence for the Lorentzian lattice shown in Figure 2.1b. (a) The mean and standard deviation of the number of points. (b) The histogram of the number of points for different volumes.

does not have this problem because it eventually goes to itself. It is then reasonable to expect a better number–volume correspondence in this case.

We have investigated the N-V correspondence for various Lorentzian lattices using simulations. Figure 2.2 shows the result of one such analysis on the lattice shown in Figure 2.1b. The setup is as follows: we consider 1000 different causal diamonds with the same volume V, whose centres and shapes vary randomly throughout the lattice[5]. For each realization, the number of lattice points inside the causal diamond is counted, leading to a distribution of the number of points for a given volume V. This procedure is then repeated for different volumes. As it can be seen from Figure 2.2, the Lorentzian lattice shown in Figure 2.1b realizes the number–volume correspondence with much less noise than Poisson sprinkling for macroscopic volumes. In fact, Figure 2.2b shows that the dispersion about the mean is barely growing with volume at all. The same exercise with the integer lattice results in a huge dispersion, much larger than that of Poisson, which is to be expected.

[5] We made sure to include "stretched out" causal diamonds, such as the black diamond shown in Figure 2.1a, as they are responsible for the poor realization of the number–volume correspondence in the integer lattice.

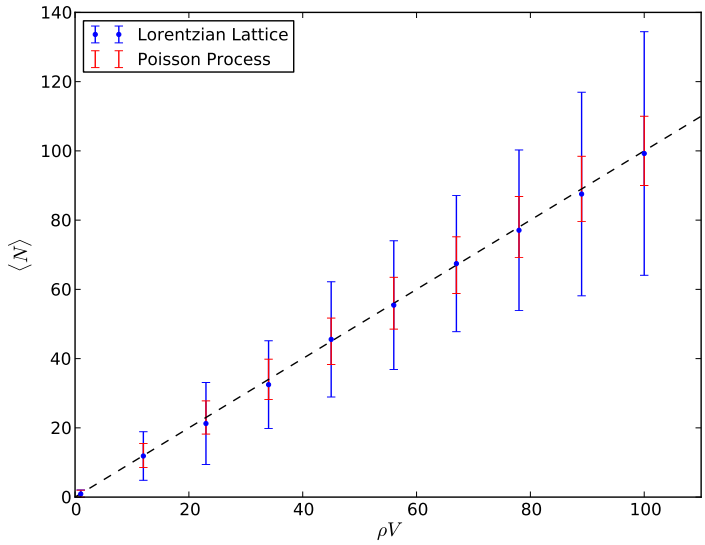


Figure 2.3: The number–volume correspondence for the $2+1$ dimensional integer lattice. For a given volume V, 200 different causal diamonds with volume V and randomly varying shapes are chosen. The mean and standard deviation of the number of points (blue) is compared with that of the Poisson process (red).

2.4 Higher–Dimensional Lorentzian Lattices

What about Lorentzian lattices in $3+1$ dimensions? Would they also realize the number–volume correspondence better than Poisson sprinkling? What is quite surprising is that the integer lattice *is* a Lorentzian lattice in both $2+1$ and $3+1$ dimensions [25].[6] We know from the $1+1$ dimensional integer lattice, however, that a boost along any spatial coordinate direction would create huge voids in any higher-dimensional integer lattice. Therefore, one would expect a poor number–volume realization in this case. We have confirmed this intuition for the $2+1$ dimensional integer lattice using simulations similar to those discussed previously (see Figure 2.3).

What makes $1+1$ dimensional Minkowski space special is that boosts can only be performed along one spatial direction. If a lattice is invariant under the action of a boost with velocity $v = \tanh\phi_*$, it is also left invariant when $v = \tanh(n\phi_*)$, where n is any

[6] In $2+1$, for instance, the following boosts take the integer lattice to itself: $v_x = v_y = 2/3$ and $v_x = 18/35, v_y = 6/7$.

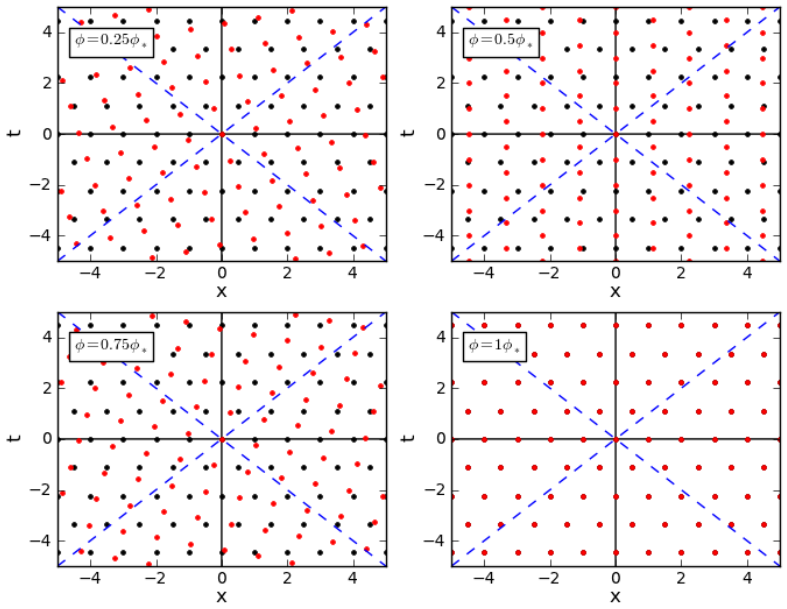


Figure 2.4: Various boosts of the Lorentzian lattice in Figure 2.1b, which is shown here with black dots. The red dots are active boosts of the lattice with velocity $v = \tanh(\phi = x\phi_*)$, where $\phi_* = \tanh^{-1}(\sqrt{5}/3)$ and x=$0.25, 0.5, 0.75, 1$ is used in different figures. The dashed blue lines correspond the lightcones of the origin, i.e. $t = \pm x$.

integer. Because of this periodicity, a Lorentzian lattice does not "change" too drastically under the action of an arbitrary boost. This is demonstrated in Figure 2.4.

A higher-dimensional lattice in Minkowski space would enjoy the same property if for *any* given spatial direction, one can find a boost (in that direction) which takes the lattice to itself. If this is not true, i.e. if there is a direction along which no boost leaves the lattice invariant, a boosted observer in that direction would see a non-uniform lattice with large voids, and therefore a poor realization of the number–volume correspondence. Intuitively, it is hard to imagine such a lattice could exist, as there are infinitely many directions along which one can boost (as opposed to just one in the case of $1 + 1$ dimensions). In what follows, we present a formal proof of this fact.

Theorem 2. *No lattice in D–dimensional Minkowski space, with $D > 2$, enjoys the following property: given any spatial direction, there exists a boost in that direction which takes the lattice to itself.*

Proof. We shall prove the theorem by contradiction. Let $\Lambda(\hat{n}, v)$ be a boost along the spatial direction $\hat{n}$ with speed v. We assume there exists a lattice generated by D linearly-independent vectors $\xi_{(d)}$ $(d \in \{0, 1, 2, \cdots, D-1\})$, with the property that for any given spatial direction $\hat{n}$, there exists $v(\hat{n})$ such that $\Lambda(\hat{n}, v(\hat{n}))$ takes the lattice to itself. Then, for any spatial direction $\hat{n} + \delta\hat{n}$ infinitesimally away from $\hat{n}$, there should also be a boost $\Lambda(\hat{n} + \delta\hat{n}, v(\hat{n} + \delta\hat{n}))$ which keeps the lattice invariant. Let us explore the consequences of this fact.

A boost $\Lambda(\hat{n}, v)$ can be written as

$$\Lambda(\hat{n}, v) = R^{-1}(\hat{n})\beta_x(v)R(\hat{n}), \tag{2.8}$$

where $R(\hat{n})$ is the rotation which takes $\hat{n}$ to the unit vector on the positive x–axis and $\beta_x(v)$ is a boost along the positive x–direction with magnitude v. Let δR and $\delta\beta_x$ denote the change in R and β_x under an infinitesimal change in the direction $\hat{n} \to \hat{n} + \delta\hat{n}$ and magnitude of boost $v \to v + \delta v$, respectively:

$$R(\hat{n} + \delta\hat{n}) = R(\hat{n}) + \delta R(\hat{n}), \tag{2.9}$$
$$\beta_x(v + \delta v) = \beta_x(v) + \delta\beta_x(v). \tag{2.10}$$

To first order in δR, it can be shown that $R(\hat{n} + \delta\hat{n})^{-1} = R^{-1}(\hat{n}) - R^{-1}(\hat{n})\delta R(\hat{n})R^{-1}(\hat{n}) + H.O$, from which it follows that $\Lambda(\hat{n} + \delta\hat{n}, v + \delta v) = \Lambda(\hat{n}, v) + \delta\Lambda(\hat{n}, v) + H.O$, where

$$\delta\Lambda(\hat{n}, v) = R^{-1}(\hat{n})\delta\beta_x(v)R(\hat{n}) + [\Lambda(\hat{n}, v), R^{-1}(\hat{n})\delta R(\hat{n})]. \tag{2.11}$$

As is shown in Appendix A.2, a lattice generated by D linearly-independent vectors $\xi_{(d)}$ is invariant under the action of a Lorentz transformation $\Lambda(\hat{n}, v)$ when all components of the matrix $A(\hat{n}, v) = C(\hat{n}, v)B^{-1}$ are integers, where

$$B_{(d)}{}^{(d')} \equiv \xi_{(d)} \cdot \xi_{(d')}, \qquad C_{(d)}{}^{(d')}(\hat{n}, v) \equiv \Lambda(\hat{n}, v)\xi_{(d)} \cdot \xi_{(d')}. \tag{2.12}$$

We have assumed there exists a lattice with the following property: for every $\hat{n}$, there exists $v(\hat{n})$ such that for all $\hat{n}$, all components of $A(\hat{n}, v(\hat{n}))$ are integers. To first order in $\delta\hat{n}$, $C(\hat{n} + \delta\hat{n}, v(\hat{n} + \delta\hat{n})) = C(\hat{n}, v(\hat{n})) + \delta C + H.O$, where

$$\delta C_{(d)}{}^{(d')}(\hat{n}, v(\hat{n})) = \delta\Lambda(\hat{n}, v(\hat{n}))\xi_{(d)} \cdot \xi_{(d')}. \tag{2.13}$$

Finally,

$$A(\hat{n}, v(\hat{n})) \to A(\hat{n}, v(\hat{n})) + \delta A(\hat{n}, v(\hat{n})), \qquad \delta A(\hat{n}, v(\hat{n})) = \delta C(\hat{n}, v(\hat{n}))B^{-1}. \tag{2.14}$$

Since by assumption all components of $A(\hat{n}+\delta\hat{n}, v(\hat{n}+\delta\hat{n}))$ should remain integers, we ought to have $\delta A(\hat{n}, v(\hat{n})) = 0$ to first order. If any component $\delta A_{(d)}{}^{(d')}(\hat{n}, v(\hat{n}))$ is non-zero, we can always pick $\delta\hat{n}$ small enough so that $|\delta A_{(d)}{}^{(d')}(\hat{n}, v(\hat{n}))| < 1$, which would in turn imply that $A_{(d)}{}^{(d')}(\hat{n}, v(\hat{n}))$ is not an integer.

Because B is invertible, $\delta A(\hat{n}, v(\hat{n})) = 0$ is equivalent to $\delta C(\hat{n}, v(\hat{n})) = 0$, which as we will now show leads to a contradiction when $D > 2$. Consider first $2+1$ dimensional Minkowski space. In this case, there is only one angle of rotation θ and the rotation matrix takes the form[7]

$$R(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix}. \tag{2.15}$$

Under an infinitesimal change of the angle of rotation $\theta + \delta\theta$, the rotation matrix changes to first order in $\delta\theta$ by

$$\delta R(\theta) = \delta\theta \begin{pmatrix} 0 & 0 & 0 \\ 0 & -\sin\theta & \cos\theta \\ 0 & -\cos\theta & \sin\theta \end{pmatrix}. \tag{2.16}$$

For every spatial direction θ, there should exist $\phi(\theta)$ such that a boost with magnitude $v = \tanh(\phi)$ along that direction takes the lattice to itself. As usual

$$\beta_x(\phi(\theta)) = \begin{pmatrix} \cosh\phi(\theta) & \sinh\phi(\theta) & 0 \\ \sinh\phi(\theta) & \cosh\phi(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{2.17}$$

and to first order in $\delta\theta$ we have

$$\delta\beta_x = \delta\theta \frac{d\phi}{d\theta} \begin{pmatrix} \sinh\phi & \cosh\phi & 0 \\ \cosh\phi & \sinh\phi & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{2.18}$$

For convenience, we may take $\theta = 0$. Let $\phi(0) \equiv \phi_0$ and $\frac{d\phi}{d\theta}(0) \equiv \dot{\phi}_0$. (Of course, we are interested in $\phi_0 \neq 0$, since $\phi_0 = 0$ is no boost at all.) In this case, it may be verified that

$$\delta\Lambda = \delta\theta \left(\dot{\phi}_0 D + E\right), \tag{2.19}$$

where

$$D = \begin{pmatrix} \sinh\phi_0 & \cosh\phi_0 & 0 \\ \cosh\phi_0 & \sinh\phi_0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad E = \begin{pmatrix} 0 & 0 & \sinh\phi_0 \\ 0 & 0 & \cosh\phi_0 - 1 \\ \sinh\phi_0 & \cosh\phi_0 - 1 & 0 \end{pmatrix}. \tag{2.20}$$

[7] As usual, for a point (x, y) in the x-y plane, θ is defined by $\tan\theta = y/x$.

As argued before, we ought to have $\delta C = 0$, which is equivalent to

$$\left(\dot{\phi}_0 D + E\right) \xi_d \cdot \xi_{d'} = 0, \qquad \forall \qquad d, d' \in \{0, 1, 2\}. \tag{2.21}$$

Since $\{\xi_d\}$ are linearly independent, this is equivalent to

$$\left(\dot{\phi}_0 D + E\right) V \cdot W = 0, \qquad \forall \qquad V, W \in \mathbb{R}^3. \tag{2.22}$$

Taking $V = W$ to be the unit vector on the y-axis (namely $V = W = (0, 0, 1)$), it can be shown that (2.22) is true only when $\phi_0 = 0$, which is a contradiction.

This proof generalizes trivially to higher dimensions, since boosts confined to the $x - y$ plane would lead to the same conclusion. □

Therefore, given any Lorentzian lattice in D–dimensional Minkowski space, with $D > 2$, there would be many boosted observers for whom the lattice looks highly irregular. As argued previously, this indicates a poor realization of the number–volume correspondence. This is not the case for Poisson sprinkling, because of its random and uncorrelated nature. No inertial observer is likely to see large macroscopic voids because a boosted random lattice is itself a random lattice. Concretely, the Poisson process does not pick out *any* preferred frame in the sense that one cannot find a measurable map from sprinklings to spacetime directions [24]. This suggests that Poisson sprinkling may be the best way of realizing the number–volume correspondence in $3 + 1$ dimensions. Theorem 2.5 shows this is the case if one requires the correspondence to hold even for arbitrarily small regions. However, based on the results and arguments presented thus far, our expectation is that Poisson sprinkling realizes the number–volume correspondence with the least noise even when only spacetime regions with macroscopically large volumes are considered. Below we formulate this expectation as a conjecture:

Conjecture 1. *Let ξ be a point process whose realizations are points of a $3+1$-dimensional smooth Lorentzian manifold (M, g). Let N_S be the random variable which counts the number of points in a causal interval $S \subset M$: it takes on a value $n \in \{0, 1, 2, \dots\}$ with probability $P_S(n)$. Assume also that ξ realizes the number–volume correspondence on average $\forall$ S: $\langle N_S \rangle = V_S$, where V_S is the spacetime volume of S. Then, $\nexists$ $V_* > 0$ such that for all causal intervals S with volume $V_S > V_*$, the following holds:*

$$\langle (N_S - V_S)^2 \rangle \leq \alpha V_S \quad \textit{where} \qquad 0 \leq \alpha < 1. \tag{2.23}$$

2.5 Conclusions

Causal set theory maintains that all information about the continuum spacetime of general relativity is contained microscopically in a partially order and locally finite set. Discreteness allows one to count elements, which is thought to provide information about scale: a spacetime region with volume V should contain about ρV causal set elements. In this chapter, we proved a theorem which shows that this number–volume correspondence is best realized via Poisson sprinkling for arbitrarily small volumes. Quite surprisingly, we also showed that $1 + 1$-dimensional Lorentzian lattices provide a much better number–volume correspondence than Poisson sprinkling for large volumes. We presented evidence, however, that this feature should not persist in $3 + 1$ dimensions and conjectured that the Poisson process should indeed provide the best number–volume correspondence for macroscopically large spacetime regions.

In the next Chapter, we address how to define a wave propagation on a causal set that reproduces local evolution law on large scales.

Chapter 3

Generalized Causal Set d'Alembertians

3.1 Introduction

Causal set theory postulates that the fundamental structure of spacetime is that of a locally finite partially ordered set [12].[1] Its marriage of discreteness with causal order implies that physics cannot remain local at all scales. To appreciate why this should be, let us consider how one might define a notion of "closeness" in a causal set, confining ourselves to causal sets C which are obtained by randomly selecting points from a Lorentzian manifold M and endowing the selected points with the causal relations inherited from the manifold[2]. Given such a causet, any intrinsically defined notion of closeness between two elements of C will reflect their Lorentzian distance in the embedding spacetime. But a small Lorentzian distance between two points of M does *not* mean that they are confined to a small neighbourhood within M. Rather, the second point can be "arbitrarily distant" from the first, as long as it is located near to the lightcone of the latter. Thus, an element of C will in-

[1] Characterized mathematically, this is a set C endowed with a binary relation $\prec$ such that for all $x, y, z \in C$ the following axioms are satisfied: (1) transitivity: $x \prec y \ \& \ y \prec z \Rightarrow x \prec z$; (2) irreflexivity: $x \not\prec x$; (3): local finiteness: $|\{y \in C | x \prec y \prec z\}| < \infty$. Thus a causal set (causet) is in a certain sense both Lorentzian [in virtue of (1) and (2)] and discrete [in virtue of (3)].

[2] This process is known as *Poisson sprinkling*: Given a spacetime M, let the discrete subset of points, C, be one particular realization of a Poisson process in M, and let the elements of C retain the causal relations they have when regarded as points of M. In order that the resulting precedence relation on C approximately encode the metric of M, one must exclude spacetimes with closed causal curves, for example by requiring M to be globally hyperbolic.

evitably possess *very many* "nearest neighbours", no matter how that notion is formalized. In this manner, the concept of locality provided by the topology of a continuous spacetime manifold is lost.

This nonlocality manifests itself concretely when one seeks to describe the wave propagation of a scalar field on a causal set by defining a discrete counterpart of the d'Alembertian operator, $\Box$. For the aforementioned reasons, it seems impossible to proceed in analogy with what one does when, for example, one discretizes the Laplacian operator in a Riemannian spacetime. Nevertheless, a non-local operator was suggested in [26] which on average reproduces $\Box$ in the appropriate continuum limit for $1+1$ dimensional Minkowski space $\mathbb{M}^2$ (i.e. for causets derived by sprinkling $\mathbb{M}^2$). The expression introduced in [26] was generalized to $D = 4$ dimensions in [27] and recently to arbitrary D in [28, 29].

We shall denote a discrete causal set d'Alembertian designed for $\mathbb{M}^D$ by $B_\rho^{(D)}$, where ρ (dimensionally an inverse spacetime volume) is a volume-scale that controls the extent of the non-locality. In the case of causal sets which are well-approximated by D-dimensional Minkowski space $\mathbb{M}^D$, averaging $B_\rho^{(D)}$ over all such causets (i.e. averaging over all sprinklings of $\mathbb{M}^D$ in the sense of footnote 2) leads to a *non-local* and retarded continuum operator $\Box_\rho^{(D)}$ defined in $\mathbb{M}^D$. We shall refer to this operator as the *continuum causal set d'Alembertian*. Its crucial property is that it reproduces the usual d'Alembertian in the limit of zero non-locality scale: $\Box_\rho^{(D)}\phi \to \Box\phi$ as $\rho \to \infty$ for test-functions ϕ of compact support.

Although the causet operator $B_\rho^{(D)}$ is necessarily nonlocal, one might expect that the range of its nonlocality could be confined to the discreteness scale itself. In other words, one might expect that $\rho \sim \ell^{-4}$, ℓ being the — presumably Planckian — discreteness length. However, one can also cite reasons why one might need to have $\rho \ll \ell^{-4}$, leading to a more long-range nonlocality.[3] Although these reasons are not conclusive, let us accept them provisionally. A natural question then arises: might such a "mesoscopic" nonlocality show up at energy-scales accessible by current experiments?

Ideally, one would address this question in the fully discrete setting, but it seems much easier to begin with the continuum version of the same question by asking what changes

[3] The issue here concerns the behavior of $B_\rho^{(D)}$ for one particular sprinkling versus its behavior after averaging over all sprinklings. The latter converges to $\Box$ as $\rho \to \infty$ but the former incurs fluctuations which grow larger as $\rho \to \infty$ and which therefore will be sizable if ρ is the sprinkling density, ℓ^{-4}. Which behavior is relevant physically? In full quantum gravity some sort of sum over different causets will be involved, including in particular a sum over sprinklings. Such a sum differs from a simple average and might or might not damp out the fluctuations, or they might cancel in other ways. But if neither of these things happens, the only way out [26] would be to choose ρ small enough that the necessary averaging will occur within each individual causet.

when the local operator $\Box$ is replaced by the nonlocal operator $\Box_\rho^{(D)}$. In this Chapter, we make a start on answering this question by analysing the "spectral properties" (Fourier transform) of a family of continuum operators $\Box_\rho^{(D)}$. In Section 3.2, we discuss the continuum operators corresponding to the original 2D [26] and 4D [27] causet d'Alembertians, and in Section 3.3 we generalize the discussion to an infinite family of operators parametrized by a set of coefficients, $\{a, b_n\}$, for which we derive explicit equations that ensure the usual flat space d'Alembertian is recovered in the infrared limit. Based on the UV behaviour of these operators (which we determine for all dimensions and coefficients $\{a, b_n\}$), we propose a genuinely Lorentzian perturbative regulator for quantum field theory (QFT). Finally, we address the question of whether or not the evolution defined by the (classical) equation $\Box_\rho^{(D)}\phi = 0$ is stable. We devise a numerical method to test for stability and present strong evidence that the original $4D$ causal set d'Alembertian is unstable in this sense, while its 2D counterpart is stable.

Throughout the paper we use the metric signature $(-+++\cdots)$ and set $\hbar = c = 1$.

3.2 The Original 2D and 4D Causet d'Alembertians

In this Section we discuss the original continuum causet d'Alembertians for dimensions two [26] and four [27]. Let us start by establishing some terminology. Given any two elements x, y of a causal set C, we define the *order interval* $\mathrm{Int}(x, y)$ between them as the set of all elements common to the (exclusive) future of x and the (exclusive) past of y: $\mathrm{Int}(x, y) = \{z \in C | x \prec z \prec y\}$. Notice that in our convention, $\mathrm{Int}(x, y)$ does not include x or y. An element $y \prec x$ is then considered a past nth neighbour of x if $\mathrm{Int}(y, x)$ contains n elements. For instance, y is a 0th neighbour of x if $\mathrm{Int}(y, x)$ is empty, a first neighbour if $\mathrm{Int}(y, x)$ contains one element, and so on (see Figure 3.1 for an example). We denote the set of all past nth neighbours of x by $I_n(x)$.

Throughout the paper, we will only consider causal sets which are obtained by Poisson sprinklings of Minkowski space at density ρ.

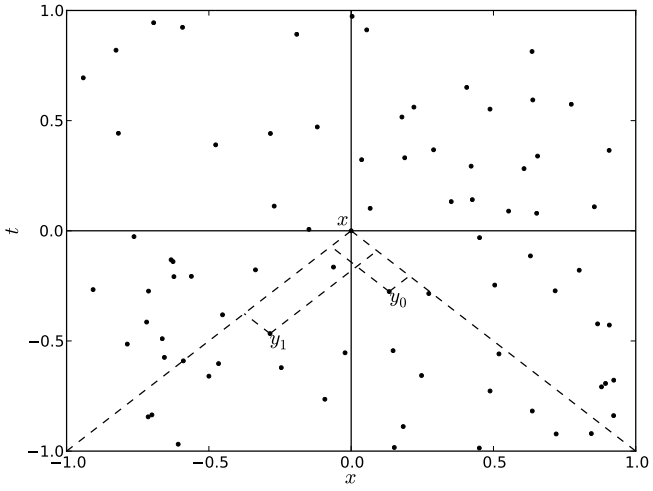


Figure 3.1: A Poisson sprinkling of $1+1$ Minkowski space at density $\rho = 80$. Here y_0 is a 0th neighbour of x because there are no elements which are both to the future of y_0 and the past of x. Similarly, y_1 is a first neighbour of x. The contributions of the points y_0 and y_1 to $\rho^{-1}(B_\rho^{(2)}\Phi)(x)$ are $b_0^{(2)}\Phi(y_0)$ and $b_1^{(2)}\Phi(y_1)$, respectively. The continuum limit, or rather average, of $(B_\rho^{(2)}\Phi)(x)$ can be understood as follows: fix the point x, keep sprinkling at density ρ and compute $(B_\rho^{(2)}\Phi)(x)$ for every sprinkling. The average of all these values is equal to $(\Box_\rho^{(2)}\Phi)(x)$.

3.2.1 2D

The original causet d'Alembertian for dimension 2, which we denote by $B_\rho^{(2)}$, acts on a scalar field $\Phi(x)$ on the causal set in the following way [26]:

$$\rho^{-1}(B_\rho^{(2)}\Phi)(x) = a^{(2)}\Phi(x) + \sum_{n=0}^{2} b_n^{(2)} \sum_{y\in I_n(x)} \Phi(y), \tag{3.1}$$

where

$$a^{(2)} = -2, \qquad b_0^{(2)} = 4, \qquad b_1^{(2)} = -8, \qquad b_2^{(2)} = 4. \tag{3.2}$$

Figure 3.1 illustrates how $B_\rho^{(2)}$ is defined, given a Poisson sprinkling of 2D Minkowski space $\mathbb{M}^2$. The continuum operator $\Box_\rho^{(2)}$ is obtained by averaging $B_\rho^{(2)}$ over all such Poisson sprinklings at density ρ:

$$\rho^{-1}(\Box_\rho^{(2)}\Phi)(x) = a^{(2)}\Phi(x) + \rho \sum_{n=0}^{2} \frac{b_n^{(2)}}{n!} \int_{J^-(x)} e^{-\rho V(x-y)}[\rho V(x-y)]^n \Phi(y)\, d^2y\,. \tag{3.3}$$

Here $J^-(x)$ denotes the causal past of x, and $V(x-y)$ is the spacetime volume enclosed by the past lightcone of x and the future lightcone of y. Note that $\Box_\rho^{(2)}$ is a *retarded* operator, in the sense that (3.3) uses information only from the causal past of x.

The operator $\Box_\rho^{(2)}$ can be studied by analysing its action on plane waves. Due to translation symmetry of Minkowski space,[4] any plane wave $e^{ip\cdot x}$ is an eigenfunction of $\Box_\rho^{(2)}$ (provided that the integrals in (3.3) converge, so that the left hand side is well defined):

$$\Box_\rho^{(2)} e^{ip\cdot x} = g_\rho^{(2)}(p) e^{ip\cdot x}, \tag{3.4}$$

where $p\cdot x \equiv \eta_{\mu\nu}p^\mu x^\nu$ and $\eta_{\mu\nu} = \mathrm{diag}(-1,1)$. Interestingly enough, $g_\rho^{(2)}(p)$ in this case can be expressed in closed form:[5]

$$\rho^{-1} g_\rho^{(2)}(p) = -Z e^{Z/2} \mathrm{E}_2(Z/2), \tag{3.5}$$

where $\mathrm{E}_2(z)$ is a generalized exponential integral function (see e.g. 8.19 of [30]) and

$$Z \equiv \rho^{-1} p\cdot p. \tag{3.6}$$

[4] This is why the volume V in (3.3) is a function only of the difference, $x-y$.

[5] This formula is derived in Appendix B.3, using the general formalism developed in Section 3.3.

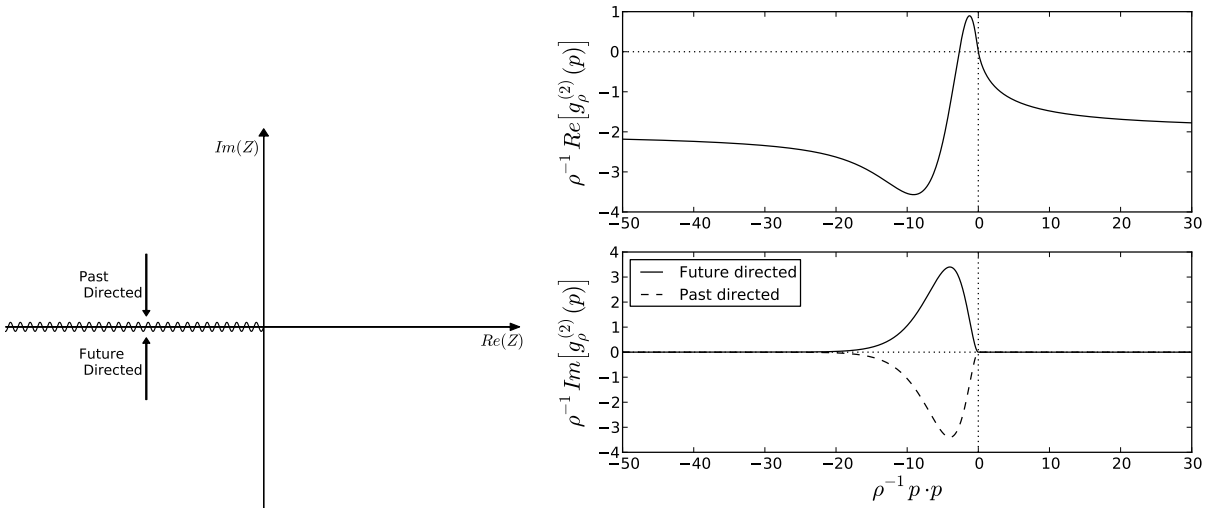


Figure 3.2: (a) The principal branch of $\rho^{-1}g_\rho^{(2)}(p)$, which (for real p) depends only on $Z = \rho^{-1}p\cdot p$, and on $\mathrm{sgn}(p^0)$ when p is timelike. (b) The spectrum $g_\rho^{(2)}(p)$ of the original 2D continuum causet d'Alembertian for real momenta p. For spacelike momenta ($p\cdot p > 0$), $g^{(2)}(p)$ is real. For timelike momenta, it is complex with an imaginary part whose sign is opposite for past-directed and future-directed momenta.

Here, as illustrated in Figure 3.2, $\mathrm{E}_2(z)$ assumes its principal value, with a branch cut along the negative real axis. For real and spacelike momenta ($Z > 0$), $g^{(2)}$ is real. For real and timelike momenta ($Z < 0$), its value above/below the branch cut corresponds to past/future-directed momentum-vectors. There, $g^{(2)}_\rho$ is complex and changes to its complex conjugate across the cut. That the spectrum is different for past and future-directed momenta should come as no surprise, given that $\Box^{(2)}_\rho$ is retarded by definition. We will see in Section 3.3 that these features persist in all dimensions and for a much broader class of causet d'Alembertians.

The infrared (IR) and ultraviolet (UV) behaviours of $g^{(2)}_\rho(p)$ are easily deduced from the asymptotic forms of $E_2(Z)$ (see e.g. 8.11.2, 8.19.1, and 8.19.8 of [30]):

$$\rho^{-1} g^{(2)}_\rho(p) \xrightarrow{Z\to 0} -Z + \cdots \tag{3.7}$$

$$\rho^{-1} g^{(2)}_\rho(p) \xrightarrow{Z\to \infty} -2 + \frac{8}{Z} + \cdots . \tag{3.8}$$

The first of these two equations shows that the usual d'Alembertian $\Box$ is indeed reproduced in the limit of zero non-locality. The second equation, on the other hand, reveals a UV behaviour quite unlike that of the usual d'Alembertian; in Section 3.3.2 it will lead us to propose a new regularization scheme for quantum field theory.

An important question is whether the evolution defined by $\Box^{(2)}_\rho \Phi = 0$ is stable or not. To a large extent this is answered by the fact that the only zero of $g^{(2)}_\rho(p)$ occurs at $Z = \rho^{-1} p\cdot p = 0$. To demonstrate this, we note that $g^{(2)}_\rho(p)$ has the following representation (see e.g. 8.19.1 and 8.6.4 of [30]):

$$\rho^{-1} g^{(2)}_\rho(p) = -Z f(Z), \qquad f(Z) \equiv \int_0^\infty \frac{te^{-t}}{t + Z/2}\, \mathrm{d}t. \tag{3.9}$$

It therefore suffices to prove that $f(Z)$ has no zeros when $Z \neq 0$. But the imaginary part of $f(Z)$ is

$$\mathrm{Im}(f(Z)) = -\frac{\mathrm{Im}(Z)}{2} \int_0^\infty \frac{te^{-t}}{\left[t + \frac{\mathrm{Re}(Z)}{2}\right]^2 + \left[\frac{\mathrm{Im}(Z)}{2}\right]^2}\, \mathrm{d}t. \tag{3.10}$$

Because the integral that multiplies $-\mathrm{Im}(Z)/2$ in (3.10) is strictly positive, $Zf(Z)$ could vanish only for real Z. Obviously, it does vanish for $Z = 0$, but elsewhere on the real axis, it remains nonzero, as illustrated in Figure 3.2b.

What we have just proven is that a plane wave solves the equation $\Box^{(2)}_\rho \Phi = 0$ iff it solves the equation $\Box\Phi = 0$. To the extent that the general solutions of these two wave equations

can be composed of plane waves, they therefore share the same space of solutions. This, of course, is an important result in itself. But it also, a fortiori, answers the stability question in the affirmative, since we know that the evolution corresponding to $\Box$ is stable.

If there remains any doubt about stability or about the fact that both $\Box\Phi = 0$ and $\Box^{(2)}_\rho\Phi = 0$ yield the same evolution, it springs from a possible uncertainty about boundary conditions. In the usual situation (that of the ordinary d'Alembertian $\Box$), one understands how to relate a general solution to its initial data on an arbitrary Cauchy surface, and when Φ falls off suitably at infinity, its total energy is defined and conserved. From energy conservation, stability also follows — relative to the given choice of boundary conditions. On the other hand in the case of $\Box^{(2)}_\rho$, a connection between solutions and Cauchy data remains to be found, as does a better understanding of appropriate falloff conditions. But absent some such boundary condition there is nothing to exclude complex momenta p that lead to exponential growth in time, e.g. an imaginary multiple of a real lightlike vector.

For these reasons, we would like to discuss stability from a slightly different angle, which also will be helpful when we come to deal with the 4D case. Quite generally, instabilities tend to be associated with exponentially growing "modes" (in this case plane waves). Let us then *assume* that we can take this as our criterion of (in)stability. And to exclude the kind of "fake instability" mentioned above, let us also require any putative unstable mode, $\Phi(x) = e^{ip\cdot x}$, to be bounded at spatial infinity in at least one Lorentz frame. (Unfortunately we cannot say "in all Lorentz frames", since for a plane wave, exponential growth in time induces exponential growth in space via a Lorentz boost.) We might hope that the condition just formulated is equivalent to the following more natural one: consider only solutions of $\Box^{(2)}_\rho\Phi(x) = 0$ which have compact support on every Cauchy hypersurface (compact spatial support in every frame.)

Be that as it may, if this criterion is accepted, then we can establish stability very simply in the present case, because an unstable mode, $\Phi(x) = e^{ip\cdot x}$, is then precisely one such that p possesses a future-directed timelike imaginary part: $p = p_R + ip_I$ with $p_I \cdot p_I < 0$ and $p^0_I > 0$. This, however, is impossible for $Z = 0$, as one sees from the equation $0 = p \cdot p = p_R \cdot p_R - p_I \cdot p_I + 2ip_R \cdot p_I$, whose right-hand side has a strictly positive real part when p_I is timelike. For logical completeness, we should also observe that (3.5) is valid for all complex p whose imaginary parts are timelike and future-directed. (For more general complex momenta, the integral defining $\Box^{(2)}_\rho\Phi$ might not converge, a circumstance that, depending once again on the choice of falloff conditions, might or might not impinge on the claimed identity between our solutions and those of the ordinary wave equation.)

3.2.2 4D

The causet d'Alembertian for dimension 4, has the same general form as that for $\mathbb{M}^2$, but with different coefficients [27] :

$$\rho^{-\frac{1}{2}}(B_\rho^{(4)}\Phi)(x) = a^{(4)}\Phi(x) + \sum_{n=0}^{3} b_n^{(4)} \sum_{y\in I_n(x)} \Phi(y), \tag{3.11}$$

where

$$a^{(4)} = -\frac{4}{\sqrt{6}}, \qquad b_0^{(4)} = \frac{4}{\sqrt{6}}, \qquad b_1^{(4)} = -\frac{36}{\sqrt{6}}, \qquad b_2^{(4)} = \frac{64}{\sqrt{6}}, \qquad b_3^{(4)} = -\frac{32}{\sqrt{6}}. \tag{3.12}$$

The continuum average $\Box_\rho^{(4)}$ then also takes a similar form:

$$\rho^{-\frac{1}{2}}(\Box_\rho^{(4)}\Phi)(x) = a^{(4)}\Phi(x) + \rho\sum_{n=0}^{3} \frac{b_n^{(4)}}{n!} \int_{J^-(x)} e^{-\rho V(x-y)}[\rho V(x-y)]^n \Phi(y) d^4y. \tag{3.13}$$

We will show in Section 3.3.1 that the "spectrum" of $\Box_\rho^{(4)}$, as defined by $\Box_\rho^{(4)} e^{ip\cdot x} = g_\rho^{(4)}(p)e^{ip\cdot x}$, is given by

$$\rho^{-1/2} g_\rho^{(4)}(p) = a^{(4)} + 4\pi Z^{-1/2} \sum_{n=0}^{3} \frac{b_n^{(4)}}{n!} C_4^n \int_0^\infty s^{4n+2} e^{-C_4 s^4} K_1(Z^{1/2}s)\, \mathrm{d}s, \tag{3.14}$$

where K_1 is a modified Bessel function of the second kind and

$$Z \equiv \rho^{-1/2} p\cdot p, \qquad C_4 = \frac{\pi}{24}. \tag{3.15}$$

All functions in (3.14) assume their principal values with branch cuts along the negative real axis. Many properties of the 2D function $g_\rho^{(2)}(p)$ carry over to $g_\rho^{(4)}(p)$. For timelike p, the value of $g_\rho^{(4)}(p)$ above/below the branch cut corresponds to past/future-directed momenta, and it changes to its complex conjugate across the cut. Also, $g_\rho^{(4)}$ is real for spacelike momenta. Figure 3.3b shows the behaviour of $g_\rho^{(4)}(p)$ for real momenta.

The IR and UV behaviours of $g_\rho^{(4)}(p)$, which are derived in Sections 3.3.2 and 3.3.3, are given by

$$\rho^{-1/2} g_\rho^{(4)}(p) \xrightarrow{Z\to 0} -Z + \cdots \tag{3.16}$$

$$\rho^{-1/2} g_\rho^{(4)}(p) \xrightarrow{Z\to \infty} -\frac{4}{\sqrt{6}} + \frac{32\pi}{\sqrt{6}Z^2} + \cdots . \tag{3.17}$$

Again, the IR behaviour confirms that the usual d'Alembertian is reproduced in the limit of zero non-locality. The UV limit has the form of a constant plus a term proportional to p^{-4}. The inverse of $g_\rho^{(4)}(p)$, which defines the retarded Green's function in Fourier space, takes exactly the same form in the UV:

$$\frac{\rho^{1/2}}{g_\rho^{(4)}(p)} \xrightarrow{Z\to\infty} -\frac{\sqrt{6}}{4} - \frac{2\pi\sqrt{6}}{Z^2} + \cdots . \tag{3.18}$$

In any QFT based on $\Box_\rho^{(4)}$, the propagator associated with internal lines in Feynman diagrams would presumably have the same UV behaviour. Subtracting the constant term from the propagator (which corresponds to subtracting a δ-function in real space) would then render all loops finite. This procedure could be the basis of a genuinely Lorentzian regularization and renormalization scheme for QFT. We will discuss these things more generally in Sections 3.3.3 and 3.3.4.

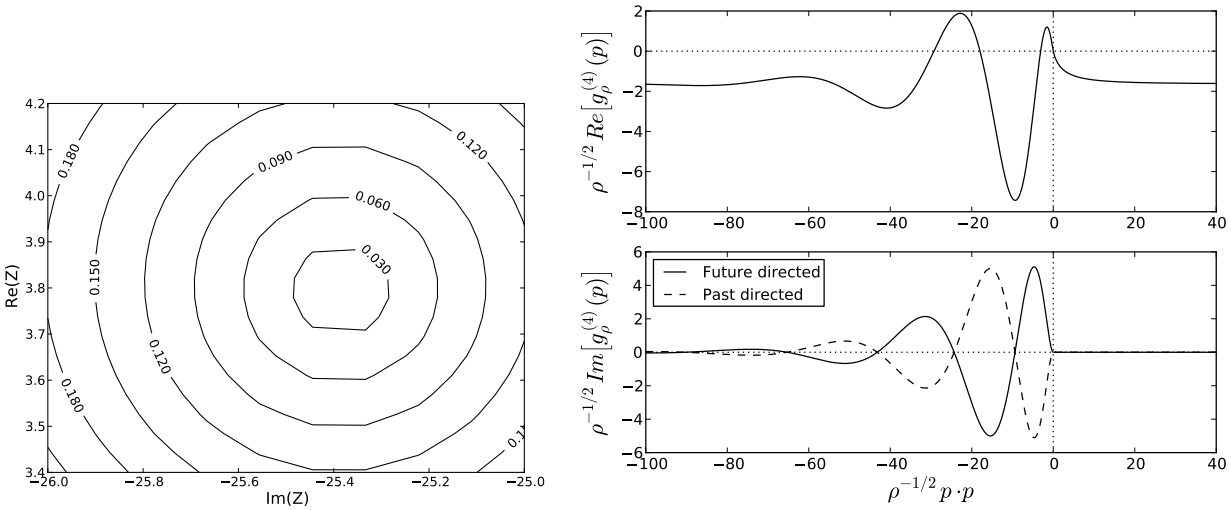
We have only been able to address the question of stability by numerical means in this case, and we refer the reader to Section 3.3.5. It turns out that $g_\rho^{(4)}(p)$ does in fact have unstable modes in the sense that there exist complex momentum-vectors p which satisfy $g_\rho^{(4)}(p) = 0$, and whose imaginary parts are timelike and future-directed. Such a mode corresponds to a complex zero of $g_\rho^{(4)}$ in the complex Z-plane, and Figure 3.3a shows one such zero (the other one being its complex conjugate).

3.3 The Generalized Causet Box (GCB) Operators

The key property of the causet d'Alembertians introduced in the previous Section is that they reproduce $\Box$ in the continuum-averaged (averaged over all sprinklings) and local ($\rho \to \infty$) limit. In this Section, we explore a larger family of operators $B_\rho^{(D)}$ which share the same property. We place the following conditions on $B_\rho^{(D)}$:

1. **Linearity**: when $B_\rho^{(D)}$ acts on a scalar field Φ, the result at an element x of the causet should be a linear combination of the values of Φ at other elements y (possibly including x itself). This is a natural requirement because $\Box$ itself is linear.

2. **Retardedness**: $(B_\rho^{(D)}\Phi)(x)$ should depend only on $\Phi(y)$, with y in the causal past of x. This requirement allows for a consistent evolution of a partial solution specified on any "downward closed" subset of the causet.



(a) (b)

Figure 3.3: (a) An unstable zero of $g_\rho^{(4)}(p)$. Contours of constant $|\rho^{-1/2} g_\rho^{(4)}|$ are plotted as a function of the real and imaginary parts of $Z = \rho^{-1/2} p \cdot p$. (b) Spectrum $g_\rho^{(4)}(p)$ of the original 4D causet d'Alembertian for real momenta p. For spacelike momenta ($p \cdot p > 0$), $g^{(4)}(p)$ is real. For timelike momenta, it contains also an imaginary part whose sign is opposite for past-directed and future-directed momentum-vectors.

3. **Label invariance**: $B_\rho^{(D)}$ should be invariant under relabellings of causal set elements. This is the discrete analogue of general covariance.

4. **Neighbourly democracy**: all nth neighbours of x should contribute to $(B_\rho^{(D)}\Phi)(x)$ with the same coupling.

Considering all these requirements, $(B^{(D)}\Phi)(x)$ can be expressed in the following general form

$$\rho^{-\frac{2}{D}}(B_\rho^{(D)}\Phi)(x) = a\Phi(x) + \sum_{n=0}^{L_{max}} b_n \sum_{y\in I_n(x)} \Phi(y), \tag{3.19}$$

where $\{a, b_n\}$ are dimensionless coefficients and $I_n(x)$ is the set of all nth neighbours to the past of x (see beginning of Section 3.2). This is a straightforward generalization of (3.1) and (3.11), where we have now allowed ourselves up to L_{max} neighbours. We will soon see that recovering $\Box$ requires keeping a *minimum* number of layers: e.g. $L_{max} \geq 2$ in 2D and $L_{max} \geq 3$ in 4D. The original 2D and 4D proposals are then the minimal cases in this sense.

