Cosmic Atoms: from Causal Sets to Clusters

by

Siavash Aslanbeigi

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, 2014

Author's Declaration

I hereby declare that I am the sole author of this 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.

Abstract

This thesis is a collection of works which seek to address certain aspects of the following puzzles: quantum gravity, initial conditions of the universe, the cosmological constant problem, and baryonic processes in clusters of galaxies.

The causal set theory approach to quantum gravity postulates that the fundamental structure of spacetime is a partially ordered and locally finite set. Within this framework, we formulate a statistical correspondence for judging when a causal set is well approximated by a Lorentzian manifold. To that end, we prove a theorem which shows that the number-volume correspondence, if required to hold even for arbitrarily small regions, is best realized via Poisson sprinkling. Surprisingly, we also show that 1+1 dimensional Lorentzian lattices provide a much better N–V correspondence than Poisson sprinkling for macroscopically large volumes.

We then study phenomenological implications of wave propagation on causal sets. We introduce a family of generalized d'Alembertian operators $\Box_\rho$ in D-dimensional Minkowski spacetime which are manifestly Lorentz-invariant, retarded, and non-local, the extent of the nonlocality being governed by a single parameter ρ. These operators arise as averages of matrix operators meant to describe the propagation of a scalar field in a causal set. We compute their action on plane waves and show that their spectrum $g(p)$ (p being the momentum-vector) contains also an imaginary part for timelike p, 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. We also address the question of whether or not the evolution defined by $\Box_\rho$ is stable, finding evidence that the original 4D causal set d'Alembertian is unstable, while its 2D counterpart is stable.

Motivated by the construction of a quantum field theory on a background causal set, we proceed to define a distinguished "ground state" or "vacuum" for a free scalar quantum field in a globally hyperbolic region of an arbitrarily curved spacetime. This state, which we call the Sorkin-Johnston (SJ) state, is defined for all compact regions and for many noncompact ones. In a static spacetime we find that the SJ vacuum coincides with the usual ground state. We then determine it for a massive free scalar field in $D = d+1$ dimensional de Sitter space. In cases where it is well-defined, the SJ vacuum always corresponds to one of the de Sitter-invariant α-vacua, with the value of α depending on (i) whether the mass of the field is in the complementary or principal series, (iii) whether it is evaluated on the complete de Sitter manifold or its Poincaré half-space, and (iii) whether the spacetime dimension is even or odd. We determine the SJ state also for a radiation-filled, spatially homogeneous

and isotropic cosmos, and show that the super-horizon correlations are approximately the same as those of a thermal state. Finally, we illustrate the inherent non-locality of our prescription with the example of a spacetime which sandwiches a region with curvature in-between flat initial and final regions.

The cosmological constant "problem" refers to the huge discrepancy between the observed value of the cosmological constant and our expectations from the vacuum energy of quantum fields. We explore a modified-gravity proposal which addresses this puzzle by decoupling the quantum vacuum from space-time geometry via an incompressible fluid, known as *Gravitational Aether*. We discuss classical predictions of this theory along with its compatibility with cosmological and experimental tests of gravity. We argue that deviations from General Relativity (GR) in this theory are sourced by pressure or vorticity. In particular, the theory predicts that the gravitational constant for radiation is 33% larger than that of non-relativistic matter, which we show is consistent at the $\sim 2\sigma$ level with (most pre-*Planck*) cosmic microwave background (CMB), Ly-α forest, and light element abundance observations. It is further shown that all Parametrized Post-Newtonian (PPN) parameters have the standard GR values aside from the anomalous coupling to pressure ζ_4. A more subtle prediction of this model (assuming irrotational aether) is that the so-called gravitomagnetic effect is 33% larger than the GR prediction. This is consistent with current limits from LAGEOS and Gravity Probe B at $\sim 2\sigma$ level.

Finally, we present a statistically-optimal and model-independent method to extract the pressure profile of hot gas in the intracluster medium (ICM). Using the thermal Sunyaev-Zeldovich effect, we constrain the mean pressure profile of the ICM by appropriately considering *all* primary CMB and instrumental noise correlations, while using the maximum resolution and sensitivity of all frequency channels. As a first application, we analyze CMB maps of WMAP 9-year data through a study of the Meta-Catalogue of X-ray detected Clusters of galaxies (MCXC). We constrain the universal pressure profile out to $4R_{500}$ with 15σ confidence, though our measurements are only significant out to R_{200}. Using a temperature profile constrained from X-ray observations, we measure the mean gas mass fraction out to R_{200}. Within statistical and systematic uncertainties, our constraints are compatible with the cosmic baryon fraction and the expected gas fraction in halos. While *Planck* multi-frequency CMB data are expected to reduce statistical uncertainties by a factor of ~ 20, we argue that systematic errors in determining mass of clusters dominate the uncertainty in gas mass fraction measurements at the level of ~ 20 percent.

Acknowledgements

The past four years have been undoubtedly some of the most memorable years of my life. It gives me immense pleasure to thank the Perimeter Institute for Theoretical Physics and the University of Waterloo for hosting me during this time.

Since coming to graduate school, I have learned from many people. First and foremost, I would like to thank my advisor, Niayesh Afshordi. Much of the research in this thesis was suggested by him, directly or indirectly. I am thankful to him for not only allowing me, but also encouraging me to pursue whatever intrigued me, even when he was not particularly interested in those topics. I have always admired his passion for physics, broad range of interests, and impressive ability to simplify difficult problems. I am forever indebted to him for all that I have learned from him throughout our many discussions.

For all intents and purposes, Rafael Sorkin has been my co-supervisor during the past four years. It is because of him that I became interested in causal set theory, an approach to quantum gravity he pioneered. His broad knowledge of physics and mathematics, his careful way of thinking and reasoning, and more generally, his world-views have made a deep impression on me. If anyone is to blame for the long list of appendices in this thesis, aside from myself, it is probably him, as I have learned from him not to leave logical gaps between arguments. I would like to thank him for all that he has taught me.

I would also like to thank my good friend and collaborator Mehdi Saravani, with whom I have spent countless hours discussing physics and matters of life. He has impressed me time and again with his sharp intellect and is responsible for many of the breakthroughs in our work together. It gives me pleasure to also acknowledge my other collaborators, particularly Michel Buck and Guilhem Lavaux. I am especially indebted to Guilhem Lavaux for teaching me Bayesian statistics, CMB template fitting, and Python.