The continuum-average $\Box_\rho^{(D)}$ of $B_\rho^{(D)}$ acts on a scalar field $\Phi(x)$ in the following way:

$$\rho^{-2/D}(\Box_\rho^{(D)}\Phi)(x) = a\Phi(x) + \rho \sum_{n=0}^{L_{max}} \frac{b_n}{n!} \int_{J^-(x)} e^{-\rho V(x,y)}[\rho V(x,y)]^n \Phi(y) d^D y. \tag{3.20}$$

Here as before, $J^-(x)$ denotes the causal past of x, while $V(x,y)$ is the spacetime volume enclosed by the past light cone of x and the future light cone of y.

The occurrence of the factor $e^{-\rho V}$ in (3.20) shows that the parameter ρ (which dimensionally is an energy-density) functions as a kind of "nonlocality scale" controlling the distance over which the operator $\Box_\rho^{(D)}$ acts. As our definitions stand so far, this nonlocality-scale directly reflects the fundamental discreteness-scale, because (3.20) was derived under the assumption that ρ was the sprinkling-density in $\mathbb{M}^D$. However it turns out that one can decouple the two scales by tweaking the definition (3.19) in such a way as to produce a more general causet operator whose sprinkling-average reproduces the same continuum operator (3.20), even when ρ is smaller than the sprinkling density. With this operator, the nonlocality can extend over a much greater distance than that of the fundamental discreteness-scale. Although modifying $B_\rho^{(D)}$ in this way has the disadvantge of introducing a second, independent length scale, it allows one to overcome a potential difficulty pointed out in [26], namely that (3.19) with fixed coefficients leads to fluctuations in $(B_\rho^{(D)}\Phi)(x)$ which grow with ρ, rather than diminishing. We have provided the definition

of this "tweaked" operator and the derivation of its continuum average in Appendix B.4; but henceforth, we will concern ourselves exclusively with the continuum operator $\Box_\rho^{(D)}$, without worrying about its relationship with any underlying discreteness. Correspondingly, *ρ will henceforth denote a non-locality-scale with no necessary relation to any discreteness scale.*

3.3.1 Spectrum

That any plane wave $e^{ip\cdot x}$ is an eigenfunction of $\Box_\rho^{(D)}$ in $\mathbb{M}^D$ follows from translational symmetry: $V(x,y) = V(x-y)$. It can be shown in fact that

$$\Box_\rho^{(D)}\ e^{ip\cdot x} = g_\rho^{(D)}(p)\ e^{ip\cdot x}, \tag{3.21}$$

$$\rho^{-2/D} g_\rho^{(D)}(p) = a + \sum_{n=0}^{L_{max}} \frac{(-1)^n \rho^{n+1}}{n!}\, b_n\, \frac{\partial^n}{\partial \rho^n} \chi(p,\rho), \tag{3.22}$$

$$\chi(p,\rho) = \int_{J^+(0)} e^{-\rho V(y)}\ e^{-ip\cdot y}\ d^D y, \tag{3.23}$$

where $V(y) = V(O,y)$ is the spacetime volume enclosed by the past light cone of y and the future light cone of the origin:

$$V(y) = C_D\ |y\cdot y|^{D/2}, \qquad C_D = \frac{\left(\frac{\pi}{4}\right)^{\frac{D-1}{2}}}{D\Gamma(\frac{D+1}{2})}\ . \tag{3.24}$$

Evaluating $\chi(p,\rho)$ amounts to computing the Laplace transform of a retarded, Lorentz-invariant function, which has been done in [31]. It follows from their result that

$$\chi(p,\rho) = 2(2\pi)^{D/2-1}(p\cdot p)^{\frac{2-D}{4}} \int_0^\infty s^{D/2} e^{-\rho C_D s^D} K_{\frac{D}{2}-1}(\sqrt{p\cdot p}\, s)\, \mathrm{d}s, \tag{3.25}$$

where K_ν is the modified Bessel function of the second kind. All functions in (3.25) assume their principal values, with a branch cut along the negative real axis. This result is valid for all p whose imaginary part is timelike and future-directed, i.e. $p_I \cdot p_I < 0$ and $p_I^0 > 0$, where $p = p_R + ip_I$ and the Lorentzian norm is given by $p\cdot p = p_R\cdot p_R - p_I\cdot p_I + 2ip_R\cdot p_I$. For momenta satisfying these conditions, the integral that defines $\chi(p,\rho)$, and consequently

$\Box_\rho^{(D)} e^{ip\cdot x}$, is absolutely convergent. Plugging (3.25) into (3.22) we find

$$\boxed{\rho^{-2/D} g_\rho^{(D)}(p) = a + 2(2\pi)^{D/2-1} Z^{\frac{2-D}{4}} \sum_{n=0}^{L_{max}} \frac{b_n}{n!} C_D^n \int_0^\infty s^{D(n+1/2)} e^{-C_D s^D} K_{\frac{D}{2}-1}(Z^{1/2} s)\, \mathrm{d}s,} \tag{3.26}$$

where Z is a dimensionless quantity defined by

$$Z \equiv \rho^{-\frac{2}{D}} p \cdot p. \tag{3.27}$$

For real $p = p_R$, $g_\rho^{(D)}(p)$ can be defined by first adding a small future-pointing and timelike imaginary part p_I^ϵ to p_R, and then taking the limit as p_I^ϵ shrinks:

$$g_\rho^{(D)}(p_R) := \lim_{\epsilon \to 0^+} g_\rho^{(D)}(p_R + i p_I^\epsilon), \qquad p_I^\epsilon \cdot p_I^\epsilon = -\epsilon^2. \tag{3.28}$$

When p_R is timelike, this amounts to changing $Z = \rho^{-\frac{2}{D}} p_R \cdot p_R$ on the right hand side of (3.26) to $Z + i\epsilon$ for past-directed, and $Z - i\epsilon$ for future-directed p_R. This is illustrated in Figure 3.2a. Because of the appearance of $Z^{1/2}$ in (3.26) and the fact that $K_\nu(\bar{z}) = \overline{K_\nu(z)}$, it follows for timelike p that

$$g_\rho^{(D)}(-p) = \overline{g_\rho^{(D)}(p)}. \tag{3.29}$$

Therefore, $g_\rho^{(D)}(p)$ *differs for past- and future-directed timelike* p. This is to be expected, since requiring $\Box_\rho^{(D)}$ to be retarded builds in a direction of time. For spacelike momenta ($Z > 0$), $g_\rho^{(D)}(p)$ is real, as follows from the fact that $K_\nu(z)$ is real when ν is real and $\mathrm{ph}(z) = 0$ [30].

3.3.2 IR Behaviour

We want to choose the coefficients a and b_n so that the usual d'Alembertian operator is recovered in the limit of zero non-locality:

$$\lim_{\rho \to \infty} \Box_\rho^{(D)} \phi = \Box \phi. \tag{3.30}$$

This requirement is equivalent to demanding

$$g_\rho^{(D)}(p) \xrightarrow{Z \to 0} -p \cdot p\ . \tag{3.31}$$

In Appendix B.1, we derive equations for a and b_n which guarantee this behaviour for an arbitrary spacetime dimension D. We expand $Z^{\frac{2-D}{4}} K_{\frac{D}{2}-1}(Z^{1/2}s)$ on the right hand side of (3.26) about $Z = 0$, and arrange a, b_n so that the terms which grow faster than Z vanish, while the coefficient of the term proportional to Z is -1. We state the main results here and refer the reader to Appendix B.1 for the details.

In **even dimensions**, letting $D = 2N + 2$ with $N = 0, 1, 2, \ldots$, the equations that need to be satisfied are

$$\sum_{n=0}^{L_{max}} \frac{b_n}{n!}\Gamma(n + \frac{k+1}{N+1}) = 0, \qquad k = 0, 1, \ldots, N+1 \tag{3.32a}$$

$$a + \frac{2(-1)^{N+1}\pi^N}{N!D^2C_D} \sum_{n=0}^{L_{max}} b_n \psi(n+1) = 0, \tag{3.32b}$$

$$\sum_{n=0}^{L_{max}} \frac{b_n}{n!}\Gamma(n + \frac{N+2}{N+1})\psi(n + \frac{N+2}{N+1}) = \frac{2(-1)^N(N+1)!}{\pi^N} D^2 C_D^{\frac{N+2}{N+1}}, \tag{3.32c}$$

where $\psi(n)$ is the digamma function. Equations (3.32a) and (3.32c) determine b_n, after which (3.32b) fixes a. The minimum number of terms required to solve these equations is determined by $L_{max} \geq N + 2$. In 2D and 4D in particular, keeping this minimum number of terms leads to the solutions (3.2) and (3.12), respectively.

In **odd dimensions**, letting $D = 2N + 1$ with $N = 0, 1, 2, \ldots$, the equation are

$$\sum_{n=0}^{L_{max}} \frac{b_n}{n!}\Gamma(n + \frac{2k+2}{2N+1}) = 0, \qquad k = 0, 1, \ldots, N \tag{3.33a}$$

$$a + \frac{(-1)^N \pi^{N+\frac{1}{2}}}{DC_D\Gamma(N+\frac{1}{2})} \sum_{n=0}^{L_{max}} b_n = 0, \tag{3.33b}$$

$$\sum_{n=0}^{L_{max}} \frac{b_n}{n!}\Gamma(n + \frac{2N+3}{2N+1}) = \frac{4(-1)^{N-1}\Gamma(N+\frac{3}{2})}{\pi^{N+\frac{1}{2}}} DC_D^{\frac{2N+3}{2N+1}}. \tag{3.33c}$$

Similarly to the even case, Equations (3.33a) and (3.33c) determine b_n, after which (3.33b) fixes a. The minimum number of terms is determined by $L_{max} \geq N + 1$.

3.3.3 UV Behaviour and the Retarded Green's Function

The UV behaviour of $g_\rho^{(D)}(p)$, as derived in Appendix B.2, is

$$\rho^{-2/D} g_\rho^{(D)}(p) \xrightarrow{Z\to\infty} a + 2^{D-1}\pi^{\frac{D}{2}-1}\Gamma(D/2)\, b_0\, Z^{-\frac{D}{2}} + \cdots . \tag{3.34}$$

Thus, $g_\rho^{(D)}(p)$ behaves as a constant plus a term proportional to $(p\cdot p)^{-D/2}$. Let us explore the consequences of this fact for the retarded Green's function $G_R(x,y)$ associated with $\Box_\rho^{(D)}$, which satisfies the usual equation

$$\Box_\rho^{(D)} G_R(x,y) = \delta^{(D)}(x-y), \tag{3.35}$$

subject to the boundary condition $G_R(x,y) = 0 \;\forall\; x \not\geq y$.

Of course, translation invariance implies $G_R(x,y) = G_R(x-y)$. The Fourier transform $\tilde{G}_R(p)$ of $G_R(x-y)$ is given by the reciprocal of $g_\rho^{(D)}(p)$:

$$G_R(x-y) = \int \frac{d^Dp}{(2\pi)^D}\tilde{G}_R(p)e^{ip\cdot(x-y)=} \int \frac{d^Dp}{(2\pi)^D}\frac{1}{g_\rho^{(D)}(p)}e^{ip\cdot(x-y)}. \tag{3.36}$$

Figure 3.4a shows the path of integration in the complex p^0 plane. When $g_\rho^{(D)}(p)$ has no zero in complex plane apart from at $p\cdot p = 0$, this choice of contour ensures that G_R is indeed retarded. As we will argue in the next section, the presence of such zeros implies that evolution defined by $\Box_\rho^{(D)}$ is unstable. Therefore, we shall ignore these cases for our current discussion.

The behaviour of $G_R(x-y)$ in the coincidence limit $x \to y$ is determined by the behaviour of $\tilde{G}_R(p)$ at large momenta:

$$\rho^{2/D}\tilde{G}_R(p) \xrightarrow{Z\to\infty} \frac{1}{a} - 2^{D-1}\pi^{\frac{D}{2}-1}\Gamma(D/2)\,\frac{b_0}{a^2}\,Z^{-\frac{D}{2}} + \cdots \tag{3.37}$$

Here we have assumed $a \neq 0$. When $a = 0$, $\tilde{G}_R(p)$ scales as p^D for large momenta, a badly divergent UV behaviour. Therefore we will confine ourselves to cases where $a \neq 0$.

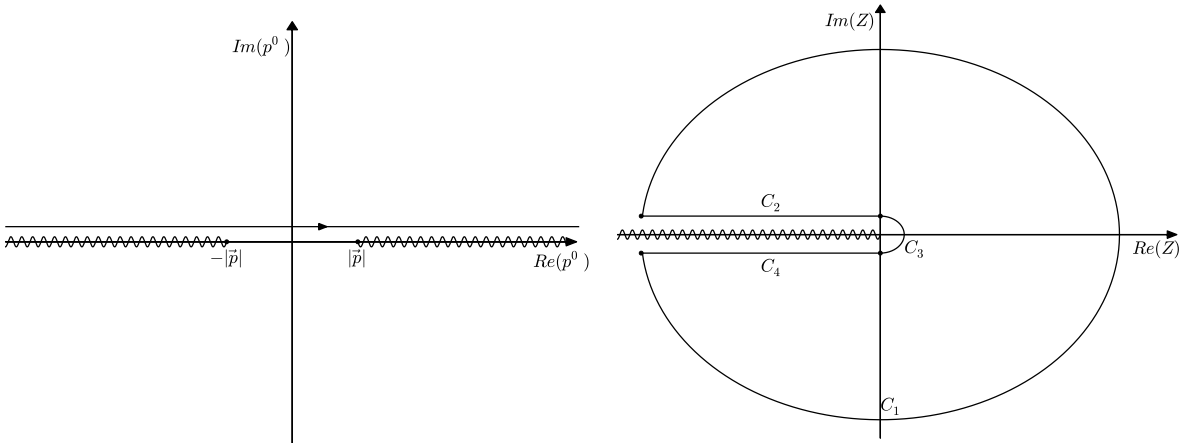
The constant term $\frac{1}{a}$ represents a δ-function in real space. The other terms in the series have the form $\int d^Dp\; p^{-nD}, n = 1, 2, \cdots$, and it can be shown that they are all finite. It then looks like subtracting $\frac{1}{a}\delta^{(D)}(x-y)$ from $\rho^{2/D}G_R(x-y)$ must result in a completely smooth function in the coincidence limit, and we will now show this is indeed the case.

Although $D = 4$ is the dimension of greatest interest, the proof which we shall present is valid in all even dimensions. Let us define

$$\rho^{2/D}G(x-y) \equiv \rho^{2/D}G_R(x-y) - \frac{1}{a}\delta^{(D)}(x-y). \tag{3.38}$$

Our task is then to show $G(x-y)$ is a smooth function at $x = y$. It follows from (3.36) that

$$\rho^{2/D}G(x-y) = \int \frac{d^Dp}{(2\pi)^D}\left[\frac{1}{\rho^{-2/D}g_\rho^{(D)}(p)} - \frac{1}{a}\right]e^{ip\cdot(x-y)}. \tag{3.39}$$



(a) (b)

Figure 3.4: (a) The integration path in the complex p^0 plane which defines the retarded Green's function. (b) The contour of integration used for counting the unstable modes of $\Box_\rho^{(D)}$. The direction of integration is taken to be counter-clockwise.

Because $G_R(x-y)$ is retarded by definition,

$$\int \frac{d^D p}{(2\pi)^D} \frac{1}{g_\rho^{(D)}(p)} e^{ip\cdot(x-y)} = 0, \qquad x \not\succ y \ . \tag{3.40}$$

From this it follows for all $x \succ y$ that

$$\int \frac{d^D p}{(2\pi)^D} \frac{1}{\overline{g_\rho^{(D)}}(p)} e^{ip\cdot(x-y)} = \int \frac{d^D p}{(2\pi)^D} \frac{1}{g_\rho^{(D)}(p)} e^{-ip\cdot(x-y)} \overset{x\succ y}{=} 0 \ , \tag{3.41}$$

where the first equality is obtained by changing $p \to -p$ and then using (3.29), and the second equality is a direct consequence of (3.40) with x and y interchanged. Returning to (3.39), and subtracting zero in the form of (3.41), we obtain

$$G(x-y) \overset{x\succ y}{=} \int \frac{d^D p}{(2\pi)^D} \left[\frac{1}{g_\rho^{(D)}(p)} - \frac{1}{\overline{g_\rho^{(D)}}(p)} \right] e^{ip\cdot(x-y)} \tag{3.42}$$

$$= \int_{p^2<0} \frac{d^D p}{(2\pi)^D} \left[\frac{1}{g_\rho^{(D)}(p)} - \frac{1}{\overline{g_\rho^{(D)}}(p)} \right] e^{ip\cdot(x-y)} \ , \tag{3.43}$$

where the second equality is true because $g_\rho^{(D)}(p)$ is real for space-like momenta. (Note that the $\frac{1}{a}$ term contributes only when $x = y$.) In what follows, we let

$$\rho^{-2/D} g_\rho^{(D)}(p) \equiv \tilde{g}(Z), \tag{3.44}$$

as given in the right hand side of (3.26).

The integral in (3.43) can be divided into two integrals over $p^0 > 0$ and $p^0 < 0$. For a fixed sign of p^0, $g_\rho^{(D)}(p)$ is only a function of $p \cdot p$, making (3.43) the Laplace transform of a Lorentz-invariant function. Similarly to how we derived (3.25), we use the result of [31] to compute $G(x-y)$:

$$\begin{aligned}\rho^{2/D} G(x-y) \overset{x\succ y}{=} & \frac{2}{\pi(2\pi)^{D/2}} \int_0^\infty d\xi\ \xi^{D/2} \\ & \times \mathrm{Re}\left[\left(\sqrt{s_\epsilon^2}\right)^{1-\frac{D}{2}} K_{\frac{D}{2}-1}(\sqrt{s_\epsilon^2}\xi) \left(\frac{1}{\tilde{g}(-\xi^2+i\epsilon)} - \frac{1}{\overline{\tilde{g}(-\xi^2+i\epsilon)}} \right) \right],\end{aligned} \tag{3.45}$$

where $s_\epsilon^2 = -(t_x - t_y + i\epsilon)^2 + |\vec{r}_x - \vec{r}_y|^2$ and ϵ is a small positive number which should be taken to zero at the end of calculations. When $x-y$ is timelike and future-directed, we can let $\sqrt{s_\epsilon^2} = -i\tau_{xy}$ where $\tau_{xy} > 0$. Using properties of Bessel functions (see e.g. 10.27.9 of [30]), (3.45) can be simplified into the following form for even D:

$$\begin{aligned}\rho^{2/D} G(x-y) &\overset{x\succ y}{=} \frac{-i(-1)^{\frac{D}{2}} \tau_{xy}^{1-\frac{D}{2}}}{(2\pi)^{D/2}} \int_0^\infty d\xi\ \xi^{D/2} \left(\frac{1}{\tilde{g}(-\xi^2+i\epsilon)} - \frac{1}{\overline{\tilde{g}(-\xi^2+i\epsilon)}} \right) J_{\frac{D}{2}-1}(\tau_{xy}\xi) \\ &= \frac{2(-1)^{1+\frac{D}{2}} \tau_{xy}^{1-\frac{D}{2}}}{(2\pi)^{D/2}} \int_0^\infty d\xi\ \xi^{D/2} \frac{\mathrm{Im}\left[\tilde{g}(-\xi^2+i\epsilon)\right]}{\left|\tilde{g}(-\xi^2+i\epsilon)\right|^2} J_{\frac{D}{2}-1}(\tau_{xy}\xi).\end{aligned} \tag{3.46}$$

Using $(x/2)^{1-D/2} J_{\frac{D}{2}-1}(x) \xrightarrow{x\to 0} \Gamma(D/2)^{-1}$ (see e.g. 10.2.2 of [30]) and the fact that $\mathrm{Im}\left[\tilde{g}(-\xi^2+i\epsilon)\right]$ is exponentially damped for large ξ (see Appendix B.2.1), it can be verified that

$$\lim_{x\to y} \rho^{2/D} G(x-y) = \frac{2^{2-\frac{D}{2}}(-1)^{1+\frac{D}{2}}}{(2\pi)^{\frac{D}{2}}\Gamma(\frac{D}{2})} \int_0^\infty d\xi\ \xi^{D-1} \frac{\mathrm{Im}\left[\tilde{g}(-\xi^2+i\epsilon)\right]}{\left|\tilde{g}(-\xi^2+i\epsilon)\right|^2}. \tag{3.47}$$

Thus $G(x-y)$ approaches a constant in the coincidence limit.[6] Strictly speaking, the discussion above only analyzes the behavior of $G(x-y)$ as τ_{xy} *approaches* 0, and consequently it does not exclude the presence of terms which blow up discontinuously on the light cone, such as $\delta(\tau_{xy}^2)$. However, a similar treatment for the case where $x - y \neq 0$ is null rather than timelike removes this loophole.

[6] One can understand intuitively why $G_R(x-y)$ is the sum of a δ-function with a bounded remainder

3.3.4 A Possible Regularization Scheme for Quantum Field Theoy

As was shown in the previous Section, changing the usual d'Alambertian to the nonlocal operator $\Box_\rho^{(D)}$ makes the coincidence limit more divergent, rather than smoothing it out as one might have initially expected. But it does so in an interesting way: all the divergences have now been absorbed into one δ-function at $x = y$. This feature has a natural application as a regularization tool for quantum field theory. In any QFT based on $\Box_\rho^{(D)}$, one would expect the propagator associated with internal lines in Feynman diagrams to have the same UV behaviour as (3.37). Subtracting the constant term in (3.37) (which corresponds to subtracting a δ-function in real space) would then render all loops finite. This would be a genuinely Lorentzian regulator, with no need for Wick rotation. It would also be physically motivated, with the "UV completion" being understood as a theory on the causal set. It would be interesting to apply this technique to the renormalization of some well-understood scalar field theories.

3.3.5 Stability

Is the evolution defined by $\Box_\rho^{(D)}$ stable? As we discussed in Section 3.2.1, instabilities are in general associated with "unstable modes", and we agreed to use this as our criterion of instability for purposes of this paper. More specifically, we took such a mode to be a plane-wave $\Phi(x) = e^{ip\cdot x}$ satisfying the equation of motion $\Box_\rho^{(D)}\Phi(x) = 0$, with the wave-vector p possessing a future-directed timelike imaginary part (i.e. $p = p_R + ip_I$ where $p_I \cdot p_I < 0$ and $p_I^0 > 0$).

The necessary and sufficient condition for avoiding unstable modes is then

$$\tilde{g}(Z) \neq 0 \ , \qquad \forall \ Z \neq 0, \tag{3.48}$$

where $\tilde{g}(Z)$ is defined in (3.44). Let us argue why this is the case. First observe that plane solutions of our wave-equation correspond exactly with zeros of $\tilde{g}(Z)$. If the above condition is verified, then the only such zero is at $Z = 0$, just as for the usual d'Alembertian. But we know (as is also easy to demonstrate ab initio) that there are no unstable modes

by noticing that (up to an overall numerical factor) our nonlocal d'Alembertian operator has the form $1 - S$, where the '1' corresponds to the first term in (3.3) or (3.20) and the remainder S is given by an integral-kernel which is both bounded and retarded. The inverse operator G_R would then be $G_R = (1-S)^{-1} = 1 + SG_R = 1 + S + S^2 + S^3 \cdots$, a series that should converge sufficiently near to $x = y$. Since the operator 1 is represented by a term of $\delta(x-y)$ in $G_R(x-y)$, one sees that $G_R(x-y)$ is the sum of a δ-function with a term involving only smooth bounded functions.

in the usual case. Conversely, when the above condition is violated for some complex $Z \neq 0$, it is always possible to find a corresponding p with a timelike and future-directed imaginary part which satisfies $p \cdot p = \rho^{\frac{2}{D}} Z$. To see this, we let $p = p_R + i p_I$ and take $p_R =< \pi_R^0, \vec{\pi}_R >$ and $p_I =< \pi_I, \vec{0} >$ with $\pi_I > 0$. This is always possible because p_I is timelike and future-directed. The equations that need to be satisfied are

$$p_R \cdot p_R - p_I \cdot p_I = \rho^{\frac{2}{D}} \mathrm{Re}(Z), \qquad 2 p_R \cdot p_I = \rho^{\frac{2}{D}} \mathrm{Im}(Z). \tag{3.49}$$

Substituting for p_I leads to

$$\pi_R^0 = \frac{\rho^{\frac{2}{D}} \mathrm{Im}(Z)}{-2\pi_I}, \qquad |\vec{\pi}_R|^2 = \rho^{\frac{2}{D}} \mathrm{Re}(Z) + \frac{\rho^{\frac{4}{D}} \mathrm{Im}(Z)^2}{4\pi_I^2} - \pi_I^2. \tag{3.50}$$

This system of equations always has a solution. In fact, there is a whole family of such unstable modes parametrized by π_I. Note however that the condition $|\vec{\pi}_R|^2 > 0$ puts an upper bound on the value of π_I, and therefore on the growth rate of such an instability.

We have thus reduced the question of whether or not $\Box_\rho^{(D)}$ has unstable modes to the question of whether $\tilde{g}(Z)$ has zeros other than $Z = 0$ in the complex plane. We can answer this question by counting the zeros of $\tilde{g}(Z)$ with the aid of the "argument principle" of complex analysis:

$$\frac{1}{2\pi i} \oint_C \frac{\tilde{g}'(Z)}{\tilde{g}(Z)} dZ = N - P, \tag{3.51}$$

where N and P are the number of zeros and poles, respectively, inside of the closed contour C, which we choose as shown in Figure 3.4b. The number of poles inside C is zero because all terms appearing in $\tilde{g}(Z)$ are finite in that region (at least when L_{max} is finite). As shown in Figure 3.4b, the path of integration C comprises four pieces: C_2 and C_4 run from $-\infty$ to 0 a distance ϵ above and below the negative real axis respectively, C_3 is a semicircle of radius ϵ about the origin, and C_1 is (almost) a circle whose radius should be taken to infinity. For large Z we have from (3.34),

$$\frac{\tilde{g}'(Z)}{\tilde{g}(Z)} \xrightarrow{Z \to \infty} \frac{-D 2^{D-1} \pi^{\frac{D}{2}-1} \Gamma(D/2)}{2a} b_0 Z^{-\frac{D}{2}-1} + \cdots, \tag{3.52}$$

and it follows that

$$\int_{C_1} \frac{\tilde{g}'(Z)}{\tilde{g}(Z)} dZ = 0. \tag{3.53}$$

(We remind the reader of our standing assumption that $a \neq 0$. See the remarks following (3.37).) On the other hand the IR behaviour, $\tilde{g}(Z) \xrightarrow{Z\to 0} -Z$, leads to

$$\int_{C_3} \frac{\tilde{g}'(Z)}{\tilde{g}(Z)} dZ = i\pi. \tag{3.54}$$

Also, because $\tilde{g}(x + i\epsilon) = \overline{\tilde{g}(x - i\epsilon)}$ for $x < 0$:

$$\int_{C_2+C_4} \frac{\tilde{g}'(Z)}{\tilde{g}(Z)} dZ = 2i \int_{C_2} \mathrm{Im}\left[\frac{\tilde{g}'(Z)}{\tilde{g}(Z)}\right] dZ. \tag{3.55}$$

Performing this last integral will allow us to determine whether $\Box_\rho^{(D)}$ has unstable modes or not.

Given a choice of the parameters a and b_n, the last integral can be computed numerically. In the minimal 4D case discussed in Section 3.2.2, we find that $\Box_\rho^{(4)}$ has precisely two "unstable zeros". (Notice that because $\tilde{g}(\bar{Z}) = \overline{\tilde{g}(Z)}$, if Z is a zero of $\tilde{g}(Z)$, so also is $\bar{Z}$.) We have located these zeros numerically, as shown in Figure 3.3a. With different choices of the parameters $\{a, b_n\}$, the number of zeros can change, but we have not been able to find any choice that would make $\Box_\rho^{(4)}$ stable. It would be interesting to find an analytical method to check for stability.

3.4 Summary and Remarks

We have defined an infinite family of scalar-field operators on causal sets which we dubbed Generalized Causet Box (GCB) operators. For causal sets made by sprinkling D-dimensional Minkowski space $\mathbb{M}^D$, these operators reproduce the usual d'Alembertian $\Box = \nabla_\mu \nabla^\mu$ when one averages over all sprinklings and takes the limit of infinite sprinkling-density ρ. If, on the other hand, one averages over all sprinklings while holding ρ fixed, one obtains an integral operator $\Box_\rho^{(D)}$ in $\mathbb{M}^D$ which is manifestly Lorentz-invariant, retarded, and nonlocal, with the degree of nonlocality set by ρ. In the present Chapter, we have been concerned primarily with these continuum operators, whose nonlocality can be regarded as a "mesoscopic" residue of the underlying causal set discreteness.

The GCB operators $B_\rho^{(D)}$ and their continuum averages $\Box_\rho^{(D)}$ are parametrized by a set of coefficients, and we derived the equations in these coefficients which ensure that $\Box$ is recovered in the infrared limit. The minimal solutions of these equations turned

out to reproduce the original operators proposed in [26]. We also computed the Fourier transform of $\Box_\rho^{(D)}$, or equivalently its "spectrum of eigenvalues" obtained by applying it to an arbitrary plane wave. For spacelilke momenta the spectrum is real. For timelike momenta it contains also an imaginary part, which changes sign under interchange of past with future. The UV behaviour of the spectrum differs from that of $\Box$ in a way which led us to propose a genuinely Lorentzian, perturbative regulator for quantum field theory.

We also studied the question of whether the evolution defined by the continuum-averaged GCB operators is stable. This is of interest in relation to nonlocal field theories based on $\Box_\rho^{(D)}$; it can also serve as an indicator of the stability or instability of the corresponding causet operator $B_\rho^{(D)}$. The continuum-average of the minimal 2D causal set d'Alembertian was shown to be stable by a direct proof. In 4D we did not settle the question analytically, but we devised a numerical diagnostic that applies to all the operators $\Box_\rho^{(D)}$, and which disclosed a pair of unstable modes when applied to the minimal 4D causal set d'Alembertian. Are any of the continuum-averaged GCB operators stable in $3+1$ dimensions? We were not able to find any, but there are an infinite number of such operators and a definitive search could only be conducted by analytical means.[7] Finally, it bears repeating that there might be more reliable indicators of instability than simply the existence of an exponentially growing plane-wave solution, which a priori tells us nothing about the behavior of solutions of limited spatial extent. For that reason, it would be worthwhile to analyze directly the late-time behavior of the Green function $G_R(x-y)$ which is inverse to $\Box_\rho^{(D)}$. If it were bounded that would imply stability, and if it grew exponentially, that would imply instability.

Our results also suggest other problems for further work. It would be interesting, for example, to work out the continuum-averaged GCB operators in curved spacetimes. It was found in [27] that the minimal 4D operator has the following limit as $\rho \to \infty$: $\Box_\rho^{(4)}\Phi \to \Box\Phi - \frac{1}{2}R\Phi$, where R is the Ricci scalar. (In fact one obtains the same limit in all dimensions D [29].) Would this feature persist for all of the GCB operators? This feature has also been used to define an action-functional for causal sets [27]. A final question then is whether the instability found above has any consequences for this causal set action?

In the next Chapter, we review quantizing a nonlocal field theory of this type and explain how this could potentially address one of the problems in fundamental physics, namely the origin of dark matter.

[7] Also interesting would be an unstable operator whose corresponding growth-time was either very large (cosmological) or very small (Plankian). In the former case, the instability would be irrelevant physically, in the latter case it might still be compatible with stability of the corresponding discrete evolution. We were not able to find any such operator in 4D either.

Chapter 4

Dark Matter From Spacetime Nonlocality

4.1 Introduction

The nature of dark matter is one of the most important problems in modern physics. Almost a century after it was hypothesized, though, our understanding of it is still limited to its gravitational signature on luminous matter. It is often assumed that dark matter is a new weakly interacting particle which is just hard to detect. However, so far there has been no conclusive direct or indirect detection in accelerators or cosmological/astrophysical settings. In what follows, we propose that dark matter is not yet another new particle in nature, but that it is a remnant of quantum gravitational effects on known fields. We arrive at this possibility in an indirect and surprising manner: by considering retarded, nonlocal, and Lorentzian evolution for quantum fields. Concretely, we study the consequences of replacing the d'Alembertian $\Box$ with an operator $\widetilde{\Box}$ which is Lorentz invariant, reduces to $\Box$ at low energies, and defines a retarded evolution: $(\widetilde{\Box}\phi)(x)$ only depends on $\phi(y)$, with y is in the causal past of x. Why is this type of evolution interesting, what does it have to do with quantum gravity, and how does it lead to a proposal for the nature of dark matter?

The causal set theory approach to quantum gravity postulates that the fundamental structure of spacetime is that of a locally finite and partially ordered set [12]. Its marriage of discreteness with causal order implies that physics cannot remain local at all scales. This nonlocality manifests itself concretely, for instance, when one seeks to describe the wave propagation of a scalar field on a causal set. As we showed in Chapter 3 coarse-graining the quantum gravitational degrees of freedom leads to a nonlocal field theory described by an

operator exactly of the type $\widetilde{\Box}$ [32, 33, 34, 35, 36]. There are reasons to suspect that this type of nonlocality is not necessarily confined to the Planck scale, and that it may have nontrivial implications for physics at energy scales accessible by current experiments (see [37, 38] and references therein for implications of nonlocality in the context of cosmology). It is then only natural to wonder what a quantum field theory built upon $\widetilde{\Box}$ would look like, especially that it may contain information about the fundamental structure of spacetime.

Studying $\widetilde{\Box}$ is also interesting from a purely field-theoretic perspective, since it forces us to relax one of the core assumptions of quantum field theory: locality. Most nonlocal and Lorentzian quantum field theories studied in the literature consider modifications of the type $\Box \to f(\Box)$. In this paper, we consider explicitly retarded operators, which are more generic and have more interesting properties as a result. For instance, the Fourier transform of $\widetilde{\Box}$ is generically complex, which is a direct consequence of retarded evolution. In fact, this feature is at the heart of our proposal for the nature of dark matter. It is also worth mentioning that quantizing a field theory of the type described here is non-trivial due to the absence of a local action principle. This presents a technical challenge, from which one may gain deeper insight into quantization schemes.

What is the relation between a quantum field theory based on $\widetilde{\Box}$ and dark matter? Upon quantizing a free massless scalar field $\phi(x)$ with the classical equation of motion $\widetilde{\Box}\phi(x) = 0$, we find *off-shell modes* in the mode expansion of the quantized field operator $\widehat{\phi}(x)$. These are modes which do not satisfy any dispersion relation, unlike in usual local quantum field theory (LQFT) where every Fourier mode with four-momentum p is an on-shell quanta, i.e. it satisfies $p \cdot p = 0$.[1] This is equivalent to the statement that the quantized field operator does not generically satisfy the classical equation of motion: $\widetilde{\Box}\widehat{\phi}(x) \neq 0$. Note that an off-shell mode of a massless scalar field has an effective mass, and can be thought of as a massive quanta in itself. We show that the off-shell modes can exist in "in" and "out" states of scattering, and are different from virtual particles which exist as intermediate states in Feynman diagrams. When considering the interacting theory, we find an extremely surprising result: the cross-section of any scattering process which contains one or more off-shell particle(s)[2] in the "in" state is zero. That is to say, *on-shell quanta can scatter and produce off-shell particles, but once produced, off-shell particles no longer interact.* It is this behaviour that makes these off-shell particles a natural candidate for dark matter. The phenomenological story would be that dark matter particles were produced in the early universe in this fashion: as off-shell modes of quantum fields. This feature of the theory can be traced back to the fact that $\widetilde{\Box}$ defines an explicitly retarded

[1] We use a signature of $-+++$ for the Minkowski metric $\eta_{\mu\nu}$. Also, $p_1 \cdot p_2 \equiv \eta_{\mu\nu} p_1^\mu p_2^\nu$.

[2] In the quantum theory, an off-shell particle is 1-particle quantum state with a well-defined (non-zero) mass and momentum, i.e. a massive eigenstate of Hamiltonian and momentum operator.

evolution, which as mentioned previously, may be a remnant of quantum gravitational degrees of freedom.

This Chapter is organized as follows. In Section 4.2, we start by setting forth a series of axioms which any non-local, retarded, and Lorentzian modification of $\Box$ at high energies should satisfy. In Section 4.3, we argue there is no action principle for the theory of interest, which forces us to carefully study, in Section 4.4, what quantization scheme should be used. There, we argue that canonical quantization and the Feynman path-integral approach do not work, and explain why the Schwinger-Keldysh (also known as the double path integral or in-in) formalism provides the appropriate framework. Sections 4.5 and 4.6 describe the interacting theory, where we work out the modified Feynman rules, find S-matrix amplitudes, and compute cross-sections for various examples and comment on the time reversibility of the theory. Although a *continuum superposition* of off-shell particles can in principle scatter into on-shell modes, we argue why this is unlikely to happen. Extension to massive scalar fields is discussed in 4.7. Section 5.5 concludes the paper.

4.2 Modified d'Alembertian: Definition

In this section we study generic spectral properties of non-local and Lorentzian modifications of the d'Alembertian $\Box$. We focus on a class of operators $\widetilde{\Box}$ which defines an explicitly retarded evolution: $(\widetilde{\Box}\phi)(x)$ depends only on $\phi(y)$ with y in the causal past of x. As we will see, such operators have interesting features which are absent in modifications of the type $f(\Box)$.[3] We start by setting forth a series of axioms which a non-local, retarded, and Lorentzian modification of $\Box$ at high energies should satisfy:

1. **Linearity**:
$$\widetilde{\Box}(a\phi + b\psi) = a\widetilde{\Box}\phi + b\widetilde{\Box}\psi, \qquad a, b \in \mathbb{C}, \tag{4.1}$$
where ϕ and ψ are complex scalar fields and $\mathbb{C}$ denotes the set of complex numbers.

2. **Reality**: for any real scalar field ϕ, $\widetilde{\Box}\phi$ is also real. Note that reality and linearity imply for any complex scalar field ϕ that
$$(\widetilde{\Box}\phi^*) = (\widetilde{\Box}\phi)^*, \tag{4.2}$$
where $*$ denotes complex conjugation.

[3] $f(x)$ being an analytic function of x.

3. **Poincare-invariance**: evolution defined by $\widetilde{\Box}$ is Poincare-invariant. Consider a scalar field $\phi(x)$ which transforms to $\phi'(x) = \phi(\Lambda^{-1}x)$ under a Poincare transformation $x \to \Lambda x$. We require $\widetilde{\Box}$ to be invariant under the action of Λ:

$$(\widetilde{\Box}\phi')(x) = (\widetilde{\Box}\phi)(\Lambda^{-1}x). \tag{4.3}$$

Taking Λ to be a spacetime translation $\Lambda(x) = x+a$, one finds that the eigenfunctions of $\widetilde{\Box}$ are plane waves. To see this, let $\phi(x) = e^{ip\cdot x}$ and define $\psi(x) \equiv (\widetilde{\Box}\phi)(x)$. It then follows from (4.3) that

$$e^{-ip\cdot a}\psi(x) = \psi(x-a), \tag{4.4}$$

where we have used the linearity condition. Solutions to the above equation are plane waves:

$$\psi(x) = \widetilde{\Box}e^{ip\cdot x} = B(p)e^{ip\cdot x}, \tag{4.5}$$

where $B(p)$ is any function of the wave-vector p. Therefore, it follows from translational invariance that $e^{ip\cdot x}$ is an eigenfunction of $\widetilde{\Box}$ with the corresponding eigenvalue $B(p)$. Taking Λ to be a Lorentz transformation, it can be shown that $B(p)$ can only depend on the the Lorentzian norm of p, i.e. $p\cdot p \equiv \eta_{\mu\nu}p^\mu p^\nu$, and whether or not p is future or past directed, i.e. $\mathrm{sgn}(p^0)$:

$$B(p) = B(\mathrm{sgn}(p^0), p\cdot p). \tag{4.6}$$

Combining (4.5) and (4.2) we find $B(-p) = B^*(p)$, which using (4.6) is equivalent to

$$B(-\mathrm{sgn}(p^0), p\cdot p) = B(\mathrm{sgn}(p^0), p\cdot p)^*. \tag{4.7}$$

For a spacelike wave-vector p^μ, it is always possible to find a coordinate system in which $p^0 = 0$. As a result, $B(p)$ is real for spacelike p. For timelike momenta, however, $B(p)$ *may be complex and its imaginary part changes sign when* $p^0 \to -p^0$.

Most nonlocal modifications of $\Box$ considered in the literature are of the form $f(\Box)$, in which case $B(p)$ is only a function of $p\cdot p$. In this paper we focus on a class of nonlocal operators for which $B(p)$ does depend on $\mathrm{sgn}(p^0)$, and find many interesting consequences as a result.

4. **Locality at low energies**: since $\Box$ provides a good description of nature at low energies, we require $\widetilde{\Box} \to \Box$ in this regime. In other words, expanding $B(\mathrm{sgn}(p^0), p\cdot p)$ for "small" values of $p\cdot p$, we require the leading order behaviour to be that of $\Box$:

$$B(p) \xrightarrow{p\cdot p\to 0} -p\cdot p. \tag{4.8}$$

Note that by a "small" value of $p\cdot p$, we mean in comparison to a scale which can be interpreted as the non-locality scale, implicitly defined through $\widetilde{\Box}$.

5. **Stability**: we require that evolution defined by $\widetilde{\Box}$ is stable. This condition implies that $B(p)$, when analytically continued to the complex plane of p, only has a zero at $p \cdot p = 0$ [33].

6. **Retardedness**: $(\widetilde{\Box}\phi)(x)$ only depends on $\phi(y)$, with y is in the causal past of x.

Let us briefly consider a class of operators which satisfy all the aforementioned axioms. We shall let Λ denote the nonlocality energy scale and define

$$\Lambda^{-2}(\widetilde{\Box}\phi)(x) = a\phi(x) + \Lambda^4 \int_{J^-(x)} f(\Lambda^2 \tau_{xy}^2)\phi(y) d^4y, \tag{4.9}$$

where a is a dimensionless real number, $J^-(x)$ denotes the causal past of x, and τ_{xy} is the Lorentzian distance between x and y:

$$\tau_{xy}^2 = (x^0 - y^0)^2 - |\mathbf{x} - \mathbf{y}|^2. \tag{4.10}$$

Examples of such operators have arisen in the causal set theory program [32, 33, 34, 35, 36] (see the previous Chapter). This operator is clearly linear, real, Poincare-invariant and retarded. It is shown in Appendix C.1 that there are choices of a and f for which $\widetilde{\Box}$ is also stable and has the desired infrared behaviour (4.8). One such choice is

$$f(s) = \frac{4}{\pi}\delta(s - \epsilon) - \frac{e^{-s/2}}{4\pi}(24 - 12s + s^2), \qquad a = -2, \tag{4.11}$$

where ϵ is an infinitesimally small positive number.

The eigenvalues $B(p)$ of $\widetilde{\Box}$ take the form (see [33])

$$\Lambda^{-2}B(p) = \lim_{\epsilon \to 0^+} g((p + ip_\epsilon) \cdot (p + ip_\epsilon)/\Lambda^2), \tag{4.12}$$

$$g(Z) = a + 4\pi Z^{-\frac{1}{2}} \int_0^\infty f(s^2) s^2 K_1(Z^{1/2}s) ds, \tag{4.13}$$

where p_ϵ is an infinitesimally small ($p_\epsilon \cdot p_\epsilon = -\epsilon^2$), timelike, and future-directed ($p_\epsilon^0 > 0$) wave-vector. The analytic structure of $B(p)$ is shown in Figure 4.1. Figure 4.2 shows the behaviour of $B(p)$ as a function of $p \cdot p$ and $\mathrm{sgn}(p^0)$ for the choice of f and a given in (4.11).

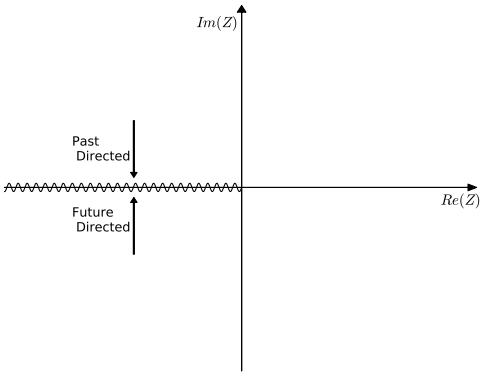


Figure 4.1: Analytic structure of $B(p)$ in the complex plane of $Z = p \cdot p/\Lambda^2$

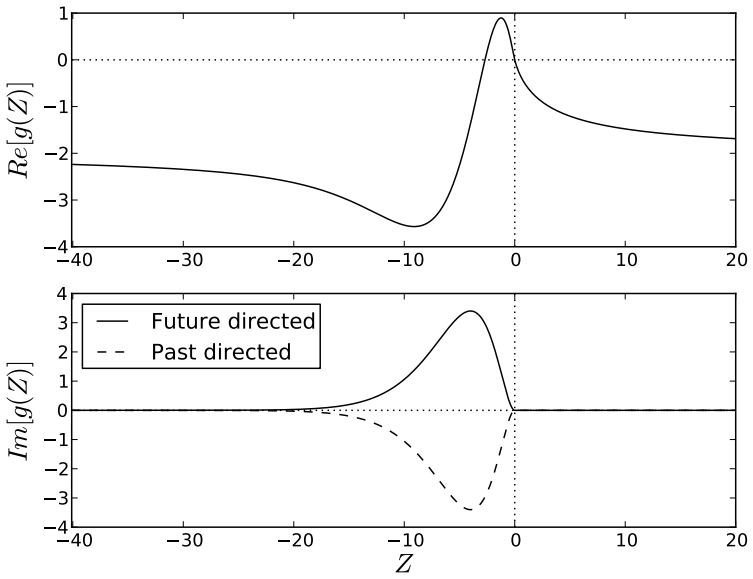


Figure 4.2: The Fourier transform $B(p) = g(p \cdot p/\Lambda^2)$ of $\widetilde{\Box}$ defined in (4.9), where a and f are given by (4.11).

4.3 Classical Theory

How would such non-local and retarded evolution manifest itself? To get a start on answering this question, we modify the evolution of a massless scalar field ϕ coupled to a source $J(x)$ via $\Box \to \widetilde{\Box}$:

$$\Box\phi(x) = J(x) \to \widetilde{\Box}\phi(x) = J(x). \tag{4.14}$$

It is worth noting that the solutions of $\widetilde{\Box}\phi(x) = 0$ are identical to those of $\Box\phi(x) = 0$. This follows from requiring a stable evolution for $\widetilde{\Box}$ (see [33]). As we will see in Section 4.3.2, however, the story changes when $J(x) \neq 0$.

4.3.1 Absence of an action principle

It is natural to ask whether an action principle exists for ϕ, whose variation would produce the non-local equation of motion $\widetilde{\Box}\phi(x) = J(x)$. One might propose to substitute $\Box$ with $\widetilde{\Box}$ in the action of a massless scalar field:

$$S[\phi] = \int d^4x \left(\frac{1}{2}\phi(x)\widetilde{\Box}\phi(x) - J(x)\phi(x)\right). \tag{4.15}$$

Requiring $S[\phi]$ to be stationary with respect to first order variations in ϕ we find [4]

$$\frac{1}{2}(\widetilde{\Box} + \widetilde{\Box}^T)\phi(x) = J(x), \tag{4.19}$$

where $\widetilde{\Box}^T$ is defined in Fourier space via

$$\widetilde{\Box}^T e^{ip\cdot x} = B(p)^* e^{ip\cdot x}. \tag{4.20}$$

[4] To see this, it is instructive to express the action in Fourier space. Define the Fourier transform $f(p)$ of $f(x)$ via

$$f(x) = \int \frac{d^4p}{(2\pi)^4} f(p) e^{ip\cdot x}. \tag{4.16}$$

Then, it can be shown that

$$S = \int \frac{d^4p}{(2\pi)^4}\left[\phi(p)^* \frac{1}{4}(B(p) + B(p)^*)\phi(p) - \phi(p)^* J(p)\right]. \tag{4.17}$$

Requiring S to be stationary with respect to first order variations $\phi(p)$ we find

$$\frac{1}{2}(B(p) + B(p)^*)\phi(p) = J(p). \tag{4.18}$$

In the case of the retarded operator (4.9), for instance, $\widetilde{\Box}^T\phi(x)$ is the right hand side of (4.9) with the domain of integration changed to the causal *future* of point x. Therefore, (4.15) does not lead to a retarded equation of motion.

Due to the absence of a local Lagrangian description, quantizing a massless scalar field theory built upon $\widetilde{\Box}$ is non-trivial. We shall address this problem in Section 4.4, where we argue that the the Schwinger-Keldysh quantization scheme can still be used to obtain the desired non-local quantum field theory.

4.3.2 Green's function

The Green's functions of $\Box$ and $\widetilde{\Box}$ are quite different, especially in the ultraviolet (UV) where their spectra differ. One important difference is that $\widetilde{\Box}$, unlike $\Box$, has a unique inverse. Since $\widetilde{\Box}$ is a retarded operator by definition, it only has a retarded Green's function. Recall that $\Box$ has both a retarded $G^R(x,y)$ and advanced $G^A(x,y)$ Green's function:

$$\Box_x G^{R,A}(x,y) = \delta^{(4)}(x-y), \tag{4.21}$$

which satisfy the following "boundary conditions": $G^R(x,y)$ vanishes unless $x \succ y$ (x is in the causal future of y), and $G^A(x,y)$ vanishes unless $y \succ x$. The two Green's functions are related to one another via $G^A(x,y) = G^R(y,x)$. In the case of $\widetilde{\Box}$, Green's function is unique (just the retarded one) and switching the arguments of the retarded Green's function *does not* produce another Green's function. Let us show why this is.

Let $\widetilde{G}(x,y)$ denote the Green's function associated with $\widetilde{\Box}$:

$$\widetilde{\Box}_x \widetilde{G}(x,y) = \delta^{(4)}(x-y), \tag{4.22}$$

Note that $\widetilde{G}(x,y)$ can be expressed as

$$\widetilde{G}(x,y) = \int \frac{d^4p}{(2\pi)^4} \frac{1}{B(p)} e^{ip\cdot(x-y)}. \tag{4.23}$$

The path of integration in the complex p^0 plane is shown in Figure 4.3. This comes from the fact that $\widetilde{\Box}$ is a retarded operator, so $B(p)$ analytically continued to the complex p^0 plane takes its value above the cut. When $B(p)$ has no zeros in complex plane apart from at $p\cdot p = 0$, which is guaranteed by the stability requirement, this choice of contour ensures

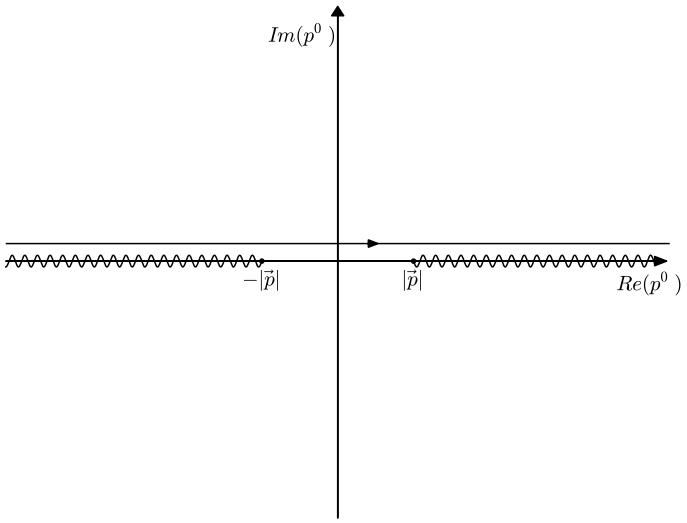


Figure 4.3: The integration path in the complex p^0 plane which defines the retarded Green's function associated with $\widetilde{\Box}$.

that $\widetilde{G}(x,y) \equiv \widetilde{G}^R(x,y)$ is indeed retarded. Switching the arguments of $\widetilde{G}^R(x,y)$, we find

$$\widetilde{G}^R(y,x) = \int \frac{d^4p}{(2\pi)^4} \frac{1}{B(p)} e^{ip\cdot(y-x)} \tag{4.24}$$

$$= \int \frac{d^4p}{(2\pi)^4} \frac{1}{B(-p)} e^{ip\cdot(x-y)} \tag{4.25}$$

$$= \int \frac{d^4p}{(2\pi)^4} \frac{1}{B(p)^*} e^{ip\cdot(x-y)}, \tag{4.26}$$

where in the second line we have changed integration variables from p to $-p$. Then

$$\widetilde{\Box}_x \widetilde{G}^R(y,x) = \int \frac{d^4p}{(2\pi)^4} \frac{B(p)}{B(p)^*} e^{ip\cdot(x-y)} \neq \delta^{(4)}(x-y), \tag{4.27}$$

since $B(p)$ is generically complex. As we will see in the sections to come, the fact that $\widetilde{\Box}$ has a unique inverse plays a crucial role in the quantum theory of $\widetilde{\Box}$.

4.4 Quantum Theory

We wish to construct a quantum theory of a massless scalar field ϕ whose classical limit reproduces the retarded evolution induced by $\widetilde{\Box}$. The quantization scheme which we believe is most suited in this case is the Schwinger-Keldysh (or double path integral) formalism. In what follows, we will first review the usual paths to quantization (i.e. canonical quantization and the Feynman path integral) and show why they fail in the case of a non-local and retarded operator like $\widetilde{\Box}$. The goal of these discussions is to make clear why we choose the Schwinger-Keldysh formalism to construct a quantum field theory based on $\widetilde{\Box}$.

4.4.1 Canonical quantization

Let us consider the canonical quantization of a free massless scalar field ϕ. The typical route to quantization is as follows: start from an action principal for ϕ, derive the Hamiltonian in terms of ϕ and its conjugate momentum, impose equal-time commutation relations, and finally specify the dynamics via the Heisenberg equation. There is an equivalent approach, however, which defines the theory with no reference to an action principle, using the Klein-Gordon equation supplemented by the so-called Peierls form of the commutation relations:

$$\Box\widehat{\phi}(x) = 0 \tag{4.28}$$

$$[\widehat{\phi}(x), \widehat{\phi}(y)] = i\Delta(x,y), \tag{4.29}$$

where $\Delta(x,y)$ is the Pauli-Jordan function:

$$\begin{aligned}\Delta(x,y) &= G^R(x,y) - G^A(x,y) \\ &= G^R(x,y) - G^R(y,x).\end{aligned} \tag{4.30}$$

It is well known that (4.29) is entirely equivalent to, but more explicitly covariant than, the more commonly seen equal-time commutation relations (see e.g. Section C.2 of [16]). Since $\Delta(x,y)$ is the difference of two Green's functions, it satisfies the equation of motion:

$$\Box_x\Delta(x,y) = 0. \tag{4.31}$$

This is why (4.28) and (4.29) are consistent with one another: both the left and right hand side of (4.29) vanish when $\Box_x$ is applied.

It is tempting to build the quantum theory of $\widetilde{\Box}$ in a similar fashion:

$$\widetilde{\Box}\widehat{\phi}(x) = 0 \tag{4.32}$$

$$[\widehat{\phi}(x), \widehat{\phi}(y)] = i\widetilde{\Delta}(x,y) \equiv i(\widetilde{G}^R(x,y) - \widetilde{G}^R(y,x)). \tag{4.33}$$

In this case, however, $\widetilde{\Delta}(x,y)$ does not satisfy the equation of motion ($\widetilde{\Box}_x\widetilde{\Delta}(x,y) \neq 0$) because $\widetilde{G}^R(y,x)$ is not a Green's function of $\widetilde{\Box}$ (see Section 4.3 and (4.27)). Therefore, the equation of motion (4.32) is not consistent with the commutation relations (4.33).

It is worth noting that the root of this inconsistency is that the Fourier transform $B(p)$ of $\widetilde{\Box}$ is complex, which in turn follows from the fact that $\widetilde{\Box}$ is retarded by definition. In Section 4.4.3 we will arrive at a consistent quantum theory via the Schwinger-Keldysh formalism, using which we also build a Hilbert space representation of the theory. There we will see that the equation of motion (4.32) is given up in favour of the commutation relations (4.33). As it turns out, the degree to which (4.32) is violated depends on the imaginary part of $B(p)$.

4.4.2 Feynman path integral

The Feynman path integral formalism requires a local Lagrangian description for the scalar field ϕ. As was argued in Section 4.3.1, however, this is not viable if one requires a retarded equation of motion. Therefore, the Feynman path integral formalism is also not suitable for quantizing this theory.

4.4.3 Schwinger-Keldysh formalism

The Schwinger-Keldysh formalism has a natural way of incorporating a retarded operator. In this approach an amplitude (called the decoherence functional $\mathcal{D}(\phi^+,\phi^-)$) is assigned to a pair of paths (ϕ^+,ϕ^-), which are constrained to meet at the final time ($\phi^+(t_f,\mathbf{x}) = \phi^-(t_f,\mathbf{x})$). The decoherence functional for a free massless scalar field takes the form

$$\mathcal{D}(\phi^+,\phi^-) = \text{Exp}\left[i\int d^4x \frac{1}{2}\phi^q\Box^R\phi^{cl} + \frac{1}{2}\phi^{cl}\Box^A\phi^q + \frac{1}{2}\phi^q\Box^K\phi^q\right], \tag{4.34}$$

where

$$\phi^{cl} \equiv \frac{1}{\sqrt{2}}\left(\phi^+ + \phi^-\right), \tag{4.35}$$

$$\phi^q \equiv \frac{1}{\sqrt{2}}\left(\phi^+ - \phi^-\right). \tag{4.36}$$

In (4.34), $\Box^R$ is the retarded d'Alembertian, $\Box^A = (\Box^R)^\dagger$ is the advanced d'Alembertian, and $\Box^K$ is an anti-Hermitian operator which contains information about the initial wave

function [39].[5] Any source term $J(x)$ can be included by adding $-J\phi^+ + J\phi^- = -\sqrt{2}J\phi^q$ to the integrand.

Any n-point function in this theory is given by

$$\begin{aligned}&\langle\ \phi^{(\alpha_1)}(x_1)\cdots\phi^{(\alpha_n)}(x_n)\rangle\\&\quad=\int D\phi^+D\phi^-\phi^{(\alpha_1)}(x_1)\cdots\phi^{(\alpha_n)}(x_n)\mathcal{D}(\phi^+,\phi^-),\end{aligned}\tag{4.37}$$

where $\alpha_i\in\{+,-,q,cl\}$. These correlation functions are related to the correlation functions in Hilbert space representation by the following rule:

$$\begin{aligned}&\langle\ \phi^+(x_1)\cdots\phi^+(x_n)\phi^-(y_1)\cdots\phi^-(y_m)\rangle\\&\quad=\langle 0|\widetilde{T}\left[\widehat{\phi}(y_1)\cdots\widehat{\phi}(y_m)\right]T\left[\widehat{\phi}(x_1)\cdots\widehat{\phi}(x_n)\right]|0\rangle\end{aligned}\tag{4.38}$$

where $T(\widetilde{T})$ is the (anti) time-ordered operator, and $|0\rangle$ is the vacuum state of the free theory.

In order to come up with a quantum theory for a non-local retarded operator, we replace $\Box^R$ with $\widetilde{\Box}$ in (4.34) (and $\Box^K$ with $\widetilde{\Box}^K$).[6]

Classical limit

Before going any further, let us take a look at the classical limit of this theory. Performing Gaussian integrals (in the presence of a source term), we get

$$\langle\phi^{cl}(x)\rangle=\frac{1}{\sqrt{2}}\int d^4y\ \widetilde{G}^R(x,y)J(y),\tag{4.39}$$

$$\langle\phi^q(x)\rangle=0,\tag{4.40}$$

resulting in

$$\langle\phi^+(x)\rangle=\langle\phi^-(x)\rangle=\int d^4y\ \widetilde{G}^R(x,y)J(y).\tag{4.41}$$

It shows that in the classical limit where the field is represented by its expectation value, there is no difference between ϕ^+ and ϕ^- and both satisfy the retarded equation of motion $\widetilde{\Box}\phi=J$.

[5] The retarded and advanced d'Alembertians are defined via $G^{R,A}(\Box^{R,A}f)=f$ for all suitable test functions f, where $G^{R,A}$ are the integral operators associated with the retarded and advanced Green's functions $G^{R,A}(x,y)$.

[6] We still need to determine $\widetilde{\Box}^K$. This has been done in 4.4.3.

Green's functions

Let us consider the two point correlation functions of this theory in the absence of any source

$$-i\ \langle\phi^{cl}(x)\phi^{q}(y)\rangle = \widetilde{G}^{R}(x,y) \tag{4.42}$$

$$-i\ \langle\phi^{q}(x)\phi^{cl}(y)\rangle \equiv \widetilde{G}^{A}(x,y) = \widetilde{G}^{R}(y,x) \tag{4.43}$$

$$\begin{aligned} -i\ \langle\phi^{cl}(x)\phi^{cl}(y)\rangle &\equiv \widetilde{G}^{K}(x,y) \\ &= -\int d^4z d^4w\ \widetilde{G}^{R}(x,z)\widetilde{B}^{K}(z,w)\widetilde{G}^{A}(w,y) \end{aligned} \tag{4.44}$$

$$-i\ \langle\phi^{q}(x)\phi^{q}(y)\rangle = 0 \tag{4.45}$$

where $\widetilde{B}^{K}(x,y)$ is the kernel of $\widetilde{\Box}^{K}$. [7] Using the definition of ϕ^{q} and ϕ^{cl}, we get

$$-i\langle\phi^{+}(x)\phi^{+}(y)\rangle = \frac{1}{2}\left[\widetilde{G}^{K}(x,y) + \widetilde{G}^{R}(x,y) + \widetilde{G}^{A}(x,y)\right], \tag{4.46}$$

$$-i\langle\phi^{-}(x)\phi^{-}(y)\rangle = \frac{1}{2}\left[\widetilde{G}^{K}(x,y) - \widetilde{G}^{R}(x,y) - \widetilde{G}^{A}(x,y)\right], \tag{4.47}$$

$$-i\langle\phi^{-}(x)\phi^{+}(y)\rangle = \frac{1}{2}\left[\widetilde{G}^{K}(x,y) + \widetilde{G}^{R}(x,y) - \widetilde{G}^{A}(x,y)\right]. \tag{4.48}$$

Note that if this theory has an equivalent representation in terms of field operator in a Hilbert space, then the above mentioned terms correspond to time-ordered two point function, anti time-ordered two point function and two point function respectively (see (4.38)).

We require that the theory describes a free scalar field in flat space-time at its ground state. As a result, all n-point correlation functions of this theory must be translation invariant,

$$\langle\phi^{(\alpha_1)}(x_1)\cdots\phi^{(\alpha_n)}(x_n)\rangle = \langle\phi^{(\alpha_1)}(x_1+y)\cdots\phi^{(\alpha_n)}(x_n+y)\rangle. \tag{4.49}$$

This condition requires that all operators $\widetilde{\Box}$, $\widetilde{\Box}^{\dagger}$ and $\widetilde{\Box}^{K}$ must be translation invariant. Consequently, we get

$$\widetilde{\Box}^{K}e^{ip\cdot x} = \widetilde{B}^{K}(p)e^{ip\cdot x}, \tag{4.50}$$

$$\widetilde{G}^{K}(x,y) = -\int \frac{d^4p}{(2\pi)^4}\widetilde{G}^{R}(p)\widetilde{B}^{K}(p)\widetilde{G}^{A}(p)e^{ip\cdot(x-y)} \tag{4.51}$$

[7] If $\delta_y(x) \equiv \delta^{(4)}(x-y)$, then $\widetilde{B}^{K}(x,y) \equiv (\widetilde{\Box}^{K}\delta_y)(x)$. With this definition, $(\widetilde{\Box}^{K}\phi)(x) = \int d^4y\widetilde{B}^{K}(x,y)\phi(y)$.

Note that $\widetilde{\Box}^K$ is an anti-Hermitian operator. It means $\widetilde{B}^K(p)$ is a total imaginary number (and $\widetilde{G}^K(p) \equiv -\widetilde{G}^R(p)\widetilde{B}^K(p)\widetilde{G}^A(p)$ is also total imaginary since $\widetilde{G}^R(p)\widetilde{G}^A(p)$ is real.)

Fixing $\widetilde{G}^K$

From here on, we assume that there is a Hilbert space representation of this theory with a Hamiltonian evolution. We will justify this assumption later by finding the representation itself. In Appendix C.2 we show that this assumption leads to the following relation, when the quantum system is in its ground state

$$\widetilde{G}^K(p) = \mathrm{sgn}(p^0)\left[\widetilde{G}^R(p) - \widetilde{G}^A(p)\right]. \tag{4.52}$$

Note that (4.52) is nothing but the fluctuation dissipation theorem (FDT) at zero temperature. This fixes the eigenvalues of $\widetilde{\Box}^K$ as follows:

$$\widetilde{B}^K(p) = 2i\mathrm{Im}B(p)\mathrm{sgn}(\mathrm{p}^0) \tag{4.53}$$

Hilbert space representation

We wish to find an equivalent Hilbert space representation in terms of a field operator $\widehat{\phi}(x)$ for this theory. As we mentioned earlier, (4.48) is the two point function of such a representation,

$$W(x,y) \equiv \langle 0|\widehat{\phi}(x)\widehat{\phi}(y)|0\rangle = \langle \phi^-(x)\phi^+(y)\rangle, \tag{4.54}$$

where $|0\rangle$ is the ground state. If we use (4.48) and (4.52), we arrive at

$$W(x,y) = \int \frac{d^4p}{(2\pi)^4}\frac{2\mathrm{Im}[B(p)]\theta(p^0)}{|B(p)|^2}e^{ip\cdot(x-y)}, \tag{4.55}$$

where we call $\widetilde{W}(p) \equiv \frac{2\mathrm{Im}[B(p)]\theta(p^0)}{|B(p)|^2}$. Since $W(x,y)$ is a positive operator, $\mathrm{Im}[B(p)]\theta(p^0)$ must be a non-negative number. So, we further *assume*

$$\mathrm{sgn}\left(\mathrm{Im}[B(p)]\right) = \mathrm{sgn}(p^0). \tag{4.56}$$

Once this condition is satisfied, the field operator $\widehat{\phi}(x)$ and ground state $|0\rangle$, defined to be

$$\widehat{\phi}(x) = \int \frac{d^4p}{(2\pi)^2}\sqrt{\widetilde{W}(p)}\left(\widehat{a}_p e^{ip\cdot x} + \widehat{a}_p^\dagger e^{-ip\cdot x}\right), \tag{4.57}$$

$$[\widehat{a}_p, \widehat{a}_q] = \delta^{(4)}(p-q), \tag{4.58}$$

$$\widehat{a}_p|0\rangle = 0 \quad \forall p, \tag{4.59}$$

yield the desired correlation functions.

Note that a_p is only defined for time-like future-directed p, because otherwise $\widetilde{W}(p)$ is zero in the field expansion. It means that all time-like future-directed (positive energy) momenta contribute to the field expansion (4.57).

Hamiltonian

By definition, time evolution operator is the operator that evolves $\widehat{\phi}(x)$ in time,

$$\widehat{\phi}(t,\mathbf{x}) = \widehat{U}(t,t_0)\widehat{\phi}(t_0,\mathbf{x})\widehat{U}^\dagger(t,t_0). \tag{4.60}$$

It can be directly checked that

$$\widehat{U}(t,t_0) = e^{-i\widehat{H}_0(t-t_0)}, \tag{4.61}$$

$$\widehat{H}_0 = \int d^4p\ p^0 \widehat{a}_p^\dagger \widehat{a}_p, \tag{4.62}$$

gives the right time evolution.

State $|0\rangle$ defined in (4.59) is the *ground state* of this Hamiltonian. Excited states (n-particle states) can be built by acting $a^\dagger$'s on $|0\rangle$,

$$|p_1 \cdots p_n\rangle = \widehat{a}_{p_1}^\dagger \cdots \widehat{a}_{p_n}^\dagger |0\rangle. \tag{4.63}$$

The excited state $|p\rangle$ represents a particle with energy p^0 and momentum $\mathbf{p}$[8] where p^0 is independent of $\mathbf{p}$[9]. This shows that the theory contains a continuum of massive particles with positive energy. The existence of a continuum of massive particles in the context of Causal Set theory also has been pointed out in [40], although their result is rather different in some other aspects.

Comparison to local evolution

At this point, it would be illustrative to consider the result of this formalism for LQFT. In this case

$$B(p) = B_{local}(p) = (p^0 + i\epsilon)^2 - |\mathbf{p}|^2, \tag{4.64}$$

[8]Momentum operator $\widehat{\mathbf{P}} \equiv \int d^4p\ \mathbf{p}\ \widehat{a}_p^\dagger \widehat{a}_p$ is the generator of spacial translation.

[9]Note that these states are different from the usual states $|\mathbf{p}\rangle$ used in LQFT which describe a particle with momentum $\mathbf{p}$ and energy $|\mathbf{p}|$.

where ϵ is a small positive number taken to zero at the end of calculation. The two point function is given by

$$\widetilde{W}(p) = 2\frac{\epsilon p^0}{(p^2)^2 + (\epsilon p^0)^2}\theta(p^0) = 2\pi\delta(p^2)\theta(p^0). \tag{4.65}$$

As a result,

$$W(x,y) = \int \frac{d^4p}{(2\pi)^4} 2\pi\delta(p^2)\theta(p^0)e^{ip\cdot(x-y)}, \tag{4.66}$$

$$\widehat{\phi}(x) = \int \frac{d^4p}{(2\pi)^2}\sqrt{2\pi\delta(p^2)\theta(p^0)}\left(\widehat{a}_p e^{ip\cdot x} + \widehat{a}_p^\dagger e^{-ip\cdot x}\right). \tag{4.67}$$

Two point function and field expansion are exactly the ones we expected. Only on-shell particles ($p\cdot p = 0$) contribute to the field expansion.

Here, we see one important difference between local and retarded non-local evolution. In the local case, only on-shell modes ($p\cdot p = 0$) contribute to the field expansion. As a result, excited states of the theory consist of all *on-shell* particles. In non-local retarded case (where generically $\mathrm{Im}[B(p)] \neq 0$), off-shell modes ($p\cdot p \neq 0$) also contribute to the field expansion. Consequently, one expect the existence of off-shell modes in "in" and "out" state of scatterings in the interacting theory.

Let us investigate properties of $\widetilde{W}(p)$ for a generic non-local retarded operator. First of all, it is only non-zero for time-like future-directed momenta. This means that only time-like future-directed momenta contribute to the field expansion and can exist in "in" and "out" state (particles with time-like momentum and positive energy).

Considering that $B(p)$ is only zero at $p\cdot p = 0$, $\widetilde{W}(p)$ is a finite number for all $p\cdot p \neq 0$ (we will see the significance of this result in 4.6.2). On the other hand, since in the subspace of on-shell modes $\widetilde{\Box}$ operator is exactly the same as $\Box$, we conclude that $\widetilde{W}(p) = 2\pi\delta(p^2)\theta(p^0)$ for $p\cdot p = 0$. Therefore, $\widetilde{W}(p)$ consists of a divergent part at $p\cdot p = 0$ and a finite part for $p\cdot p \neq 0$. This means that there are two different contributions to the field expansion (4.57), one from on-shell modes that is the same as (4.67) and one from off-shell modes which only exists in the case of non-local retarded evolution

$$\begin{aligned}\widehat{\phi}(x) \;&= \int \frac{d^4p}{(2\pi)^2}\sqrt{2\pi\delta(p^2)\theta(p^0)}\left(\widehat{a}_p e^{ip\cdot x} + \widehat{a}_p^\dagger e^{-ip\cdot x}\right)\\ &+ \int_{p^2\neq 0} \frac{d^4p}{(2\pi)^2}\sqrt{\widetilde{W}(p)}\left(\widehat{a}_p e^{ip\cdot x} + \widehat{a}_p^\dagger e^{-ip\cdot x}\right).\end{aligned} \tag{4.68}$$

4.4.4 Sorkin–Johnston quantization

The Sorkin-Johnston (SJ) proposal defines a unique vacuum state for a free massive scalar field in an arbitrarily curved spacetime [41]. This proposal is a continuum generalization of Johnston's formulation of a free quantum scalar field theory on a background causal set [42]. As is the case for $\widetilde{\Box}$, canonical quantization does not admit an obvious generalization for a causal set. The SJ quantization scheme uses only the retarded Green's function $G_R(x,y)$ to arrive at the quantum theory. Since $\widetilde{\Box}$ also admits a retarded Green's function, one can apply the SJ prescription to arrive at a free quantum field theory of the massless scalar field we have been considering. In what follows, we will show that the SJ proposal applied to $\widetilde{\Box}$ produces the same free quantum theory as the Schwinger Keldysh formalism, provided condition (4.56) is met.

Consider the corresponding integral operator of the kernel $i\Delta(x,y) = G_R(x,y) - G_R(y,x)$:

$$(i\Delta f)(x) = \int i\Delta(x,y) f(y) d^4y. \tag{4.69}$$

It can be shown that $i\Delta$ is Hermitian, which implies it has real eigenvalues, and that its non-zero eigenvalues come in positive and negative pairs:

$$(i\Delta T_{\mathbf{p}})(x) = \lambda^2_{\mathbf{p}} T_{\mathbf{p}}(x) \qquad \rightarrow \qquad (i\Delta T^*_{\mathbf{p}})(x) = -\lambda^2_{\mathbf{p}} T^*_{\mathbf{p}}(x). \tag{4.70}$$

We have assumed here that the eigenfunctions $T_{\mathbf{p}}$ form an orthonormal basis of L^2, which can always be achieved since $i\Delta$ is Hermitian. The Sorkin-Johnston proposal is then to define the two-point function to be the positive part of $i\Delta(x,y)$ in the following sense:

$$\langle 0|\widehat{\phi}(x)\widehat{\phi}(y)|0\rangle = \sum_{\mathbf{p}} \lambda^2_{\mathbf{p}} T_{\mathbf{p}}(x) T^*_{\mathbf{p}}(y). \tag{4.71}$$

Taking $G_R(x,y)$ to be the retarded Green's function of $\widetilde{\Box}$ (see (4.23) and (4.26)), we find

$$i\Delta e^{ip\cdot x} = \frac{2\mathrm{Im}(B(p))}{|B(p)|^2} e^{ip\cdot x}, \tag{4.72}$$

which using the SJ formalism then leads to the two-point function

$$\langle 0|\widehat{\phi}(x)\widehat{\phi}(y)|0\rangle = \int \frac{d^4p}{(2\pi)^4} \frac{2\mathrm{Im}(B(p))}{|B(p)|^2} \theta(\mathrm{Im}(B(p))) e^{ip\cdot x}. \tag{4.73}$$

If condition (4.56) is satisfied, this two-point function is at that derived from the Schwinger-Keldysh formalism (see (4.55) and (4.56)). It is reassuring that two different paths to quantization, at least at the free level, lead to the same theory.

4.5 Interacting Field Theory

Let us now consider the interacting theory. We introduce the interaction in the Hilbert space representation by adding a potential term to the free Hamiltonian as follows:

$$\widehat{H}(t) = \widehat{H}_0 + \int d^3\mathbf{x} V(\widehat{\phi}(t,\mathbf{x})). \tag{4.74}$$

Starting with a general initial wave function, one is able to find the final state of the system by solving Heisenberg equation of motion in principle. However, in practice this is a very hard task to do. So, we try to find the S-matrix amplitudes perturbatively.

In order to do so, we can use the available machinery of LQFT, and move to the interaction picture. Time evolution in the interaction picture is given by

$$\widehat{U}_I = Te^{-i\int d^4x V(\widehat{\phi}_I)} \tag{4.75}$$

where $\widehat{\phi}_I$ is the field in the interaction picture given by (4.57). Perturbative expansion of $\widehat{U}_I$ yields S-matrix amplitudes. Performing the calculations to find the S-matrix, we come up with modified Feynman rules for this theory. We explain these modifications in the following two examples.

4.5.1 Example 1: 2-2 Scattering $p_1p_2 \to q_1q_2$ in $\frac{\lambda}{4!}\phi^4$ theory

Scattering amplitude $S_{q_1q_2,p_1p_2}$ is given by

$$S_{q_1q_2,p_1p_2} = \langle q_1q_2|Te^{-i\int d^4x \frac{\lambda}{4!}\widehat{\phi}_I^4}|p_1p_2\rangle. \tag{4.76}$$

To first order in λ, it yields

$$\begin{aligned} S_{q_1q_2,p_1p_2} &= -i\frac{\lambda}{4!}\int d^4x \ \langle q_1q_2|\widehat{\phi}_I^4(x)|p_1p_2\rangle \\ &= \frac{-i\lambda}{(2\pi)^4}\sqrt{\widetilde{W}(p_1)\widetilde{W}(p_2)\widetilde{W}(q_1)\widetilde{W}(q_2)}\delta^{(4)}(\sum p - \sum q), \end{aligned} \tag{4.77}$$

where we have substituted for $\widehat{\phi}_I$ from (4.57). It is interesting to note that (4.77) is time reversal invariant.

In the transition from local to retarded non-local propagation, here we see the first change in the scattering amplitudes. The values assigned to each external line have changed from $\sqrt{2\pi\delta(p^2)\theta(p^0)}$ to $\sqrt{\widetilde{W}(p)}$. Note that here the scattering amplitude is computed in the basis of 4-momentum $|p\rangle$ which is different from 3-momentum basis $|\mathbf{p}\rangle$ of LQFT.

4.5.2 Example 2: 2-2 Scattering $p_1p_2 \to q_1q_2$ in $\frac{\lambda}{3!}\phi^3$ theory

In this case, $S_{q_1q_2,p_1p_2}$ is given by

$$S_{q_1q_2,p_1p_2} = \langle q_1q_2|Te^{-i\int d^4x \frac{\lambda}{3!}\widehat{\phi}_I^3}|p_1p_2\rangle. \tag{4.78}$$

To second order in λ, it yields

$$\begin{aligned} S_{q_1q_2,p_1p_2} &= \frac{1}{2}(\frac{-i\lambda}{3!})^2 \int d^4x\ d^4y\ \langle q_1q_2|T\widehat{\phi}_I^3(x)\widehat{\phi}_I^3(y)|p_1p_2\rangle \\ &= \frac{-i\lambda^2}{(2\pi)^8}\sqrt{\widetilde{W}(p_1)\widetilde{W}(p_2)\widetilde{W}(q_1)\widetilde{W}(q_2)}\delta^{(4)}(\sum p - \sum q) \\ &\quad \times \left[\widetilde{G}^F(p_1+p_2) + \widetilde{G}^F(p_1-q_1) + \widetilde{G}^F(p_1-q_2)\right]. \end{aligned}$$

$\widetilde{G}^F(p) = \frac{\theta(p^0)}{B(p)} + \frac{\theta(-p^0)}{B^*(p)}$ is the time-ordered two point function (4.46) in Fourier space. In the transition from local to non-local operator, here we see another change in the scattering amplitude. The values assigned to each internal line have changed to the new value for the Feynman propagator $\widetilde{G}_F(p)$.

From these examples, it is obvious how scattering amplitudes can be computed in this theory. For any Feynman diagram only the values assigned to external lines and internal lines have changed. Note that the amplitude of some diagrams in LQFT is zero, as a result of energy-momentum conservation, while in this theory they are not. For example in LQFT $\lambda\phi^3$ theory, the amplitude assigned to diagram 4.4 is zero, because the sum of two (non-parallel) null vectors cannot be a null vector. However, in this theory there is a continuum of massive particles, and for example two on-shell particles can interact and produce one off-shell particle.

4.6 From Scattering Amplitude to Transition Rate

At this point, we want to find the rate of a process using the S-matrix amplitudes. In 4.6.2 we have shown that if one (or more) of the incoming particles is off-shell, then the differential transition rate of such scattering is zero. It means that in order to have a non-zero transition rate (and cross-section), all of the incoming particles must be on-shell. This is the most distinctive property of off-shell particles: cross-section of any scattering with off-shell particles is zero.

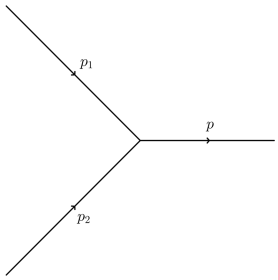


Figure 4.4: The amplitude of this diagram in LQFT is zero, because of the energy momentum conservation; two massless particles cannot produce a massless particle. However, in our theory there is a continuum of massive particles and the amplitude of this scattering is generically non-zero.

For now consider the scattering from state $|\alpha\rangle = |p_1 \cdots p_{N_\alpha}\rangle$ to $|\beta\rangle = |q_1 \cdots q_{N_\beta}\rangle$ where all the incoming particles are on-shell, $p_i^2 = 0$. Assuming that the interactions happen inside a box with volume V (see [43]), differential transition rate is given by

$$\begin{aligned} d\Gamma = \; & 2\pi^{N_\alpha+1}\left[\frac{(2\pi)^3}{V}\right]^{N_\alpha-1}\frac{1}{E_{\mathbf{p_1}}\cdots E_{\mathbf{p_{N_\alpha}}}}\delta^{(4)}(\sum p_i - \sum q_i) \\ & \times|\widetilde{M}_{\beta\alpha}|^2 d^4q_1\cdots d^4q_{N_\beta}, \end{aligned} \tag{4.79}$$

where $E_{\mathbf{p_i}} = |\mathbf{p_i}|$ and

$$S_{\beta\alpha} = -2\pi i\delta^{(4)}(\sum p_i - \sum q_i)\sqrt{\widetilde{W}(p_1)\cdots\widetilde{W}(p_{N_\alpha})}\widetilde{M}_{\beta\alpha}. \tag{4.80}$$

In the case of 2-2 scattering, the differential cross section is given by

$$d\sigma = \frac{d\Gamma}{\frac{u}{V}} = \frac{\pi^2(2\pi)^4}{E_{\mathbf{p_1}}E_{\mathbf{p_2}}u}\delta^{(4)}(\sum p_i - \sum q_i)|\widetilde{M}_{\beta\alpha}|^2 d^4q_1 d^4q_2, \tag{4.81}$$

where

$$u = \frac{\sqrt{(p_1.p_2)^2 - p_1^2p_2^2}}{p_1^0p_2^0} \tag{4.82}$$

is the speed of particle 1 in the frame of reference of particle 2 (and vice versa) and $\frac{u}{V}$ is the flux of incoming particles.

4.6.1 $p_1p_2 \to q_1q_2$ cross section in $\frac{\lambda}{4!}\phi^4$

As an example, we will find the cross section of $p_1p_2 \to q_1q_2$ where $p_i^2 = 0$. Using (4.77) and the definition (4.80), to first order in λ

$$\widetilde{M} = \frac{\lambda}{(2\pi)^5}\sqrt{\widetilde{W}(q_1)\widetilde{W}(q_2)}. \tag{4.83}$$

As a result, cross section is given by

$$d\sigma = \frac{\lambda^2}{4\,(2\pi)^4\,|p_1 \cdot p_2|}\widetilde{W}(q_1)\widetilde{W}(q_2)\delta^{(4)}(p_1 + p_2 - q_1 - q_2)d^4q_1 d^4q_2. \tag{4.84}$$

Let us constraint the outgoing particles to be only on-shell $q_i^2 = 0$. In this case $\widetilde{W}$ functions in (4.84) pick up a delta function and one can check that (4.84) for outgoing on-shell particles results in the usual cross section of $\lambda\phi^4$ in LQFT. However, if we constraint (at least) one of the outgoing particles to be off-shell with a *fixed* mass, the cross section becomes zero. Cross section over outgoing off-shell particles is only non-zero when the integration over continuum mass is also performed. We see the significance of this in the next section when considering the scattering of off-shell particles. Due to the contribution of off-shell states, the total cross section (4.84) is increased compared to the local theory.

4.6.2 Off-shell particles and cross section

In order to calculate the cross section of any scattering involving incoming off-shell particles, we make use of the fact that off-shell particles can be thought as a continuum of massive particles.

This can be done by expressing the two-point function as a sum over massive two point functions:

$$W(x,y) = \int_0^\infty d\mu^2\rho(\mu^2)\int\frac{d^4p}{(2\pi)^4}2\pi\theta(p^0)\delta(p^2+\mu^2)e^{ip\cdot(x-y)}, \tag{4.85}$$

where $\rho(-p^2) = \frac{\widetilde{W}(p)}{2\pi}$ for $p^0 > 0$. Note from (4.68) that $\rho(\mu^2) = \delta(\mu^2) + \widetilde{\rho}(\mu^2)$ where $\widetilde{\rho}$ is a finite function. In other words,

$$\begin{aligned} W(x,y) &= \int\frac{d^4p}{(2\pi)^4}2\pi\theta(p^0)\delta(p^2)e^{ip\cdot(x-y)} \\ &+ \int_0^\infty d\mu^2\widetilde{\rho}(\mu^2)\int\frac{d^4p}{(2\pi)^4}2\pi\theta(p^0)\delta(p^2+\mu^2)e^{ip\cdot(x-y)}. \end{aligned} \tag{4.86}$$

In order to make everything more similar to LQFT, we discretize the mass parameter to get

$$W(x,y) = \int \frac{d^4p}{(2\pi)^4} 2\pi\theta(p^0)\delta(p^2)e^{ip\cdot(x-y)} \\ + \sum_{j=1}^{\infty} \Delta\mu^2 \widetilde{\rho}(\mu_j^2) \int \frac{d^4p}{(2\pi)^4} 2\pi\theta(p^0)\delta(p^2+\mu_j^2)e^{ip\cdot(x-y)}, \tag{4.87}$$

where $\mu_j^2 = j\Delta\mu^2$. (4.87) is the same as (4.86) in the limit $\Delta\mu^2 \to 0$.

The following field operator will yield the above two point function

$$\widehat{\phi}(x) = \int \frac{d^3\mathbf{p}}{(2\pi)^{3/2}} \frac{1}{\sqrt{2|\mathbf{p}|}} \left(\widehat{a}_{\mathbf{p},0} e^{ip\cdot x} + c.c\right)\big|_{p^0=|\mathbf{p}|} \\ + \sum_{j=1}^{\infty} \sqrt{\Delta\mu^2\widetilde{\rho}(\mu_j^2)} \int \frac{d^3\mathbf{p}}{(2\pi)^{3/2}\sqrt{2E_{\mathbf{p},\mu_j}}} \left(\widehat{a}_{\mathbf{p},\mu_j} e^{ip\cdot x} + c.c\right)$$

where

$$E_{\mathbf{p},\mu} = \sqrt{\mathbf{p}^2+\mu^2} \tag{4.88}$$

$$\left[\widehat{a}_{\mathbf{p},\mu_i}, \widehat{a}^{\dagger}_{\mathbf{q},\mu_j}\right] = \delta^{(3)}(\mathbf{p}-\mathbf{q})\delta_{\mu_i,\mu_j} \tag{4.89}$$

$$\widehat{a}_{\mathbf{p},\mu}|0\rangle = 0 \tag{4.90}$$

and state $|\mathbf{p},\mu\rangle \equiv \widehat{a}^{\dagger}_{\mathbf{p},\mu}|0\rangle$ is a one particle state with momentum $\mathbf{p}$, mass μ and energy $E_{\mathbf{p},\mu}$.

From now on, consider a concrete example of 2-2 scattering with $\frac{\lambda}{4!}\widehat{\phi}^4$ interaction and incoming particles with definite mass and momentum. The idea behind this proof can be generalized to more complicated examples. Up to first order in λ

$$\langle\ \mathbf{p_1}, m_1; \mathbf{p_2}, m_2|\widehat{S}|\mathbf{q_1}, \mu_1; \mathbf{q_2}, \mu_2\rangle = \\ -\frac{i\lambda}{(2\pi)^2}\delta^{(4)}\left(\sum p - \sum q\right)\sqrt{\prod_{i=1}^{2} \frac{(\Delta\mu^2)^2\rho(\mu_i^2)\rho(m_i^2)}{4E_{\mathbf{q_i},\mu_i}E_{\mathbf{p_i},m_i}}}. \tag{4.91}$$

In (4.91), if any of the particles was on-shell (say $\mu_1 = 0$), we should set $\Delta\mu^2\rho(\mu_1^2) = 1$, otherwise ρ is replaced by $\widetilde{\rho}$.

The differential cross section is given by

$$d\sigma = (2\pi)^{-2} \ \lambda^2 \frac{(\Delta\mu^2)^4 \rho(\mu_1^2)\rho(\mu_2^2)\rho(m_1^2)\rho(m_2^2)}{16u E_{\mathbf{p_1},m_1} E_{\mathbf{p_2},m_2} E_{\mathbf{q_1},\mu_1} E_{\mathbf{q_2},\mu_2}} \\ \delta^{(4)}(p_1 + p_2 - q_1 - q_2) d^3\mathbf{p_1} d^3\mathbf{p_2}. \tag{4.92}$$

In order to get the total cross section, we should also sum over the mass parameter in the phase space of outgoing particles. In the (mass) continuum limit this means

$$\sum \Delta\mu^2 \rho(m_i^2) \to \int dm_i^2 \rho(m_i^2) \tag{4.93}$$

which absorbs two factor of $\Delta\mu^2$ in (4.92); however, there are two remaining factors of $\Delta\mu^2$. If the incoming particles (even one of them) are off-shell, since $\rho(\mu^2)$ is a finite number, in the limit $\Delta\mu^2 \to 0$, the cross section becomes zero. This means that the (total) transition rate of scattering with off-shell particles with *fixed* mass is zero. The cross section is only non-zero when both of the incoming particles are on-shell.

This is, in fact, consistent with what we have found in the previous section. There, we have shown that the transition rate of on-shell $\to$ off-shell is non-zero, only when the integration over mass of the off-shell particles is performed. In fact, scattering transition rate of on-shell particles to off-shell particles with fixed masses is zero. Since the theory is time reversal invariant, this suggests that the scattering transition rate of off-shell particles with fixed masses must be zero too; consistent with what we have found here.

This also means that an initial state with a suitable continuum superposition of off-shell masses can scatter into on-shell modes (time reverse of the process of on-shell scattering into off-shell). However, as we argue in the next Section, these states are fine-tuned and generally we do not expect to find the system in these superpositions.