I have learned various interesting facts about physics, math, and other matters of life from having conversations with Achim Kempf, Latham Boyle, Matthew Johnson, Fay Dowker, Erich Poppitz, Amir Hajian, Ghazal Geshnizjani, David Rideout, Dionigi Benincasa, Kendrick Smith, Lee Smolin, Bill Unruh, Kurt Hinterbichler, Paul Mcfadden, John Moffat, Laurant Freidel, Farbod Kamiab, Yasaman Yazdi, Peter Lunts, Heidar Moradi, Joao Caetano, Nima Doroud, Dalimil Mazac, Razieh Pourhasan, Anton Van Niekerk, Luis Lehner, Avery Broderick, Daniel Carrasco Guariento, Mohammad Shalaby, Nosiphiwo Zwane, Uzair Hussain, Erfan Shalchian, and Ali Feizmohammadi.

I would like to thank Dan Lynch and the Black Hole Bistro staff of the Perimeter Institute for their amazing food and services.

A special thanks is due to my good friend and roommate Farbod Kamiab, with whom I have enjoyed many late-night conversations about physics, history, politics, music, and life.

I am indebted to my siblings Sara, Parisa, and Siamak Aslanbeigi for their love and support. I am eternally grateful to my parents, Aram Kavoosi and Shahram Aslanbeigi, to whom I dedicate this thesis.

Finally, I would like to thank my fiancé Ana, who has patiently tolerated our distance apart. Her continuous love and support has been incredibly important to me.

Dedication

To my parents, Aram Kavoosi and Shahram Aslanbeigi.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The past four years, during which the material presented in this thesis was prepared, have been incredibly exciting for fundamental physics. The Standard Model Higgs boson was observed by the ATLAS detector at the Large Hadron Collider, with a measured mass of $126.0 \pm 0.4(\text{stat}) \pm 0.4(\text{sys})$ GeV [1]. It is somewhat reassuring for the purposes of this thesis,[1] and cosmological models in general, that a fundamental scalar field exists in nature. There have also been important advances in observational cosmology. We now have all-sky measurements of the temperature anisotropies of the cosmic microwave background (CMB) by the *Planck* satellite. These measurements have unprecedented precision and span a much wider range of frequencies (30–857 GHz) compared to the Wilkinson Microwave Anisotropy Probe (WMAP) CMB data [116]. One of the most notable findings of the *Planck* team was the absence of any detectable primordial non-Gaussianity [119]. Previous constraints came from 9-year WMAP data, which obtained (as an example) $f_{NL}^{\text{loc}} = 37.2 \pm 19.9$ for the primordial local bispectrum amplitude [25]. *Planck*'s measurement of the same quantity is $f_{NL}^{\text{loc}} = 2.7 \pm 5.8$, shrinking the error bar by a factor of ~ 3.5. In Chapters 5 and 6 of this thesis, we will encounter two other cases where *Planck*'s CMB measurements make a decisive difference.

Another recent exciting development has been the detection of the so-called B-mode component of the CMB polarization, as claimed by the BICEP2 experiment [28]. This "curl" component of the polarization pattern cannot be generated by linear gaussian density perturbations and is considered an indirect detection of primordial gravitational waves. The BICEP2 team has obtained $r = 0.20^{+0.07}_{-0.05}$, where r is the ratio of the amplitude of primordial gravitational waves to that of the scalar fluctuations. Let us assume for the

[1] Chapters 3 and 4 consider only scalar fields.

moment that the BICEP2 experiment has indeed measured primordial gravitational waves. In the context of the simplest inflationary models, this measurement fixes the energy scale of inflation to be $\sim 10^{16}$ GeV (see e.g. [98]). That is, with the BICEP2 result, we would have probed an energy scale of $\sim 10^{16}$ GeV, only 5 orders of magnitude short of the Planck scale. Furthermore, the inflationary predictions are in the context of linearized quantum gravity, meaning that we would have seen evidence for quantum fluctuations of spacetime itself. We need not be married to the inflationary paradigm to appreciate that a detection of primordial gravitational waves could potentially guide us towards the correct theory of the early universe, and hopefully quantum gravity. For instance, in the causal set theory approach to quantum gravity (which is the subject of Chapters 2 and 3 of this thesis), it is hoped that the currently probed initial conditions of the universe are remnants of the quantum gravity era. Therefore, depending on whether or not tensorial perturbations are suppressed primordially, one may choose to pursue a different route towards the quantum dynamics of causal sets. Unfortunately, it is not clear at the moment whether the BICEP2 experiment has really detected primordial gravitational waves or not. The trouble seems to be that the detected signal can also be explained by significant dust polarization (see e.g. [63, 103]). The *Planck* CMB polarization data and future observations should eventually be able to settle this debate.

As a final example of an exciting moment in the past four years, I cannot help but mention the claim that OPERA had detected superluminal neutrinos [2].[2] Interestingly enough, by this point I had spent some time working out Lorentz-*preserving* phenomenological implications of causal set theory, one of whose core assumptions is that Lorentz symmetry is fundamental. Had the OPERA result not have been due to experimental error, I may have had no reason to include Chapters 2 and 3 in this thesis.

As fruitful as the recent years have been, we still face many unanswered questions. A long-lasting mystery is the problem of quantum gravity: a theory which reduces to general relativity on the one hand, and the standard model of particle physics on the other hand, in certain appropriate limits. It is hoped that such a theory would provide insights into important physical questions such as the nature of black holes and the big bang, where general relativity breaks down. The initial conditions of the universe pose another important (and probably related) puzzle. What set of initial conditions should one consider as "natural"? To what extent is it possible to do away with "fine tuning"? At lower energies and larger scales, we face the fact that (optically) visible matter makes up only $\sim 5\%$ of the energy budget of the universe [118]. We have learned a lot about dark matter through its gravitational interaction, but have yet to detect it directly or indirectly. The other dark component of the universe, i.e. dark energy (or more appropriately any

[2] See version 1 on the arXiv:1109.4897v1.

mechanism which can explain the recent accelerated expansion of the universe), is another long-standing puzzle, often referred to as the "cosmological constant problem".

This thesis is a collection of works which seek to address certain aspects of the aforementioned problems. The topics covered can be broadly categorized as causal set quantum gravity, quantum field theory in curved spacetimes, modified gravity, and cluster cosmology. In the next section, I will provide an overview of the material presented in all chapters, as well as how they connect and fit in the big picture. I have tried to make every chapter as self-contained as possible, so that the reader could explore each one in isolation. For the convenience of the reader, I have also provided details of (almost all) derivations and calculations in the appendices. In some cases, the appendices also include derivations of standard results which are quoted but not properly documented in the literature.