4.6.3 Off-shell $\to$ on-shell scattering: continued

In the previous Section, we showed that the transition rate of scattering with off-shell particle(s) is zero. However, a suitable continuum superposition of off-shell particles can scatter non-trivially. In this section, we want to explain this point to a greater extent and argue that it is unlikely to find the system in these superpositions. We will not go through the detail of calculations since it is not essential to our argument in this secion.

We make use of the following toy model theory that mimics many properties of the proposed nonlocal theory:

$$\mathcal{L} = \frac{1}{2}\psi_0 \Box \psi_0 + \sum_{i=1}^{N} \frac{1}{2}\psi_{m_i}(\Box - m_i^2)\psi_{m_i} - \lambda\psi^4, \tag{4.94}$$
$$\psi \equiv \psi_0 + \sum_{i=1}^{N} \frac{g_i}{\sqrt{N}}\psi_{m_i}.$$

This is a theory of one massless scalar field (playing the role of on-shell modes) in addition to N massive scalar fields (playing the role of off-shell modes) and we are interested in $N \to \infty$ limit of the theory (λ and g_i's are coupling constants and do not scale with N). The advantage of working with this theory is that while its behaviour is very similar to the non-local theory, (4.94) is a local quantum field theory and possibly more comprehensible to the reader. The interaction term in (4.94) is designed in a way that interactions with massive (off-shell) fields are suppressed by a factor of $\sqrt{N}$ and in $N \to \infty$ limit their interactions become negligible. On the other hand, the number of off-shell fields goes to infinity. In what follows, we explain that this theory imitates many properties of off-shell and on-shell particles in the non-local theory.

First, let us define the following quantities: $\sigma^{\vec{p}_1\vec{p}_2}_{m_1m_2\to\mu_1\mu_2}$ is the scattering cross section of two particles with masses and momenta $m_1, \vec{p}_1$ and $m_2, \vec{p}_2$ into two particles with masses μ_1 and μ_2 (ψ_{μ_1} and ψ_{μ_2}) and $\sigma^{\vec{p}_1\vec{p}_2}_{m_1m_2}$ is the total scattering cross section of two particles with masses and momenta $m_1, \vec{p}_1$ and $m_2, \vec{p}_2$.

Consider the scattering of two ψ_0 particles into two final particles. If we restrict the two final particles to be massive (off-shell fields with *fixed* masses), then the scattering cross section in $N \to \infty$ limit goes to zero. However, if we sum over all massive final states (all off-shell particles), the total cross section is non-zero. In fact, for different final states the corresponding cross sections scales with N as follows:

$$\begin{aligned}
\sigma^{\vec{p}_1\vec{p}_2}_{00\to00} &\propto N^0, \\
\sigma^{\vec{p}_1\vec{p}_2}_{00\to0m} &\propto \frac{1}{N}, \quad m \neq 0, \\
\sigma^{\vec{p}_1\vec{p}_2}_{00\to m_1m_2} &\propto \frac{1}{N^2}, \quad m_1, m_2 \neq 0.
\end{aligned}$$

While the interactions with individual massive fields are suppressed, the number of massive states scales with N. In this way, the *total* scattering cross section of two initial massless particles into two massive final states, summed over all masses, is finite and non-zero (the same scaling works for scattering into one massless and one massive particles).

On the other hand, any scattering with (at least) one massive initial state result into zero cross section. For example, the following total scattering cross sections (summed over all final states) scale with N as

$$\sigma_{0m}^{\vec{p}_1\vec{p}_2} \propto \frac{1}{N}, \quad m \neq 0, \tag{4.95}$$

$$\sigma_{m_1m_2}^{\vec{p}_1\vec{p}_2} \propto \frac{1}{N^2}, \quad m_1, m_2 \neq 0, \tag{4.96}$$

and they vanish in $N \to \infty$ limit.

As we showed, massive particles in this theory (4.94) mimic the properties of off-shell states in the non-local theory; they can be produced by the scattering of massless states, while the reverse process (scattering of massive states into massless) does not happen.

However, the theory is (obviously) time reversal invariant and massive $\to$ massless scatterings must take place. This is indeed true, but as we demonstrate here the initial massive state that scatters non-trivially must be a superposition of different masses. Consider state γ, a superposition of M different masses, scatters off a massless particle. Then, the total transition probability $\Gamma_{0\gamma}$ scales as

$$\Gamma_{0\gamma} < A\frac{M}{N} \tag{4.97}$$

Where A have no dependence on M and N (see Appendix C.3 for proof). This transition probability is non-zero in $N \to \infty$ limit, only when M also scales with N.

So, massive $\to$ massless scattering indeed happens. However, the massive state that scatters non-trivially must be a superposition of (infinitely) many different masses and in this sense is fine-tuned. It is similar to an egg that smashes into pieces upon falling on the ground; the reverse process of pieces assembling an egg can in principle happen, but it is very unlikely.

In this sense, we expect the off-shell to on-shell scattering in the non-local theory to be negligible. In principle this transition can happen, but it is very implausible. The essence of our reasoning in this Section is based on thermodynamical arguments and although it is not a complete proof, we hope that we have provided enough evidence to show that off-shell $\to$ on-shell scattering is very unlikely. Definitely, further quantitative studies are needed to augment (or disprove) our claim. Perhaps, a good starting point is to consider the toy model theory (4.94), since it shares a lot of properties of the non-local theory.

4.7 Extension to Massive Scalar Fields

Throughout this Chapter, we only considered the modification of a massless scalar field. But what about massive scalar fields? One may suggest to replace $\Box$ with $\widetilde{\Box}$ in the equation of motion of a massive scalar field as follows

$$(\widetilde{\Box} - M^2)\phi(x) = J(x) \tag{4.98}$$

and follow similar steps of quantization. However, this method does not work. If M is a real number, then there is no mode satisfying (4.98) in the absence of J. In other words, there is no on-shell modes.

Another way is to choose M to be a complex number such that for a time-like future directed momentum p, $B(p) = M^2$. In this case, the mass of on-shell mode is given by $m^2 \equiv -p^2$. However, $\widetilde{\Box} - M^2$ is no longer a real operator and the solution to (4.98) generically cannot be real.

The extension to massive scalar fields can be done by considering the following observation. All of the properties in massless case can be read from the analytic structure of $B(p)$ in Figure 4.1. Massless modes are on-shell because there is a simple zero at $p^2 = 0$ and there are off-shell modes for time-like momenta because there is a cut for time-like momenta in 4.1.

In this way, the extension to massive case seems much simpler. $\Box - m^2$ must be replaced with $\widetilde{\Box}_m$ whose eigenvalues $B_m(p)$ satisfy the followings:

1. There is only one simple zero at $p^2 = -m^2$. Also $\lim_{p^2+m^2\to 0} \frac{B_m(p)}{p^2+m^2} = -1$ to get the correct local limit.

2. The cut must be only on momenta with higher masses $p^2 < -m^2$. Otherwise, in the quantum theory, there are off-shell modes with masses smaller than m which makes the on-shell mode unstable (on-shell modes can always decay into off-shell modes with less mass).

3. $\mathrm{Im}B_m(p) \geq 0$ for $p^0 > 0$.

Conditions 4 and 5 in Section 4.2 and (4.56) must be replaced by the above-mentioned conditions. One easy way to come up with such an operator is to make use of the existing operator $B(p)$ in the massless case, and consider it as a function of p^2 and $sgn(p^0)$. Then,

$$B_m(p) = B(p^2 + m^2, sgn(p^0)). \tag{4.99}$$

has all the desired properties (this also has been shown in [40]).

4.8 Conclusion

In this Chapter, we studied the physical consequences of a causal non-local evolution of a massless scalar field. We started by modifying the d'Alembertian to a causal non-local operator at high energies. Quantization of a free field showed that the field represents a continuum of massive particles. In fact, there were two sets of modes: on-shell modes (massless particles) and off-shell modes (massive particles).

The Feynman rules for the perturbative calculation of S-matrix amplitudes were discussed. The most important result (in our opinion) is the fact that the cross section of any scattering with off-shell particles is zero. This suggests that although these modes exist and probably can be detected by other means, there is no way of detecting them through scattering experiments. This property opens up the possibility that dark matter particles might be just the off-shell modes of known matter. Finally, we extended this formalism to massive scalar fields.

Throughout this Chapter we only considered scalar field theories, but how about other types of fields? Extension to other types of fields, such as vector field, is not as straightforward as for scalar fields. Incorporating gauge symmetry in the theory is another important issue. Whether causal Lorentzian evolution can be extended to vector fields (and other types fields rather than scalars) can be the subject of future studies.

In the next Chapter, we explore phenomenological implications of the proposed dark matter candidate in the context of cosmology.

Chapter 5

Off-shell Dark Matter:
A Cosmological relic of Quantum Gravity

5.1 Introduction

A vast range of observations in Astrophysics and Cosmology have now provided concrete evidence for the existence of cosmological cold dark matter (CDM), which appears to make up the majority of mass density in our universe (only second to the mysterious dark energy). Rotation curves of galaxies (e.g. [44]), gravitational lensing (e.g. [45]), and Cosmic Microwave Background (CMB) [1, 46] all indicate that General Relativity with ordinary (or known) matter is not consistent with observations. It is worth noting that, unlike dark energy, evidence for the existence of CDM ranges from cosmological to galactic (i.e. six orders of magnitude) in physical scale.

Since all the observational evidence for CDM is through its gravitational interactions, it has been tempting to explore a modification of Einstein gravity as a substitute (e.g. [47, 48, 49, 50]). However, given the range of observational data matched by CDM (in particular, the precision measurements of CMB anisotropy power spectrum [1, 46]) it has become nearly impossible to fit the data with any modified gravity alternative (which does not have an effective built-in dark matter component) [51].

As a result, the most popular approach has been to consider CDM as a new (beyond Standard Model) weakly interacting particle. There are strong evidences that CDM particle has to be (at most) weakly interacting with the Standard Model, as otherwise it should have been detected by now, through various astrophysical or terrestrial probes (see, e.g.

[52]). It also has to be sufficiently cold, as there is no evidence for a thermal cut-off in the cosmological matter power spectrum, down to sub-Mpc scales [53]. It is quite remarkable that a simple assumption of adding a non-relativistic (and non-interacting) dark matter is compatible with all the cosmological observations.

Here, we study a rather different approach which has been proposed in the previous Chapter, and we shall refer to as off-shell dark matter (OfDM). In this proposal, CDM originates from considering quantum gravitational effects on the evolution of fields. These effects manifest themselves through modifying the evolution law of quantum fields to a non-local evolution described by a causal non-local operator $\widetilde{\Box}$ which substitutes the role of d'Alembertian.

Let us outline some features of this model. First, this non-local modification results in the appearance of a new set of modes (or excitations) associated to each field. In fact, modification of a field with mass M leads to two sets of modes:

1. Modes with mass M, called on-shell.
2. A continuum of massive modes with mass higher than M, called off-shell.

We call the original mass of the field (M) "intrinsic mass". In other words, intrinsic mass is the mass of the on-shell modes (or the least value mass of the excitations).

The important property that differentiates these two sets of modes and points to the direction of dark matter is the following: *transition rate of any scattering including (even) one off-shell mode in the initial state is zero.* This property makes off-shell modes a natural candidate for CDM, simply because they cannot be detected through non-gravitational scattering experiments [54]. In fact, they can be produced by scattering of "on-shell" particles, but they do not scatter, annihilate or decay. As such, the only way to detect these particles is through their gravitational signatures.

In the next Section, we will review the important features of this model. Section 5.3 is dedicated to the production of OfDM in the context of inflation and reheating. We will discuss the effect of OfDM on matter power spectrum in Section 5.4. I, Section 5.5 concludes the paper.

5.2 Review of OfDM

Let us start this Section by the following question: If off-shell modes of matter can be produced by the scattering of on-shell modes, while the reverse does not happen, shouldn't

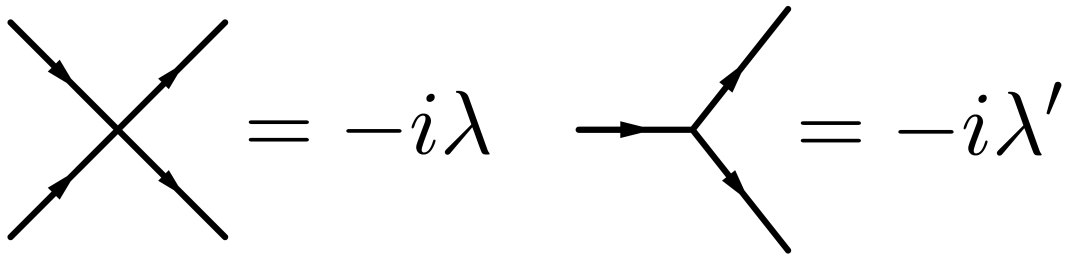


Figure 5.1: A simple annihilation process (on left) and decay process (on right).

we see any signature of this in scattering experiments, for example in Large Hadron Collider (LHC)? In other words, whenever we perform scattering experiments, a part of the incoming energy must transfer to off-shell modes and become undetectable. Shouldn't we have already seen this effect by now?

In order to answer this question, consider a simple annihilation or decay process (Figure 5.1). First, let us define the following quantities: σ_{1F} (Γ_{1F}) is the cross-section (rate) of producing one off-shell particle and one on-shell particle and σ_O (Γ_O) is the cross-section (rate) of producing purely on-shell particles. If we assume that the energy of the process is much higher than the intrinsic mass of the out states, $E_{\rm CM} \gg M$ (as we will see later, this is the relevant regime for dark matter production), following the results in [54], we arrive at[1]

$$\frac{\Gamma_{1F}}{\Gamma_O} = \frac{\sigma_{1F}}{\sigma_O} = \frac{\int d^4p_1 d^4p_2 2\pi\delta_+(p_1^2)\widetilde{W}(p_2)\delta^4(q-p_1-p_2)}{\int d^4p_1 d^4p_2 2\pi\delta_+(p_1^2)2\pi\delta_+(p_2^2)\delta^4(q-p_1-p_2)} \tag{5.1}$$

where q is the incoming energy-momentum and $\widetilde{W}(p)$ is given in terms of the spectrum of

[1] $\delta_+(p^2) \equiv \delta(p^2)\theta(p^0)$

non-local operator $\widetilde{\Box}$

$$\widetilde{W}(p) = \frac{2\text{Im } B(p)}{|B(p)|^2}\theta(p^0), \tag{5.2}$$

$$\widetilde{\Box} e^{ip\cdot x} = B(p) e^{ip\cdot x}. \tag{5.3}$$

Note that $\widetilde{W}(p)$ is the two point correlation function (or Wightman function) of the field in the momentum space

$$\langle 0|\hat{\psi}(x)\hat{\psi}(y)|0\rangle = \int \frac{d^4p}{(2\pi)^4}\widetilde{W}(p)e^{ip\cdot(x-y)} \tag{5.4}$$

Equation (5.1) can be simplified further if we assume that the energy scale of the scattering[2] $-q^2 \equiv E_{CM}^2$ is much lower than the non-locality scale Λ defined through $\widetilde{\Box}$. In this regime,

$$B(q) = -q^2 + \mathcal{O}\left(\frac{q^4}{\Lambda^2}\right) \tag{5.5}$$

$$\text{Im } B(q) = a\frac{q^4}{\Lambda^2} + \mathcal{O}\left(\frac{q^6}{\Lambda^4}\right). \tag{5.6}$$

For $a \neq 0$ [3], Λ can be redefined to set $a = \frac{1}{2}$.

With this assumption, we can make use of the Taylor expansion of $\widetilde{W}$

$$\widetilde{W}(q) = \frac{1}{\Lambda^2} + \mathcal{O}\left(\frac{q^2}{\Lambda^4}\right), \quad M^2 \ll -q^2 \ll \Lambda^2, \tag{5.7}$$

to finally get (to the leading order)

$$\frac{\Gamma_{1F}}{\Gamma_O} = \frac{\sigma_{1F}}{\sigma_O} = \frac{1}{4\pi}\left(\frac{E_{\text{CM}}}{\Lambda}\right)^2, \tag{5.8}$$

where $E_{\text{CM}} \ll \Lambda$ is the centre of mass energy of the incoming particle(s). Note that for a decay process, E_{CM} is the mass of the decaying particle. Although, we derived (5.8) for simple interactions of Figure 5.1, it is generally correct (up to order one corrections)

[2] Throughout this paper we are using $(-+++)$ signature for the metric.

[3] Another possibility would be that $a = 0$. In that case, the leading term to the imaginary part of B comes in 6^{th} order. We will not pursue this possibility in this paper.

as long as $E_{\rm CM}$ is much higher than the intrinsic mass of the intermediate particle(s) in Feynman diagrams.

Now, let us define σ_{2F} (Γ_{2F}) to be the cross section (rate) of producing two off-shell particles in the out state (Figure 5.1). Then,

$$\frac{\Gamma_{2F}}{\Gamma_O} = \frac{\sigma_{2F}}{\sigma_O} = \frac{\int d^4p_1 d^4p_2 \widetilde{W}(p_1)\widetilde{W}(p_2)\delta^4(q-p_1-p_2)}{\int d^4p_1 d^4p_2 2\pi\delta_+(p_1^2)2\pi\delta_+(p_2^2)\delta^4(q-p_1-p_2)} = \frac{1}{48\pi^2}\left(\frac{E_{CM}}{\Lambda}\right)^4 \tag{5.9}$$

As we see, adding one more off-shell particle in the final state suppresses the cross section by another factor of $\left(\frac{E_{CM}}{\Lambda}\right)^2$. So, the rate of two off-shell particles production is suppressed by a factor of $\left(\frac{E_{CM}}{\Lambda}\right)^2$ compared to one off-shell particle production.

Before going any further, let us discuss the typical mass of the off-shell particle produced in Figure 5.1. For one off-shell particle production, the mass distribution of the produced off-shell particle is given by

$$P_{1F}(m) = N\int d^4p_1 d^4p_2 \delta_+(p_1^2)\widetilde{W}(p_2)\delta^{(4)}(q-p_1-p_2)\; m\delta(p_2^2+m^2), \tag{5.10}$$

Where N is the normalization factor. Using (5.7) it reduces to

$$P_{1F}(m) = \frac{4m}{E_{\rm CM}^2}\left(1-\frac{m^2}{E_{\rm CM}^2}\right) \quad 0<m<E_{\rm CM}, \tag{5.11}$$

assuming that the off-shell particle is intrinsically massless (or its mass much smaller than E_{CM}). For two off-shell particles production, the mass distribution is given by

$$P_{2F}(m) = N'\int d^4p_1 d^4p_2 \widetilde{W}(p_1)\widetilde{W}(p_2)\delta^{(4)}(q-p_1-p_2)\; m\delta_(p_2^2+m^2), \tag{5.12}$$

which reduces to

$$P_{2F}(m) = \frac{48m}{E_{CM}^2}\left(\frac{1}{4}-\frac{1}{4}(\frac{m}{E_{CM}})^4-(\frac{m}{E_{CM}})^2 Sinh^{-1}[\frac{E_{CM}^2-m^2}{2mE_{CM}}]\right) \tag{5.13}$$

In both cases the typical mass of the produced off-shell particles is $\sim E_{CM}/2$.

Now, we can estimate how likely it is to produce off-shell particles in LHC experiments. If we set $\Lambda \sim M_{\rm P} \equiv \frac{1}{\sqrt{8\pi G}} \sim 10^{18}$ GeV and $E_{\rm CM} \sim 1$ TeV (LHC energy scale), we realize

that the rate of producing off-shell particles in LHC is 10^{-31} lower than the rate of a normal scattering happening. In other words, out of 10^{31} scatterings in LHC, on average one results into the production of an undetectable particle (off-shell mode), explaining why OfDM could be well-hidden from high energy physics experiments.

However, in the cosmological history of the universe, much higher energy scales are reachable and off-shell dark matter production is most efficient. In other words, through cosmological history, a part of the energy in the on-shell sector has been transferred to off-shell sector (while the reverse does not happen) and we detect this energy gravitationally as dark matter. The main purpose of this study to investigate the production of OfDM in the early universe and its observational consequences.

In summary:

- Whenever a scattering happens, there is a chance of producing dark matter particles which is given by (5.8) and (5.9). Furthermore, the probability of producing two dark matter particles in one scattering is much lower than producing only one.
- Dark matter production is most efficient at high (center of mass) energy scatterings. Therefore, most of the dark matter is produced during the stages in the cosmological history where the universe is dense (lots of scatterings) and hot (high energies), i.e. early universe.

Before ending this section, let us discuss the physical range for the non-locality scale Λ. If Λ comes from quantum gravitational effects or fundamental discreteness of spacetime [55, 33, 56], we expect it to be around $M_{\rm P}$. On the other hand, *a priori*, Λ can be much smaller than $M_{\rm P}$, even as low as ~ 10 TeV, as suggested in large extra dimension models that are constructed to address the hierarchy problem (e.g., [57]), or by the cosmological non-constant problem [58]. However, in this paper we assume $\Lambda \gg H_{\rm inf}$, i.e. the non-locality scale is much larger than the Hubble scale during inflation. Otherwise, it would not be consistent to use the standard results of slow-roll inflation when $\Lambda \lesssim H_{\rm inf}$, since the effect of non-locality on the evolution of inflaton or metric could not be neglected.

5.3 Off-shell Dark Matter Production

What are the processes in the early universe that are relevant for OfDM production? First of all, we consider inflation as a starting point in the universe. Whatever happened before inflation is diluted by the exponential expansion of the universe and is not relevant for our

discussion. Furthermore, the effect of non-locality on the inflationary predictions can be neglected in the $H_{\rm inf} \ll \Lambda$ regime. After inflation, we consider two major processes that produce dark matter particles: inflaton decay to standard model particles (reheating) and radiation self interaction in the universe.

5.3.1 Reheating

In this Section, we consider the simplest reheating model: inflation (ϕ field) decays through the effective interaction $g\phi\psi\bar{\psi}$, where ψ represents standard model fields or an intermediate field[4] that decays into standard model particles later.

Decay of inflaton into (on-shell) standard model particles makes the radiation fluid of the universe (their kinetic energy is much higher than their masses.) As we mentioned earlier, however, inflation cannot only decay into on-shell particles; it also has to decay into off-shell particles (off-shell dark matter). Based on (5.8), decay rate into dark matter compared to the decay rate into radiation is suppressed by a factor of

$$f = \frac{1}{4\pi}\left(\frac{m_\phi}{\Lambda}\right)^2 \ll 1, \tag{5.14}$$

where m_ϕ is the mass of inflaton at the end of inflation. As a result, after inflation there are three major constituents of the universe:

1. Inflaton field (ϕ): This field can be treated as a non-relativistic matter after inflation when $m \gg H$ [59]. Inflaton energy density (ρ_ϕ) is the dominant energy density of the universe after inflation and it perturbatively decays into radiation (decay rate Γ) and dark matter (decay rate $f\Gamma$). We later comment on why the coherent decay of inflaton can be ignored.

2. Radiation: This includes all (on-shell) ψ particles. Since the decay rate of inflaton into radiation is much bigger than the decay rate into dark matter, radiation energy density (ρ_r) will dominate the energy density of the universe after the decay of inflaton field.

3. Dark matter: This includes all off-shell ψ particles. As we argue later, dark matter acts as a non-relativistic matter and its energy density is the last one to become dominant.

[4] In this case we assume that the mass of ψ field is much smaller than the inflaton's.

This system of three fluids satisfies the following equations:

$$\dot{\rho}_\phi + 3H\rho_\phi = -(1+f)\Gamma\rho_\phi \tag{5.15}$$
$$\dot{\rho}_r + 4H\rho_r = \Gamma\rho_\phi \tag{5.16}$$
$$\dot{\rho}_{\phi\to DM} + 3H\rho_{\phi\to DM} = f\Gamma\rho_\phi \tag{5.17}$$

with Friedmann equation, where $H = \frac{\dot{a}}{a}$ is the Hubble parameter, a is the scale factor of the universe and $\rho_{\phi\to DM}$ is the contribution to dark matter energy density from inflaton decay.[5]

Let us define the fraction of total dark matter energy density from inflaton decay

$$x = \frac{\rho_{\phi\to DM}}{\rho_{DM}}, \tag{5.18}$$

where ρ_{DM} is the total dark matter energy density. Solving the system of differential equations, we arrive at [60]

$$T_{rh} = x\frac{T_{eq}}{f}, \tag{5.19}$$

where T_{rh} is the reheating temperature (temperature of radiation at the time of inflaton-radiation equality) and T_{eq} is the temperature at the matter-radiation equality.

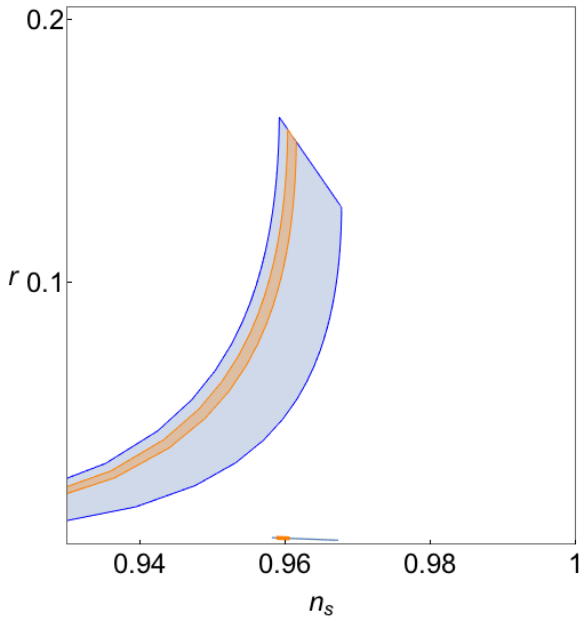
Since $T_{eq} \simeq 0.75$ eV, Equation (5.19) fixes the reheating temperature for a given mass of inflaton.[6] This can be used, for example, to constraint spectral index (n_s) and tensor to scalar ratio (r) of a given inflationary potential by using the following equation:

$$N_e = 67 - \ln\left(\frac{k}{a_0H_0}\right) + \frac{1}{4}\ln\left(\frac{V}{M_{\rm P}^4}\right) + \frac{1}{4}\ln\left(\frac{V}{V_e}\right) + \frac{1}{12}\ln\left(\frac{\rho_{th}}{V_e}\right) - \frac{1}{12}\ln g_{th} \tag{5.20}$$

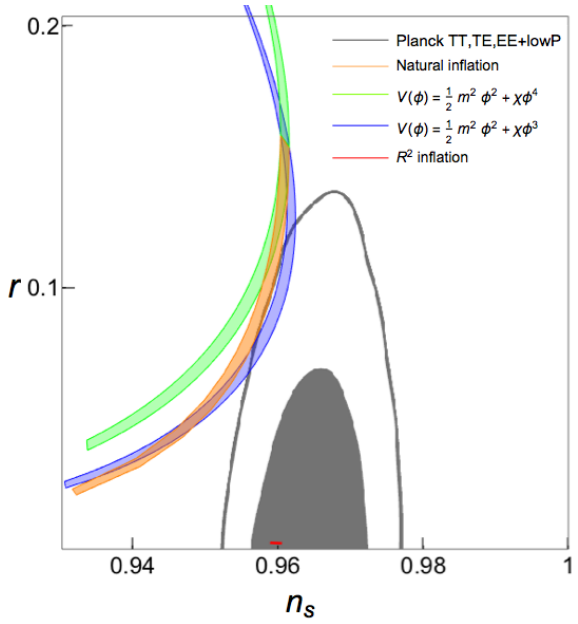
where N_e is the number of e-foldings that mode k is superhorizon during inflation, V_e is the potential energy at the end of inflation, $\rho_{th} \sim g_{th}T_{rh}^4$ is the radiation energy density at reheating temperature, a_0H_0 is the present Hubble radius, V is the potential energy when mode k crosses the horizon during inflation, g_{th} is the number of effective bosonic degrees

[5] Annihilation of radiation into OfDM barely changes the radiation energy density, which is why it has been ignored in (5.16).

[6] We will show later that x is very close to 1.



(a) Blue regions show the prediction of natural and R^2 inflation for $k = 0.002$ Mpc^{-1} with $T_{rh} = 10$ MeV-10^{15}GeV. Orange regions show the prediction of the same models with the constraint coming from OfDM model for $\Lambda = 0.1M_{\rm P} - M_{\rm P}$.



(b) Prediction of n_s and r for different inflationary potentials at $k = 0.002$Mpc^{-1}. Each region represents the prediction with the assumption of OfDM with $\Lambda = 0.1M_{\rm P} - M_{\rm P}$. The shaded region (curve) show the 68% (95%) constraints from CMB observations [3].

Figure 5.2: Predictions of spectral index, n_s, and tensor to scalar ratio, r, for a number of inflationary potentials with OfDM constraint (5.19).

of freedom at reheating temperature and we have assumed pressureless effective equation of state for inflaton during reheating [61].

Figure 5.2a shows how the predicted regions for the Natural [62] and R^2 [63] inflations have shrunk significantly in the n_s, r plane as a result of fixing the reheating temperature. A similar constraint can be found for other inflationary potentials, e.g. Figure 5.2b shows the prediction of OfDM model for a number of inflationary models.

We shall next review and justify the assumptions we made in the above calculations.

Coherent decay of inflaton

The coherent decay of inflaton is negligible if the following condition is satisfied [60, 59]

$$\frac{\Gamma}{m_\phi} \ll \left(\frac{m_\phi}{M_{\rm P}}\right)^2 . \tag{5.21}$$

Using $\Gamma \sim \frac{T_{rh}^2}{M_P}$ and (5.19), this reduces to

$$10^{-18} \left(\frac{\Lambda}{M_p}\right)^4 \left(\frac{10^{-5} M_P}{m_\phi}\right)^7 \ll 1 \tag{5.22}$$

Non-relativistic dark matter

The mass distribution of dark matter particles is given in (5.11). When a dark matter particle is produced, its energy is below $E_{\rm CM}$, while, according to (5.11), masses of the 98% of the dark matter particles are above $0.1 E_{\rm CM}$. In other words, upon production, dark matter particles are relativistic (but not highly relativistic) and through the expansion of the universe they soon become non-relativistic. This justifies our earlier assumption to model dark matter particles as a non-relativistic fluid.

5.3.2 Radiation self-interaction

How much dark matter is produced as a result of radiation self interaction? Here we find an upper bound on the amount of dark matter production through self interaction of radiation. Let us assume a simple annihilation process, such as in Figure 5.1, and ignore the intrinsic mass of the particles. Ignoring the intrinsic mass of the particles is consistent with finding

an upper limit for the dark matter production, since we are allowing for more dark matter production by ignoring the intrinsic masses (more phase space volume to produce OfDM). The average mass of the produced dark matter particles is

$$\int dm\ mP_{1F}(m) = \frac{8}{15}E_{\rm CM}, \tag{5.23}$$

and the cross section of producing one dark matter particle is[7]

$$\sigma_{1F} = \frac{\sigma_O}{4\pi}\left(\frac{E_{\rm CM}}{\Lambda}\right)^2 = \frac{\lambda^2}{128\pi^2\Lambda^2}. \tag{5.24}$$

Since this contribution to dark matter has been produced at very high energies (lower bound on reheating temperature is $T_{rh} > 5$ MeV), it will be highly redshifted today. As a result, current energy density of dark matter is the same as its mass density (see Section 5.3.1). The comoving mass density of the produced dark matter particles through radiation self interaction is given by

$$d\rho_{rad\to DM}/dt = a^3(t)\int \frac{d^3p_1}{(2\pi)^3}\frac{d^3p_2}{(2\pi)^3}g_1 n(\vec{p_1})g_2 n(\vec{p_2})\langle m\sigma_{1F}v_{rel}\rangle, \tag{5.25}$$

where t is the cosmological time, $n(\vec{p}) = \frac{1}{e^{|\vec{p}|/T}\pm 1}$ is the occupation number of incoming on-shell states at temperature T, g is the degeneracy factor, v_{rel} is the relative velocity of the incoming particles and $\vec{p}_i$'s are the momenta of the incoming particles. It is clear that (5.25) results into a bigger comoving mass density when we choose bosonic occupation number.

Using (5.23), (5.24), $v_{rel} \lesssim 2$ and performing the integrals over momenta in (5.25), we arrive at

$$d\rho_{rad\to DM}/dt \lesssim g_1 g_2 \frac{8\lambda^2}{45(2\pi)^6}\Gamma^2[3.5]\zeta^2[3.5]a^3(t)\frac{T^7}{\Lambda^2}, \tag{5.26}$$

where Γ and ζ are Gamma and Zeta function respectively.

Perturbative calculations are valid only if $\lambda < 1$. If we consider this condition in (5.26) and sum over all constituent of the radiation fluid, we arrive at

$$\rho_{rad\to DM} < 4\times 10^{-5}\int dt\ g^2 a^3(t)\frac{T^7}{\Lambda^2}, \tag{5.27}$$

[7]This is again consistent with finding the upper bound, since the cross section of two off-shell production is much smaller.

where g is the total number of degrees of freedom in the radiation fluid.

During reheating (by solving (5.15-5.17))

$$t \propto a^{3/2}, \quad T^4 \propto \rho_{rad} \propto a^{-3/2}. \tag{5.28}$$

Substituting these values back in (5.27), we realize that the annihilation of radiation into dark matter is most efficient at the end of reheating. The same manipulation shows that the annihilation of radiation into dark matter during radiation era happens at the beginning of radiation era and is of the same order.

Let us now work out how much dark matter will be produced in radiation era (after reheating). During radiation era

$$t = \sqrt{\frac{45}{2\pi^2 g}} \frac{M_{\rm P}}{T^2}. \tag{5.29}$$

Combining this, with Eq. (5.27), and the results of Sec. (5.3.1), we find:

$$\begin{aligned} \frac{\rho_{rad\to DM}}{\rho_{DM}} &< 10^{-5} \times \frac{g^{3/2} M_{\rm P} T_{rh}^2}{T_{eq}\Lambda^2} \\ &\sim 10^{-3} \times \frac{g^{3/2} M_{\rm P} T_{eq}\Lambda^2}{m_\phi^4} \\ &\sim 10^{-7} \left(\frac{g}{124}\right)^{3/2} \left(\frac{\Lambda}{M_{\rm P}}\right)^2 \left(\frac{m_\phi}{10^{-5} M_{\rm P}}\right)^{-4}, \end{aligned} \tag{5.30}$$

where we used $g \simeq 124$ for standard model of particle physics.

Therefore, for $\Lambda \sim M_{\rm P}$ and high scale inflation $m_\phi \approx 10^{-5} M_{\rm P}$, the production of O$f$DM due to radiation self-interaction is much smaller than the contribution from inflaton decay (in effect $x = 1$). However, $\rho_{rad\to DM}$ can become important in scenarios with lighter inflaton, i.e. if $m_\phi \lesssim 10^{-7}(M_{\rm P}\Lambda)^{1/2}$.

So far we have studied the predictions of this model in the context of inflation. As we showed earlier, this model effectively fixes the reheating temperature of the universe. By constraining the reheating temperature, we can narrow the predictions of (n_s, r) for a given inflationary potential, by fixing the number of e-foldings. However, the predictions for (n_s, r) are model dependent and vary with the inflationary potential. Conversely, one can use the observational constraints on (n_s, r) as a way to fix the non-locality scale Λ, in the context of a given inflationary model.

5.4 Cold OfDM

In principle, OfDM particles with very low masses can be produced in scatterings. These low mass particles can behave like hot dark matter at different stages in the evolution of the universe. Let us estimate an upper bound on the fraction of hot OfDM particles at a given redshift.

An off-shell dark matter particle with mass m has energy $E_m = \frac{E_{\rm CM}^2+m^2}{2E_{\rm CM}}$ and momentum $p_m = \frac{E_{\rm CM}^2-m^2}{2E_{\rm CM}}$, where $E_{\rm CM}$ is the energy of the process producing the dark matter particle.[8] At redshift z, this particles is relativistic if $p_m \frac{1+z}{1+z_{pr}} \gtrsim m$, where z_{pr} is the redshift at the time of production.

Given the mass distribution of OfDM particles and assuming that most of the dark matter particles are produced at the time of reheating (as we discussed in previous sections), we can find the fraction of hot dark matter particles (Ω_h). This has been shown in Figure 5.3a. Only a small fraction of OfDM is hot at $z < 1000$, which makes it a good candidate for CDM. This result is not surprising since, as we mentioned earlier, even at the time of production these particles are not highly relativistic.

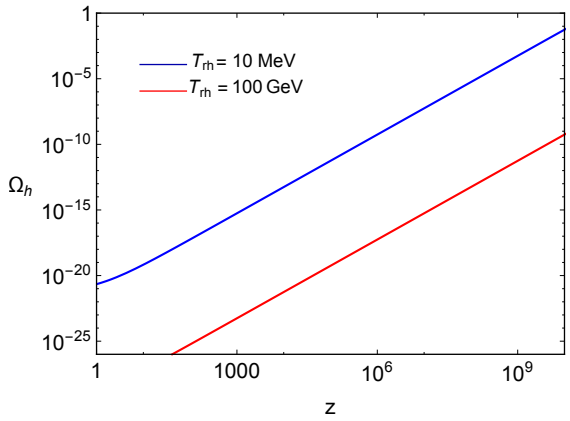
Let us work out the distribution of free streaming distance λ_{fs}. This is given by

$$\lambda_{fs} = u \int \frac{dt}{\sqrt{a^4 + u^2 a^2}} \tag{5.31}$$

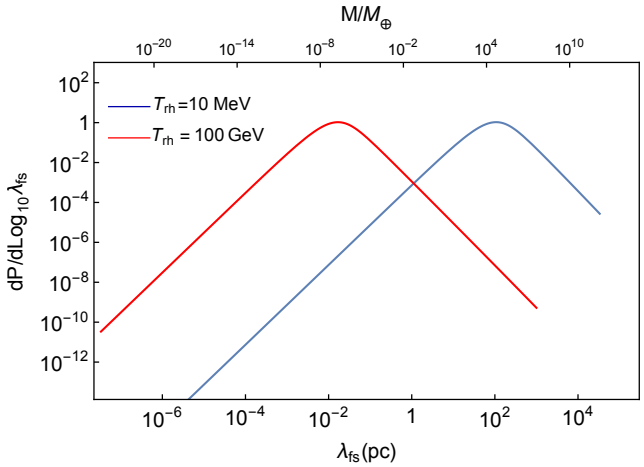
where $u = a_{pr}\frac{v}{\sqrt{1-v^2}}$ and $v = \frac{p_m}{E_m}$ is the velocity of dark matter particle with mass m at the time of production. Assuming $a_{pr} = a_{rh}$, Equation (5.31) gives the free streaming distance in terms of m and T_{rh}. This equation can be used further to derive the probability distribution of λ_{fs}, since the probability distribution of m (5.11) is known. The result has been shown in Figure 5.3b. Since the velocity distribution of OfDM particles is different from Maxwell-Boltzmann distribution, probability distribution of λ_{fs} in this model is different from ordinary thermal WIMP scenario. In particular, it has a much shallower power-law (rather than gaussian) cut-off at large λ_{fs}'s. This leads to a different matter power spectrum (on small-scales) which can, in principle, be a probe to distinguish these two models. Figure 5.4 shows the matter transfer function $T(k)$.

In Figure 5.4 two effects has been considered: Growth in matter fluctuations due to an early era of matter domination (inflaton dominated era) and free streaming effect. Early

[8] This comes from conservation of energy-momentum in the rest frame of incoming particle(s). Here, we have ignored mass of the on-shell particle produced together with OfDM particle.



(a) The fraction of off-shell dark matter particles, produced at the time of reheating, that remain relativistic down to a given redshift.



(b) Distribution of free streaming distance of OfDM for different reheating temperatures. The top axis shows the characteristic halo mass associated with the free streaming scale, in units of Earth mass.

Figure 5.3

matter era result into amplification of matter fluctuations for modes that enter the horizon during reheating. This amplification is roughly $\propto \frac{k^2}{\ln(k)}$ [60]. On the other hand, free streaming effect result into the decrease in the matter power spectrum on small scales $\propto k^{-2}$. The combination of the two effects has been shown in Figure 5.4. On small scales, transfer function drops as $(\ln k)^{-1}$ which is to be contrasted with a much steeper gaussian cut-off in thermal scenarios.

Future gravitational probes of dark matter structure on small scales can potentially test this prediction for matter power spectrum on $10^{-1} - 10^{-3}$ pc scales [64, 65, 66].

5.5 Summary & Conclusion

In this Chapter, we laid out some features of off-shell dark matter model. This model is motivated by considering the effect of nonlocality on the evolutions of fields which manifests itself by introducing a new set of excitations. The new excitations, named as off-shell modes, cannot be detected through scattering experiments that makes them a natural

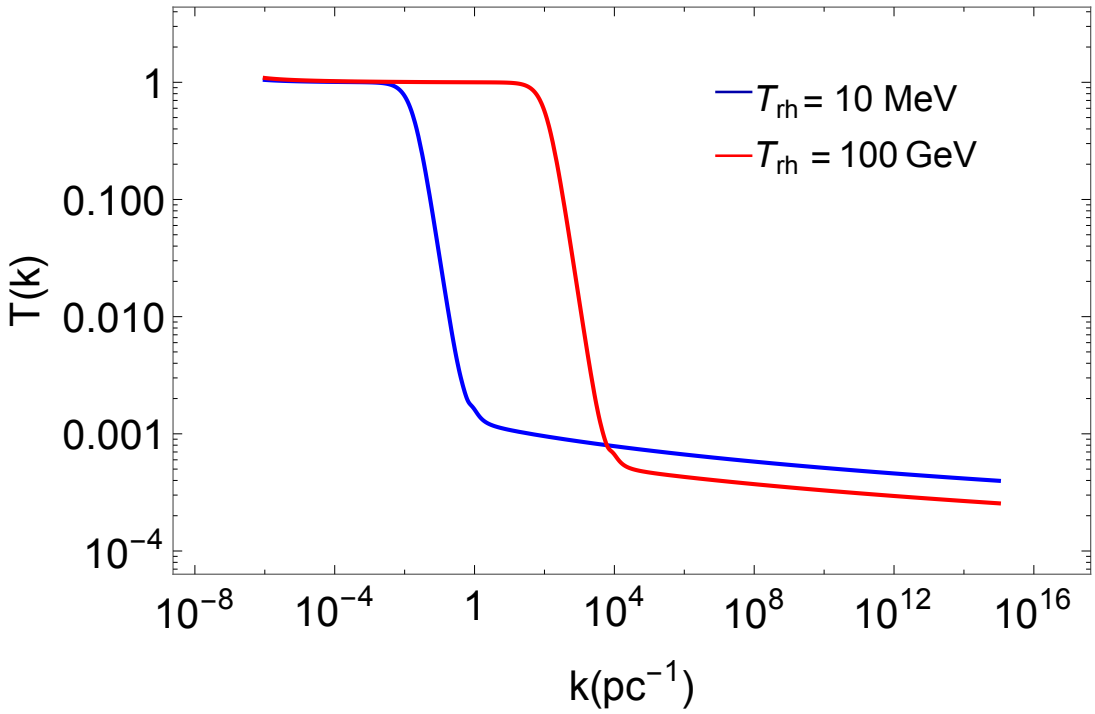


Figure 5.4: Matter transfer function due to the growth in early matter era and free streaming effect. Instead of an exponential cut-off for large k in thermal scenarios, there is $\propto (\ln k)^{-1}$ drop in OfDM scenario.

candidate for dark matter. So, if OfDM makes up the observed cosmological dark matter, we would not be able to detect dark matter particles directly.

However, OfDM particles can be produced in scattering experiments and this is one way to indirectly confirm their existence by detecting missing energy in scatterings. The probability of missing energy is given by (5.8) and (5.9). Scattering experiments with enough precisions to detect this missing energy could be a possible way to test this model.

We discussed predictions of OfDM model in the context of cosmology and showed that it is intertwined with the physics of inflation and reheating. For a very simple reheating model, we showed that OfDM particles are generically produced in the era of reheating and through the decay of inflaton. Since OfDM particles are not interacting with other particles (and themselves), they do not have a thermal distribution. We calculated their distribution function in our simple reheating model and showed that it leads to much shallower suppression of matter power spectrum on small scales compared to a gaussian cutoff of thermal dark matter candidates. This in principle could be another way to test the model via the observations probing matter power spectrum in sub-pc scales.

We end this Chapter by noting the following theoretical aspects of OfDM which are yet to be explored:

1. Throughout this Chapter we assumed that off-shell modes of a nonlocal field gravitate like ordinary (on-shell) matter, i.e. an off-shell mode with mass m gravitates like a normal particle with the same mass. This assumption, which seems reasonable, is yet to be verified through a consistent coupling of nonlocal quantum field theories to gravity.

2. So far, the quantization of this type of nonlocal field theory has only been done for scalars. But how about spinor or gauge fields? This is especially important in the case of gauge theories which govern all interactions in the Standard Model of particle physics. There are (at least) two obvious ways to proceed here:

   (a) One can define a nonlocal version of gauge transformations to keep gauge invariance. This presumably implies that scattering processes have to include pairs of on-shell modes, or otherwise charge conservation would be violated. In the case of our phenomenological reheating model in Section 5.3.1, it means that the inflaton field has to first decay into a neutral field which later decays into standard model particles, otherwise Equation (5.14) is not applicable.

   (b) Gauge invariance is broken at a Planck suppressed level, similarly to the violation of diffeomorphism invariance in Horava-Lifhsitz gravity [13]. In this

case, one should look for (possibly dangerous) physical consequences of breaking gauge invariance.