1.1 Outline

Our journey starts with the problem of quantum gravity in Chapter 2. One way to hope for a deeper understanding of spacetime is to question the initial assumptions of our current theories, especially those which are suspected to lead to trouble down the road. A core assumption of most known physical theories, including general relativity (GR), is that spacetime is continuous. We argue that the continuum description of spacetime should be given up on physical grounds, mentioning amongst other reasons the problematic infinities of general relativity (i.e. singularities) and quantum field theory. On our way to abandoning the continuum, we suggest that causal order information should be retained because it does not depend too heavily on the assumption of continuity *and* contains a great deal of information: the topology, differential structure, and metric of a Lorentzian geometry can be recovered from its causal order up to a local conformal factor [73, 94]. We point out that the missing information about absolute scale can be *recovered* in a discrete setting, through the counting of elements. This leads us to causal set theory, which postulates that the fundamental structure of spacetime is that of a partially ordered and locally finite set, or in short a *causal set* (causet).

We then proceed to explore certain aspects of the question: when is a causal set well-approximated by a Lorentzian manifold? In particular, we begin by seeking a formal notion of encoding volume information through counting of causal set elements: a spacetime region with volume V should contain "about" ρV causal set elements, where $\rho \sim l_p^{-4}$ sets the fundamental discreteness scale. We provide a statistical formulation of this so-called *number–volume (N–V) correspondence*. The attitude in the causal set program is that the N–V correspondence is best realized (i.e. with the least noise) through *Poisson sprinkling*,

which is a procedure for generating a causal set from a Lorentzian manifold (M, g): randomly select points from (M, g) at a given density ρ using the Poisson process, and endow the selected points with their causal order information from (M, g). The main reason for this expectation is that, unlike in a Riemannian setting, *random* lattices preserve Lorentz invariance much better than regular lattices [33]. To that end, we prove a theorem which shows that the N–V correspondence is best realized via Poisson sprinkling for arbitrarily small volumes. Unfortunately, this result has little physical significance because 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 sprinkling, our result would have to be generalized to the case of larger volumes. Quite surprisingly, though, we present a counter-example: $1 + 1$ dimensional Lorentzian lattices. A Lorentzian lattice is a regular lattice which is invariant under a discrete subgroup of the Lorentz group. We show that $1+1$ dimensional Lorentzian lattices provide a much better N–V correspondence than Poisson sprinkling for large volumes. However, we present evidence that this feature should not persist in $3+1$ dimensions and conjecture that the Poisson process does indeed provide the best N–V correspondence for macroscopically large spacetime regions.

An intriguing observation about causal sets is that their marriage of discreteness with causal order implies that physics cannot remain local at all scales. To appreciate why this should be, consider how one might define a notion of "closeness" in a causal set, confining ourselves to causal sets C which are obtained from Poisson sprinklings of a Lorentzian manifold M. 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 inevitably 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. Chapter 3 seeks to explore phenomenological implications of this observation.

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 (locally) discretizes the Laplacian operator in a Riemannian spacetime. Nevertheless, a non-local operator was suggested in [142] 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 $\mathbb{M}^2$). In this proposal, one starts with a discrete operator B_ρ on the causal set, 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 $\mathbb{M}^2$, averaging B_ρ over all such causets (i.e. over all Poisson sprinklings of $\mathbb{M}^2$) leads to a *non-local*, *retarded*, and manifestly *Lorentz-invariant* continuum operator $\Box_\rho$, which has the crucial property that it reproduces the usual d'Alembertian in the limit of zero non-locality scale: $\Box_\rho\phi \to \Box\phi$ as $\rho \to \infty$. The expression introduced in [142] was generalized to $D = 4$ dimensions in [24] and recently to arbitrary D in [56, 67].

A natural question then arises: might such a nonlocality in the wave propagation of matter fields 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 when the local operator $\Box$ is replaced by the nonlocal operator $\Box_\rho$. We make a start on answering this question in Chapter 3 by analysing the "spectral properties" (Fourier transform) of $\Box_\rho$ in D–dimensional Minkowski space $\mathbb{M}^D$. Along the way, we generalize the original proposals [142, 24] to an infinite family of "Generalized Causet Box (GCB) operators" which are 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 (for all dimensions). We compute the action of all Poisson-averaged GCB operators on plane waves, or equivalently their Fourier transforms $g(p)$ [p being the momentum-vector]. Quite remarkably, 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*. In real space, this UV behaviour corresponds to a delta-function contribution to the propagator, meaning that the nonlocality has made the UV divergences *worse*, which is the opposite of what one would have expected. We use this behaviour to raise the possibility of a genuinely Lorentzian perturbative regulator for quantum field theory in $\mathbb{M}^D$. Finally, we address the question of whether or not the evolution defined by the (classical) equation $\Box_\rho\phi = 0$ 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.

So far we have described the propagation of a *classical* scalar field on the causal set. What about a *quantum* field, though? After a moment of thought, it becomes clear that the standard canonical quantization procedure in the continuum has no natural extension to the causal set. For instance, it has proven difficult to define an intrinsic notion of a spacelike hypersurface on the causal set. This immediately rules out the possibility of posing equal-time commutation relations. More generally, the entire quantization process in the continuum— as usually conceived — boils down to selecting an appropriate subspace of the solution space of the Klein-Gordon equation. But that way of organizing the problem seems to break down in the case of a causal set. There, the notion of "approximate solution"

seems to be the best that is available, and one therefore requires a different starting point. In [84], Steven Johnston found such a starting point in (the discrete analog of) the retarded Green's function. On that basis a complete counterpart of the quantum field theory of a free scalar field was built up, and a unique "vacuum state" was derived.

It was then noticed by Rafael Sorkin that Johnston's proposal can be generalized to quantum fields on continuum curved spacetimes, in order to tackle a long-lasting problem in the framework "quantum field theory in curved spacetime": the absence of a unique notion of vacuum. Chapter 4 describes this so-called Sorkin-Johnston (SJ) proposal, which defines a *distinguished* vacuum state for a free scalar quantum field in an *arbitrarily* curved spacetime. The availability of a distinguished vacuum could be welcome especially in relation to cosmology, where great interest attaches to the question whether certain sorts of states can be regarded as "natural" for the early universe.

We start by showing that the SJ state is well-defined for all compact globally-hyperbolic spacetimes. We then prove that for static spacetimes, the SJ state coincides with the usual ground state. This is reassuring because in static spacetimes a natural choice of vacuum exists, namely the state which minimizes the Hamiltonian (which in this case is time independent).

We then proceed to determine the SJ vacuum for a massive free scalar field in $D = d+1$ dimensional de Sitter space. In those cases where the prescription is well-defined, we find that the SJ vacuum always corresponds to one of the de Sitter-invariant α-vacua. We show that the SJ vacuum depends on *(i)* whether the mass of the field is in the complementary or principal series (i.e. below or above the critical value $(D-1)/(2\ell)$, where ℓ is the de Sitter radius), *(ii)* whether it is evaluated on the complete de Sitter manifold or its Poincaré half-space, and *(iii)* whether the spacetime dimension is even or odd. For a field of mass $m < (D-1)/(2\ell)$ on the Poincaré patch, the SJ prescription cannot be applied to the entire spacetime, but only a bounded globally hyperbolic subregion of it. We find that the SJ vacuum in de Sitter space does not in general correspond to the Bunch-Davies or Euclidean state, and as a result is not always Hadamard.

Using the discrete SJ formalism on a causal set, we also determine the SJ state on a sprinkling of a causal diamond in $1+1$ dimensional de Sitter space. Our simulation shows that the mean of the discrete SJ two-point function is consistent with that of an α-vacuum and in particular with that of the Euclidean vacuum in the centre of the diamond (away from the edges) for a field of mass $m \ll \sqrt{\rho}$, where ρ is the sprinkling density. This is encouraging, since the QFT defined on causal sets by the SJ formalism seems to reproduce what one would expect: a state that respects the spacetime isometries in the appropriate "continuum limit". It is quite remarkable that the SJ formalism on the causal set allows

us to simulate quantum field theories on curved background.

Furthermore, to make a start on exploring the phenomenological implications of the SJ state in relation to cosmology, we determine it for a scalar field of mass $m \geq 0$ in a radiation-filled, spatially flat, homogeneous and isotropic cosmos. In that connection we also compute an "effective temperature" that can be defined for the super-horizon modes of the massive field. The correlations found thereby could open up a new perspective on the question of primordial fluctuations and the so-called "horizon problem".

Finally, we point out a peculiar aspect of the SJ prescription, namely its temporal nonlocality. We demonstrate this feature by the example of a spacetime which sandwiches a region with curvature in-between flat initial and final regions. In a cosmological setting, this temporal nonlocality poses a puzzle because it is not clear what portion of spacetime one should use to compute the SJ vacuum. For instance, should one consider the behaviour of late-time cosmology to determine the SJ vacuum for the early universe? Of course, the SJ proposal is not the only setting in which there is a connection between late-time cosmology and the physics of the very small. The so-called cosmological constant problem, which is the subject of Chapter 5, is perhaps the most puzzling manifestation of this sort of connection.

The discovery of recent acceleration of cosmic expansion was one of the most surprising findings in modern cosmology [130, 112]. The standard cosmological model drives this expansion with a cosmological constant, which is consistent with nearly all current cosmological observations. There is, however, something unsettling about the extremely small (in Planck units) but nonzero value of the cosmological constant. This becomes even more puzzling when we consider the vacuum energy of quantum fields $\langle T_{\mu\nu}\rangle = -\langle\rho_V\rangle g_{\mu\nu} + \cdots$, as they too act like a cosmological constant. Of course, $\langle\rho_V\rangle$ is formally divergent. In practical calculations, one needs to use a regulator which preserves general covariance, such as dimensional regularization, to even arrive at $\langle T_{\mu\nu}\rangle = -\langle\rho_V\rangle g_{\mu\nu} + \cdots$. Naively speaking though, if we believe general relativity is the correct description of gravity up to an energy scale E_Λ, we would *expect* $\langle\rho_V\rangle \sim E_\Lambda^4$. Taking the cut-off E_Λ to be on the the order of Planck energy, we find the famous discrepancy of some 120 orders of magnitude between $\langle\rho_V\rangle$ and the observed vacuum energy density. This is the so-called cosmological constant "problem": how is it that the vacuum energy of quantum fields, each of which is expected to contribute $\sim E_\Lambda^4$, cancel out to such remarkable precision (e.g. 120 decimal places), to produce the observed value?

If our expectations from high energy physics are reasonable, there is no choice but to abandon the idea that vacuum energy should gravitate. This, however, requires modifying Einstein's theory of gravity, in which *all* sources of energy gravitate. In [3], Afshordi

proposed a novel approach in which quantum vacuum fluctuations (of linear order in the metric) are decoupled from gravity through the introduction of a fluid called the *Gravitational Aether* (GA). In Chapter 5, we explore in detail the predictions of this theory in the context of cosmology and solar-system tests of gravity. We show that deviations from general relativity can only be significant in situations with relativistic pressure or (potentially) relativistic vorticity. The most prominent prediction of this theory is that gravity should be 33% stronger in the cosmological radiation era than GR predictions. In other words, the effective gravitational coupling of relativistic matter, which we called G_R, is given in the radiation era by $G_R = \frac{4}{3}G_N$.

Comparing theoretical prediction of big bang nucleosynthesis with the observational light element abundances provides a constraint on G_N/G_R. We show that every light element abundance agrees with the gravitational aether theory to within 2σ. In fact, the ^{7}Li primordial abundance prefers the prediction of gravitational aether to that of general relativity, which is of course $G_R = G_N$.

We also study the evolution of cosmological perturbations in the gravitational aether framework. In doing so, we introduce the Generalized Gravitational Aether (GGA) theory, which interpolates between general relativity and gravitational aether through an extra gravitational coupling constant G_R. This allows us to constrain G_N/G_R using cosmic microwave background (CMB) measurements from WMAP, as well as other data sets. The gravitational aether theory is consistent at the $\sim 2\sigma$ level with all different combinations of data sets. In some cases, the gravitational aether prediction is even preferred to that of general relativity. Our analysis has recently been repeated using CMB data from the *Planck* satellite [109]. Based on their measurements, the gravitational aether prediction is ruled out at the 2.5–5σ level, depending on the combination of data sets used. In contrast, observational constraints at lower redshifts, in particular data of the Ly-α forest [138], prefer the aether prediction.

We then examine the implications for precision tests of gravity using the Parameterized post-Newtonian (PPN) formalism, and show that the only PPN parameter that deviates from its GR value is ζ_4, the anomalous coupling to pressure, that has only recently been measured in [109], thanks to the GGA formalism. Moreover, we argue that current tests of Earth's gravitomagnetic effect mildly prefer a co-rotation of aether with matter, although they are consistent with an irrotational aether at 2σ level.

Finally, we point out a curious fact about the behaviour of quantum corrections of matter fields in the gravitational aether theory. We show that the gravitational aether theory is not stable against quantum corrections in the matter sector, in the sense that any bare equation of state for aether gets renormalized to the value 1/3.

Another interesting implication of this theory is for the cosmic baryon fraction. As we increase the gravity due to radiation, we need to increase the dark matter density to keep the redshift of equality constant, since it is well constrained by the CMB power spectrum (see *e.g.,* [87]). This implies that the total matter density should be bigger by a factor of 4/3 (Fig. 5.2). Given that baryon density is insensitive to this change, the cosmic baryon fraction will decrease by a factor of 3/4, i.e. from 17% [87] to 13%. This leads us to the final chapter of this thesis, where we introduce a model-independent and statistically optimal method for measuring the pressure profile of galaxy clusters, which has immediate implications for the cosmic baryonic budget.