Chapter 6

Dynamical Emergence of Universal Horizons

6.1 Introduction

General relativity (GR) has been our best classical theory of gravity that is compatible with a wide variety of experiments. 4D diffeomorphism invariance is the fundamental gauge symmetry of GR, resulting in the absence of a preferred frame. Despite this, there are good reasons to consider a fundamental preferred frame. Here, we list a few of the interesting alternative theories of gravity that invoke this property:

1. One reason to consider theories with a preferred frame is purely phenomenological. An example is Einstein-Aether theory [15]. The preferred frame is built into the theory via a unit time-like vector u^μ. The action is the Einstein-Hilbert action plus all possible terms containing first order derivatives of u^μ. This yields several free parameters that can be constrained/detected experimentally (e.g. [67, 68]). In particular, these constraints imply that aether disturbances should propagate (super)luminally [69].

2. Another theory with a preferred frame is Gravitational Aether theory [17], which is an attempt to solve the (old) cosmological constant problem by simply subtracting the trace of the energy-momentum tensor from the right hand side of Einstein's equations. This ensures that the zero point energy of quantum field theory does not gravitate. But in order to satisfy the Bianchi identities a new term (a symmetric

tensor) must be added to the right hand side of Einstein's equations. The Bianchi identities then relate this term to the trace of matter energy-momentum tensor, via energy-momentum conservation. The new term is assumed to have the form of a perfect fluid (or the gravitational aether). In the limit of zero energy density (incompressibility), no new dynamical degree of freedom appears and the Bianchi identities completely fix the evolution of the aether, whose four-velocity introduces a preferred direction of time.

3. Cosmological dynamical scalar fields generically introduce a preferred frame. An example of this type of theory is K-essence, constructed so that the scalar field develops a negative pressure once the matter dominated era begins. Its associated dynamical behaviour is then deemed responsible for the accelerating cosmic expansion of our universe, whilst avoiding anthropic arguments [18]. However, any such model that solves such problems necessarily has perturbations propagating faster than speed of light [70].

4. Pushing K-essence to its limit, cuscuton theory is a scalar field theory with *infinite* sound speed [19]. Cuscuton action is given by

$$S = \int d^4x\sqrt{-g}(\mu^2\sqrt{\partial_\nu\phi\partial^\nu\phi} - V(\phi)). \tag{6.1}$$

   This theory is the same as the low energy limit of (non-projectable) Hořava-Lifshitz gravity for quadratic potential $V(\phi)$ [71]. The relation between parameters of cuscuton and λ parameter of Hořava-Lifshitz gravity is as follows

$$\mu^2 = -V''(\phi) = \frac{\lambda - 1}{16\pi G_N}. \tag{6.2}$$

   They also have the same solution as Einstein-Aether theory when aether is hypersurface-orthogonal and $c_2 = \lambda - 1$ is the only non-vanishing term in Einstein-Aether action.

   Constant field surfaces of cuscuton define a preferred time direction because signals propagate instantaneously on these surfaces. A constant field surface also has constant density and constant mean curvature. As a result, in a cosmological space-time, cuscuton can be considered as global time.

5. Shape dynamics is an alternative theory of gravity whose fundamental symmetry is scale invariance [72]. It has been shown that shape dynamics and GR produce the same solutions in regions of space-time that admit a CMC slicing [73]. Whether shape dynamics predicts a different solution (or even no solution) where there is no CMC slicing is still an open question.

6. Finally, Hořava-Lifshitz gravity [13, 14] is a potentially renormalizable theory of gravity that breaks 4D diffemorphism invariant at high energies. However the non-projectable version of the theory reduces in the low energy limit to the Einstein-Hilbert action together with a scalar field with infinite sound speed (cuscuton) [71].

Spherically symmetric black hole solutions in Hořava-Lifshitz gravity have been studied in [74], and it has been shown that the Schwarzschild metric is a solution to the equations of motion for large black holes (whose curvature radius is much bigger than the Planck length). These solutions are the same as the spherically symmetric black hole solutions of Einstein-Aether theories [75], since spherical symmetry requires the aether vector field to be hypersurface-orthogonal. As long as the effect of the cuscuton on geometry is negligible, the Schwarzschild metric remains a black hole solution in Hořava-Lifshitz gravity. However, the behavior of the cuscuton is important to the causal structure of spacetime, simply because its sound speed is greater than the speed of light.

In this Chapter, we investigate this issue. Our motivation is to study black hole solutions in theories with a preferred time direction. We specifically consider the causal structure of black hole solutions in Hořava-Lifshitz (or cuscuton) gravity. Although spherically symmetric black holes in Hořava-Lifshitz (and Einstein-Aether) gravity are close to the Schwarzschild solution, they possess a new feature: they contain a trapped surface forbidding the escape of any signal, no matter how fast its propagation speed. This new type of horizon has been called a "Universal" horizon, as it is universal to all signals with arbitrary speed. Previously demonstrated for *static* spherically symmetric systems [76, 77, 78] (and [79] for stationary solutions), we investigate here the collapse of a spherical thin shell and show how a universal horizon emerges in a dynamical setting. Also, unlike previous studies considering only asymptotically flat background, we have done our study of universal horizon in asymptotically cosmological solution. We note that the dynamical formation of a similar additional trapped surface in K-essence models was also recently demonstrated, though there was numerical evidence of a breakdown of the initial value problem [80, 81].

The structure of this Chapter is as follows. We start by reviewing the propagation of a scalar field (with a general action) in a general background space-time. We then show how perturbations of the scalar field propagate through space-time, and derive the "propagation cone" of perturbations at any given point. The propagation cone (sometimes called the sound cone) is an analogue of the light cone for the scalar field perturbations. For a scalar field, we find that the propagation cone depends on the constant background field surfaces. In Section 6.3, we explicitly derive the equation of motion of background field and propagation cone in the limit where the sound speed is very large. Section

6.4 contains the solution for a collapsing spherical thin shell space-time. We show that constant field surfaces are well behaved as long as the shell's radius (R) is bigger than Schwarzschild radius ($2M$). However, when R approaches the critical value $R_c < 1.5M$ constant field surfaces start to stack up around the $r = 1.5M$ surface. This behaviour shows that the field perturbations cannot escape from inside $r = 1.5M$ to infinity even though they propagate almost instantaneously. As a result, there exists a horizon for these perturbations (universal horizon) at $r = 1.5M$. This result is in agreement with the previous study of universal horizon in the infinite speed Einstein-Aether [82]. Section 6.5 covers the emergence of universal horizon.

6.2 Introduction to Signal Propagation

Consider a scalar field ϕ with the following action

$$S = \int d^4x\sqrt{-g}\mathcal{L}\left(X,\phi\right), \tag{6.3}$$

where $X = \frac{1}{2}g^{\mu\nu}\nabla_\mu\phi\nabla_\nu\phi$ and $g_{\mu\nu}$[1] is the spacetime metric. We have restricted the Lagrangian $\mathcal{L}$ to depend only on the field and its first derivative. The energy-momentum tensor

$$T_{\mu\nu} = \mathcal{L}_{,X}\nabla_\mu\phi\nabla_\nu\phi - g_{\mu\nu}\mathcal{L}, \tag{6.4}$$

is the same as that of a perfect fluid $T_{\mu\nu} = (\rho + p)\, u_\mu u_\nu - p\; g_{\mu\nu}$ with

$$u_\mu = \frac{\nabla_\mu\phi}{\sqrt{\nabla_\alpha\phi\nabla^\alpha\phi}}, \tag{6.5}$$

$$p = \mathcal{L}, \tag{6.6}$$

$$\rho = 2X\mathcal{L}_{,X} - \mathcal{L}, \tag{6.7}$$

provided that $X > 0$ (so that the fluid has a rest frame). Henceforth we assume that $X > 0$, which has two advantages. Not only can the scalar field be understood as a perfect fluid (as noted above) but together with null energy condition (which requires that $\mathcal{L}_{,X} \geq 0$) this assumption implies that the spacetime is *stably causal* [83].

Variation of the action (6.3) with respect to ϕ yields the following equations of motion (for the derivation of the following equations (6.8)-(6.14) see [83])

$$\tilde{G}^{\mu\nu}\nabla_\mu\nabla_\nu\phi + 2X\mathcal{L}_{,X\phi} - \mathcal{L}_{,\phi} = 0, \tag{6.8}$$

[1] metric signature $(+---)$

where $\tilde{G}^{\mu\nu} = \mathcal{L}_{,X} g^{\mu\nu} + \mathcal{L}_{,XX} \nabla^\mu \phi \nabla^\nu \phi$.

To see how a ϕ-signal propagates in this spacetime, consider a small field perturbation $\pi(x)$ on some background field $\phi_0(x)$ (neglecting the geometric back-reaction). These perturbations satisfy the following hyperbolic equation

$$\frac{1}{\sqrt{-G}} \partial_\mu \left(\sqrt{-G} G^{\mu\nu} \partial_\nu \pi \right) + M_{eff}^2 \pi = 0, \tag{6.9}$$

where

$$G^{\mu\nu} = \frac{c_s}{\mathcal{L}_{,X}^2} \tilde{G}^{\mu\nu}, \;\; \left(G^{-1}\right)_{\mu\nu} G^{\nu\rho} = \delta_\mu^\rho, \tag{6.10}$$

$$\sqrt{-G} = \sqrt{-det\left(G^{-1}\right)_{\mu\nu}}, \tag{6.11}$$

$$M_{eff}^2 = \frac{c_s}{\mathcal{L}_{,X}^2} \left(2X\mathcal{L}_{,X\phi\phi} - \mathcal{L}_{,\phi\phi} + \frac{\partial \tilde{G}^{\mu\nu}}{\partial \phi} \nabla_\mu \nabla_\nu \phi_0 \right), \tag{6.12}$$

$$c_s^2 = \frac{1}{1 + 2X \frac{\mathcal{L}_{,XX}}{\mathcal{L}_{,X}}}. \tag{6.13}$$

The quantity c_s is the propagation speed of the field perturbation π in the field rest frame (co-moving frame).

Equation (6.9) is a Klein-Gordon equation with the effective metric

$$\left(G^{-1}\right)_{\mu\nu} = \frac{\mathcal{L}_{,X}}{c_s} \left(g_{\mu\nu} - c_s^2 \frac{\mathcal{L}_{,XX}}{\mathcal{L}_{,X}} \nabla_\mu \phi_0 \nabla_\nu \phi_0 \right), \tag{6.14}$$

which determines the propagation of perturbations. Indeed, $\left(G^{-1}\right)_{\mu\nu}$ defines a "propagation" cone at any point in spacetime, through

$$\left(G^{-1}\right)_{\mu\nu} v^\mu v^\nu = 0. \tag{6.15}$$

Using the above equations, we get

$$g_{\mu\nu} v^\mu v^\nu = \left(1 - c_s^2\right) \left(g_{\mu\nu} u^\mu v^\nu\right)^2 \tag{6.16}$$

showing that for superluminal perturbations ($c_s > 1$), the vector v must be space-like with respect to the metric $g_{\mu\nu}$, consistent with our expectation that the influence cone is wider than the light cone for superluminal propagation. It also shows that the propagation cone at any point depends on the background field through the vector field u^μ.

From now on we will focus on a scalar field with the following Lagrangian

$$\mathcal{L} = aX^n - V(\phi), \tag{6.17}$$

where a and n are constants. So, Equation (6.13) yields

$$c_s^2 = \frac{1}{2n-1}. \tag{6.18}$$

Note that $\frac{1}{2} < n < 1$ and $n > 1$ respectively correspond to superluminal and subluminal propagation. The fluid also becomes incompressible (i.e. infinite speed of sound) at $n = \frac{1}{2}$. For $n < \frac{1}{2}$ the sound speed becomes imaginary, which is a sign of instability. We are interested in the superluminal case $\frac{1}{2} < n < 1$. The propagation cone is then given by

$$g_{\mu\nu}v^\mu v^\nu = \frac{2\,(n-1)}{2n-1}\left(g_{\mu\nu}u^\mu v^\nu\right)^2. \tag{6.19}$$

The right hand side of equation (6.19) is negative for $\frac{1}{2} < n < 1$, implying v is space-like. Normalizing $g_{\mu\nu}v^\mu v^\nu = -1$, we get

$$\left(u_\mu v^\mu\right)^2 = -\frac{2n-1}{2\,(n-1)}. \tag{6.20}$$

Note that as n approaches $\frac{1}{2}$, v becomes orthogonal to the velocity vector u. It means that the propagation cone becomes almost tangent to constant field surfaces, for which perturbations propagate (almost) parallel to the constant background field surfaces.

In summary, in order to determine how ϕ-signals (perturbations) propagate through spacetime, we first solve Equation (6.8) for the background field ϕ_0. Equation (6.20) then determines the propagation cone at any point of spacetime. Henceforth, we shall focus on fields with a very large sound speed, for which the values of n are very close to $\frac{1}{2}$.

6.3 Background Field and Propagation Cone

Equation (6.8) (together with (6.17)) yields

$$anX^{n-1}\left(g^{\mu\nu} + (n-1)\frac{\nabla^\mu\phi\nabla^\nu\phi}{X}\right)\nabla_\mu\nabla_\nu\phi + V'(\phi) = 0. \tag{6.21}$$

where $n = \frac{1}{2}(1+\epsilon^2)$, $\epsilon \ll 1$. In order to find the propagation cone for small perturbations, we need to solve (6.20) and (6.21), which we shall do as a power series in ϵ. Considering first the equations for $n = \frac{1}{2}$ (zeroth order in ϵ), we have

$$\frac{a\sqrt{2}}{2}\nabla_\mu u^\mu + V'(\phi) = 0, \tag{6.22}$$

$$u_\mu v^\mu = 0, \tag{6.23}$$

where we have used

$$\nabla_\mu u^\mu = \frac{1}{\sqrt{\nabla_\alpha\phi\nabla^\alpha\phi}}\left(g^{\mu\nu} - \frac{\nabla^\mu\phi\nabla^\nu\phi}{\nabla_\alpha\phi\nabla^\alpha\phi}\right)\nabla_\mu\nabla_\nu\phi, \tag{6.24}$$

to get (6.22) (which is the cuscuton equation of motion). For values of $n \gtrsim \frac{1}{2}$ (slightly greater than $\frac{1}{2}$), up to first order in ϵ (6.20) and (6.21) become

$$\frac{a}{2\sqrt{X}}\left(g^{\mu\nu} - \frac{\nabla^\mu\phi\nabla^\nu\phi}{2X}\right)\nabla_\mu\nabla_\nu\phi + V'(\phi) = 0, \tag{6.25}$$

$$u_\mu v^\mu = \epsilon. \tag{6.26}$$

Since equation (6.25) is the same as (6.22), up to first order in ϵ the field ϕ satisfies the cuscuton equation of motion. Also the propagation cone is determined by equation (6.26). Because of the key role of the cuscuton field in the discussion, we will explain some of its properties; this will help us to solve (6.25).

6.3.1 Cuscuton Characteristics

Cuscuton is a scalar field with an infinite speed of sound ($n = \frac{1}{2}$). Its energy-momentum tensor can be expressed in the form of a perfect fluid $T_{\mu\nu} = (\rho + p)u_{\mu\nu} - pg_{\mu\nu}$, where

$$\rho = V(\phi), \tag{6.27}$$

$$p = a\sqrt{X} - V(\phi), \tag{6.28}$$

$$u_\mu = \frac{\nabla_\mu\phi}{\sqrt{\nabla_\alpha\phi\nabla^\alpha\phi}}, \tag{6.29}$$

and it satisfies the equation of motion (6.22).

Equation (6.29) shows that constant field surfaces are the same as co-moving surfaces, as the field's velocity vector u^μ is the normal vector to constant field surfaces. Since the

mean curvature K (the trace of the extrinsic curvature) of constant field surfaces is the divergence of the normal vector to the surface, we have

$$K = u^{\alpha}{}_{;\alpha} = -\frac{\sqrt{2}}{a}V'(\phi), \tag{6.30}$$

where we have used (6.22) in the second equality. Equation (6.30) shows that mean curvature K is constant on a constant field surface. It means that constant cuscuton field surfaces are CMC (constant mean curvature) surfaces. Consequently in order to find constant cuscuton field surfaces we only need to find the CMC surfaces of the background spacetime.

6.4 CMC Surfaces of Spherically Collapsing Thin Shell of Dust Spacetime

As we mentioned in the previous Section, we only need to find CMC surfaces of the background spacetime to determine the propagation cone. We consider here a collapsing shell of spherically symmetric dust as the background spacetime and derive its CMC surfaces.

Assuming that the thin shell is located at $r = R(t)$, it divides spacetime into two regions with the following metrics:

$$\begin{aligned} \text{I} : ds^2 &= A^2(t)dt^2 - dr^2 - r^2 d\Omega^2, \ r < R(t) \\ \text{II} : ds^2 &= f(r)dt^2 - \frac{dr^2}{f(r)} - r^2 d\Omega^2, \ r > R(t), \end{aligned}$$

in which $f(r) = 1 - \frac{2M}{r}$, and we have ignored the gravitational back reaction of the cuscuton field. The shell radius satisfies the following geodesic equation

$$\dot{R} = -f(R)\sqrt{1 - \frac{f(R)}{e^2}}, \tag{6.31}$$

where e is a constant of motion and $\cdot = \frac{d}{dt}$. The function $A(t)$ can be found by matching the line elements at $r = R(t)$

$$A = f(R)\sqrt{1 + \frac{2M}{e^2 R}}. \tag{6.32}$$

In order to find CMC surfaces in this spacetime, we need to find CMC surfaces in each region and match them at $r = R(t)$. If $t_{CMC} = T(r)$ is a CMC surface with constant mean curvature K, the normal vector u_μ (in region II) to this surface will be

$$u_\mu = \frac{1}{N}\nabla_\mu(t - T(r)) = \frac{1}{N}(1, -T'(r), 0, 0), \tag{6.33}$$

where N is the normalization factor

$$N^2 = \frac{1}{f(r)} - f(r)T'^2(r) \tag{6.34}$$

that we choose to be positive. As a result

$$u^\mu = \frac{1}{N}(\frac{1}{f(r)}, f(r)T'(r), 0, 0) \tag{6.35}$$

and $K = \nabla_\mu u^\mu$ yields

$$u^r = \frac{K}{3}r - \frac{B}{r^2}, \tag{6.36}$$

where B is an integration constant. This constant may vary from one CMC surface to another. Comparing (6.35) with (6.36) yields

$$T'(r) = \frac{u^r}{f(r)\sqrt{f(r) + (u^r)^2}} = \frac{\frac{K}{3}r - \frac{B}{r^2}}{f(r)\sqrt{f(r) + \left(\frac{K}{3}r - \frac{B}{r^2}\right)^2}}, \tag{6.37}$$

with the following unit normal to the CMC surface

$$u^r_{II} = \frac{K}{3}r - \frac{B}{r^2}, \tag{6.38}$$

$$u^t_{II} = \frac{1}{f(r)}\sqrt{(\frac{K}{3}r - \frac{B}{r^2})^2 + f(r)}. \tag{6.39}$$

Similar calculations yield the following unit normal

$$u^r_I = \frac{K}{3}r, \tag{6.40}$$

$$u^t_I = \frac{\sqrt{(\frac{Kr}{3})^2 + 1}}{A}. \tag{6.41}$$

to the CMC surface in the first region.

6.4.1 Finding B

Our next task is to obtain B. Its value can be fixed by matching the two CMC solutions on the shell's surface. We construct a set of orthonormal basis vectors $\{n, e_i\}$, $i \in \{1, 2, 3\}$, where the e_i's form a complete orthonormal basis for the shell's surface and n is the unit vector normal to the surface.

The value of B must be chosen such that $K = \nabla_\mu u^\mu$ remains non-singular on the surface of the shell. The previous derivations for u in region I and II are only valid inside each region and not on the surface.

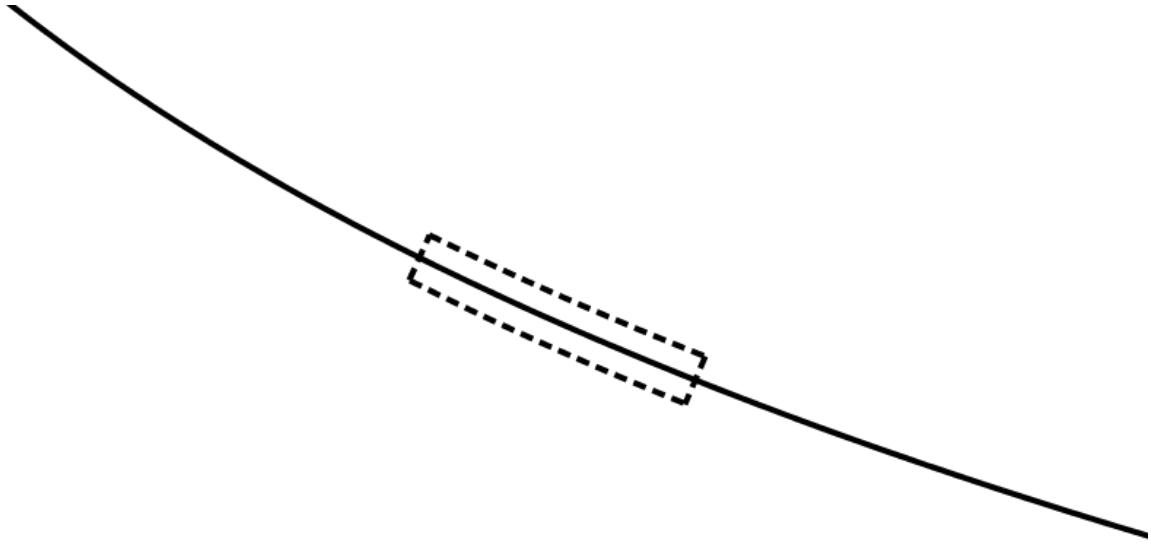


Figure 6.1: Space-time diagram showing $r = R(t)$. The region inside the dashed lines is V. Two sides of this region, normal to shell's surface, are much smaller than the other sides, so their contribution to the R.H.S of (6.42) is negligible.

However, non-singularity of K can be imposed by Gauss's law. Consider a small space-time volume V (Fig. 6.1); using Gauss's law we find

$$\int_V K \, dV = \int_V \nabla_\mu u^\mu \, dV = \int_{\partial V} u_\mu n_V^\mu \, dS, \tag{6.42}$$

where n_V is the normal vector to the region V.

For sufficiently small V only sides parallel to the shell's surface contribute to the right-hand side of (6.42) and $n_V = \pm n$. The left-hand side approaches zero as $V \to 0$. Hence

$u \cdot n \equiv u_\mu n^\mu$ must remain continuous across the surface. Since K is non-singular, we find that $u_I \cdot n = u_{II} \cdot n$, where the equality must be imposed on the shell.

Moreover, in order to have a smooth CMC surface, we demand that the projection of u onto the surface of the shell remain continuous, implying $u_I \cdot e_i = u_{II} \cdot e_i$. Although this smoothness condition results in three equations, two of them are trivial because of spherical symmetry. The non-trivial equation reads

$$g_{tt} u_I^t dt + g_{rr} u_I^r dR = g_{tt} u_{II}^t dt + g_{rr} u_{II}^r dR \tag{6.43}$$

This equation can be solved (analytically) for B in terms of R and K, and it has two different solutions. The previous condition (non-singularity) picks one of them.

It can be easily shown that imposing smoothness condition requires that $u \cdot n$ either remains continuous or flips sign across shell's surface. The correct value of B is the one that does not change the sign of $u \cdot n$.

In the following, we will explicitly derive CMC surfaces in two cases.

6.4.2 $K > 0$

Using (6.31) and (6.37), we get

$$\begin{aligned} t_{CMC}(r) &= L(K) \\ &\quad - \int_r^{K^{-1}} \mathrm{d}x \frac{\frac{K}{3}x - \frac{B}{x^2}}{f(x)\sqrt{f(x) + \left(\frac{K}{3}x - \frac{B}{x^2}\right)^2}}, \end{aligned} \tag{6.44}$$

$$t_{shell}(r) = -\int_{r_0}^{r} \mathrm{d}x \frac{1}{f(x)\sqrt{1 - \frac{f(x)}{e^2}}} + t_0, \tag{6.45}$$

where t_0 and r_0 are constants (determining the initial position of shell), $L(K)$ is another integration constant (determining the behavior of CMC surfaces at large radii) and (6.44) is only valid for $r > R$. Note that $t = t_{shell}(r)$ and $r = R(t)$ describe the same surface.

$L(K)$ can be fixed by matching cuscuton solutions to cosmological ones at large distances, remembering that a CMC surface is also a cuscuton constant density surface.

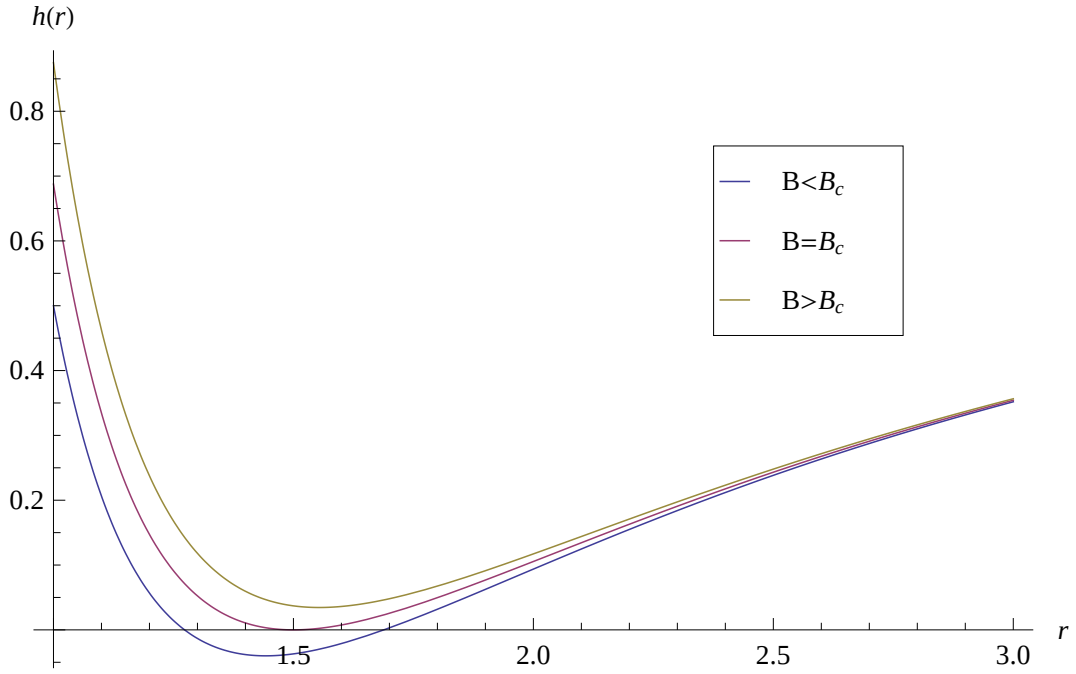


Figure 6.2: The function $h(r, R)$ in terms of r for different values of B (setting $M = 1$). h is always positive when $B > B_c$. It has double root at $r = 1.5M$ when $B = B_c$ and becomes negative for $B < B_c$.

Taking the derivative with respect to K in (6.44), we obtain

$$\frac{1}{\dot{K}} \equiv \left(\frac{\partial t}{\partial K}\right)_{r\sim K^{-1}} \approx L'(K) + \frac{1}{K^2}\frac{\frac{1}{3} - K^2 B}{\sqrt{1 + (\frac{1}{3} - K^2 B)^2}}$$
$$\approx L'(K) + \frac{1}{K^2}\frac{1}{\sqrt{10}} \tag{6.46}$$

where we have used $MK \ll 1$ (as the Schwarzschild horizon $2M$ is much smaller than the cosmological horizon K^{-1}) and $K^2 B \ll 1$ (since we expect to have a homogeneous cuscuton field on cosmological scales (6.36)). Knowing $\dot{K}$ from cosmology, we can fix $L(K)$.

Once all the constants (t_0, r_0 and $L(K)$) are fixed, we require $t_{CMC}(R) = t_{shell}(R)$. This equation together with (6.43) for each value of R gives the corresponding value of K and B and completely fixes the CMC surfaces. Equation (6.44) can be expressed also in

the following form

$$t_{CMC}(r) = t_{shell}(R) + \int_R^r \mathrm{d}x \frac{\frac{K}{3}x - \frac{B}{x^2}}{f(x)\sqrt{f(x) + \left(\frac{K}{3}x - \frac{B}{x^2}\right)^2}}, \ r > R \tag{6.47}$$

Notice that (6.47) is a meaningful equation only if for all $r > R$, $h(r, R) \equiv f(r) + \left(\frac{K}{3}r - \frac{B(R)}{r^2}\right)^2$ remains positive (R only fixes the value of B). Clearly, this condition is satisfied for $r > 2M$. For $r < 2M$, if we neglect $\frac{K}{3}r$ term (as it is much smaller than $\frac{B}{r^2}$), there is a critical value $B_c = \frac{\sqrt{27}}{4}M^2$ above which $h(r)$ is always positive. For $B = B_c$, function $h(r)$ has a double root at $r = \frac{3}{2}M$ (Fig. 6.2). This argument shows that the behavior of CMC surfaces depend heavily on how B changes with R (Fig. 6.3).

6.4.3 $K = 0$

In this case, we get

$$t_{CMC}(r) = -\int_{r_0'}^r \mathrm{d}x \frac{\frac{B}{x^2}}{f(x)\sqrt{f(x) + \frac{B^2}{x^4}}} + t_0', \ r > R, \tag{6.48}$$

where r_0' and t_0' are constants. In the case of maximal surfaces ($K = 0$), (6.43) gives $B = B(R)$ (for $K = 0$), and (6.48) together with $t_{CMC}(R) = t_{shell}(R)$ determine the CMC surfaces for radii larger than R. Consequently,

$$t_{CMC}(r) = -\int_R^r \mathrm{d}x \frac{\frac{B(R)}{x^2}}{f(x)\sqrt{f(x) + \frac{B(R)^2}{x^4}}} + t_{shell}(R), \ r > R, \tag{6.49}$$

6.5 Emergence of the Universal Horizon

As we mentioned earlier, an observer inside $r = 1.5M$ cannot send any signal outside this radius, after some stage in collapse, even using superluminal ϕ-signals that propagate almost instantaneously. We shall demonstrate this for two cases.

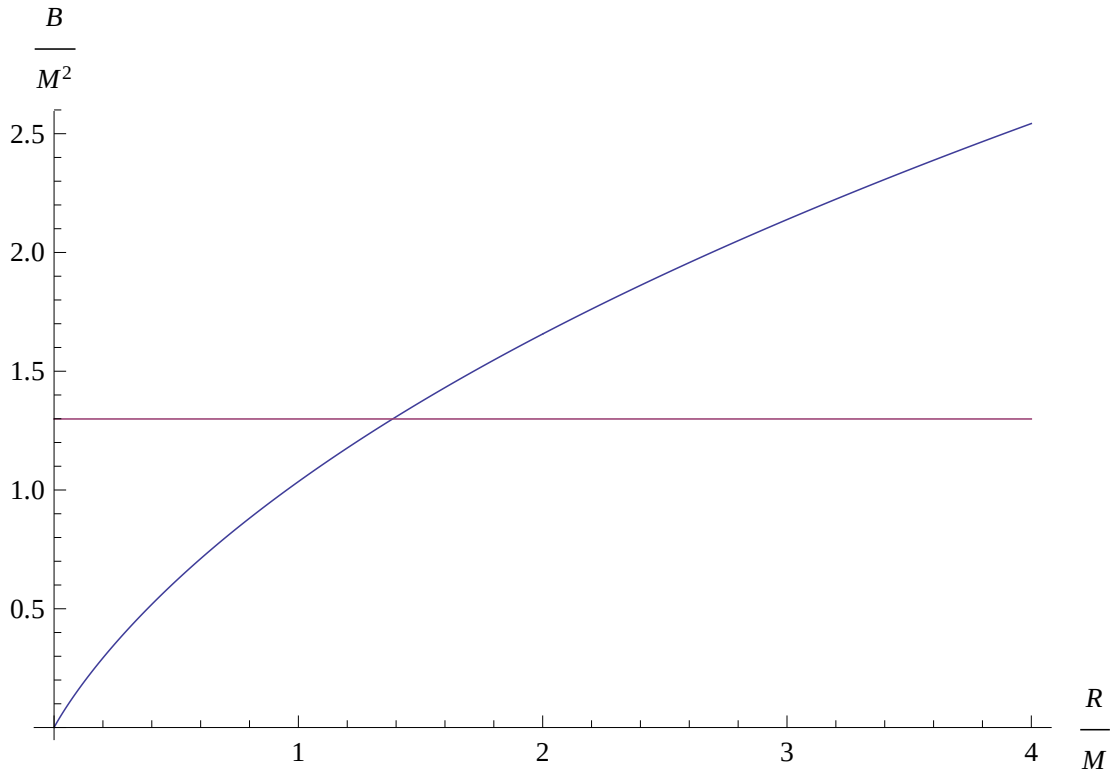


Figure 6.3: B as a function of R for $K = 0$ and $e = 1$. The horizontal line shows $B = B_c$. The radius at which $B(R) = B_c$ is called R_c. Note that $R_c < 1.5M$.

6.5.1 $V(\phi) = 0$

If we set $V(\phi) = 0$ in (6.30), we get

$$K = u^{\alpha}{}_{;\alpha} = 0, \tag{6.50}$$

implying that constant field surfaces are maximal surfaces. As we showed earlier, in order to determine the signal propagation in this background, we need to find normal vector u^{μ} to constant field surfaces. Then Equation (6.26) determines the sound cone at any point of spacetime. Using (6.38) and (6.39), we get

$$u^r_{II} = -\frac{B}{r^2}, \tag{6.51}$$

$$u^t_{II} = \frac{1}{f(r)}\sqrt{\frac{B^2}{r^4} + f(r)}, \tag{6.52}$$

where B is given by (6.43) ($K = 0$). As shown in Appendix D.1, there is always a radius $R_c \leq 1.5M$ for which the corresponding value of B is B_c, $B(R_c) = B_c$. It means that when shell's radius approaches R_c, the value of B becomes closer to B_c, and the t-component of the normal vector (u^t_{II}) at $r = 1.5M$ approaches zero. Note that R_c must be smaller than $1.5M$; otherwise the t-component of the normal vector would be u^t_I.

On the other hand, equation (6.26) at $r = 1.5M$ yields

$$-\frac{1}{3}u^t_{II}v^t + 3u^r_{II}v^r = \epsilon. \tag{6.53}$$

Consequently, (for a fixed value of ϵ) when R reaches R_c the first term in the above equation becomes negligible. As a result $v^r < 0$ (because $B > 0$) and the propagation cone becomes tilted toward the center. As a result, no signal can escape $r \leq 1.5M$. The surface $r = 1.5M$ is the "Universal Horizon" as no signal (even with infinite propagation speed) can escape from within.

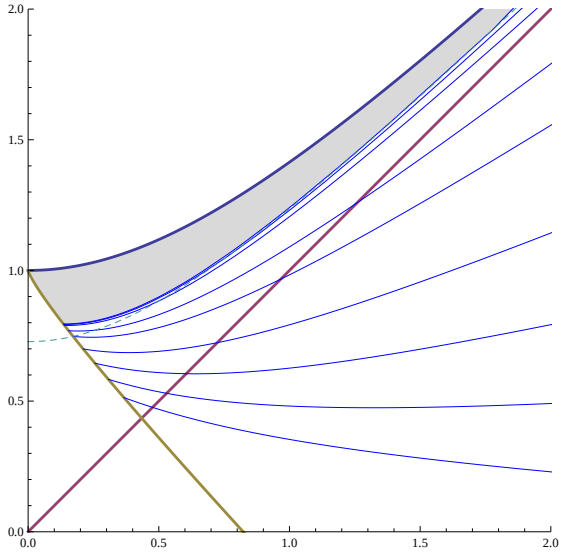
Maximal surfaces (surfaces of constant field) have been shown in Fig. 6.4 in Schwarzschild and Kruskal[2] coordinates. Close to R_c, maximal surfaces tend to stay very close to $r = 1.5M$.

6.5.2 $V(\phi) \neq 0$

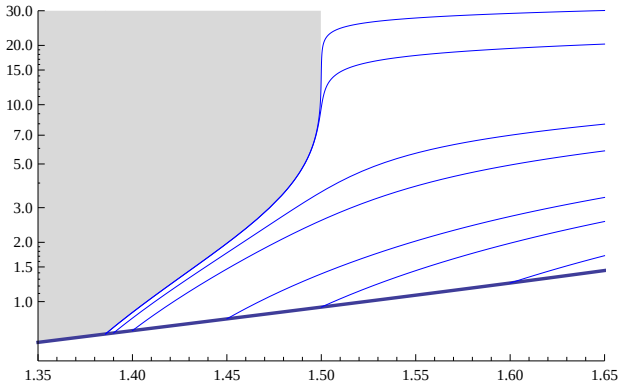
The case $V(\phi) \neq 0$ is almost the same as $V(\phi) = 0$, as long as $MK \ll 1$. This can be argued as follows. We are interested in the regions where the shell radius is of order M. As a result, solutions to (6.43), as long as $MK \ll 1$, are the same as $K = 0$. Hence a small non-zero value of K will not change the shape of the CMC surfaces at small radii (radii of the order of M), and it only affects the shape of CMC surfaces at large radii (cosmological scale). However, our derivation of the universal horizon (in the previous section) only depends on maximal surfaces inside Schwarzschild radius. Consequently, we can apply the same argument to a small non-zero value of K.

[2]Kruskal coordinates:

$$v = \left|\frac{r}{2M} - 1\right|^{\frac{1}{2}} e^{\frac{r}{4M}} \left[\sinh(\frac{t}{4M})\theta(r-2M) + \cosh(\frac{t}{4M})\theta(2M-r)\right],$$
$$u = \left|\frac{r}{2M} - 1\right|^{\frac{1}{2}} e^{\frac{r}{4M}} \left[\cosh(\frac{t}{4M})\theta(r-2M) + \sinh(\frac{t}{4M})\theta(2M-r)\right],$$



(a) Constant field surfaces in Kruskal coordinates. Thick blue, yellow and brown curves respectively represent $r = 0$, the shell's surface and $r = 2M$. Blue curves represent constant field surfaces and the dotted green curve is $r = 1.5M$. We see that after some point constant field surfaces tend to stay close to $r = 1.5M$.



(b) Constant field surfaces in Schwarzschild coordinates. The thick blue curve is the shell's surface, and blue curves are constant field surfaces. We see that after some point they tend to stay close to $r = 1.5M$.

Figure 6.4: Constant field surfaces for $e = 1$ and $M = 1$ in Kruskal and Schwarzschild coordinates. Grey area shows the region causally disconnected from infinity.

6.5.3 Is the Universal Horizon Singular?

Until now, we have ignored the effect of the cuscuton on the background geometry. This assumption is valid if the cuscuton's pressure and density remain small.

Consider a quadratic potential $V(\phi) = \lambda\phi^2$ for the cuscuton. Using the cuscuton's EOM (6.30) we find $K \propto \phi$, implying that the field value is suppressed by the Hubble parameter ($K = 3H$ where H is Hubble parameter) . As a result the cuscuton's energy density $\rho = V(\phi)$ is also Hubble-suppressed. Since the cuscuton pressure is given by the time derivative of ϕ (6.28), it is suppressed by $\dot{H}$. This argument shows that the cuscuton energy density and pressure are small. However, since constant field surfaces tend to stack around $r = 1.5M$ upon formation of the universal horizon, the above argument may not be applicable in this limit. We show here that upon forming a universal horizon, this surface is non-singular. In fact the cuscuton's pressure and density remain small when shell radius approaches R_c.

According to (6.27) the density always remains finite for a well-behaved potential and boundary condition. As an example, for $V(\phi) = \lambda\phi^2$ and matter dominated cosmology, ρ remains small (in fact, it decreases with time and approaches zero). Since the pressure is given by (6.28)

$$p = a\sqrt{X} - V(\phi), \tag{6.54}$$

we need to show that $2X = g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi$ remains finite in the limit $R \to R_c$.

In the following we assume that $V(\phi)$ is at least quadratic in ϕ. Differentiating (6.30), we get

$$\partial_\mu K = -\frac{\sqrt{2}}{a}V''(\phi)\partial_\mu\phi. \tag{6.55}$$

As a result,

$$\begin{aligned} 2X &= \frac{a^2}{2\left(V''(\phi)\right)^2} g^{\mu\nu}\partial_\mu K\partial_\nu K \\ &= \frac{a^2}{2\left(V''(\phi)\right)^2}\left[\frac{1}{f(r)}\left(\partial_t K\right)^2 - f(r)\left(\partial_r K\right)^2\right]. \end{aligned} \tag{6.56}$$

Using the identity

$$\left(\frac{\partial K}{\partial r}\right)_t = -\left(\frac{\partial t}{\partial r}\right)_K \left(\frac{\partial K}{\partial t}\right)_r, \tag{6.57}$$

together with (6.37), equation (6.56) yields

$$2X = \frac{a^2}{2\left(V''(\phi)\right)^2} \frac{\left(\frac{\partial K}{\partial t}\right)_r^2}{f(r) + \left(\frac{K}{3}r - \frac{B}{r^2}\right)^2}. \tag{6.58}$$

In order to have a singularity the following term

$$\left[f(r)+\left(\frac{K}{3}r-\frac{B}{r^2}\right)^2\right]\left(\frac{\partial t}{\partial K}\right)_r^2 \tag{6.59}$$

must approach zero. Let us provisionally assume that

$$\left(\frac{\partial t}{\partial K}\right)_r=\frac{1}{\dot{K}}-\int_r^{K^{-1}}\mathrm{d}x\frac{\frac{x}{3}-\frac{dB/dK}{x^2}}{\left[f(x)+\left(\frac{K}{3}x-\frac{B}{x^2}\right)^2\right]^{\frac{3}{2}}} \tag{6.60}$$

does not approach zero. The other term $f(r)+\left(\frac{K}{3}r-\frac{B}{r^2}\right)^2$ can reach zero upon formation of universal horizon ($R\to R_c$, $B\to B_c$) at $r=1.5M$. However, $\left(\frac{\partial t}{\partial K}\right)_{r=1.5M}$ is also diverging at the same limit. Considering (6.59) in the limit $R\to R_c$ and $r\to 1.5M$ we find

$$\begin{aligned}&\left[f(r)+\left(\frac{K}{3}r-\frac{B}{r^2}\right)^2\right]\left(\frac{\partial t}{\partial K}\right)_r^2\\&\sim(r-1.5M)^2\times\frac{1}{(r-1.5M)^4}\sim\frac{1}{(r-1.5M)^2}\end{aligned} \tag{6.61}$$

showing that not only (6.59) does not reach zero, it rather diverges as $r\to 1.5M$. Hence $\rho=-p$ at $r=1.5M$ in the limit of the formation of the universal horizon.

The behaviour of $\left(\frac{\partial t}{\partial K}\right)_r$ heavily depends on the value of dB/dK. Note that if $\left(\frac{\partial t}{\partial K}\right)_r$ becomes zero at some point, it means that two CMC surfaces crossed each other. As a result, it seems that as long as we are able to find solutions for the cuscuton, $\left(\frac{\partial t}{\partial K}\right)_r$ never vanishes. In the following, we will prove (by contradiction) that for $\dot{K}<0$, this term does not vanish.

If $\left(\frac{\partial t}{\partial K}\right)_r=0$ at some point, dB/dK must be positive; otherwise the integrand in (6.60) will be always positive and $\left(\frac{\partial t}{\partial K}\right)_r<0$.

On the other hand, $\left(\frac{\partial t}{\partial K}\right)_r=0$ means that two CMC surface have crossed each other (as in Fig. 6.5). Note that these surfaces must cross and cannot be tangent to each other because $\left(\frac{\partial t}{\partial K}\right)_r$ changes sign near its zero. It means that $R_1\leq R_2$ (where R_i is the shell radius corresponding to surface i), and consequently $B(R_1)\leq B(R_2)$, valid at least when the shell's radius is close to R_c.

If we assume that these surfaces are infinitesimally close to each other, then

$$dB/dK=\frac{B(R_1)-B(R_2)}{K_1-K_2}\leq 0. \tag{6.62}$$

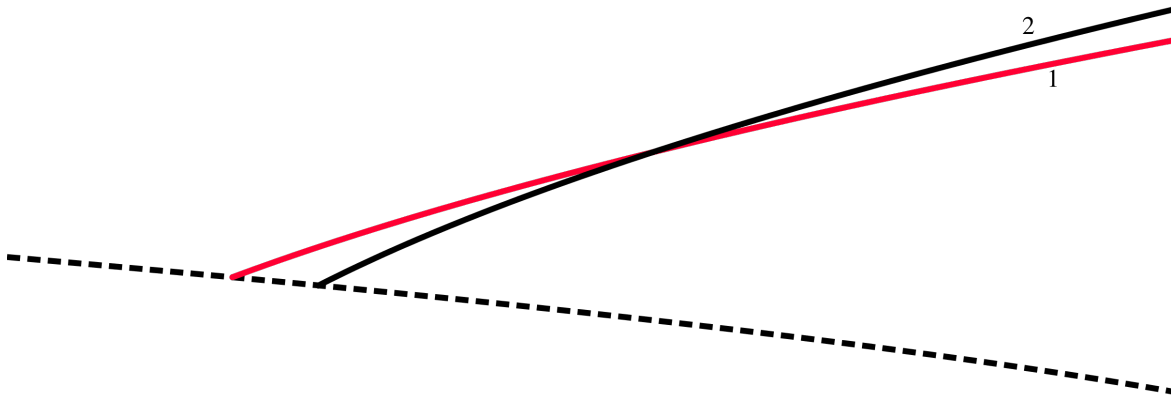


Figure 6.5: Space-time diagram showing two crossing CMC surfaces and the shell's surface (dashed line).