Clusters of galaxies have long been recognized as remarkable laboratories to test cosmological theories. They are the largest gravitationally bound objects in the universe, thought to have formed from the eventual collapse of initially overdense dark matter perturbations. Their abundance and large scale properties are sensitive to the expansion and initial conditions of the universe, making them excellent tools to constrain cosmological models. On smaller scales, physics of clusters is dominated by complex baryonic processes such as gas cooling, star formation, and feedback from supernovae and active galactic nuclei. In order to use clusters as standard probes of the geometry and dynamics of the universe, it is necessary to reliably model these processes and distinguish amongst different feedback mechanisms [88, 13].

The main baryonic budget of clusters is a hot plasma of ionized hydrogen and helium in the intracluster medium (ICM), making it the natural target for studying the complex astrophysical processes at play. The thermal Sunyaev-Zeldovich (tSZ) effect [151] is an important probe of the ICM: as the CMB photons inverse-Compton scatter off of the hot electrons in the ICM, their blackbody spectrum is distorted. An important feature of the tSZ effect is that it is directly proportional to the integrated pressure of free electrons along the line of sight, which makes it a powerful probe of the ICM in the outskirts, where X-ray emission is extremely faint.

Resolving the tSZ signal for individual clusters requires high resolution CMB measurements, which have become available only in the recent years. In fact, the only all-sky CMB survey with high enough resolution *and* sensitivity to detect individual SZ clusters is *Planck*. Even with *Planck*'s sensitivity, it is necessary to combine the tSZ signal from many clusters to meaningfully constrain physical quantities of interest, such as baryonic mass fraction [121]. Luckily, there is fairly strong evidence from X-ray observations and numerical simulations that clusters are self-similar to a good approximation (see e.g. [107, 16]). This fact justifies combining SZ signatures of many clusters to obtain constraints on the *mean* ICM properties.

In Chapter 6, we present a statistically-optimal and model-independent method to extract the pressure profile of hot gas in the intracluster medium (ICM). Using the thermal Sunyaev-Zeldovich effect, we constrain the mean pressure profile of the ICM by appropriately considering *all* primary cosmic microwave background (CMB) and instrumental noise correlations, while using the maximum resolution and sensitivity of all frequency channels. As a first application, we analyze CMB maps of WMAP 9-year data through a study of the Meta-Catalogue of X-ray detected Clusters of galaxies (MCXC). We constrain the universal pressure profile out to $4R_{500}$[3] with 15σ confidence, though our measurements are only significant out to R_{200}. Using a temperature profile constrained from X-ray observations, we measure the mean gas mass fraction out to R_{200}. Within statistical and systematic uncertainties, our constraints are compatible with the cosmic baryon fraction and the expected gas fraction in halos. While *Planck* multi-frequency CMB data are expected to reduce statistical uncertainties by a factor of ~ 20, we argue that systematic errors in determining mass of clusters dominate the uncertainty in gas mass fraction measurements at the level of ~ 20 percent.

1.2 Published Material

The research presented in this thesis has either been published or submitted for publication. All these papers have been a collaborative effort. Chapter 2 is based on [134], which has been submitted for publication to Classical and Quantum Gravity. Chapter 3 is based on [20]. Chapter 4 is based on the two papers [5, 17]. Finally, Chapters 5 and 6 are based on [19] and [18], respectively.

[3] The quantity R_Δ is defined as the radius up to which the matter density is Δ times the critical mass-density of the universe.

Chapter 2

Causal Sets and their Continuum Approximation

2.1 Introduction

What is the fundamental structure of spacetime? Our best answer currently comes from the general theory of relativity: spacetime is a smooth Lorentzian manifold whose dynamics is governed by the Einstein field equations. Although general relativity (GR) has been remarkably successful, it is well known that it cannot be the whole story. Because of the presence of spacetime singularities, for instance, GR cannot describe the physics at the big bang or the centre of black holes. Moreover, general relativity is a classical theory and it should ultimately be replaced by some appropriate "quantum version". It is widely believed that such a theory of "quantum gravity" would resolve spacetime singularities and tell us more about the big bang and black holes. Over the years, many approaches to quantum gravity have been proposed and pursued. This chapter focuses on one such proposal: causal set theory.

One way to hope for a deeper understanding of spacetime is to question the initial assumptions of the current theory, especially those which are suspected to lead to trouble down the road. A core assumption of most known physical theories, including general relativity, is that spacetime is continuous. In this picture, spacetime events can be *arbitrarily* close to one another. This is certainly a mathematically convenient assumption, but should it be re-examined on physical grounds? Interestingly enough, this question seems to have had troubled Einstein himself [144]:[1]

[1] This is a translation of (part of) a letter which Einstein had written to Walter Dällenbach in Nov.

> But you have correctly grasped the drawback that the continuum brings. If the molecular view of matter is the correct (appropriate) one, i.e., if a part of the universe is to be represented by a finite number of moving points, then the continuum of the present theory contains too great a manifold of possibilities. I also believe that this too great is responsible for the fact that our present means of description miscarry with the quantum theory. The problem seems to me how one can formulate statements about a discontinuum without calling upon a continuum (space-time) as an aid; the latter should be banned from the theory as a supplementary construction not justified by the essence of the problem, which corresponds to nothing real. But we still lack the mathematical structure unfortunately. How much have I already plagued myself in this way!

There are compelling reasons to question the continuum description of spacetime. The problematic infinities of general relativity (i.e. singularities) and quantum field theory (QFT) are caused by the lack of a minimum length cut-off in the local degrees of freedom. Although the renormalisation procedure ameliorates these problems in QFT, they return in naive attempts to quantise gravity. Moreover, given what we know about the uncertainty principle of quantum mechanics, resolving arbitrarily close spacetime events is simply impossible: measuring very small distances (e.g. order of the Planck length) requires inputting such an enormous amount of energy that a black hole would eventually form and swallow the entire laboratory whole. Finally, to mention a more practical problem, there are technical difficulties with defining the path integral in the continuum, since it is necessary to integrate over a continuous history space. If spacetime is discrete, the path integral becomes a sum, thus eliminating the problem.