Note that the crossing point can be on the shell, which in this case $dB/dK = 0$. This is in contradiction with the fact that dB/dK must be positive. As a result, we have shown that for negative values of $\dot{K}$, $\left(\frac{\partial t}{\partial K}\right)$ never becomes zero.

In Appendix D.2, we have derived an approximate formula for dB/dK which further shows that $\left(\frac{\partial t}{\partial K}\right)$ does not become zero.

6.6 Summary and Conclusions

Constant field hypersurfaces of a cuscuton field represent the preferred time slicing in the low energy limit of Hořava-Lifshitz gravity. In this Chapter, we have demonstrated that, although the space-time geometry of a large black hole in Hořava-Lifshitz gravity is very similar to Schwarzschild geometry, the causal structure is completely different.

Nevertheless, we showed that as a black hole forms, there still exists an event horizon for signals with arbitrarily large speed. No matter how fast a signal propagates, it cannot escape from inside this *universal horizon* (i.e. grey area in Fig. 6.4), as seen in static solutions previously [76, 77, 78]. If this was not the case, one could have imagined that

signals originating from regions close to the singularity could in principle propagate outside, leading to a *naked singularity*, and rendering the classical theory unpredictive. Instead, the emergence of a universal horizon during the formation of the black hole implies that a version of *Cosmic Censorship* might still hold here. These results are also consistent with earlier studies of the gravitational collapse of K-essence matter [80, 81], in which sonic horizons could form inside luminal horizons, and gravitational collapse in the context of Einstein-aether theories [84].

This causal structure has an additional interesting property. The universal horizon relates the finite time coordinate inside black hole to future infinity (outside). As a result, any observer falling into black hole will hit the universal horizon at a finite proper time, prior to which she can, in principle, see events that happen outside black hole at arbitrarily late times (if she can see arbitrarily superluminal signals in the preferred frame). For example, the black hole itself radiates its mass through Hawking evaporation, with quantum mechanical effects becoming important when the black hole radiates a substantive portion of its mass. While, for an observer outside a massive black hole, this takes a long time ($\sim M^3$), an observer falling into black hole can detect these quantum mechanical effects (via superluminal contact with outside) within a much shorter time ($\sim M$), just before hitting the universal horizon. In this sense, the universal horizon can be considered the causal future boundary of the classical space-time.

This realization could also be intimately related to the claim that, while the universal horizon in our spherically symmetric system is regular, it is unstable to aspherical perturbations that might change it to a singular surface [76]. On the other hand, [82, 85] argue that, similar to the ordinary null horizons, universal horizons may radiate particles, and a fixed temperature and entropy can be assigned to them. However, it seems that the instability of universal horizon by aspherical perturbations is incompatible with the derivation of horizon temperature.

Another important issue regarding Lorentz violating theories is their apparent tension with generalized 2nd law. It has been argued that in a theory with two different fields A and B with different speeds c_A and c_B where each has its own horizon around black hole, one can violate generalized 2nd law with building a perpetual motion machine (see for example [86, 87]).

In the next Chapter, we consider the case of charged and spinning black holes.

Chapter 7

Cosmic Censorship in Lorentz Violating Theories of Gravity

7.1 Introduction

Theory of general relativity (GR) has been successful in describing a wide range of phenomena, from solar system to cosmological scales. In addition to being consistent with various experiments, the mathematical elegance of the theory is very appealing. Diffeomorphism invariance, at the core of GR, gives a straightforward constructive way of building the theory. In fact, GR is the simplest diffeomorphism invariant theory for metric.

From observational point of view, there is no reason to abandon this theory. GR is compatible with a wide variety of experimental constraints[1]. On the other hand, many attempts have shown so far that modifying GR is a tricky task, and one often faces physically unacceptable results, e.g. the appearance of Boulware-Deser ghost in massive gravity [88] and ghost degrees of freedom in quadratic gravity [89].

However, studying non-GR theories of gravity is still valuable, and the main reason stems from quantizing gravity. GR, while being a very successful classical theory, has failed to cope with quantum mechanics. Therefore, one approach to quantum gravity has been to abandon diffeomorphism invariance, as e.g., done in the celebrated Horava-Liftshitz gravity [13]. However, the strict empirical constraints on violations of equivalence principle

[1] Although there have been various attempts to solve the problems of dark matter and dark energy with GR modifications, simple solutions to these problem in the context of GR exist. In other words, there is no apparent observational contradiction with GR which necessitates GR modifications.

requires observable deviations from Lorentz symmetry to be limited to high energies, which provides a challenge for the construction of these theories (e.g., [90]).

In different examples of theories with broken Lorentz invariance, superluminal degrees of freedom appear (see [14, 15]). The existence of superluminal excitations (SLE) points out that a different causal structure exists in these theories compared to GR, even when the back-reaction of these excitations on the geometry is negligible. This property is especially of significance in the black hole (BH) solutions. While potentially SLE can escape the traditional Killing horizon of a BH and make the classical theory unpredictable [91], it has been shown in many examples [92, 93, 78, 94, 95, 79, 77, 76] that a notion of horizon (called universal horizon) still exists in these theories. Moreover, universal horizon (UH) thermally radiate and satisfies the first law of horizon thermodynamics[2] [96, 82, 85, 97] (but also see [98]). Studying the notion of universal horizon and its temperature and entropy is important since it guides us to better understanding the structure of UV theory.

Following our discussion from the previous Chapter, here we study the universal horizon formation in dynamical or stationary spacetimes with an inner Killing horizon, in the limit of infinite sound speed for excitations (i.e. *incompressible* limit). In order to do so, we make use of the fact that surfaces of global time (defined by the background field), in the incompressible limit of Lorentz violating theories, coincide with constant mean curvature (CMC) surfaces of the spacetime (see Section 6.3). Furthermore, the backreaction of the incompressible field on the spacetime geometry is negligible as long as the event horizon is much smaller than the cosmological horizon [99]. In the next Section, we show how the universal horizon forms in a dynamical setting, in the collapse of a charged shell, and we derive a formula for the radius of the universal horizon in terms of the charge. In Section 7.3.1, we propose a geometric definition for universal horizon. This allows us to study the universal horizon for spinning black holes. In 7.3.2 we show that there are three axisymmetric surfaces which satisfy the conditions of a universal horizon. As we show, this means that two families (with infinite numbers) of axi-symmetric universal horizons in Schwarzschild case exist. Section 7.4 concludes this Chapter.

7.2 Universal Horizon in Dynamic Reissner–Nordstrom Geometry

We start this section by finding CMC slicing of dynamic Reissner–Nordstrom (RN) geometry. As we mentioned earlier, CMC surfaces of this spacetime are the constant global time

[2] so far only for spherically symmetric solutions

surfaces of the background incompressible field, and they define the new causal structure imposed by this field. Once we derive the CMC slicing, we focus on the (universal) horizon formation in this geometry.

7.2.1 CMC Surfaces in a Dynamic Reissner–Nordstrom Geometry

In order to examine the formation of the universal horizon in a dynamic Reissner–Nordstrom geometry, one must first describe surfaces of constant mean curvature for a collapsing charged massive spherical shell. An examination of CMC surfaces has been similarly looked at in the restricted case of maximal surfaces ($K = 0$) [100]. The dynamics of the collapse itself is well known and described by Israel [101]. Describing the metric in the standard way has the geometry inside the shell as flat and the RN outside. We write this geometry as:

$$\begin{aligned} \mathrm{d}s^2 &= f_-(r)\mathrm{d}t_-^2 - f_-(r)^{-1}\mathrm{d}r^2 - r^2\mathrm{d}\Omega^2 \qquad && (r < R) \\ \mathrm{d}s^2 &= f_+(r)\mathrm{d}t_+^2 - f_+(r)^{-1}\mathrm{d}r^2 - r^2\mathrm{d}\Omega^2 \qquad && (r > R) \end{aligned}$$

where $f_-(r) = 1$ and $f_+(r) = 1 - \frac{2M}{r} + \frac{Q^2}{r^2}$ in $G = c = 1$ units. The parameters are the gravitational mass M and shell's charge Q. For simplicity we will often use relative charge $q = Q/M$. While the spherical coordinates are shared between the inner and outer regions, the time coordinates t_- and t_+ correspond to the Minkowski and RN time respectively.

Let the family of spacelike CMC surfaces be denoted by $\Sigma_K(t_g)$ where t_g is a global time coordinate that is constant for each surface. The timelike normal vector to this surface is labelled v^μ. The CMC condition implies $\nabla_\mu v^\mu = K$, resulting in:

$$\frac{\partial}{\partial t_\pm} v^{t_\pm} + \frac{1}{r^2}\frac{\partial}{\partial r} r^2 v^r = K \tag{7.1}$$

If we denote $B \equiv -r^2 v^r$ and use the normalization condition $v_\mu v^\mu = 1$ then:

$$r^2 v^{t_\pm} = \pm f_\pm(r)^{-1}\sqrt{B^2 + f_\pm r^4}. \tag{7.2}$$

For now we use the '+' case so that for $v_t > 0$ for $r \gg M$. Additional explanation and the cases where '−' is relevant will be seen in Section 7.2.4. Combining this result with (7.1) we get

$$\frac{B}{f_\pm(r)\sqrt{h(r,B)}}\frac{\partial}{\partial t_\pm}B - \frac{\partial}{\partial r}B = Kr^2 \tag{7.3}$$

with $h(r,B) = B^2 + f_\pm r^4$. The characteristic equations of (7.3) are simply:

$$\frac{\mathrm{d}t_\pm}{\mathrm{d}s} = \frac{B}{f_\pm(r)\sqrt{h(r,B)}}, \quad \frac{\mathrm{d}r}{\mathrm{d}s} = -1, \quad \text{and} \quad \frac{\mathrm{d}B}{\mathrm{d}s} = Kr^2, \tag{7.4}$$

for some parameter s. Using the second equation of (7.4) to integrate the first and third equations results in:

$$t_\pm = (t_\pm)_0 - \int_{r_0}^{r} \frac{B\mathrm{d}r}{f_\pm(r)\sqrt{h(r,B)}}, \quad \text{and} \quad B = \frac{K}{3}(r^3 - r_0^3) + B_0, \tag{7.5}$$

where $(t_\pm)_0$, r_0, and B_0 are integration constants. In order to fix these constants, we examine the internal and external cases separately.

1) Inside the shell: If $r_0 = 0$ then $B(r=0) = B_0$. If $B_0 \neq 0$ this would lead to a contradiction, as $v^r = \frac{-B}{r^2}$ should be finite in the flat geometry. Therefore with $r_0 = 0$, the equation reduces to:

$$t_- = (t_-)_0 + \int_0^r \frac{Kr^3\mathrm{d}r}{3\sqrt{(\frac{Kr^3}{3})^2 + r^4}} \text{ and } B = \frac{K}{3}r^3 \tag{7.6}$$

2) Ouside the shell: Let $r_0 = R((t_+)_0)$, we can determine B_0 by looking at the boundary between the flat and RN spaces. Projecting the vector v^μ along the shell should give us continuous observable values. The shell timelike path comes from $S = R(t_\pm) - r = 0$ which creates the unit normal vector and tangent vector labelled as n^μ and u^μ respectively. If we choose the sign of the normalization factors such that $u^r < 0$, the vectors take the form of:

$$n_-^\mu = \frac{g^{\mu\nu}}{N_-}(\nabla_-)_\nu S = \frac{1}{N_-}(\frac{\mathrm{d}R}{\mathrm{d}t_-}, 1, 0, 0), \tag{7.7}$$

$$u_-^\mu = \frac{1}{N_-}(1, \frac{\mathrm{d}R}{\mathrm{d}t_-}, 0, 0), \tag{7.8}$$

$$N_-^2 = 1 - (\frac{\mathrm{d}R}{\mathrm{d}t_-})^2, \tag{7.9}$$

inside the shell, while outside takes the from of:

$$n_+^\mu = \frac{g^{\mu\nu}}{N_+}(\nabla_+)_\nu S = \frac{1}{N_+}(f_+^{-1}\frac{\mathrm{d}R}{\mathrm{d}t_+}, f_+, 0, 0), \tag{7.10}$$

$$u_+^\mu = \frac{1}{N_+}(1, \frac{\mathrm{d}R}{\mathrm{d}t_+}, 0, 0), \tag{7.11}$$

$$N_+^2 = \frac{f_+^2 - (\frac{\mathrm{d}R}{\mathrm{d}t_+})^2}{f_+}. \tag{7.12}$$

We wish to find functions $C(R)$ and $D(R)$, such that:

$$v_-^\mu = Cn_-^\mu + Du_-^\mu. \tag{7.13}$$

From inside the shell $v^\mu(R) = (1, 0, 0, 0)$ which means $C = \frac{-1}{N_-}\frac{\mathrm{d}R}{\mathrm{d}t_-}$ and $D = \frac{1}{N_-}$. Requiring projections (C and D) to be the same from outside, we get:

$$B_0 = -R^2(C(n_+)^r + D(u_+)^r) = \frac{R^2}{N_+N_-}\left(\frac{\mathrm{d}R}{\mathrm{d}t_+} - f_+\frac{\mathrm{d}R}{\mathrm{d}t_-}\right). \tag{7.14}$$

So, if we specify the dynamics of the shell $\frac{dR}{dt_\pm}$, all the parameters are fixed.

The description of the radial velocity comes from Israel and De La Cruz [101]:

$$\left(\frac{\mathrm{d}R}{\mathrm{d}t_-}\right)^2 = 1 - \frac{R^2}{(\epsilon R - b)^2}, \tag{7.15}$$

$$\left(\frac{\mathrm{d}R}{\mathrm{d}t_+}\right)^2 = f_+^2 - \frac{f_+^3 R^2}{(\epsilon R - b - \frac{m}{\epsilon})^2}, \tag{7.16}$$

where $\epsilon = \frac{M}{\mathcal{M}}$ and $b = \frac{M(\epsilon^2 q^2 - 1)}{2\epsilon}$ with $\mathcal{M}$ denoting the total rest mass. We can use (7.15) and (7.16) to reduce $N_+ = \frac{Rf_+}{\epsilon R - b - \frac{M}{\epsilon}}$ and $N_- = \frac{R}{\epsilon R - b}$. Note that N_+ changes signs to enforce $u^r < 0$, becoming negative only when $\frac{\mathrm{d}R}{\mathrm{d}t_\pm}$ flips signs. These choices simplifies B_0 to:

$$B_0 = \frac{M}{\epsilon}\sqrt{(\epsilon R - b)^2 - R^2}. \tag{7.17}$$

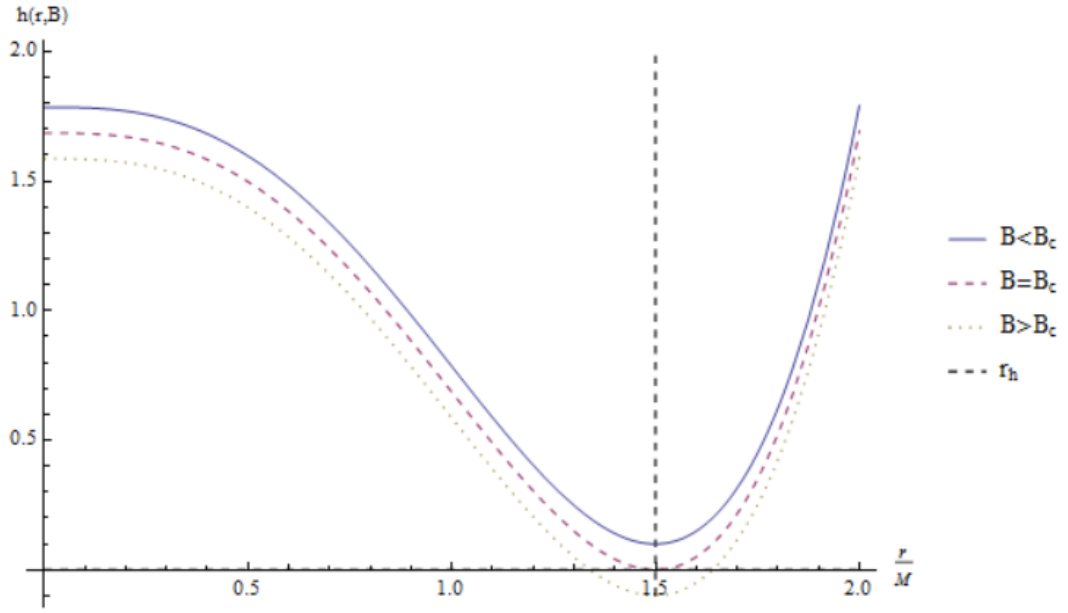


Figure 7.1: $h(r, B)$ with sub-critical, post-critical and critical B for $Q = 0$.

7.2.2 Horizon Formation

Following the analysis of previous Chapter, we examine the properties of t_+. The behaviour of t_+ heavily depends on $h(r, B)$. While B is large, which corresponds to large R, $h(r, B)$ is never vanishing. However when a critical value B_c is reached, $h(r, B_c)$ has double root at a particular value of r labelled r_h (see Figure 7.1). Something interesting will occur when r_h is larger than the radius of the shell for which B_c occurs named R_{lc} or radius of *last contact* ($B(R_{lc}) = B_c$). A signal sent out from the shell at R_{lc} will proceed out to r_h but takes infinitely long time to ever reach this radius. In fact, signals sent just outside R_{lc} will form an envelope around r_h staying at this radius longer and longer, as R_{lc} is approached, before escaping to infinity (see Figure 7.3). The values of B_c and r_h can be found by finding the solutions to $h(r, B) = \frac{\partial h(r,B)}{\partial \mathrm{r}} = 0$. We examine this equation in two different cases.

Case 1: $K = 0$

Equations for the double root reduce to:

$$r_h^4 - 2Mr_h^3 + Q^2 r_h^2 + B_c^2 = 0 \tag{7.18}$$
$$2r_h^3 - 3Mr_h^2 + Q^2 r_h = 0 \tag{7.19}$$

to which the solutions with non-negative real B_c are the trivial $r_h = B_c = 0$ and

$$r_h = \frac{3M}{4} + \frac{M}{4}\sqrt{9 - 8q^2} \tag{7.20}$$
$$B_c = r_h\sqrt{-r_h^2 + 2mr_h - Q^2} = r_h^2\sqrt{-f_+(r_h)}. \tag{7.21}$$

It is of interest to note that the UH is always between the inner and outer Killing horizons of the metric (see Figure 7.2).

Case 2: $K \neq 0$

Equations for the double root are written as:

$$\frac{K^2}{9}r_h^6 + r_h^4 + (\frac{2KB_0}{3} - 2M)r_h^3 + Q^2 r_h^2 + B_0^2 = 0,$$
$$\frac{K^2}{3}r_h^5 + 2r_h^3 + (kB_0 - 3M)r_h^2 + Q^2 r_h = 0.$$

The non-trivial solution for B_c is:

$$B_c = r_h\sqrt{-r_h^2 + 2Mr_h - Q^2} - \frac{Kr_h^3}{3} = r_h^2\sqrt{-f_+(r_h)} - \frac{Kr_h^3}{3}, \tag{7.22}$$

however the solution for r_h can be at best expressed perturbatively in K. To linear order the solution is:

$$r_h = r_h^0 - K\frac{(r_h^0)^3\sqrt{-f_+(r_h^0)}}{M\sqrt{9 - 8q^2}} + \mathcal{O}(K^2), \tag{7.23}$$

where $r_h^0 = \frac{3M}{4} + \frac{M}{4}\sqrt{9 - 8q^2}$. Assuming that the expansion of the background field is negligible (for example fixed by cosmology, as $K = 3\times$Hubble constant), in the region of interest $0 < r < 2M$ the effect of terms containing K are insignificant. From here we set $K = 0$ as its effect will only come into play when looking at the causal structure when R is very large.

The last unknown of the universal horizon is where it begins before it asymptotes to r_h. To solve for R_{lc} we use that $B(R_{lc}) = B_c$ and it results in either:

$$R_{lc} = \frac{b}{2} - \frac{B_c^2}{2bM^2}, \tag{7.24}$$

for $\epsilon = 1$ or

$$R_{lc} = \frac{\epsilon b + \sqrt{\epsilon^2(\epsilon^2-1)B_c^2/M^2 + b^2}}{\epsilon^2 - 1}, \tag{7.25}$$

for $\epsilon > 1$.

We take the larger of the solutions to the quadratic, as the first instance of B_c will create the behaviour desired.

7.2.3 Inside the Universal Horizon

The foliation can be extended for $B < B_c$, however some subtleties arise. Denote the unit tangent vector of the CMC surfaces at the point the surface intersects the shell as s^μ which in components can be written as:

$$s^\mu = \frac{1}{N_s}(T'_{cmc}(R), 1, 0, 0), \tag{7.26}$$

$$N_s^2 = \frac{1 - f_+^2 T'^2_{cmc}}{f_+} = \frac{R^4}{h(R, B(R))}, \tag{7.27}$$

where the last equality comes from using the derivative of (7.5). In general $h(r, B(R))$ is a quartic that can not easily be factored, however when restricted to the surface of the shell it can be factored to:

$$h(R, B(R)) = \left(R^2 - MR + \frac{Mb}{\epsilon}\right)^2. \tag{7.28}$$

Thus one can write the normalization factor as:

$$N_s = \frac{R^2}{R^2 - MR + \frac{Mb}{\epsilon}}. \tag{7.29}$$

As a result, between the zeroes of $1/N_s$ at $M\frac{1\pm\sqrt{1-4b/\epsilon}}{2}$ we get $s^r < 0$. In particular this means that rather than increasing in r the CMC surfaces that intersect between these two roots will have a strictly decreasing r coordinate. Moreover it is precisely at these points

where s^t switches signs corresponding to the second solution for T'_{cmc} which comes from the '-' solution to $r^2 v^{t+}$.

The second complication occurs when R does not lie between the roots of N_s while still being less than R_{lc}. Here the CMC surface increases its radial coordinate initially only to encounter a zero of $h(r, B(R))$ at r_{turn}. The integral for t_+ can carried out since T'_{cmc} only depends on the square root of $h(r, B(R))$ and, by construction, r_{turn} is only a first order zero of $h(r, B(R))$. After this point T'_{cmc} switches signs and the r coordinate begins decreasing, flipping the direction of the integration taking the surface from r_{turn} to 0.

Now that these subtleties are understood, we are ready to examine the structure of the complete foliation.

7.2.4 Foliation Structure

We will break up the discussion into several sections. For all our analysis we consider the shell to be dropped from infinity thus $\epsilon \geq 1$ [101], in particular the inward velocity of the shell at infinity is exactly $\sqrt{\epsilon^2 - 1}$. There are 4 cases of interest:

Case 1: $Q = 0$

When restricted to Schwarzschild, $b = \frac{-M}{2\epsilon}$ making it strictly negative. In particular the value of ϵ is only relevant to the radius of last contact, and so without losing any depth of examination we set $\epsilon = 1$. Figures 7.3 and 7.4a illustrate the foliations created by the CMC surfaces for $R > R_{lc}$ and the final well defined CMC surface that creates the universal horizon in Schwarzchild and Kruskal-Szekeres coordinates. Once a conformal compactification has been performed the casual structure is clear in Figure 7.4b with the additional sub-UH CMC surfaces (which end in singularity, rather than the space-like infinity i^0).

Case 2: $Q \neq 0$ & $b \leq 0$

For ϵ small enough such that the numerator of b remains negative, the shell is unable to rebound before collapsing to a singularity. In Schwarzchild and the charged generalization of Kruskal-Szekeres remain almost identical in their analogous charts. The causal structure in Figure 7.5 reveals the distinction from case 1. The collapse ends in the coordinates, colloquially called the *parallel universe*.

Case 3: $Q \neq 0$ **&** $0 < b < b/\epsilon + 1/\epsilon^2$

For the range of ϵ such that $0 < b < \frac{M}{\epsilon^2}$, the shell rebounds at the radius of $\frac{b}{\epsilon-1}$ but in the parallel coordinates which distinguishes it from the next case. In particular this means that $t(R)$ has a stationary point between r_+ and r_-. Figure 7.6 shows the collapse and the corresponding causal structure for this case.

Case 4: $Q \neq 0$ **&** $b \geq b/\epsilon + 1/\epsilon^2$

Subsequently for $b > \frac{M}{ep}$ shell rebounds at the radius of $\frac{b}{\epsilon-1}$ in the original coordinate charts or exactly where the original and parallel coordinates meet . In particular this means that $t(R)$ has a stationary point at or inside r_-. The Schwarzchild and Kruskal-Szekeres coordinates are again nearly indistinguishable from case 1 except when the placement of R_{lc} requiring that the UH being in a second charts however this does not reveal any new structure. Figure 7.7 and 7.8 represents paths within this case, $b = b/\epsilon + 1/\epsilon^2$ and $b > b/\epsilon + 1/\epsilon^2$ respectively. In this final case $R_{lc} < r_-$ resulting in the the UH piercing the r_-. Nevertheless, the singularity and the parallel interior horizon, which is considered to be unstable [102] is still hidden within the UH.

7.2.5 Censorship in Reissner-Nordstrom

Ultimately, in all the above cases, the UH shields any singularity from being probed even using superluminal signals and preserves a sense of cosmic censorship in Lorentz violating theories. It is apparent from the structure that every CMC is terminated at i^0, i^+, $i^{-\prime}$,$i^{0\prime}$ or the singularity. We would posit an analogous, although informally made, statement to the original *weak* cosmic censorship conjecture: the set of points which can be connected to i^0 with CMC surfaces (an analogous property of being in the causal past of $\mathcal{I}^+$) is distinct from the set of points which can be connected to the singularity. Moreover, the boundary between these two sets will exactly be the universal horizon.

Even though we have plotted the maximal foliation of spacetime beyond the universal horizon, one can argue that the self-consistent evolution of the Lorentz-violating theory (including, e.g., backreaction or quantum effects, which we have ignored) stops at the universal horizon, which can be viewed as the boundary of classical spacetime (see Sec. VI in [99] for more discussions). To see this more explicitly, note that the CMC surfaces beyond universal horizon end in singularity. Therefore, arbitrarily fast communication along these surfaces leads to a breakdown of initial value formulation in this region. The

fact that both curvature singularity and the (potentially unstable) inner Killing horizon [102] lie beyond the universal horizon, further suggests a notion of *strong* cosmic censorship.

Note that in the cases where a parallel universe exists, a second UH acts as a white hole horizon for superluminal signals.

7.3 Spinning Black Holes: A Tale of Three Horizons

7.3.1 Geometrical Definition of Universal Horizon

In the previous Section, we discussed the formation of universal horizon in a dynamic RN geometry. Before moving on to the spinning black hole case, it would be illuminating to acquire more intuition about the geometric nature of the universal horizon. We start by asking the following question: is there a way of finding the universal horizon in the final geometry (after collapse completed) without knowing the details of collapse?

Let's consider the Schwarzschild case ($Q = 0$ collapse). CMC surfaces in the thin shell collapse geometry describe the surfaces of constant global time. As we discussed earlier, as long as we are interested in the behaviour of these surfaces near the black hole (small radii) we can treat them as maximal surfaces ($K = 0$). Maximal surfaces in this geometry inside the Schwarzschild radius asymptote to $r = \frac{3}{2}M$ before escaping to infinity. This suggests that $r = \frac{3}{2}M$ itself should be a maximal surface. In fact, one can simply verify that $r = r_*$ is a maximal (space-like) surface in Schwarzschild spacetime, only if $r_* = \frac{3}{2}M$.

This observation suggests a geometrical definition for (asymptotic) universal horizon in stationary spacetimes; it is a maximal space-like hypersurface which is invariant under the flow of time-like Killing vector. Let's discuss each element of this definition.

First of all, UH has to be a maximal surface as we described earlier. It also has to be space-like, since it describes a constant global time surface. Secondly, it is invariant under time translation, as it is the asymptotic surface of the maximal slicing.

We will show later explicitly that this definition does not pick a unique hypersurface. However, this should not be surprising, since the position of universal horizon depends on the behaviour of the background incompressible fluid (which defines the global time), unlike the Killing horizon where its position is independent of the behaviour of the background fields [99, 83]. Again, let's consider Schwarzschild spacetime. If we use the given definition of UH *with the additional assumption of spherical symmetry*, there is a unique solution of $r = \frac{3}{2}M$. However, there are many non-spherical UHs in the same geometry.

Before moving to the spinning case, let's find the spherical UH of RN geometry using the definition given above. Assume $r = r_h$ to be the universal horizon and v_μ the unit normal vector to this surface. Solving

$$\nabla_\mu u^\mu = 0, \tag{7.30}$$

we get

$$f'(r_h) = -4f(r_h), \tag{7.31}$$

with $f(r) = 1 - \frac{2M}{r} + \frac{Q^2}{r^2}$. Eq. (7.31) has a unique solution, which coincides with our previous result for universal horizon (7.20), and is plotted in Figure (7.2). One can also directly check that (7.31) is equivalent to system of equations (7.18) and (7.19).

7.3.2 Universal Horizon in Kerr geometry

In this Section, we find the asymptotic universal horizon of Kerr metric. Given our definition above, it is a static axisymmetric[3] (space-like) maximal surface. We express the Kerr metric in the following coordinates:

$$\begin{aligned} ds^2 = &\left(1 - \frac{2mr}{\rho^2}\right) dt^2 - \frac{\rho^2}{\Delta} dr^2 - \rho^2 d\theta^2 - \frac{4mra\sin^2\theta}{\rho^2} dt d\phi \\ &- \left(r^2 + a^2 + \frac{2mra^2\sin^2\theta}{\rho^2}\right) \sin^2\theta d\phi^2 \end{aligned} \tag{7.32}$$

where

$$\rho^2 = r^2 + a^2\cos^2\theta, \tag{7.33}$$

$$\Delta = r^2 - 2mr + a^2. \tag{7.34}$$

The inner (r_-) and outer (r_+) Killing horizons are the solution to $\Delta = 0$.

UH is the surface

$$r = r_h(\theta) \tag{7.35}$$

which satisfies

$$\nabla_\mu v^\mu = 0 \tag{7.36}$$

where v^μ is the (time-like) normal vector to the universal horizon. In other words,

$$v_\mu = \frac{1}{N}(0, 1, -r_h', 0) \tag{7.37}$$

[3] we assume that the background incompressible field obeys the axial symmetry of Kerr geometry.

where $'$ is the derivative w.r.t θ and N is the normalization factor

$$N^2 = -\frac{1}{\rho^2}\left(\Delta + r_h'^2\right). \tag{7.38}$$

Equation (7.38) leads to the following conclusion: demanding UH to be a space-like surface (v^μ to be time-like) requires the UH to be positioned between the inner and the outer Killing horizons

$$N^2 > 0 \to \Delta < 0 \to r_- < r_h(\theta) < r_+. \tag{7.39}$$

Now on to finding $r_h(\theta)$: Equation (7.36) takes the form

$$\begin{aligned} &2(r_h - m) + \frac{r_h(r_h^2 - 2mr_h + a^2)}{r_h^2 + a^2\cos^2\theta} - \frac{(r_h - m)(r_h^2 - 2mr_h + a^2)}{r_h^2 - 2mr_h + a^2 + r_h'^2} \\ &= \frac{r_h'}{\tan\theta} + r_h'' - r_h'\left[\frac{a^2\sin\theta\cos\theta}{r_h^2 + a^2\cos^2\theta} + \frac{r_h' r_h''}{r_h^2 - 2mr_h + a^2 + r_h'^2}\right]. \end{aligned} \tag{7.40}$$

One way to find the solution of this differential equation is to expand $r_h(\theta)$ in powers of a

$$r_h(\theta) = m\sum_{n=0}\frac{a^n}{m^n}r^{(n)}(\theta) \tag{7.41}$$

and solve the differential equation order by order. At zero order ($a = 0$), we expect $r^{(0)} = \frac{3}{2}$. At any higher order, we find Legendre differential equation. Requiring finite solution at $\theta = 0$ and $\theta = \pi$, this gives us a unique solution at any order. Here is the solution up to the order a^4:

$$\begin{aligned} r^{(2n-1)} &= 0, \quad n \in \{1, 2, \cdots\} \\ r^{(0)} &= \frac{3}{2} \\ r^{(2)} &= -\frac{1}{36}\cos^2\theta - \frac{13}{36} \\ r^{(4)} &= \frac{49}{10692}\cos^4\theta + \frac{29}{4752}\cos^2\theta - \frac{1057}{14256}. \end{aligned}$$

Surprisingly though, upon solving (7.40) numerically, we have found two more solutions that are different from (7.41) and do not approach to $r_h = \frac{3}{2}m$ as $a \to 0$ (see Figure 7.9).

The case of $a = 0$ is interesting, since the background geometry is spherically symmetric, and yet we have found two axisymmetric UHs. Moreover, we can always perform a rotation and get two other axisymmetric UHs. This means that there are two families (with infinite number in each family) of axisymmetric UHs in Schwarzschild spacetime.

7.4 Conclusion

In the incompressible (or infinitely fast propagation speed) limit of many Lorentz-violating theories of gravity, surfaces of constant mean curvature define the preferred foliation [99] (for a careful study of theories with infinite sound speed see [94]). For such theories, one may worry that superluminal signal propagation may lead to naked singularities. In this Chapter, we have shown that a *universal* horizon always forms when a charged spherical shell collapses to form a Reissner–Nordstrom black hole. Evidence that causal horizon formation will take place in Lorentz-violating theories supports a conjecture similar to cosmic censorship in General Relativity.

We see that the universal horizon acts almost like an extension of i^+, since any observer approaching the UH will pass through all future CMC surfaces outside the UH. Consequently, the analysis conducted here is likely only valid in the classical regime (ignoring quantum effects like the evaporation of BH). As a result the region close to the UH is likely where non-classical effects will begin to become relevant. Making claims past this region may require the full UV theory.

We have also presented a geometric definition for the UH which provides a tool for finding generic solutions in non-spherically symmetric geometries. This tool is additionally valuable as the full evolution of the system up to the point of UH formation is not needed to be explored. In particular, we show how the definition can be applied to the Kerr geometry, revealing a family of solutions in a non-spherical geometry.

Slowly rotating black hole solutions of Einstein-Aether theory has been studied in [103]. Specifically, in the limit that the spin-0 mode of Aether propagates infinitely fast, slowly rotating black hole solutions of Einstein-Aether and Horava-Lifshitz are the same and they possess a universal horizon. Corrections to the location of the universal horizon does not appear in the first order of rotation parameter (Equation (104) in [103]) which is consistent with our result in the previous section.

Additional horizon solutions may be real and the result of different (possibly more generic) collapse histories; or just an artifact of our definition which does not single out the correct universal horizon. Since the spherical universal horizon in Schwarzschild case is suspected to become singular [76] by aspherical perturbations, the outermost universal horizon (that we have found) can potentially shield this singularity and save cosmic censorship. This further motivates numerical dynamical studies of *non-spherical* collapse in Lorentz-violating gravitational theories.

Gravitational dynamics within real black holes may yet have more surprises in store for us!

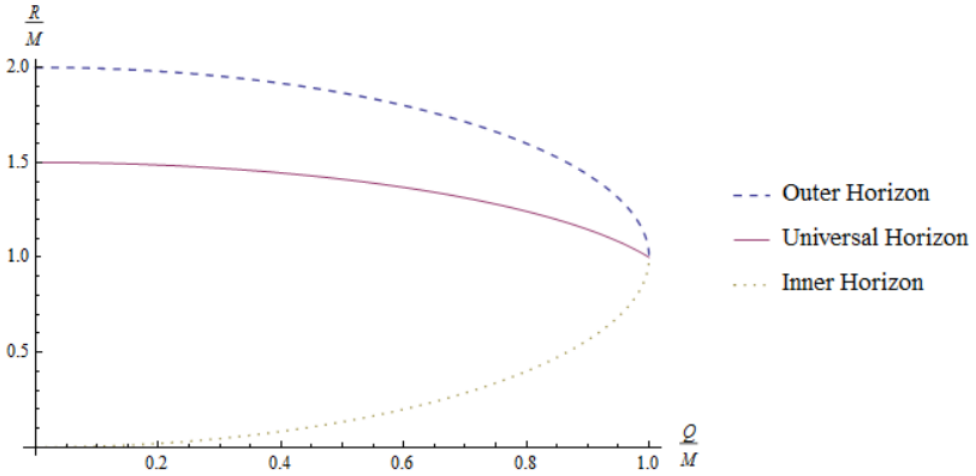


Figure 7.2: The outer, inner, and universal horizons for $K = 0$ and varying Q

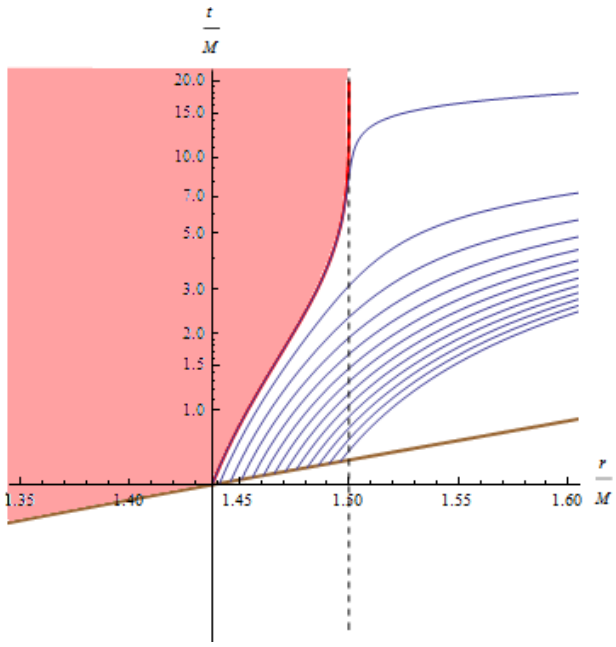
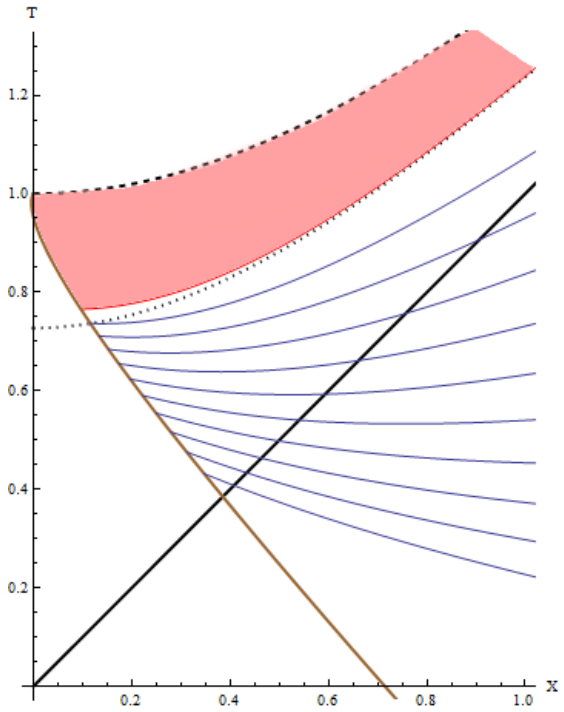
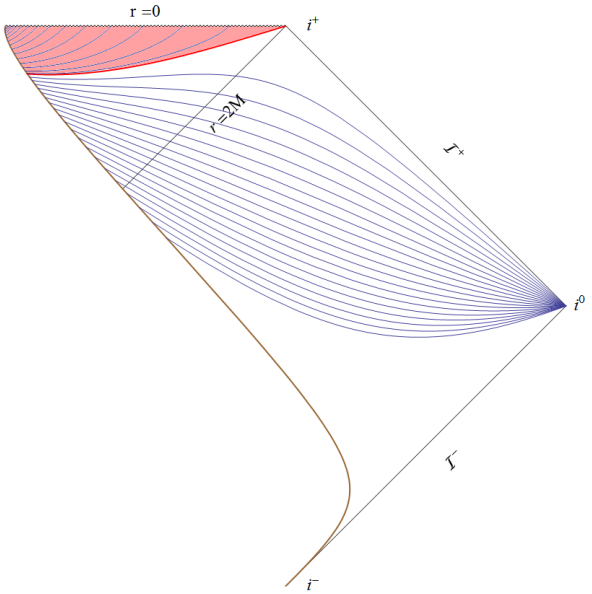


Figure 7.3: The universal horizon formation for $Q = 0$ in Schwarzschild coordinates. The blue lines represent CMC surfaces, the lowest brown line where the CMC line originate from is the shell's surface, the red line and the boundary of the shaded region is the universal horizon (UH).



(a) The universal horizon formation for $Q = 0$ in Kruskal-Szekeres coordinates. The lines/region have the same meaning as Figure 7.3, additionally ,the dotted black is the radius that UH asymptotes to, the thick black line represents the null horizons, and the dashed line the singularity.



(b) The Penrose diagram for a $Q = 0$ collapsing shell depicting the UH horizon. The lines/region have the same meaning as Figure 7.3 with the inclusion of sub-UH CMC surfaces in blue.

Figure 7.4: Surfaces for constant global time and formation of the universal horizon in Kruskal-Szekeres coordinates and Penrose diagram for $Q = 0$.

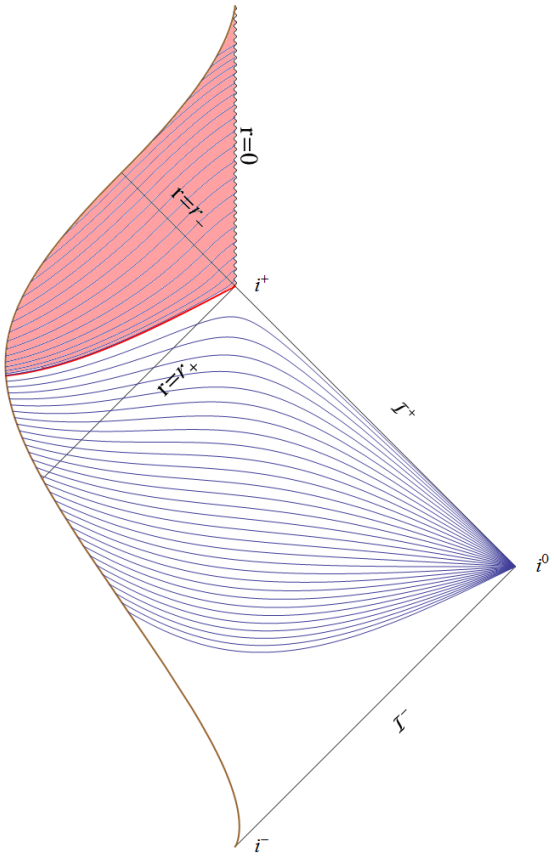


Figure 7.5: The Penrose diagram for $Q = 0.99M$ and $\epsilon = 1$ collapsing shell depicting the UH. The lines/region have the same meaning as Figure 7.3 with the inclusion of sub-UH CMC surfaces in blue.

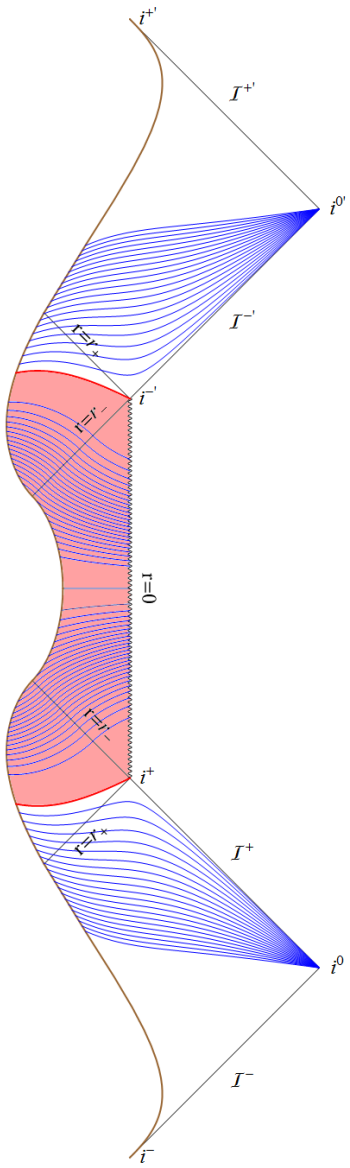


Figure 7.6: The Penrose diagram for $Q = 0.99M$ and $\epsilon = 1.1$ values. These parameters make $b/(\epsilon - 1) < b/\epsilon + 1/\epsilon^2$. The lines/region have the same meaning as Figure 7.3 with the inclusion of sub-UH CMC surfaces in blue.