If we are to abandon the continuum, however, how should we go about discretizing spacetime? Perhaps, a guiding principle could be looking for features in the usual description that do not depend too heavily on the assumption of continuity. One feature that is undoubtedly foundational to the theory of relativity is causality (i.e. the knowledge of which events could influence or be influenced by other events). As it turns out, the causal structure of spacetime is incredibly rich. It has been proven that a Lorentzian geometry can be recovered from its causal order almost completely [73, 94]. To be more precise, the topology, differential structure, and the metric of a Lorentizian geometry can be recovered from its causal order up to a local conformal factor. (We will explain how this can be done in the next section.) Although these results are grounded in the context of Lorentzian geometry, where continuity is a key assumption, nothing keeps us from retaining causal order information but giving up continuity. In fact, by discretizing spacetime, we *gain* the

1916. See references $16-17$ of [144] for details regarding translation and precise referencing.

missing information about absolute scale or volume[2] in the continuum through counting.[3] Riemann himself had appreciated this point [144]:[4]

> The question of the validity of the presuppositions of geometry in the infinitely small hangs together with the question of the inner ground of the metric relationships of space. In connection with the latter question, which probably can still be reckoned to be part of the science of space, the above remark applies, that for a discrete manifold, the principle of its metric relationships is already contained in the concept of the manifold itself, whereas for a continuous manifold, it must come from somewhere else. Therefore, either the reality which underlies physical space must form a discrete manifold or else the basis of its metric relationships must be sought for outside it, in binding forces that act on it;

and also

> Definite portions of a manifold, distinguished by a criterion or a boundary, are called quanta. Their quantitative comparison happens for discrete magnitudes through counting, for continuous ones through measurement.

It would then seem that a combination of causal order and discreteness should be enough to recover all information about the topology, differential structure, and metric of the continuum spacetime. Causal set theory is in essence nothing but a realization of this observation:

$$Order + Number = Geometry. \tag{2.1}$$

It is also interesting to notice that causal order singles out the Lorentzian signature of spacetime, $(-,+,+,+)$ in 4 dimensions, as the only one consistent with the notion of past and future. A note of historical significance is that the idea described so far (and summarized in the slogan 1) was proposed independently several times by J. Myrheim [106], G. tHooft [152] and finally L. Bombelli, J.H. Lee, D. Meyer and R. Sorkin [34].

Before defining what a causal set is, let us make the arguments presented so far more concrete.

[2] The freedom in local conformal transformations of the metric can be translated directly into not having fixed a volume element $\sqrt{-g}d^4x$.

[3] We will formalize this notion in Section 2.4.

[4] These passages have been translated into English in [144]. See reference 15 therein.

2.2 Towards Causal Sets

In the previous section, we claimed that a Lorentzian geometry (i.e. its topology, differential structure, and metric) can be recovered from its causal order up to local conformal transformations. The following two theorems give precise meaning to this claim:[5]

Malament's Theorem [94]. *If (M, g) and (M', g') are past and future distinguishing spacetimes and if $f : M \to M'$ is a causal isomorphism (i.e., a bijection where both f and f^{-1} preserve the causal order $\preceq$), then f must be a homeomorphism.*

Hawking's Theorem [73]. *Suppose (M, g) and (M', g') are spacetimes and $f : M \to M'$ is a homeomorphism where both f and f^{-1} preserve future directed continuous null geodesics. Then f is a smooth conformal isometry.*

Malament's theorem shows that the topology of a spacetime which is both past and future distinguishing is completely determined by its causal order. Before this result was proven, it was known that the Alexandroff topology (one whose base topology on M is all sets of form $I^+(p) \cap I^-(q)$) is equal to the manifold topology for all strongly causal spacetimes. Malament's theorem shows that a weaker condition of past and future distinguishing spacetimes suffices. Moreover, he demonstrates in [94] that this assertion no longer applies if this condition is relaxed any further. Hawking's theorem shows that differential structure (i.e. that f is smooth) and 9/10 components of the metric (i.e. that f is a conformal isometry) can be recovered from causal order information only. Let us give an example of how a geometrical object can be reconstructed from causal relations. Consider, for instance, reconstructing light rays in Minkowski space. Define the causal interval $I(p, q)$ between any two points p and q by $I(p, q) = J^+(p) \cap J^-(q) = \{x \in M | p \preceq x \preceq q\}$. Also, a subset $S \subset M$ is called a *chain* if for all $p, q \in S$ either $p \preceq q$ or $q \preceq p$. We can now define a light ray l to be a maximal[6] chain such that for all $p, q \in l$, $I(p, q)$ is also a chain. Similarly, it is possible to define all other geometrical objects such as spacelike lines, planes, vectors, etc [144].

Let us now turn to the other side of the story: discreteness. What should we take as the discreteness scale? On dimensional grounds, we would expect that effects of discreteness become important at about the Planck length $l_p = \sqrt{G\hbar/c^3}$. There is, however, a more satisfying argument that brings us to the same conclusion: the black hole entropy. We

[5] See Section A.1 of Appendix A for definitions relevant to these theorems. Here a spacetime (M, g) is assumed to be a connected, four-dimensional smooth manifold M without boundary, together with a smooth Lorentzian metric g.

[6] A maximal chain is one that can not be enlarged while remaining a chain.

know from Hawking's calculation of black hole radiance that the entropy associated with a black hole is $S_{BH} = k_B A/(4l_p^2)$, where k_B is the Boltzmann constant and A is the area of the event horizon. Suppose we discretize the horizon by little plackets of area l_{min}^2 , so that the total number of plackets is $N = A/l_{min}^2$. Now imagine that every placket could have a value of either 1 or 0. The total number of microstates is then $\Omega = 2^N$, and the associated entropy can be calculated from Boltzmann's formula: $S = k_B \ln \Omega = k_B \ln 2^{\frac{A}{l_{min}^2}}$. Equating this result with S_{BH}, we find $l_{min} = 2\sqrt{\ln 2} l_p$. It is important to notice that in the limit where $\hbar \to 0$, the discreteness scale also vanishes. This means that the classical limit is necessarily a continuum limit, suggesting that *spacetime discreteness is inherently quantum.*

We are now in a position to define precisely what a causal set is.

2.3 Causal Set Kinematics

A causal set (causet), as proposed in [34], is a partially ordered and locally finite set. Concretely, a causal set $(C, \prec)$ is a set C endowed with a binary relation $\prec$ such 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: $\text{card}\{y \in C | x \prec y \prec z\} < \infty$,

where 'card' stands for cardinality. The first two axioms encode causal order information. Notice that causal cycles are not permitted because $x_0 \prec x_1 \prec x_2 \prec \cdots \prec x_0$ implies $x_0 \prec x_0$ by transitivity, which contradicts irreflexivity.[7] Local finiteness formalizes the notion of discreteness, so that the causal interval between any two points contains a countably finite number of events.

Causets have many representations, the most intuitive one (perhaps) as a graph or family tree. If $x \prec y$, x is said to be an ancestor of y, and y a descendent of x. Figure. 2.1 shows an example of a causet with 5 elements represented by a so-called Hasse diagram. In this representation, causet elements are represented by vertices and causal relations that do

[7] Sometimes the irreflexivity axiom is replaced by (i) $x \prec x$ (reflexivity) and (ii) $x \prec y \ \& \ y \prec x \Rightarrow x = y$ (acyclicity).

not follow from transitivity are represented by lines. A causal set can also be represented by a matrix M, whose rows and columns are labeled by the elements and $M_{ij} = 1$ if $i \prec j$ and 0 otherwise. This representation is useful for computer codes.

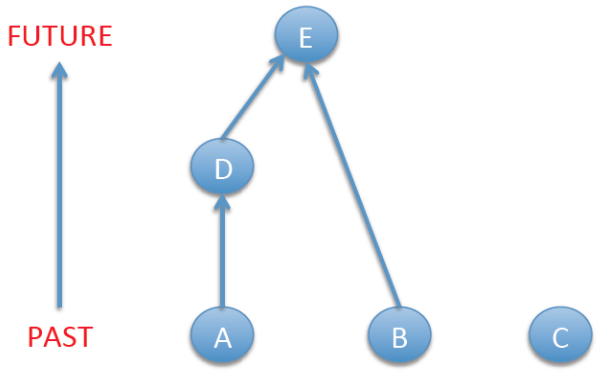


Figure 2.1: Hasse diagram of a causet with 5 elements. Only relations not implied by transitivity are drawn. For instance, since $x \prec y$ and $y \prec z$, no line is drawn between x and z.

2.4 Approximating the Continuum

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 [34, 142, 144, 55, 74]. 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 $(C, \prec)$. 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 $f(x) \in J^-(y)$ (see Section A.1 of Appendix A for the definition of $J^-(y)$). 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 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:

$$N_S \approx \rho V_S = \rho \int_S \sqrt{-g(x)} d^D x, \tag{2.2}$$

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 $M \to \infty$:

$$\langle N \rangle = \rho V, \qquad \frac{\delta N}{\langle N \rangle} = \frac{\sqrt{\langle (N - \langle N \rangle)^2 \rangle}}{\langle N \rangle} \ll 1. \tag{2.3}$$

Having the N–V formulation at hand,[8] 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 $C(M)$

[8] 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 (2.3) ensures that *almost all* regions have volumes representative of the number of embedded points in them.

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. The probability of selecting n points from a region with volume V is[9]

$$P(n) = \frac{(\rho V)^n e^{-\rho V}}{n!}. \tag{2.4}$$

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

$$\langle N \rangle_{Pois} = \rho V, \qquad \frac{\delta N_{Pois}}{\langle N \rangle_{Pois}} = \frac{1}{\sqrt{\rho V}}. \tag{2.5}$$

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), (2.3) 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 [74]. 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 not even its realizations select a preferred frame in Minkowski space [33]. 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?

The remainder of this chapter contains two results which (hopefully) shed some light on certain aspects of this question. The first result is that the number–volume correspondence,

[9] 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 (2.4) in the limit $dV \to 0$.

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 better number–volume correspondence than Poisson sprinkling for large volumes.[10] By way of providing an example, however, we will argue 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, even for spacetime regions with macroscopically large volumes.

2.5 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, $\nexists\ \xi$ such that $\forall\ S$:*

$$\langle (N_S - V_S)^2 \rangle \leq \alpha V_S \quad \textit{where} \qquad 0 \leq \alpha < 1. \tag{2.6}$$

Proof. It is shown in Section A.2 of Appendix A 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

$$\langle (N_S - V_S)^2 \rangle \geq (V_S - n_*)(n_* + 1 - V_S), \tag{2.7}$$

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 to the variance $(V_S - n_*)(n_* + 1 - V_S)$. The formal proof of this result is given in Section A.2 of Appendix A.

[10] The existence of Lorentzian lattices in $1 + 1$ dimensional Minkowski space, as well as the possibility that they might realize the number–volume correspondence better than Poisson sprinkling, was pointed out by Aron Wall to Rafael Sorkin, who then mentioned it to us.

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 (2.7) that

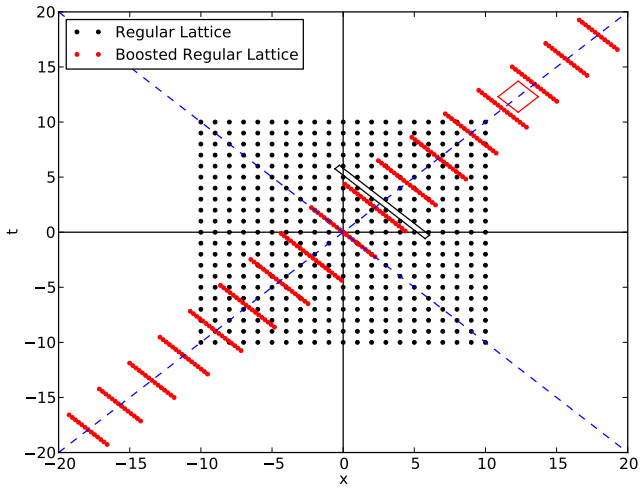
$$(V_S - n_*)(n_* + 1 - V_S) \leq \alpha V_S \qquad \forall \qquad S. \tag{2.8}$$

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

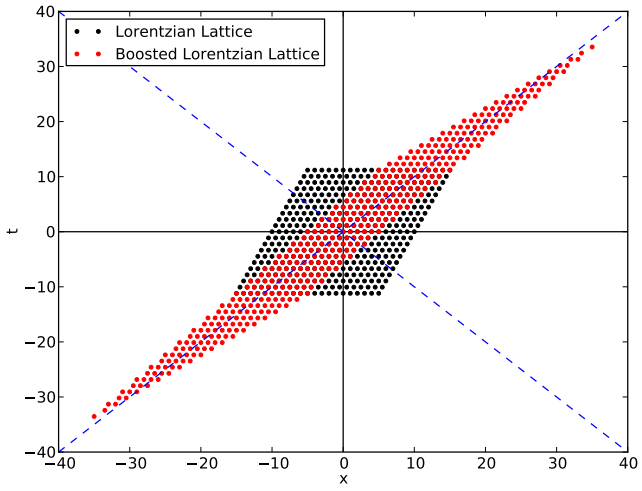
The proof of this theorem rests heavily on regions with Planckian volumes. For instance, had we required condition (2.6) for regions with $V_S > 1$, the proof would not have gone through. As we mentioned previously though, the causal set–continuum correspondence is only physically meaningful on scales 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.

2.6 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 would suggest 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 2.2a 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 2.2b: it goes to itself under the action of a discrete set of boosts. We have classified all 2D Lorentzian lattices in Section A.3 of Appendix A. In the case of the integer lattice shown in Figure 2.2a, the more it is boosted, the more irregular it becomes. A Lorentzian lattice, however, 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.



(a)



(b)

Figure 2.2: (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.

We have investigated the N–V correspondence for various Lorentzian lattices using simulations. Figure 2.3 shows the result of one such analysis on the lattice shown in Figure 2.2b. 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.[11] 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 2.3, the Lorentzian lattice shown in Figure 2.2b realizes the number–volume correspondence with much less noise than Poisson sprinkling for macroscopic volumes. In fact, Figure 2.3b 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.