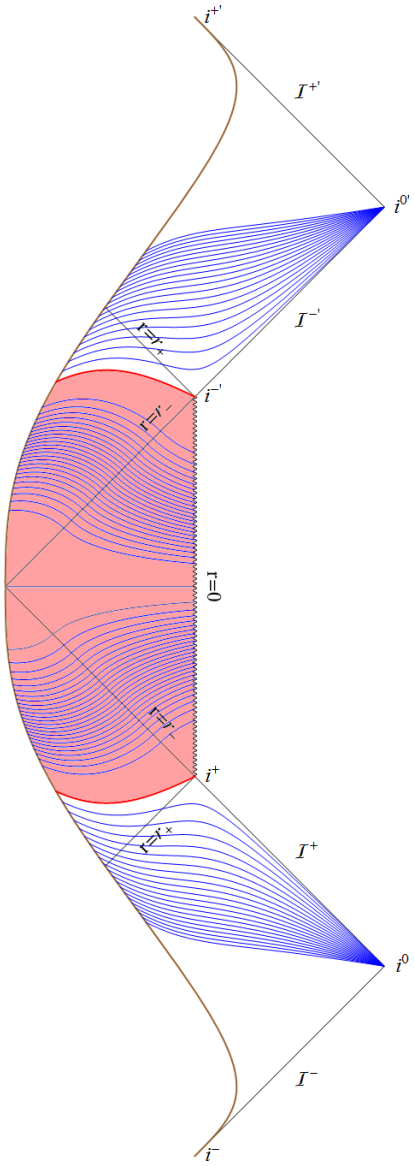


Figure 7.7: The Penrose diagram for $Q = 0.99M$ and ϵ set to make $b/(\epsilon - 1) = b/\epsilon + 1/\epsilon^2$. The lines/region have the same meaning as Figure 7.3 with the inclusion of sub-UH CMC surfaces in blue.

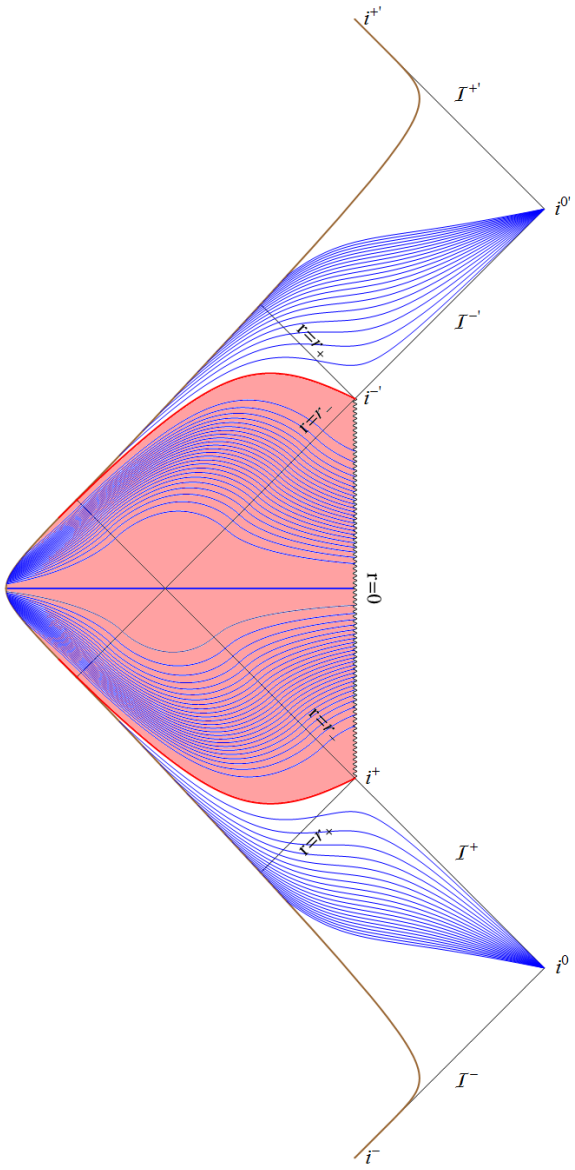
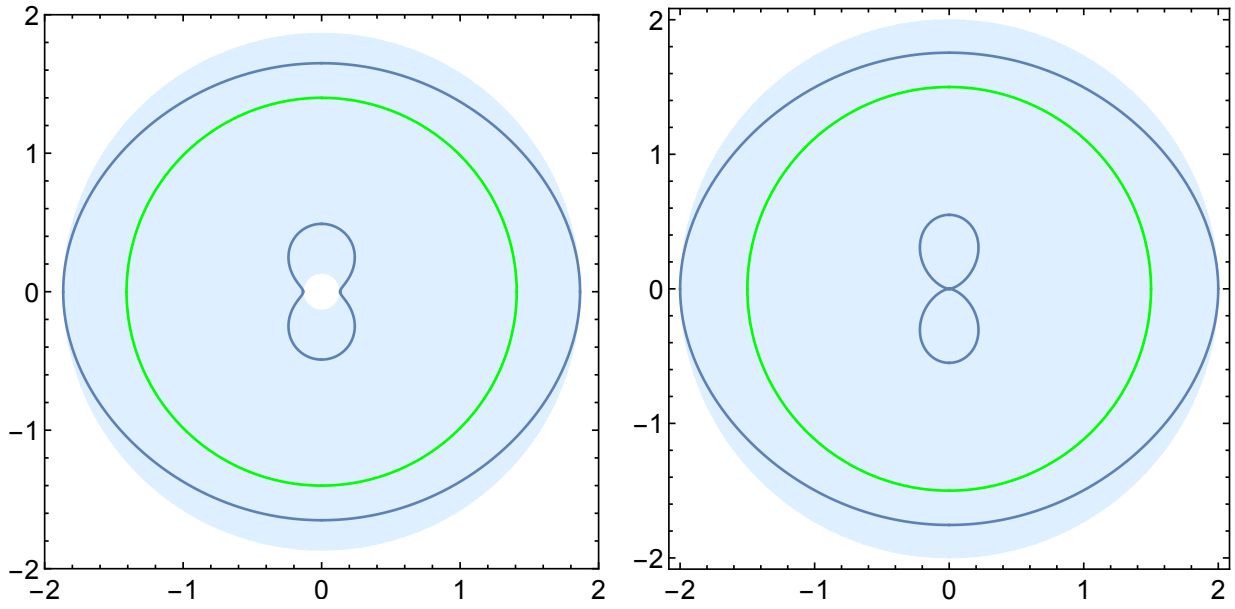


Figure 7.8: The Penrose diagram for $Q = 0.99M$ and $\epsilon = 3/2$ values. These parameters make $b/(\epsilon - 1) > b/\epsilon + 1/\epsilon^2$. Coloured lines/region have the same meaning as Figure 7.3 with the inclusion of sub-UH CMC surfaces in blue.



(a) $a = \frac{m}{2}$ (b) $a = 0$ (Schwarzschild)

Figure 7.9: Polar plot $\frac{r(\theta)}{m}$. Green curve is the UH solution of (7.41). Blue curves are additional numerical solutions to (7.40) and shaded region is the region between inner and outer Killing horizons. The outer (inner) UH is tangent to the outer (inner) Killing horizon at $\theta = \frac{\pi}{2}$.

APPENDICES

Appendix A

Supplementary material for Chapter 2

A.1 Proof of Inequality (2.6)

Theorem 3. *Let N_V be a discrete random variable which takes on a value $n \in \{0, 1, 2, \cdots\}$ with probability $P_V(n)$, and whose mean is $V > 0$:*

$$\langle N_V \rangle = \sum_{n=0}^{\infty} n P_V(n) = V. \tag{A.1}$$

N_V has the least variance when $P_V(n) = 0 \ \forall \ n \neq n_*, n_* + 1$, where n_* is the largest integer which is smaller than or equal to V. Equivalently:

$$\langle (N_V - V)^2 \rangle \geq (V - n_*)(n_* + 1 - V), \tag{A.2}$$

where the inequality is saturated for the aforementioned process.

Proof. The following three conditions must be true

$$\sum_{n=0}^{\infty} P_V(n) = 1, \tag{A.3}$$

$$\sum_{n=0}^{\infty} P_V(n) n = V, \tag{A.4}$$

$$0 \leq P_V(n) \leq 1 \qquad \forall \qquad n. \tag{A.5}$$

We denote the random variable which we claim has the least variance by N_V^m, and its probability mass function by P_V^m. It follows from (A.3) and (A.4) that

$$P_V^m(n_*) = n_* + 1 - V, \qquad P_V^m(n_* + 1) = V - n_*, \qquad \langle (N_V^m - V)^2 \rangle = (V - n_*)(n_* + 1 - V). \tag{A.6}$$

Let us now show that for any other probability mass function $P_V(n)$:

$$\sigma_V^2 \equiv \sum_{n=0}^{\infty} P_V(n)(n - V)^2 \geq (V - n_*)(n_* + 1 - V). \tag{A.7}$$

To this end, we define the following

$$A_V \equiv \sum_{n=0}^{n_*} P_V(n), \tag{A.8}$$

$$B_V \equiv \sum_{n=0}^{n_*} P_V(n)(V - n) = \sum_{n=n_*+1}^{\infty} P_V(n)(n - V), \tag{A.9}$$

where the last equality follows from (A.4). On the one hand,

$$B_V = \sum_{n=0}^{n_*} P_V(n)(V - n) \geq (V - n_*) \sum_{n=0}^{n_*} P_V(n) = A_V(V - n_*). \tag{A.10}$$

On the other hand,

$$B_V = \sum_{n=n_*+1}^{\infty} P_V(n)(n - V) \geq (n_* + 1 - V) \sum_{n=n_*+1}^{\infty} P_V(n) = (n_* + 1 - V)(1 - A_V). \tag{A.11}$$

It then follows from (A.10) and (A.11) that

$$1 - \frac{B_V}{n_* + 1 - V} \leq A_V \leq \frac{B_V}{V - n_*}, \tag{A.12}$$

which in turn implies that

$$B_V \geq (V - n_*)(n_* + 1 - V). \tag{A.13}$$

Consider now the variance:

$$\begin{aligned} \sigma_V^2 = &\sum_{n=0}^{n_*-1} P_V(n)(n - V)^2 + \sum_{n=n_*+2}^{\infty} P_V(n)(n - V)^2 \\ &+ P_V(n_*)(V - n_*)^2 + P_V(n_* + 1)(n_* + 1 - V)^2. \end{aligned} \tag{A.14}$$

For all $n \neq n_*, n_*+1$, $(n-V)^2 > |V-n|$, from which it follows that

$$\begin{aligned}\sigma_V^2 &\geq \sum_{n=0}^{n_*-1} P_V(n)(V-n) + \sum_{n=n_*+2}^{\infty} P_V(n)(n-V) \\ &\quad + P_V(n_*)(V-n_*)^2 + P_V(n_*+1)(n_*+1-V)^2 \qquad \text{(A.15)} \\ &= 2B_V + (n_*-V)(n_*+1-V)\left[P_V(n_*) + P_V(n_*+1)\right]. \qquad \text{(A.16)}\end{aligned}$$

The equality in the last line follows from recognizing that

$$\sum_{n=n_*+2}^{\infty} P_V(n)(n-V) = \sum_{n=0}^{n_*+1} P_V(n)(V-n). \qquad \text{(A.17)}$$

Finally, using the inequality (A.13):

$$\begin{aligned}\sigma_V^2 &\geq 2(V-n_*)(n_*+1-V) + (n_*-V)(n_*+1-V)\left[P_V(n_*) + P_V(n_*+1)\right] \qquad \text{(A.18)} \\ &= (V-n_*)(n_*+1-V)\left[2 - P_V(n_*) - P_V(n_*+1)\right] \qquad \text{(A.19)} \\ &\geq (V-n_*)(n_*+1-V), \qquad \text{(A.20)}\end{aligned}$$

where the last inequality follows from the fact that $P_V(n_*)+P_V(n_*+1) \leq 1$. This concludes the proof of the theorem. □

A.2 2D Lorentzian Lattices: Details

We wish to construct a lattice that is invariant under the action of a discrete subgroup of the Lorentz group. We shall work in D-dimensional Minkowski space and use the metric signature $-++\cdots$. Consider D vectors $\xi_{(d)}$, with $d \in \{0, 1, 2, \cdots, D-1\}$, which generate the lattice. In other words, any element of the lattice X can be written as

$$X = n^{(d)}\xi_{(d)}, \qquad \text{(A.21)}$$

where $n^{(d)}$ are integers and the summation over d is implicit. Let Λ be an element of the Lorentz group. We require that for all points X on the lattice, ΛX is also a point on the lattice:

$$\Lambda X = n^{(d)}\Lambda\xi_{(d)} = m^{(d)}\xi_{(d)}, \qquad \text{(A.22)}$$

where $m^{(d)}$ are integers. We may decompose $\Lambda\xi_{(d)}$ in the basis of the generators:

$$\Lambda\xi_{(d)} = A_{(d)}{}^{(d')}\xi_{(d')}, \qquad \text{(A.23)}$$

where ${A_{(d)}}^{(d')}$ are constants which depend on Λ and $\xi_{(d)}$. It then follows from (A.22) that

$$n^{(d)}{A_{(d)}}^{(d')} = m^{(d')}. \tag{A.24}$$

Therefore, ${A_{(d)}}^{(d')}$ *must be an integer for all* d *and* d' *if our lattice is to be invariant under the action of* Λ. In order to compute A, we can"dot" both sides of (A.23) by $\xi_{(d'')}$:

$$\Lambda\xi_{(d)} \cdot \xi_{(d'')} = {A_{(d)}}^{(d')}\xi_{(d')} \cdot \xi_{(d'')}. \tag{A.25}$$

Defining the matrices B and C as,

$${B_{(d)}}^{(d')} \equiv \xi_{(d)} \cdot \xi_{(d')}, \qquad {C_{(d)}}^{(d')} \equiv \Lambda\xi_{(d)} \cdot \xi_{(d')}, \tag{A.26}$$

it follows that

$$A = CB^{-1}. \tag{A.27}$$

Consider now the case of $1+1$ Minkowski space, i.e. $D = 2$. Let $\xi_{(0)}$ and $\xi_{(1)}$ be the timelike and spacelike generators:

$$\xi_{(0)} = \epsilon \begin{pmatrix} \cosh\psi \\ \sinh\psi \end{pmatrix}, \qquad \xi_{(1)} = \delta \begin{pmatrix} \sinh\theta \\ \cosh\theta \end{pmatrix}, \tag{A.28}$$

where $\epsilon, \delta > 0$. Also, since in $1+1$ we only have boosts to consider:

$$\Lambda = \begin{pmatrix} \cosh\phi & \sinh\phi \\ \sinh\phi & \cosh\phi \end{pmatrix}. \tag{A.29}$$

Defining the following quantities,

$$\gamma = \frac{\delta}{\epsilon}, \qquad \chi = \psi - \theta, \tag{A.30}$$

it follows from (A.26) that

$$B = \epsilon^2 \begin{pmatrix} -1 & \gamma\sinh\chi \\ \gamma\sinh\chi & \gamma^2 \end{pmatrix}, \qquad C = \epsilon^2 \begin{pmatrix} -\cosh\phi & \gamma\sinh(\phi+\chi) \\ \gamma\sinh(\chi-\phi) & \gamma^2\cosh\phi \end{pmatrix}. \tag{A.31}$$

Using (A.27):

$$A = \frac{1}{\cosh\chi} \begin{pmatrix} \cosh(\phi+\chi) & \frac{1}{\gamma}\sinh\phi \\ \gamma\sinh\phi & \cosh(\phi-\chi) \end{pmatrix}. \tag{A.32}$$

We need to pick ϕ, χ and γ so that all elements of A are integers. Let $k_1 - k_4$ be integers and require

$$\frac{\cosh(\phi+\chi)}{\cosh\chi} = k_1, \qquad \frac{1}{\gamma}\frac{\sinh\phi}{\cosh\chi} = k_2, \qquad \gamma\frac{\sinh\phi}{\cosh\chi} = k_3, \qquad \frac{\cosh(\phi-\chi)}{\cosh\chi} = k_4. \tag{A.33}$$

Note that

$$k_1, k_4 > 0, \qquad \mathrm{sgn}(k_2) = \mathrm{sgn}(k_3). \tag{A.34}$$

The second and third equations in (A.33) are equivalent to

$$\gamma^2 = \frac{k_3}{k_2}, \qquad \frac{\sinh^2\phi}{\cosh^2\chi} = k_2 k_3. \tag{A.35}$$

Also, the first and fourth equations in (A.33) imply

$$2\cosh\phi = k_1 + k_4, \qquad 2\sinh\phi\tanh\chi = k_1 - k_4. \tag{A.36}$$

The first equation in (A.36) fixes ϕ up to a sign, using which the second equation in (A.35) fixes χ up to a sign. Putting these together in the second equation in (A.36), we obtain the following constraint on the integers $k_1 - k_4$:

$$k_1 k_4 - k_2 k_3 = 1. \tag{A.37}$$

This equation can be satisfied for various integers, and therefore there are many Lorentzian lattices in $1+1$.

To summarize: find integers $k_1 - k_4$ that satisfy the conditions (i) $k_1, k_4 > 0$, (ii) $\mathrm{sgn}(k_2)$=$\mathrm{sgn}(k_3)$, (iii) $k_1k_4 - k_2k_3 = 1$. Then, if we let $\cosh(\phi) = \frac{k_1+k_4}{2}$, $\gamma = \sqrt{k_3/k_2}$, and $\sinh(\chi) = \frac{k_1-k_4}{2\sqrt{k_2k_3}}$, the lattice generated by $\xi_{(0)}$ and $\xi_{(1)}$ goes to itself under the action of $\Lambda(\phi)$, with $\psi, \theta, \delta, \epsilon$ satisfying (A.30). Figure 2.1b shows an example of a Lorentzian lattice with $k_1 = 2$, $k_2 = k_3 = k_4 = 1$, $\delta = 1$, and $\theta = 0$.

Appendix B

Supplementary material for Chapter 3

B.1 IR Behaviour of the GCB Operators: Details

Here we will derive the equations that the constants a and $\{b_n\}$ should satisfy in order for $\Box_\rho^{(D)}$ to have the desired IR behaviour (3.30), or equivalently (3.31), which in turn is equivalent to

$$\tilde{g}(Z) \xrightarrow{Z\to 0} -Z, \tag{B.1}$$

where $\tilde{g}(Z)$ is defined by

$$\rho^{-2/D} g_\rho^{(D)}(p) \equiv \tilde{g}(Z), \tag{B.2}$$

as given in the right hand side of (3.26).

B.1.1 Even Dimensions

Let $D = 2N + 2$ where $N = 0, 1, 2, \ldots$. Then

$$\tilde{g}(Z) = a + 2(2\pi)^N \sum_{n=0}^{L_{max}} \frac{b_n}{n!} C_D^n \int_0^\infty s^{2(N+1)n+2N+1} e^{-C_D s^D} (Z^{1/2}s)^{-N} K_N(Z^{1/2}s)\, \mathrm{d}s. \tag{B.3}$$

In order to examine the behaviour of $\tilde{g}(Z)$ as $Z \to 0$, we need to expand $(Z^{1/2}s)^{-N} K_N(Z^{1/2}s)$ in this regime. From the power series expansion of K_N (see e.g. 10.31.1 and 10.25.2 of

[30]), it follows that

$$(Z^{1/2}s)^{-N}K_N(Z^{1/2}s) = 2^{N-1}(Zs^2)^{-N}\sum_{k=0}^{N-1}\frac{\Gamma(N-k)}{k!}(-Zs^2/4)^k \tag{B.4a}$$

$$+\frac{(-1)^{N+1}}{2^{N+1}N!}\ln(Z) \tag{B.4b}$$

$$+\frac{(-1)^N}{2^{N+1}N!}\left[-2\ln(s/2)+\psi(1)+\psi(N+1)\right] \tag{B.4c}$$

$$+\frac{(-1)^{N+1}s^2}{2^{N+3}(N+1)!}Z\ln(Z) \tag{B.4d}$$

$$+\frac{(-1)^N}{2^{N+3}(N+1)!}\left[-2\ln(s/2)+\psi(2)+\psi(N+2)\right]s^2Z \tag{B.4e}$$

$$+\mathcal{O}(Z^2),$$

where $\psi(n)$ is the digamma function. Because we need the leading behaviour of $\rho^{-\frac{2}{D}}\tilde{g}(Z)$ to be $-Z$, we have only considered terms up to this order. All the terms in (B.4a) and (B.4b) diverge as $Z \to 0$, forcing us to pick the b_n such that none of them contribute to $\tilde{g}(Z)$ in the $Z \to 0$ limit. The contribution of the term (B.4d) is also unwanted and should be made to vanish by choosing b_n appropriately. This leads us to the following series of equations:

$$\sum_{n=0}^{L_{max}}\frac{b_n}{n!}C_D^n\int_0^\infty s^{2(N+1)n+2k+1}e^{-C_Ds^D}\,\mathrm{d}s = 0, \qquad k = 0,1,\ldots,N+1. \tag{B.5}$$

The integration over s can be performed (see e.g. 5.9.1 of [30]) to give us the condition reproduced above as equation (3.32a):

$$\sum_{n=0}^{L_{max}}\frac{b_n}{n!}\Gamma(n+\frac{k+1}{N+1}) = 0, \qquad k = 0,1,\ldots,N+1. \tag{B.6}$$

Requiring the contribution of the constant term (B.4c) to vanish yields

$$a+\frac{(-1)^{N+1}2\pi^N}{N!}\sum_{n=0}^{L_{max}}\frac{b_n}{n!}C_D^n\int_0^\infty s^{2(N+1)n+2N+1}e^{-C_Ds^D}\ln(s)\,\mathrm{d}s = 0. \tag{B.7}$$

We can perform the integral over s by using the formula (see e.g. 5.9.19 and 5.9.1 of [30])

$$\int_0^\infty s^\mu e^{-as^D}\ln(s)\,\mathrm{d}s = \frac{\Gamma(\frac{\mu+1}{D})}{D^2a^{\frac{\mu+1}{D}}}\left[\psi(\frac{\mu+1}{D})-\ln(a)\right], \tag{B.8}$$

leading to (3.32b):

$$a + \frac{2(-1)^{N+1}\pi^N}{D^2 C_D N!} \sum_{n=0}^{L_{max}} b_n \psi(n+1) = 0. \tag{B.9}$$

Finally, requiring the contribution of (B.4e) to reproduce the desired $-Z$ behaviour leads to

$$\sum_{n=0}^{L_{max}} \frac{b_n}{n!} C_D^n \int_0^\infty s^{2(N+1)n+2N+3} e^{-C_D s^D} \ln(s)\, \mathrm{d}s = \frac{2(-1)^N (N+1)!}{\pi^N}. \tag{B.10}$$

Performing the integral using (B.8) furnishes (3.32c):

$$\sum_{n=0}^{L_{max}} \frac{b_n}{n!} \Gamma(n + \frac{N+2}{N+1}) \psi(n + \frac{N+2}{N+1}) = \frac{2(-1)^N (N+1)!}{\pi^N} D^2 C_D^{\frac{N+2}{N+1}}. \tag{B.11}$$

B.1.2 Odd Dimensions

Let $D = 2N+1$ where $N = 0, 1, 2, \ldots$. Then

$$\tilde{g}(Z) = a + 2(2\pi)^{N-1/2} \sum_{n=0}^{L_{max}} \frac{b_n}{n!} C_D^n \int_0^\infty s^{2(N+1)n+2N} e^{-C_D s^D} (Z^{1/2} s)^{-N+1/2} K_{N-1/2}(Z^{1/2} s)\, \mathrm{d}s. \tag{B.12}$$

From the power series expansion of K_N (see 10.27.4 of and 10.25.2 of [30]), it follows that

$$(Z^{\frac{1}{2}} s)^{-N+\frac{1}{2}} K_{N-\frac{1}{2}}(Z^{\frac{1}{2}} s) = (-1)^{N-1} 2^{N-\frac{3}{2}} \pi (Z^{\frac{1}{2}} s)^{-2N+1} \sum_{k=0}^{N} \frac{(Zs^2/4)^k}{k! \Gamma(k - N + \frac{3}{2})} \tag{B.13a}$$

$$+ \frac{(-1)^N 2^{-N-\frac{1}{2}} \pi}{\Gamma(N + \frac{1}{2})} \tag{B.13b}$$

$$+ \frac{(-1)^N 2^{-N-\frac{5}{2}} \pi}{\Gamma(N + \frac{3}{2})} s^2 Z \tag{B.13c}$$

$$+ \mathcal{O}(Z^{3/2}).$$

As before, we have only kept track of terms up to Z. The contributions of all the terms in (B.13a) should be made to vanish; this leads to the equation

$$\sum_{n=0}^{L_{max}} \frac{b_n}{n!} C_D^n \int_0^\infty s^{2(N+1)n+2k+1} e^{-C_D s^D}\, \mathrm{d}s = 0, \qquad k = 0, 1, \ldots, N. \tag{B.14}$$

Performing the integral over s gives us (3.33a):

$$\sum_{n=0}^{L_{max}} \frac{b_n}{n!}\Gamma(n+\frac{2k+2}{2N+1}) = 0, \qquad k = 0, 1, \ldots, N. \tag{B.15}$$

Requiring the contribution of the constant term (B.13b) to vanish yields

$$a + \frac{(-1)^N \pi^{N+\frac{1}{2}}}{\Gamma(N+\frac{1}{2})} \sum_{n=0}^{L_{max}} \frac{b_n}{n!} C_D^n \int_0^\infty s^{2(N+1)n+2N} e^{-C_D s^D} \, \mathrm{d}s = 0 \ , \tag{B.16}$$

which is equivalent to (3.33b):

$$a + \frac{(-1)^N \pi^{N+\frac{1}{2}}}{D C_D \Gamma(N+\frac{1}{2})} \sum_{n=0}^{L_{max}} b_n = 0 \ . \tag{B.17}$$

Finally, requiring the contribution of (B.13c) to reproduce the desired $-Z$ behaviour leads to

$$\frac{(-1)^N \pi^{N+\frac{1}{2}}}{4\Gamma(N+\frac{3}{2})} \sum_{n=0}^{L_{max}} \frac{b_n}{n!} C_D^n \int_0^\infty s^{2(N+1)n+2N+2} e^{-C_D s^D} \, \mathrm{d}s = -1, \tag{B.18}$$

which furnishes (3.33c):

$$\sum_{n=0}^{L_{max}} \frac{b_n}{n!}\Gamma(n+\frac{2N+3}{2N+1}) = \frac{(-1)^{N-1} 4\Gamma(N+\frac{3}{2})}{\pi^{N+\frac{1}{2}}} D C_D^{\frac{2N+3}{2N+1}}. \tag{B.19}$$

B.2 UV Behaviour of the GCB Operators: Details

Here we derive the UV behaviour of $\Box_\rho^{(D)}$. We will make use of the following identity [31], which holds for arbitrary natural number m:

$$\frac{K_p(Z^{\frac{1}{2}} s)}{Z^{\frac{p}{2}}} = (-1)^m \left(\frac{2}{s}\right)^m \frac{d^m}{dZ^m} \left\{ \frac{K_{p-m}(Z^{\frac{1}{2}} s)}{Z^{\frac{p-m}{2}}} \right\}. \tag{B.20}$$

B.2.1 Even Dimensions

Let $D = 2N + 2$ where $N = 0, 1, 2, \ldots$, and $p = m = N$ in (B.20). It then follows that

$$Z^{-N/2} K_N(Z^{\frac{1}{2}} s) = (-1)^N \left(\frac{2}{s}\right)^N \frac{d^N}{dZ^N} K_0(Z^{\frac{1}{2}} s). \tag{B.21}$$

Substituting this in the definition of $\tilde{g}(Z)$, as given by (B.3), produces

$$\tilde{g}(Z) = a + (-1)^N 2^{2N+1} \pi^N \sum_{n=0}^{L_{max}} \frac{b_n}{n!} C_D^n \frac{d^N}{dZ^N} I_n^{(D)}(Z) \ , \tag{B.22}$$

where

$$I_n^{(D)}(Z) \equiv \int_0^\infty s^{Dn+1} e^{-C_D s^D} K_0(Z^{\frac{1}{2}} s) \, \mathrm{d}s \ . \tag{B.23}$$

It then suffices to study the behaviour of this integral as $Z \to \infty$. It follows from 10.29.4 of [30] that

$$K_0(Z^{\frac{1}{2}} s) = \frac{-1}{Z^{\frac{1}{2}} s} \frac{d}{ds} \left(s K_1(Z^{\frac{1}{2}} s) \right). \tag{B.24}$$

Plugging this relation in (B.23) and integrating by parts yields

$$I_n^{(D)}(Z) = -\frac{1}{Z^{\frac{1}{2}}} \left\{ s^{Dn+1} e^{-C_D s^D} K_1(Z^{\frac{1}{2}} s) \Big|_0^\infty - \int_0^\infty s K_1(Z^{\frac{1}{2}} s) \frac{d}{ds} (s^{Dn} e^{-C_D s^D}) \, \mathrm{d}s \right\}. \tag{B.25}$$

The first term vanishes when evaluated at ∞. When evaluated at 0, it is non-zero only when $n = 0$, because $K_1(Z^{\frac{1}{2}} s) \to Z^{\frac{1}{2}} s^{-1}$ when $s \to 0$. It then follows that

$$I_n^{(D)}(Z) = \frac{1}{Z^{\frac{1}{2}}} \left\{ \frac{\delta_{n0}}{Z^{\frac{1}{2}}} + \int_0^\infty s K_1(Z^{\frac{1}{2}} s) \frac{d}{ds} (s^{Dn} e^{-C_D s^D}) \, \mathrm{d}s \right\}. \tag{B.26}$$

From 10.29.3 of [30],

$$K_1(Z^{\frac{1}{2}} s) = \frac{-1}{Z^{\frac{1}{2}}} \frac{d}{ds} K_0(Z^{\frac{1}{2}} s). \tag{B.27}$$

Plugging this back into (B.26) and integrating once again by parts yields

$$I_n^{(D)}(Z) = \frac{1}{Z} \left\{ \delta_{n0} + \int_0^\infty K_0(Z^{\frac{1}{2}} s) \frac{d}{ds} \left[s \frac{d}{ds} (s^{Dn} e^{-C_D s^D}) \right] \, \mathrm{d}s \right\}. \tag{B.28}$$

It can be shown that

$$\lim_{Z \to \infty} \int_0^\infty K_0(Z^{\frac{1}{2}} s) \frac{d}{ds} \left[s \frac{d}{ds} (s^{Dn} e^{-C_D s^D}) \right] \, \mathrm{d}s = 0. \tag{B.29}$$

With the aid of (B.22), it then follows that for large Z,

$$\tilde{g}(Z) = a \; + \; 2^{D-1}\pi^{\frac{D}{2}-1}\Gamma(D/2)b_0\, Z^{-\frac{D}{2}} \; + \cdots \tag{B.30}$$

Notice that both these terms are real for both positive and negative Z, because $D/2$ is an integer when D is even. In order to produce the sub-leading terms, one can continue integrating by parts in (B.28). The sub-leading terms are thus also real, whence the imaginary part of $\tilde{g}(Z)$ must, for even D, decay faster than any power of Z for $Z \to \infty$. This behavior can be seen in Figures 3.2b and 3.3b.

B.2.2 Odd Dimensions

Let $D = 2N + 1$ where $N = 0, 1, 2, \ldots$, and $p = m - \frac{1}{2} = N - \frac{1}{2}$ in (B.20). It then follows that

$$Z^{\frac{1-2N}{4}} K_{N-\frac{1}{2}}(Z^{\frac{1}{2}}s) = (-1)^N \left(\frac{2}{s}\right)^N \frac{d^N}{dZ^N}\{Z^{\frac{1}{4}} K_{-\frac{1}{2}}(Z^{\frac{1}{2}}s)\}. \tag{B.31}$$

From 10.39.2 of [30], we have that

$$K_{-\frac{1}{2}}(Z^{\frac{1}{2}}s) = Z^{-\frac{1}{4}} \left(\frac{\pi}{2s}\right)^{\frac{1}{2}} e^{-Z^{\frac{1}{2}}s} \; , \tag{B.32}$$

whence

$$Z^{\frac{1-2N}{4}} K_{N-\frac{1}{2}}(Z^{\frac{1}{2}}s) = \frac{(-1)^N 2^{N-\frac{1}{2}}\pi^{\frac{1}{2}}}{s^{N+\frac{1}{2}}} \frac{d^N}{dZ^N} e^{-Z^{\frac{1}{2}}s}. \tag{B.33}$$

Substituting this into the definition of $\tilde{g}(Z)$, as given by (B.12), produces

$$\tilde{g}(Z) = a + (-1)^N 2^{2N}\pi^N \sum_{n=0}^{L_{max}} \frac{b_n}{n!} C_D^n \frac{d^N}{dZ^N} I_n^{(D)}(Z), \tag{B.34}$$

where

$$I_n^{(D)}(Z) \equiv \int_0^\infty s^{Dn} e^{-C_D s^D} e^{-Z^{\frac{1}{2}}s}\, \mathrm{d}s. \tag{B.35}$$

It then suffices to study the behaviour of this integral as $Z \to \infty$:

$$\begin{aligned} I_n^{(D)}(Z) &= -Z^{-\frac{1}{2}} \int_0^\infty s^{Dn} e^{-C_D s^D} \frac{d}{ds} e^{-Z^{\frac{1}{2}}s}\, \mathrm{d}s \\ &= -Z^{-\frac{1}{2}} \left\{ s^{Dn} e^{-C_D s^D - Z^{\frac{1}{2}}s}\Big|_0^\infty - \int_0^\infty e^{-Z^{\frac{1}{2}}s} \frac{d}{ds}(s^{Dn} e^{-C_D s^D}) \right\} \\ &= Z^{-\frac{1}{2}} \left\{ \delta_{n0} + \int_0^\infty e^{-Z^{\frac{1}{2}}s} \frac{d}{ds}(s^{Dn} e^{-C_D s^D}) \right\}. \end{aligned} \tag{B.36}$$

Again, because

$$\lim_{Z\to\infty}\int_0^\infty e^{-Z^{\frac{1}{2}}s}\frac{d}{ds}(s^{Dn}e^{-C_D s^D}) = 0, \tag{B.37}$$

we can deduce from (B.34) that

$$\tilde{g}(Z) \;=\; a \;+\; 2^{D-1}\pi^{\frac{D}{2}-1}\Gamma(D/2)b_0 Z^{-\frac{D}{2}} \;+\cdots\;. \tag{B.38}$$

B.3 Derivation of Equation (3.5)

From the general equations, (3.22) and (3.25), we have

$$\rho^{-1}g_\rho^{(2)}(p) = a^{(2)} + \rho\sum_{n=0}^{2}\frac{(-1)^n\rho^n}{n!}b_n^{(2)}\frac{\partial^n}{\partial\rho^n}\chi(p,\rho), \tag{B.39}$$

where $\{a^{(2)}, b_n^{(2)}\}$ are given in (3.2) and

$$\chi(p,\rho) = 2\int_0^\infty se^{-\rho s^2/2}K_0(\sqrt{p\cdot p}s)\,\mathrm{d}s. \tag{B.40}$$

From the relation (see e.g. 8.6.6 and 8.19.1 of [30]),

$$e^Z E_1(Z) = 2\int_0^\infty e^{-t}K_0(\sqrt{2zt})\,\mathrm{d}t, \tag{B.41}$$

it follows that

$$\chi(p,\rho) = \rho^{-1}e^{Z/2}E_1(Z/2)\;, \qquad Z = \rho^{-1}p\cdot p. \tag{B.42}$$

Furthermore, using the identities (see e.g. 8.9.14 and 8.19.12 of [30]),

$$\frac{d}{dz}\left[e^z E_p(z)\right] = e^z E_p(z)\left(1+\frac{p-1}{z}\right) - \frac{1}{z}, \tag{B.43}$$

$$pE_{p+1}(z) + zE_p(z) = e^{-z}, \tag{B.44}$$

it can be shown that

$$\rho^2\frac{\partial\chi}{\partial\rho} = e^{Z/2}E_2(Z/2) - e^{Z/2}E_1(Z/2) \tag{B.45}$$

$$\rho^3\frac{\partial^2\chi}{\partial\rho^2} = e^{Z/2}E_1(Z/2)\left[2+Z/2\right] - e^{Z/2}E_2(Z/2)\left[3+Z/2\right]. \tag{B.46}$$

Equation (3.5) results from plugging these expressions back into (B.39) and using (B.44):

$$\rho^{-1}g_\rho^{(2)}(p) = -Ze^{Z/2}\mathrm{E}_2(Z/2). \tag{B.47}$$

B.4 Damping the fluctuations

In reference [26] a prescription was given to get from the causet d'Alembertian $B_\rho^{(2)}$ of (3.1) a new operator $\tilde{B}_{\rho,\epsilon}^{(2)}$, whose fluctuations are damped, but which has the same mean over sprinklings as $B_{\tilde{\rho}}^{(2)}$ with $\tilde{\rho} = \epsilon\rho$. Here we generalize this prescription to the class of causet d'Alembertians $B_\rho^{(D)}$ defined in (3.19). (See Sections 3.2 and 3.3 for any symbol which is not defined in what follows.)

Given the causal set d'Alembertian,

$$\rho^{-2/D}(B_\rho^{(D)}\Phi)(x) = a\Phi(x) + \sum_{m=0}^{L_{max}} b_m \sum_{y\in I_m} \Phi(y), \tag{B.48}$$

we construct as follows a new operator $\tilde{B}_{\rho,\epsilon}^{(D)}$ whose effective non-locality energy-density scale is $\epsilon\rho$:

$$\tilde{\rho}^{-2/D}(\tilde{B}_{\rho,\epsilon}^{(D)}\Phi)(x) = a\Phi(x) + \sum_{n=0}^{\infty} \tilde{b}_n \sum_{y\in I_n} \Phi(y), \tag{B.49}$$

with

$$\tilde{b}_n = \epsilon(1-\epsilon)^n \sum_{m=0}^{L_{max}} \binom{n}{m} \frac{b_m \epsilon^m}{(1-\epsilon)^m}, \qquad \epsilon = \tilde{\rho}/\rho. \tag{B.50}$$

(Here, the binomial coefficient $\binom{n}{m}$ is zero by convention for $m > n$.)

Let us demonstrate that the continuum limit of $\tilde{B}_{\rho,\epsilon}^{(D)}$, which we will denote by $\tilde{\Box}_{\tilde{\rho}}^{(D)}$, is

equal to $\Box_{\tilde{\rho}}^{(D)}$:

$$
\begin{aligned}
&\tilde{\rho}^{-2/D}(\tilde{\Box}_{\tilde{\rho}}^{(D)}\Phi)(x) - a\Phi(x) \\
&= \rho \sum_{n=0}^{\infty} \frac{\tilde{b}_n}{n!} \int\limits_{J^-(x)} e^{-\rho V(x,y)} [\rho V(x,y)]^n \phi(y) dV_y \\
&= \rho\epsilon \sum_{m=0}^{L_{max}} \frac{b_m \epsilon^m}{m!} \int\limits_{J^-(x)} e^{-\rho V(x,y)} \left\{ \sum_{n=m}^{\infty} \frac{(1-\epsilon)^{n-m}}{(n-m)!} [\rho V(x,y)]^n \right\} \phi(y) dV_y \\
&= \tilde{\rho} \sum_{m=0}^{L_{max}} \frac{b_m}{m!} \int\limits_{J^-(x)} e^{-\rho V(x,y)} \left\{ \sum_{n=m}^{\infty} \frac{(1-\epsilon)^{n-m}}{(n-m)!} [\rho V(x,y)]^{n-m} \right\} [\epsilon \rho V(x,y)]^m \phi(y) dV_y \\
&= \tilde{\rho} \sum_{m=0}^{L_{max}} \frac{b_m}{m!} \int\limits_{J^-(x)} e^{-\rho V(x,y)} e^{(1-\epsilon)\rho V(x,y)} [\tilde{\rho} V(x,y)]^m \phi(y) dV_y \\
&= \tilde{\rho} \sum_{m=0}^{L_{max}} \frac{b_m}{m!} \int\limits_{J^-(x)} e^{-\tilde{\rho} V(x,y)} [\tilde{\rho} V(x,y)]^m \phi(y) dV_y. \\
&= \tilde{\rho}^{-2/D}(\Box_{\tilde{\rho}}^{(D)}\Phi)(x) - a\Phi(x).
\end{aligned}
$$

Of course, we have not proven here that the fluctuations of $\tilde{B}_{\tilde{\rho}}^{(D)}$ are actually damped. This has been confirmed numerically for the minimal 2D and 4D operators in [26] and [27]. It would be interesting to confirm it also for the full set of GCB operators in all dimensions.

Appendix C

Supplementary material for Chapter 4

C.1 Existence and Examples of $\widetilde{\Box}$

Here we will show there are operators $\widetilde{\Box}$ which satisfy all the axioms introduced in Section 4.2. In fact, we will outline a procedure for constructing such operators.

We shall consider the following operator:

$$\Lambda^{-2}(\widetilde{\Box}\phi)(x) = a\phi(x) + \Lambda^4 \int_{J^-(x)} f(\Lambda^2 \tau_{xy}^2)\phi(y) d^4y, \tag{C.1}$$

where Λ denotes the nonlocality energy scale, a is a dimension-less real number, $J^-(x)$ denotes the causal past of x, and τ_{xy} is the Lorentzian distance between x and y:

$$\tau_{xy}^2 = (x^0 - y^0)^2 - |\mathbf{x} - \mathbf{y}|^2. \tag{C.2}$$

It may be shown that

$$\widetilde{\Box} e^{ip\cdot x} = B(p) e^{ip\cdot x}, \tag{C.3}$$

$$B(p) = \Lambda^2 \widetilde{g}(p/\Lambda), \tag{C.4}$$

$$\widetilde{g}(z) = a + \int_{J^+(0)} f((y^0)^2 - |\mathbf{y}|^2) e^{-iz\cdot y} d^4y, \tag{C.5}$$

where as usual $x \cdot y = \eta_{\mu\nu} x^\mu y^\nu$. Evaluating $\widetilde{g}(z)$ amounts to computing the Laplace transform of a retarded, Lorentz invariant function, which has been done in [30]. It follows from their result that

$$\widetilde{g}(z) = g(z \cdot z), \tag{C.6}$$

$$g(Z) = a + 4\pi Z^{-\frac{1}{2}} \int_0^\infty f(s^2) s^2 K_1(Z^{1/2} s) ds, \tag{C.7}$$

where an infinitesimal time-like and future-directed imaginary part ought to be added to z on the right hand side of (C.6) (see [33] for more details).

C.1.1 IR conditions

The infrared condition (4.8) is equivalent to satisfying

$$g(Z) \xrightarrow{Z \to 0} -Z. \tag{C.8}$$

In [33], a framework is developed to determine what constraints (C.8) places on a and f, for some specific choices of f which arise in causal set theory. Generalizing that methodology in a straightforward manner, we find that (C.8) is true if and only if the following conditions are satisfied:

$$\int_0^\infty f(s^2) s^{2k+1} ds = 0, \qquad k = 0, 1, 2 \tag{C.9}$$

$$\int_0^\infty f(s^2) s^5 \ln s ds = -\frac{4}{\pi}, \tag{C.10}$$

$$a + 2\pi \int_0^\infty f(s^2) s^3 \ln s ds = 0. \tag{C.11}$$

C.1.2 From $B(p)$ to $\widetilde{\Box}$

It is often desirable to constrain the behaviour of $B(p)$, as opposed to $\widetilde{\Box}$ directly. For instance, as is argued in Section 4.4.3, the quantum theory is well behaved only when the imaginary part of $B(p)$ (for timelike and future-directed p) is always positive. The question then becomes: *are there any choices of a and f which allow for this possibility, provided the IR conditions* (C.9)–(C.11) *are satisfied*? To answer this question, we turn the problem

around. Given a choice of $B(p)$, we reconstruct a and f and then ask if the IR conditions are met.

It can be shown that for $x > 0$: (see e.g. 10.27.9 and 10.27.10 of [30])

$$g(-x^2 - i\epsilon) = g_R(-x^2 - i\epsilon) + ig_I(-x^2 - i\epsilon), \tag{C.12}$$

$$g_R(-x^2 - i\epsilon) = a + \frac{2\pi}{x}\int_0^\infty f(s^2)s^2 Y_1(xs)ds, \tag{C.13}$$

$$g_I(-x^2 - i\epsilon) = -\frac{2\pi^2}{x}\int_0^\infty f(s^2)s^2 J_1(xs)ds. \tag{C.14}$$

We can now use the following orthonormality conditions of Bessel functions (see e.g. 1.17.13 of [30]) to express f in terms of $\widetilde{g}_I$:

$$\delta(x - \widetilde{x}) = x\int_0^\infty tJ_1(xt)J_1(\widetilde{x}t)dt. \tag{C.15}$$

Doing so yields:

$$f(s^2) = f_g(s^2) + h(s^2), \tag{C.16}$$

$$f_g(s^2) = -\frac{1}{2\pi^2 s}\int_0^\infty g_I(-x^2 - i\epsilon)x^2 J_1(sx)dx, \tag{C.17}$$

where h satisfies for all x:

$$\int_0^\infty h(s^2)s^2 J_1(xs)ds = 0. \tag{C.18}$$

This means that specifying $\widetilde{g}_I(-x^2 - i\epsilon)$ fixes f up to any part for which the right hand side of (C.14) vanishes. One example of a nontrivial function which satisfies (C.18) is the delta function: $h(x) = \delta^+(x) \equiv \delta(x - \epsilon)$, where ϵ is an arbitrarily small positive real number.

We can now express the IR conditions in terms of g_I and h:

$$\int_0^\infty h(s^2)s^{2k+1}ds$$
$$-\frac{1}{2\pi^2}\int_0^\infty g_I(-x^2-i\epsilon)x^2\int_0^\infty dss^{2k}J_1(xs)=0, \tag{C.19}$$
$$\int_0^\infty h(s^2)s^5\ln sds$$
$$-\frac{1}{2\pi^2}\int_0^\infty g_I(-x^2-i\epsilon)x^2\int_0^\infty dss^4J_1(xs)\ln s=-\frac{4}{\pi}, \tag{C.20}$$
$$a\ +2\pi\int_0^\infty h(s^2)s^3\ln sds$$
$$-\frac{1}{\pi}\int_0^\infty g_I(-x^2-i\epsilon)x^2\int_0^\infty dss^2J_1(xs)\ln s=0. \tag{C.21}$$

The above integrals over s are not absolutely convergent, so use the usual trick:

$$\int_0^\infty dsJ_1(xs)e^{-\delta s}\xrightarrow{\delta\to 0}\frac{1}{x}, \tag{C.22}$$
$$\int_0^\infty dss^2J_1(xs)e^{-\delta s}\xrightarrow{\delta\to 0}\frac{3\delta}{x^4}, \tag{C.23}$$
$$\int_0^\infty dss^4J_1(xs)e^{-\delta s}\xrightarrow{\delta\to 0}\frac{-45\delta}{x^6}, \tag{C.24}$$
$$\int_0^\infty dss^2J_1(xs)\ln se^{-\delta s}\xrightarrow{\delta\to 0}-2x^{-3}, \tag{C.25}$$
$$\int_0^\infty dss^4J_1(xs)\ln se^{-\delta s}\xrightarrow{\delta\to 0}16x^{-5}. \tag{C.26}$$

Having the delta function example in mind, we shall require h to satisfy for all $k=1,2$

$$\int_0^\infty h(s^2)s^{2k+1}ds=0,\qquad \int_0^\infty h(s^2)s^{2k+1}\ln sds=0, \tag{C.27}$$

Also, we assume that the following integrals converge:

$$\left|\int_0^\infty g_I(-x^2-i\epsilon)x^{-k}dx\right|<\infty,\qquad k=1,2,3,4 \tag{C.28}$$
$$\left|\int_0^\infty g_I(-x^2-i\epsilon)x^{-k}\ln xdx\right|<\infty\qquad k=2,4. \tag{C.29}$$

The IR conditions then reduce to

$$\int_0^\infty g_I(-x^2 - i\epsilon)x \; dx = \pi^2 \int_0^\infty h(u)du, \tag{C.30}$$

$$\int_0^\infty g_I(-x^2 - i\epsilon)x^{-3}dx = \frac{\pi}{2}, \tag{C.31}$$

$$\int_0^\infty g_I(-x^2 - i\epsilon)x^{-1}dx = -\frac{\pi}{2}a. \tag{C.32}$$

Note that the only nontrivial condition to satisfy is (C.30), since (C.31) just fixes the normalization of g_I and (C.32) determines a. Note that for positive $g_I(-x^2 - i\epsilon)$ which is required by consistent quantum theory, a must be a negative number.

If h is taken to be zero, then g_I ought to change sign, which leads to a quantum theory with an unbounded Hamiltonian. We note that the class of operators which arise in causal set theory in [33] all have $h = 0$, and therefore this feature.

Let us work out a complete example in 4D. Let

$$g_I(-x - i\epsilon) = Ax^2e^{-x/2}, \qquad h(x) = \alpha\delta^+(x). \tag{C.33}$$

where A and α are real constants. It can then be shown using (C.30)–(C.32):

$$A = \frac{\pi}{2}, \qquad \alpha = \frac{4}{\pi}, \qquad a = -2. \tag{C.34}$$

It then follows from (C.17) that

$$f_g(s) = -\frac{e^{-s/2}}{4\pi}(24 - 12s + s^2). \tag{C.35}$$

Therefore:

$$f(s) = \frac{4}{\pi}\delta^+(s) - \frac{e^{-s/2}}{4\pi}(24 - 12s + s^2). \tag{C.36}$$

C.1.3 Stability from positivity of g_I

We have required that evolution defined by $\widetilde{\Box}$ should be stable. Instabilities are in general associated with "unstable modes", and in line with [33], we shall use this as our criterion of instability. More specifically, we take such a mode to be a plane-wave $e^{ip\cdot x}$ satisfying the equation of motion $\widetilde{\Box}e^{ip\cdot x} = 0$, with the wave-vector p possessing a future-directed timelike

imaginary part (i.e. $p = p_R + ip_I$ where $p_I \cdot p_I < 0$ and $p_I^0 > 0$). It is shown in [33] that the necessary and sufficient condition for avoiding unstable modes is

$$g(Z) \neq 0 \ , \qquad \forall \ Z \neq 0 \text{ and } Z \in \mathbb{C}. \tag{C.37}$$

On the other hand, we argued in 4.4.3 that for consistency reasons we need to assume $ImB(p) > 0$ for $p^0 > 0$ which implies $g(Z)$ has a positive (negative) imaginary part under (above) the cut in Figure 4.1.

Here, we show that not even stability condition and positivity of $g_I(-x^2 - i\epsilon)$ (see Appendix C.1.2) are consistent, but latter is a sufficient condition for stability. In order to prove it, we make the following assumptions:

1. $g(Z)$ has a simple zero at $Z = 0$. IR conditions on $g(Z)$ (C.8) guarantee this assumption.

2. $g(Z)$ has positive (negative) imaginary part under (above) the cut.

We prove this by counting the number of zeros of g inside contour $C = C_1 + C_2 + C_3 + C_4$ in Figure C.1.

If N and P are the number of zeros and poles of g, respectively, inside the contour C (taken to be anticlockwise), then

$$\int_C dZ \ \frac{g'(Z)}{g(Z)} = -2\pi i(N - P). \tag{C.38}$$

Let's evaluate the left hand side of (C.38) for each contour separately:

1. C_1: According to (C.7), $g(Z)$ approaches the constant value of $a < 0$ (see C.1.2) for large Z. In fact, $g(Z) \to a + \mathcal{O}(\frac{1}{Z^n})$ for some positive value of n (which depends on the function f). This means for $a \neq 0$,

$$\int_{C_1} dZ \ \frac{g'(Z)}{g(Z)} = 0. \tag{C.39}$$

2. C_2 & C_4: Since the values of g above and under the cut are complex conjugate of each other, the contribution from these diagrams can be added together to get

$$\begin{aligned} \int_{C_2+C_4} dZ \ \frac{g'(Z)}{g(Z)} &= 2i\text{Im} \int_{-\infty}^{0} dx \ \frac{g'(x + i\epsilon)}{g(x + i\epsilon)} \\ &= 2i\text{Im} \ln \left[\frac{g(0 + i\epsilon)}{g(-\infty + i\epsilon)} \right], \end{aligned} \tag{C.40}$$

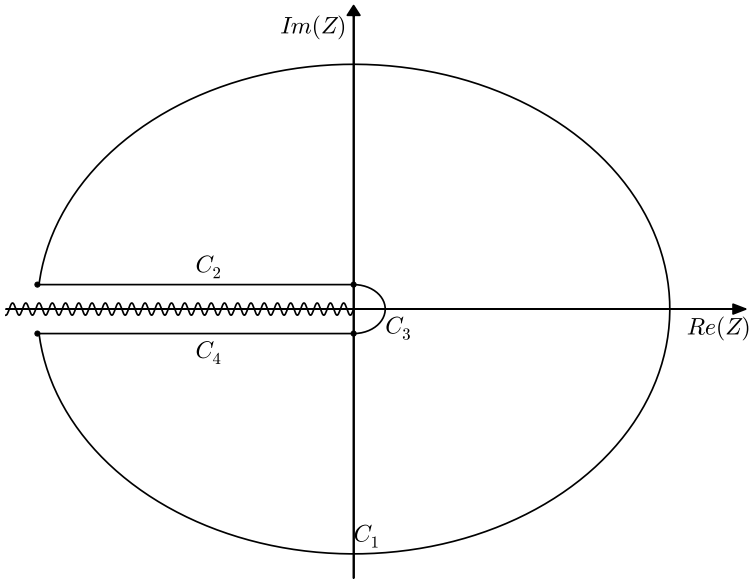


Figure C.1: The integration path in the complex Z plane. The closed contour is taken to be counterclockwise.

where ϵ is an infinitesimal positive number.

If we define $g(Z) = r_g(Z)e^{i\varphi_g(Z)}$, the right hand side of (C.40) (apart from the factor of $2i$) measures how much φ_g rotates from $Z = -\infty + i\epsilon$ to $Z = 0 + i\epsilon$. Since $\text{Im}g(x + i\epsilon) < 0$ on the whole negative real line, $\ln[g(x + i\epsilon)]$ is definable on one Riemann sheet. Combining this result with $g(-\infty + i\epsilon) = a < 0$ and $g(0 + i\epsilon) = -i\epsilon$, we get

$$\int_{C_2+C_4} dZ \, \frac{g'(Z)}{g(Z)} = i\pi. \tag{C.41}$$

3. C_3: IR conditions require that close to $Z = 0$, $g(Z) = -Z$. This means

$$\int_{C_3} dZ \, \frac{g'(Z)}{g(Z)} = \int_{C_3} \frac{1}{Z} = -i\pi. \tag{C.42}$$

Adding the values of all the contours and considering the fact that $g(Z)$ is finite everywhere ($P = 0$), we conclude that the number of zeros of g in complex plane of Z (inside contour C) is zero. Since there is no zero on the negative real line ($\text{Im}g(x + i\epsilon) \neq 0$), there is no zero of g in the complex plane of Z except the one at $Z = 0$. Therefore, stability has been proven.

C.2 FDT

Here, we present the proof of (4.52) [1]. Let's start by the following definitions

$$i\Delta(x, y) \equiv \left[\widehat{\phi}(x), \widehat{\phi}(y)\right], \tag{C.43}$$

$$G^{(1)}(x, y) \equiv \left\langle \left\{\widehat{\phi}(x), \widehat{\phi}(y)\right\} \right\rangle, \tag{C.44}$$

$$W^{+}(x, y) \equiv \left\langle \widehat{\phi}(x)\widehat{\phi}(y) \right\rangle, \tag{C.45}$$

$$W^{-}(x, y) \equiv \left\langle \widehat{\phi}(y)\widehat{\phi}(x) \right\rangle, \tag{C.46}$$

$$G^{F}(x, y) \equiv -i \left\langle T\widehat{\phi}(x)\widehat{\phi}(y) \right\rangle, \tag{C.47}$$

where $\{\}$ is anti-commutator and $\langle\rangle$ shows expectation value in a quantum state. If we define

$$G^{R}(x, y) \equiv \Delta(x, y)H(x \succ y), \tag{C.48}$$

[1] Most of the content of this appendix is taken from [104].

$$G^A(x,y) \equiv -\Delta(x,y)H(x \prec y),^{2} \tag{C.49}$$

we get the following relations

$$\begin{aligned} i\Delta(x,y) &= W^+(x,y) - W^-(x,y) \\ &= i\left[G^R(x,y) - G^A(x,y)\right], && \text{(C.50)} \\ G^{(1)}(x,y) &= W^+(x,y) + W^-(x,y), && \text{(C.51)} \\ G^A(x,y) &= G^R(y,x), && \text{(C.52)} \\ G^F(x,y) &= \frac{1}{2}\left[G^R(x,y) + G^A(x,y)\right] - \frac{i}{2}G^{(1)}(x,y). && \text{(C.53)} \end{aligned}$$

For a translational invariant system, the value of all the two point functions depend only on space-time separation. This will allow us to define the following Fourier transform with respect to time

$$\overline{A}(\omega,\mathbf{x},\mathbf{x}') \equiv \int dt\ A(t,\mathbf{x};t',\mathbf{x}')e^{-i\omega(t-t')}. \tag{C.54}$$

Now, let us assume that the quantum system is in thermal state with temperature $T = \frac{1}{\beta}$. It requires that

$$W^{\pm}(t,\mathbf{x};t',\mathbf{x}') = W^{\mp}(t+i\beta,\mathbf{x};t',\mathbf{x}'), \tag{C.55}$$

resulting in the following relation in Fourier space

$$\overline{W}^{+}(\omega,\mathbf{x},\mathbf{y}) = e^{\beta\omega}\overline{W}^{-}(\omega,\mathbf{x},\mathbf{y}). \tag{C.56}$$

Using (C.50), we get

$$\overline{W}^{+}(\omega,\mathbf{x},\mathbf{y}) = \frac{i\overline{\Delta}(\omega,\mathbf{x},\mathbf{y})}{1-e^{-\beta\omega}}, \tag{C.57}$$

$$\overline{W}^{-}(\omega,\mathbf{x},\mathbf{y}) = -\frac{i\overline{\Delta}(\omega,\mathbf{x},\mathbf{y})}{1-e^{\beta\omega}}. \tag{C.58}$$

On the other hand, since G^R and G^A are time transpose of each other, in Fourier space they are complex conjugate. As a result,

$$\begin{aligned} \mathrm{Im}\overline{G}^{F}(\omega,\mathbf{x},\mathbf{y}) &= -\frac{1}{2}\mathrm{Re}\overline{G}^{(1)}(\omega,\mathbf{x},\mathbf{y}) \\ &= -\frac{1}{2}\left[\overline{W}^{+}(\omega,\mathbf{x},\mathbf{y}) + \overline{W}^{-}(\omega,\mathbf{x},\mathbf{y})\right] \\ &= -\frac{1}{2}i\overline{\Delta}(\omega,\mathbf{x},\mathbf{y})\coth(\frac{\beta\omega}{2}) && \text{(C.59)} \end{aligned}$$

[2] where H is the Heaviside function: $H(x \succ y) = 1$ if $x \succ y$ and otherwise 0

where Im and Re are imaginary part and real part respectively and in the second line we have used the positivity of two point function W^+ (resulting that $\overline{W}^+(\omega, \mathbf{x}, \mathbf{y})$ and $\overline{W}^-(\omega, \mathbf{x}, \mathbf{y})$ are real.)

With the assumption that this field theory in Hilbert space representation has an equivalent representation in terms of double path integral, time ordered two point function is given by (4.46). In Fourier space, it reads

$$\overline{G}^F(\omega, \mathbf{x}, \mathbf{y}) = \frac{1}{2}\left[\overline{G}^K(\omega, \mathbf{x}, \mathbf{y}) + \overline{G}^R(\omega, \mathbf{x}, \mathbf{y}) + \overline{G}^A(\omega, \mathbf{x}, \mathbf{y})\right]. \tag{C.60}$$

$\overline{G}^K(\omega, \mathbf{x}, \mathbf{y})$ is a total imaginary number and $\overline{G}^R(\omega, \mathbf{x}, \mathbf{y}) + \overline{G}^A(\omega, \mathbf{x}, \mathbf{y})$ is a real number. As a result,

$$\overline{G}^K(\omega, \mathbf{x}, \mathbf{y}) = 2i\mathrm{Im}\overline{G}^F(\omega, \mathbf{x}, \mathbf{y}). \tag{C.61}$$

Combining (C.50)(C.59)(C.61) we arrive at

$$\overline{G}^K(\omega, \mathbf{x}, \mathbf{y}) = \coth(\frac{\beta\omega}{2})\left[\overline{G}^R(\omega, \mathbf{x}, \mathbf{y}) - \overline{G}^A(\omega, \mathbf{x}, \mathbf{y})\right], \tag{C.62}$$

which reduces to (4.52) at zero temperature.

C.3 Quantum Transition

We start by proving a simple theorem for any quantum system. Consider a quantum mechanical system in the (normalized) initial state $|\alpha\rangle$ evolves in time and the probability of finding the system at a later time t_f in the state $|\beta_i\rangle$ is called P_i, and assume $|\beta_i\rangle$'s are orthonormal:

$$P_i = |\langle\beta_i|U|\alpha\rangle|^2 \tag{C.63}$$

where U is the time evolution operator.

Now, consider a (normalized) state $|\beta\rangle$ as a superposition of $|\beta_i\rangle$ states:

$$|\beta\rangle = \sum_i c_i|\beta_i\rangle \tag{C.64}$$
$$\sum_i |c_i|^2 = 1.$$

Probability P of measuring the system at time t_f in the state $|\beta\rangle$ is given by

$$P = |\langle\beta|U|\alpha\rangle|^2. \tag{C.65}$$

Then,

$$\begin{aligned} P &= |\langle\beta|U|\alpha\rangle|^2 = \left|\sum_i c_i^* \langle\beta_i|U|\alpha\rangle\right|^2 \\ &\leq \left(\sum_i |c_i|^2\right)\left(\sum_i |\langle\beta_i|U|\alpha\rangle|^2\right) \\ &= \sum_i P_i \end{aligned} \tag{C.66}$$

where we have used the triangular inequality in the second line. So P is bounded from above by $\sum_i P_i$.

Now, let's get back to the scattering of a massless particle with state γ, a superposition of M different masses, in Section 4.6.3. We already have shown (see (4.95)) that Γ_{0m_i} defined as transition probability of a massless particle scattering with a massive particle (mass m_i) scales with N as

$$\Gamma_{0m_i} = \frac{A_i}{N} \tag{C.67}$$

where A_i depends on the momentum of the particles but independent of N. Using (C.66) for transition probabilities, we conclude that

$$\Gamma_{0\gamma} \leq \sum_i \frac{A_i}{N} \leq A\frac{M}{N} \tag{C.68}$$

where A is the maximum of A_i's.

Appendix D

Supplementary material for Chapter 6

D.1 Proof of the Existence of Universal Horizon

Here, we investigate the conditions for the existence of a universal horizon for $K = 0$ but arbitrary value of e. As we showed in section 6.5, a universal horizon will appear if there is a radius R_c smaller than $1.5M$ such that the corresponding value of B is B_c. In fact, this is one of the conditions for the existence of universal horizon.

Condition 1:
There exists $R_c \leq 1.5M$ such that $B(R_c) = B_c$.

We prove that this condition is always satisfied. Solving (6.43) for small values of R, we find $B(R = 0) = 0$. It can be checked easily that the minimum value of B at $R = 1.5M$ is always greater than or equal to B_c. Hence there must be $R_c \leq 1.5M$ such that $B(R_c) = B_c$. If there is more than one solution to $B(R) = B_c$ for $R \leq 1.5M$, we call the biggest one R_c.

One more condition also needs to be satisfied: solutions of CMC surfaces must be well defined before the shell radius reaches R_c. In other words, solutions in region II must be well defined all the way to R_c.

Condition 2:
There does not exist $R > R_c$ such that for some value $r > R$, $h(r, R) < 0$.

In the proof of condition 1, we showed that for $R_c \leq R \leq 1.5M$, $B(R) \geq B_c$ (otherwise there is a violation of the condition that R_c is the biggest root of $B(R) = B_c$, which is smaller than $1.5M$). This means that $h(r, R)$ is always positive for $R_c \leq R \leq 1.5M$ (Figure 6.2). As a result, condition 2 is satisfied for this range of R. Furthermore it is obvious that condition 2 is satisfied for $R > 2M$.

Finally, we prove that the condition 2 is satisfied for region $1.5M < R < 2M$ by contradiction. Assume that a radius $R_0 > 1.5M$ exists such that for some value $r_0 > R_0$, $h(r_0, R_0) < 0$. By the properties of the function h it is clear that $B(R_0) < B_c$ (otherwise h is always positive). Also, $h(r, R_0)$ has only one *minimum* at $r_{min} = (\frac{2B^2(R_0)}{m})^{1/3}$. Using $B(R_0) < B_c$, it is clear that $r_{min} < 1.5M$. Since the function h has only one minimum, we conclude

$$r_{min} < 1.5M < R_0 < r_0 \rightarrow h(R_0, R_0) < h(r_0, R_0) < 0. \tag{D.1}$$

However, it can be checked directly that $h(R, R) = 1 - \frac{2m}{R} + \frac{B^2(R)}{R^4} \geq 0$.

Consequently, the appearance of a universal horizon has been shown for arbitrary value of e.

D.2 Dependence of B on K

Here, we want to derive an analytic expression relating B to K in the limit of universal horizon formation. We will use the following approximation for this derivation.

First, we match CMC surfaces to cosmological ones deep inside the Hubble radius. This approximation let us use (6.49) to find a time coordinate for each CMC surface at large radii (large compared to the Schwarzschid radius and small compared to the cosmological horizon). Then, knowing $K = K(t)$ for a specific cosmology, we can assign a value of K to each surface. Second, we perform all calculations in the limit $R \to R_c$. In this limit the time coordinate of each CMC surfaces goes to infinity. We are interested in the leading order divergent term. In the following, when two equations are related through $\sim$, it means that they are equivalent up to their leading order.

We begin by finding t_{CMC}. According to (6.49), we have

$$\begin{aligned} t_{CMC}(r) &= t_{shell}(R) - \int_R^r \mathrm{d}x \frac{\frac{B}{x^2}}{f(x)\sqrt{f(x)+\frac{B^2}{x^4}}} \\ &= t_{shell}(R) - M\int_{R/M}^{r/M} \mathrm{d}x \frac{bx}{(x-2)\sqrt{x^4-2x^3+b^2}} \\ &= t_{shell}(R) - \\ &\quad M\int_{R/M}^{r/M} \mathrm{d}x \frac{bx}{(x-2)\left((x-1.5)^2\left(x^2+x+\frac{3}{4}\right)+b^2-b_c^2\right)^{1/2}} \end{aligned}$$

where $b \equiv \frac{B}{M^2}$ and $b_c \equiv \frac{B_c}{M^2} = \frac{\sqrt{27}}{4}$.

In the limit $R \to R_c$ ($b \to b_c$), the divergent term in the last equation comes from the integral around $x = 1.5$. Considering that $t_{shell}(R)$ limits to a constant value, we find

$$\begin{aligned} &t_{CMC}(r) \sim \\ &- M\int_{1.5-\epsilon}^{1.5+\epsilon} \mathrm{d}x \frac{bx}{(x-2)\sqrt{(x-1.5)^2\left(x^2+x+\frac{3}{4}\right)+b^2-b_c^2}} \\ &\sim 3Mb_c\int_{1.5-\epsilon}^{1.5+\epsilon} \mathrm{d}x \frac{1}{\sqrt{\frac{9}{2}(x-1.5)^2+b^2-b_c^2}}. \end{aligned} \tag{D.2}$$

Using the following identity for small y

$$\int_{-\epsilon}^{+\epsilon} \frac{\mathrm{d}x}{\sqrt{z^2x^2+y^2}} \sim -\frac{\ln(y^2)}{|z|}, \tag{D.3}$$

(D.2) yields

$$t_{CMC} \sim -\sqrt{2} M b_c \ln\left(b^2 - b_c^2\right). \tag{D.4}$$

For a specific cosmological scenario, we can relate K to B through $K = K\left(t_{CMC}\right)$. As an example, for a matter dominated cosmology $K = \frac{2}{t}$, and so

$$K = \frac{2}{t_{CMC}} \sim -\frac{\sqrt{2}}{M b_c \ln(b^2 - b_c^2)}, \tag{D.5}$$

which results in

$$\frac{dK}{dB} = \frac{1}{M^2}\frac{dK}{db} = \frac{2\sqrt{2}}{b_c M^3}\frac{b}{\left(b^2 - b_c^2\right)\left[\ln\left(b^2 - b_c^2\right)\right]^2}. \tag{D.6}$$

As a result, $dK/dB \to \infty$ as $b \to b_c$. Consequently $\left(\frac{\partial t}{\partial K}\right)_r$ does not approach zero in this limit. As another example, consider ΛCDM cosmology. At late times, we have

$$K^2 = K_\Lambda^2 + K_0^2 e^{-K_\Lambda t_{CMC}} \sim K_\Lambda^2 + K_0^2\left(b^2 - b_c^2\right)^{\sqrt{2} b_c M K_\Lambda}, \tag{D.7}$$

which again gives $dK/dB \to \infty$ ($M K_\Lambda \ll 1$).

References

[1] **Planck** Collaboration, P. A. R. Ade *et al.*, "Planck 2015 results. XIII. Cosmological parameters," arXiv:1502.01589 [astro-ph.CO].

[2] **Planck** Collaboration, P. A. R. Ade *et al.*, "Planck 2015 results. XIX. Constraints on primordial magnetic fields," arXiv:1502.01594 [astro-ph.CO].

[3] **Planck** Collaboration, P. A. R. Ade *et al.*, "Planck 2015 results. XX. Constraints on inflation," arXiv:1502.02114 [astro-ph.CO].

[4] G. Hinshaw, D. Larson, E. Komatsu, D. N. Spergel, C. L. Bennett, J. Dunkley, M. R. Nolta, M. Halpern, R. S. Hill, N. Odegard, L. Page, K. M. Smith, J. L. Weiland, B. Gold, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, E. Wollack, and E. L. Wright, "Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results," *apjs* **208** (Oct., 2013) 19, arXiv:1212.5226.

[5] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons," *Phys. Rev. Lett.* **13** (1964) 508–509.

[6] D. Buttazzo, G. Degrassi, P. P. Giardino, G. F. Giudice, F. Sala, A. Salvio, and A. Strumia, "Investigating the near-criticality of the Higgs boson," *JHEP* **12** (2013) 089, arXiv:1307.3536 [hep-ph].

[7] **Virgo, LIGO Scientific** Collaboration, B. P. Abbott *et al.*, "Observation of Gravitational Waves from a Binary Black Hole Merger," *Phys. Rev. Lett.* **116** no. 6, (2016) 061102, arXiv:1602.03837 [gr-qc].

[8] A. Einstein, "Approximative Integration of the Field Equations of Gravitation," *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* **1916** (1916) 688–696.

[9] V. A. Kostelecky and N. Russell, "Data Tables for Lorentz and CPT Violation," *Rev. Mod. Phys.* **83** (2011) 11–31, arXiv:0801.0287 [hep-ph].

[10] S. R. Coleman and S. L. Glashow, "High-energy tests of Lorentz invariance," *Phys. Rev.* **D59** (1999) 116008, arXiv:hep-ph/9812418 [hep-ph].

[11] G. V. Dunne, *Heisenberg-Euler effective Lagrangians: Basics and extensions*. 2004. arXiv:hep-th/0406216 [hep-th].

[12] L. Bombelli, J. Lee, D. Meyer, and R. D. Sorkin, "Space-time as a causal set," *Physical Review Letters* **59** (Aug., 1987) 521–524.

[13] P. Horava, "Quantum Gravity at a Lifshitz Point," *Phys. Rev.* **D79** (2009) 084008, arXiv:0901.3775 [hep-th].

[14] D. Blas, O. Pujolas, and S. Sibiryakov, "Models of non-relativistic quantum gravity: The Good, the bad and the healthy," *JHEP* **04** (2011) 018, arXiv:1007.3503 [hep-th].

[15] T. Jacobson and D. Mattingly, "Gravity with a dynamical preferred frame," *Phys. Rev.* **D64** (2001) 024028, arXiv:gr-qc/0007031 [gr-qc].

[16] S. Aslanbeigi, *Cosmic Atoms: from Causal Sets to Clusters*. PhD thesis, University of Waterloo, 2014.

[17] N. Afshordi, "Gravitational Aether and the thermodynamic solution to the cosmological constant problem," arXiv:0807.2639 [astro-ph].

[18] C. Armendariz-Picon, V. F. Mukhanov, and P. J. Steinhardt, "A Dynamical solution to the problem of a small cosmological constant and late time cosmic acceleration," *Phys.Rev.Lett.* **85** (2000) 4438–4441, arXiv:astro-ph/0004134 [astro-ph].

[19] N. Afshordi, D. J. Chung, and G. Geshnizjani, "Cuscuton: A Causal Field Theory with an Infinite Speed of Sound," *Phys.Rev.* **D75** (2007) 083513, arXiv:hep-th/0609150 [hep-th].

[20] R. D. Sorkin, "Spacetime and causal sets.," in *Relativity and Gravitation*, J. C. D'Olivo, E. Nahmad-Achar, M. Rosenbaum, M. P. Ryan, Jr., L. F. Urrutia, and F. Zertuche, eds., p. 150. 1991.

[21] R. D. Sorkin, "Causal Sets: Discrete Gravity (Notes for the Valdivia Summer School)," *ArXiv General Relativity and Quantum Cosmology e-prints* (Sept., 2003) , gr-qc/0309009.

[22] F. Dowker, "Causal Sets and Discrete Spacetime," in *Albert Einstein Century International Conference*, J.-M. Alimi and A. Füzfa, eds., vol. 861 of *American Institute of Physics Conference Series*, pp. 79–88. Nov., 2006.

[23] J. Henson, "The causal set approach to quantum gravity," *ArXiv General Relativity and Quantum Cosmology e-prints* (Jan., 2006) , gr-qc/0601121.

[24] L. Bombelli, J. Henson, and R. D. Sorkin, "Discreteness Without Symmetry Breaking:. a Theorem," *Modern Physics Letters A* **24** (2009) 2579–2587, gr-qc/0605006.

[25] A. Schild, "Discrete space-time and integral lorentz transformations," *Phys. Rev.* **73** (Feb, 1948) 414–415. http://link.aps.org/doi/10.1103/PhysRev.73.414.

[26] R. D. Sorkin, "Does Locality Fail at Intermediate Length-Scales," in *Approaches to Quantum Gravity – Towards a new understanding of space and time, edited by Daniele Oriti, Cambridge University Press* (2009) 26–43, arXiv:gr-qc/0703099.

[27] D. M. T. Benincasa and F. Dowker, "Scalar Curvature of a Causal Set," *Physical Review Letters* **104** no. 18, (May, 2010) 181301, arXiv:1001.2725 [gr-qc].

[28] F. Dowker and L. Glaser, "Causal set d'Alembertians for various dimensions," *ArXiv e-prints* (May, 2013) , arXiv:1305.2588 [gr-qc].

[29] L. Glaser, "Fixing factors for the causal set d'Alembertian in d dimensions," *ArXiv e-prints* (Nov., 2013) , arXiv:1311.1701 [math-ph].

[30] I. Thompson, "NIST Handbook of Mathematical Functions, edited by Frank W.J. Olver, Daniel W. Lozier, Ronald F. Boisvert, Charles W. Clark," *Contemporary Physics* **52** (Sept., 2011) 497–498.

[31] G. D. Domínguez and S. E. Trione, "On the Laplace transforms of retarded, Lorentz-invariant functions," *Advances in Mathematics* **31** no. 1, (Jan., 1979) 51–62.

[32] R. D. Sorkin, "Does Locality Fail at Intermediate Length-Scales," *ArXiv General Relativity and Quantum Cosmology e-prints* (Mar., 2007) , gr-qc/0703099.

[33] S. Aslanbeigi, M. Saravani, and R. D. Sorkin, "Generalized causal set d'Alembertians," *Journal of High Energy Physics* **6** (June, 2014) 24, arXiv:1403.1622 [hep-th].

[34] D. M. T. Benincasa and F. Dowker, "Scalar Curvature of a Causal Set," *Physical Review Letters* **104** no. 18, (May, 2010) 181301, arXiv:1001.2725 [gr-qc].

[35] F. Dowker and L. Glaser, "Causal set d'Alembertians for various dimensions," *Classical and Quantum Gravity* **30** no. 19, (Oct., 2013) 195016, arXiv:1305.2588 [gr-qc].

[36] L. Glaser, "A closed form expression for the causal set d'Alembertian," *Classical and Quantum Gravity* **31** no. 9, (May, 2014) 095007, arXiv:1311.1701 [math-ph].

[37] R. Woodard, "Nonlocal Models of Cosmic Acceleration," *Found.Phys.* **44** (2014) 213–233, arXiv:1401.0254 [astro-ph.CO].

[38] S. Deser and R. Woodard, "Nonlocal Cosmology," *Phys.Rev.Lett.* **99** (2007) 111301, arXiv:0706.2151 [astro-ph].

[39] A. Kamenev, "Course 3 many-body theory of non-equilibrium systems," *Les Houches* **81** (2005) 177–246.

[40] A. Belenchia, D. M. T. Benincasa, and S. Liberati, "Nonlocal Scalar Quantum Field Theory from Causal Sets," arXiv:1411.6513 [gr-qc].

[41] N. Afshordi, S. Aslanbeigi, and R. D. Sorkin, "A distinguished vacuum state for a quantum field in a curved spacetime: formalism, features, and cosmology," *Journal of High Energy Physics* **8** (Aug., 2012) 137, arXiv:1205.1296 [hep-th].

[42] S. Johnston, "Feynman Propagator for a Free Scalar Field on a Causal Set," *Physical Review Letters* **103** no. 18, (Oct., 2009) 180401, arXiv:0909.0944 [hep-th].

[43] S. Weinberg, *The Quantum Theory of Fields*. Feb., 2000.

[44] M. Persic, P. Salucci, and F. Stel, "The Universal rotation curve of spiral galaxies: 1. The Dark matter connection," *Mon.Not.Roy.Astron.Soc.* **281** (1996) 27, arXiv:astro-ph/9506004 [astro-ph].

[45] D. Clowe, M. Bradac, A. H. Gonzalez, M. Markevitch, S. W. Randall, *et al.*, "A direct empirical proof of the existence of dark matter," *Astrophys.J.* **648** (2006) L109–L113, arXiv:astro-ph/0608407 [astro-ph].

[46] **WMAP** Collaboration, G. Hinshaw *et al.*, "Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results," *Astrophys.J.Suppl.* **208** (2013) 19, arXiv:1212.5226 [astro-ph.CO].

[47] M. Milgrom, "A Modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis," *Astrophys.J.* **270** (1983) 365–370.

[48] J. D. Bekenstein, "Relativistic gravitation theory for the MOND paradigm," *Phys.Rev.* **D70** (2004) 083509, arXiv:astro-ph/0403694 [astro-ph].

[49] J. Moffat, "Gravitational theory, galaxy rotation curves and cosmology without dark matter," *JCAP* **0505** (2005) 003, arXiv:astro-ph/0412195 [astro-ph].

[50] J. Moffat, "Scalar-tensor-vector gravity theory," *JCAP* **0603** (2006) 004, arXiv:gr-qc/0506021 [gr-qc].

[51] T. Clifton, P. G. Ferreira, A. Padilla, and C. Skordis, "Modified Gravity and Cosmology," *Phys.Rept.* **513** (2012) 1–189, arXiv:1106.2476 [astro-ph.CO].

[52] J. L. Feng, "Dark Matter Candidates from Particle Physics and Methods of Detection," *Ann.Rev.Astron.Astrophys.* **48** (2010) 495–545, arXiv:1003.0904 [astro-ph.CO].

[53] M. Viel, G. D. Becker, J. S. Bolton, and M. G. Haehnelt, "Warm dark matter as a solution to the small scale crisis: New constraints from high redshift Lyman-? forest data," *Phys.Rev.* **D88** (2013) 043502, arXiv:1306.2314 [astro-ph.CO].

[54] M. Saravani and S. Aslanbeigi, "Dark Matter From Spacetime Nonlocality," *Phys. Rev.* **D92** no. 10, (2015) 103504, arXiv:1502.01655 [hep-th].

[55] R. D. Sorkin, "Does locality fail at intermediate length-scales," arXiv:gr-qc/0703099 [GR-QC].

[56] D. M. Benincasa and F. Dowker, "The Scalar Curvature of a Causal Set," *Phys.Rev.Lett.* **104** (2010) 181301, arXiv:1001.2725 [gr-qc].

[57] C. Burgess, "Supersymmetric large extra dimensions and the cosmological constant: An Update," *Annals Phys.* **313** (2004) 283–401, arXiv:hep-th/0402200 [hep-th].

[58] N. Afshordi and E. Nelson, "Cosmological Non-Constant Problem: Cosmological bounds on TeV-scale physics and beyond," arXiv:1504.00012 [hep-th].

[59] L. Kofman, A. D. Linde, and A. A. Starobinsky, "Towards the theory of reheating after inflation," *Phys.Rev.* **D56** (1997) 3258–3295, arXiv:hep-ph/9704452 [hep-ph].

[60] A. L. Erickcek and K. Sigurdson, "Reheating Effects in the Matter Power Spectrum and Implications for Substructure," *Phys.Rev.* **D84** (2011) 083503, arXiv:1106.0536 [astro-ph.CO].

[61] **Planck** Collaboration, P. Ade *et al.*, "Planck 2013 results. XXII. Constraints on inflation," *Astron.Astrophys.* **571** (2014) A22, arXiv:1303.5082 [astro-ph.CO].

[62] K. Freese, J. A. Frieman, and A. V. Olinto, "Natural inflation with pseudo - Nambu-Goldstone bosons," *Phys.Rev.Lett.* **65** (1990) 3233–3236.

[63] A. A. Starobinsky, "A New Type of Isotropic Cosmological Models Without Singularity," *Phys.Lett.* **B91** (1980) 99–102.

[64] A. L. Erickcek and N. M. Law, "Astrometric Microlensing by Local Dark Matter Subhalos," *Astrophys.J.* **729** (2011) 49, arXiv:1007.4228 [astro-ph.CO].

[65] S. Baghram, N. Afshordi, and K. M. Zurek, "Prospects for Detecting Dark Matter Halo Substructure with Pulsar Timing," *Phys.Rev.* **D84** (2011) 043511, arXiv:1101.5487 [astro-ph.CO].

[66] S. Rahvar, S. Baghram, and N. Afshordi, "Transient Weak-Lensing by Cosmological Dark Matter Microhaloes," *Phys.Rev.* **D89** no. 6, (2014) 063001, arXiv:1310.5412 [astro-ph.CO].

[67] T. Jacobson, "Einstein-aether gravity: A Status report," *PoS* **QG-PH** (2007) 020, arXiv:0801.1547 [gr-qc].

[68] K. Yagi, D. Blas, N. Yunes, and E. Barausse, "Strong Binary Pulsar Constraints on Lorentz Violation in Gravity," arXiv:1307.6219 [gr-qc].

[69] J. W. Elliott, G. D. Moore, and H. Stoica, "Constraining the new Aether: Gravitational Cerenkov radiation," *JHEP* **0508** (2005) 066, arXiv:hep-ph/0505211 [hep-ph].

[70] C. Bonvin, C. Caprini, and R. Durrer, "A no-go theorem for k-essence dark energy," *Phys.Rev.Lett.* **97** (2006) 081303, arXiv:astro-ph/0606584 [astro-ph].

[71] N. Afshordi, "Cuscuton and low energy limit of Horava-Lifshitz gravity," *Phys. Rev.* **D80** (2009) 081502, arXiv:0907.5201 [hep-th].

[72] H. Gomes, S. Gryb, and T. Koslowski, "Einstein gravity as a 3D conformally invariant theory," *Class.Quant.Grav.* **28** (2011) 045005, arXiv:1010.2481 [gr-qc].

[73] H. Gomes and T. Koslowski, "The Link between General Relativity and Shape Dynamics," *Class.Quant.Grav.* **29** (2012) 075009, arXiv:1101.5974 [gr-qc].

[74] A. Kehagias and K. Sfetsos, "The Black hole and FRW geometries of non-relativistic gravity," *Phys.Lett.* **B678** (2009) 123–126, arXiv:0905.0477 [hep-th].

[75] T. Jacobson, "Extended Horava gravity and Einstein-aether theory," *Phys.Rev.* **D81** (2010) 101502, arXiv:1001.4823 [hep-th].

[76] D. Blas and S. Sibiryakov, "Horava gravity versus thermodynamics: The Black hole case," *Phys. Rev.* **D84** (2011) 124043, arXiv:1110.2195 [hep-th].

[77] E. Barausse, T. Jacobson, and T. P. Sotiriou, "Black holes in Einstein-aether and Horava-Lifshitz gravity," *Phys. Rev.* **D83** (2011) 124043, arXiv:1104.2889 [gr-qc].

[78] E. Barausse and T. P. Sotiriou, "Slowly rotating black holes in Horava-Lifshitz gravity," *Phys. Rev.* **D87** (2013) 087504, arXiv:1212.1334.

[79] E. Babichev, V. F. Mukhanov, and A. Vikman, "Escaping from the black hole?," *JHEP* **09** (2006) 061, arXiv:hep-th/0604075 [hep-th].

[80] R. Akhoury, D. Garfinkle, and R. Saotome, "Gravitational collapse of k-essence," *JHEP* **1104** (2011) 096, arXiv:1103.0290 [gr-qc].

[81] C. D. Leonard, J. Ziprick, G. Kunstatter, and R. B. Mann, "Gravitational collapse of K-essence Matter in Painlevé-Gullstrand coordinates," *JHEP* **1110** (2011) 028, arXiv:1106.2054 [gr-qc].

[82] P. Berglund, J. Bhattacharyya, and D. Mattingly, "Mechanics of universal horizons," *Phys. Rev.* **D85** (2012) 124019, arXiv:1202.4497 [hep-th].

[83] E. Babichev, V. Mukhanov, and A. Vikman, "k-Essence, superluminal propagation, causality and emergent geometry," *JHEP* **02** (2008) 101, arXiv:0708.0561 [hep-th].

[84] D. Garfinkle, C. Eling, and T. Jacobson, "Numerical simulations of gravitational collapse in Einstein-aether theory," *Phys.Rev.* **D76** (2007) 024003, arXiv:gr-qc/0703093 [GR-QC].

[85] P. Berglund, J. Bhattacharyya, and D. Mattingly, "Towards Thermodynamics of Universal Horizons in Einstein-ther Theory," *Phys. Rev. Lett.* **110** no. 7, (2013) 071301, arXiv:1210.4940 [hep-th].

[86] S. Dubovsky and S. Sibiryakov, "Spontaneous breaking of Lorentz invariance, black holes and perpetuum mobile of the 2nd kind," *Phys.Lett.* **B638** (2006) 509–514, arXiv:hep-th/0603158 [hep-th].

[87] C. Eling, B. Z. Foster, T. Jacobson, and A. C. Wall, "Lorentz violation and perpetual motion," *Phys.Rev.* **D75** (2007) 101502, arXiv:hep-th/0702124 [HEP-TH].

[88] D. G. Boulware and S. Deser, "Can gravitation have a finite range?," *Phys. Rev.* **D6** (1972) 3368–3382.

[89] K. S. Stelle, "Renormalization of Higher Derivative Quantum Gravity," *Phys. Rev.* **D16** (1977) 953–969.

[90] N. Afshordi, "Why is High Energy Physics Lorentz Invariant?," arXiv:1511.07879 [hep-th].

[91] T. Jacobson, "Lorentz violation and Hawking radiation," in *CPT and Lorentz symmetry. Proceedings: 2nd Meeting, Bloomington, USA, Aug 15-18, 2001*, pp. 316–320. 2002. arXiv:gr-qc/0110079 [gr-qc]. http://alice.cern.ch/format/showfull?sysnb=2280342.

[92] T. P. Sotiriou, I. Vega, and D. Vernieri, "Rotating black holes in three-dimensional Ho?ava gravity," *Phys. Rev.* **D90** no. 4, (2014) 044046, arXiv:1405.3715 [gr-qc].

[93] C. Ding, A. Wang, and X. Wang, "Charged Einstein-aether black holes and Smarr formula," arXiv:1507.06618 [gr-qc].

[94] J. Bhattacharyya, M. Colombo, and T. P. Sotiriou, "Causality and black holes in spacetimes with a preferred foliation," arXiv:1509.01558 [gr-qc].

[95] E. Barausse and T. P. Sotiriou, "Black holes in Lorentz-violating gravity theories," *Class. Quant. Grav.* **30** (2013) 244010, arXiv:1307.3359 [gr-qc].

[96] A. Mohd, "On the thermodynamics of universal horizons in Einstein-Æther theory," arXiv:1309.0907 [gr-qc].

[97] J. Bhattacharyya and D. Mattingly, "Universal horizons in maximally symmetric spaces," *Int. J. Mod. Phys.* **D23** no. 13, (2014) 1443005, arXiv:1408.6479 [hep-th].

[98] F. Michel and R. Parentani, "Black hole radiation in the presence of a universal horizon," *Phys. Rev.* **D91** no. 12, (2015) 124049, arXiv:1505.00332 [gr-qc].

[99] M. Saravani, N. Afshordi, and R. Mann, "Dynamical Emergence of Universal Horizons during the formation of Black Holes," *Phys. Rev.* **D89** (2014) 084029, arXiv:gr-qc/1310.4143v2.

[100] K. Maeda, "Gravitational Collapse of Charged Dust Shell and Maximal Slicing Condition," *Progress of Theoretical Physics* **63** no. 2, (1980) 425–437.

[101] W. Israel and V. de la Cruz, "Gravitational bounce," *Il Nuovo Cimento A Series 10* **51** (1967) 744–760.

[102] E. Poisson and W. Israel, "Inner-horizon instability and mass inflation in black holes," *Phys. Rev. Lett.* **63** (1989) 1663–1666.

[103] E. Barausse, T. P. Sotiriou, and I. Vega, "Slowly rotating black holes in Einstein-æther theory," arXiv:1512.05894 [gr-qc].

[104] N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space*. Apr., 1984.

[105] M. Meiers, M. Saravani, and N. Afshordi, "Cosmic Censorship in Lorentz Violating Theories of Gravity," arXiv:1511.08969 [gr-qc].

[106] M. Saravani and S. Aslanbeigi, "On the Causal Set-Continuum Correspondence," *Class. Quant. Grav.* **31** no. 20, (2014) 205013, arXiv:1403.6429 [hep-th].

[107] K. Hinterbichler, "Theoretical Aspects of Massive Gravity," *Rev. Mod. Phys.* **84** (2012) 671–710, arXiv:1105.3735 [hep-th].