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 [135].[12] 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 2.4). What makes $1+1$ dimensional Minkowski space special is that boosts can only be performed along one direction. Therefore, a Lorentzian lattice does not "change" too drastically under the action of an arbitrary boost. This feature does not seem to persist in higher dimensions, which leads us to conclude that Lorentzian lattices in higher dimensions are not likely to realize the number–volume correspondence better than Poisson sprinkling.

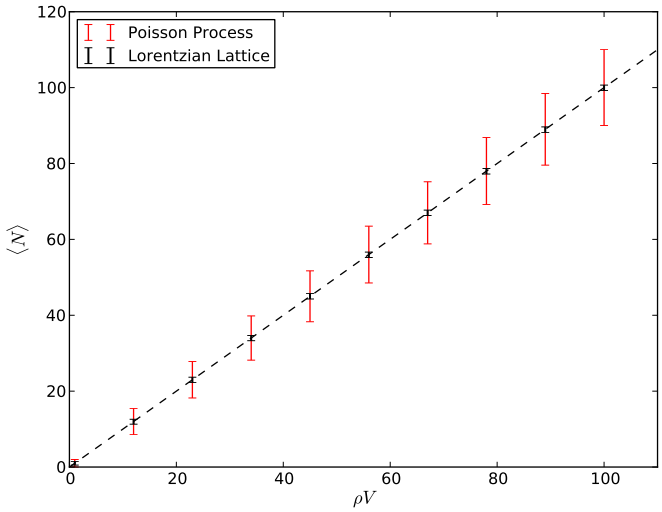
2.7 A Conjecture

Based on the results of the previous Sections, we conjecture the following:

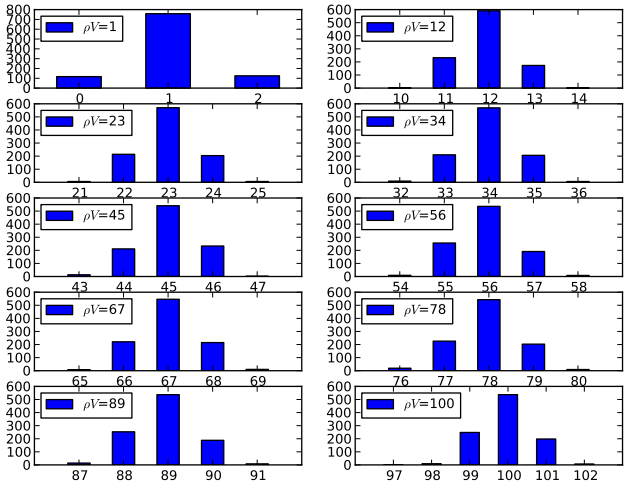
Conjecture 1. *Let ξ be a point process whose realizations are points of a $3+1$-dimensional smooth Lorentzian manifold (M, g). Let N_S be the random variable which counts the number of points in a causal interval $S \subset M$: it takes on a value $n \in \{0, 1, 2, \dots\}$ with probability*

[11] We made sure to include "stretched out" causal diamonds, such as the black diamond shown in Figure 2.2a, as they are responsible for the poor realization of the number–volume correspondence in the integer lattice.

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



(a)



(b)

Figure 2.3: The number–volume correspondence for the Lorentzian lattice shown in Figure 2.2b. (a) The mean and standard deviation of the number of points. (b) The histogram of the number of points for different volumes.

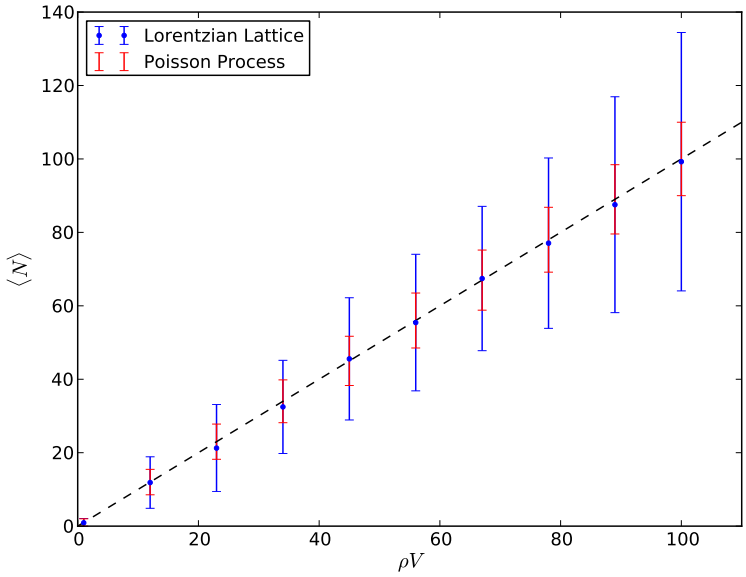


Figure 2.4: The number–volume correspondence for the $2+1$ dimensional integer lattice. For a given volume V, 200 different causal diamonds with the same volume V but randomly varying shapes are considered. The mean and standard deviation of the number of points (blue) is compared with that of the Poisson process (red).

$P_S(n)$. 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, $\nexists$ ξ and $V_* > 0$ such that for all causal intervals S with volume $V_S > V_*$:

$$\langle (N_S - V_S)^2 \rangle \leq \alpha V_S \quad \textit{where} \qquad 0 \leq \alpha < 1. \tag{2.9}$$

2.8 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, where $\rho \sim l_p^{-4}$ sets the fundamental discreteness scale. 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.

Chapter 3

Causal Set d'Alembertians and their Phenomenology

3.1 Introduction

Causal set theory postulates that the fundamental structure of spacetime is that of a locally finite partially ordered set (see e.g. Chapter 2). Thus a causal set (causet) is in a certain sense both Lorentzian and discrete. 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 from Poisson sprinklings of a Lorentzian manifold M.[1] 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 inevitably 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

[1] See Section 2.4 of Chapter 2 for the definition of Poisson sprinkling.

with what one does when, for example, one discretizes the Laplacian operator in a Riemannian spacetime. Nevertheless, a non-local operator was suggested in [145] 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 [145] was generalized to $D = 4$ dimensions in [24] and recently to arbitrary D in [56, 67].

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. over all Poisson sprinklings of $\mathbb{M}^D$) 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.[2] 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 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 [145] and 4D [24] 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

[2] 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 [145] would be to choose ρ small enough that the necessary averaging will occur within each individual causet.

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 Chapter 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 [145] and four [24]. 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 Chapter, we will only consider causal sets which are obtained by Poisson sprinklings of Minkowski space at density ρ.

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 [145]:

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

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

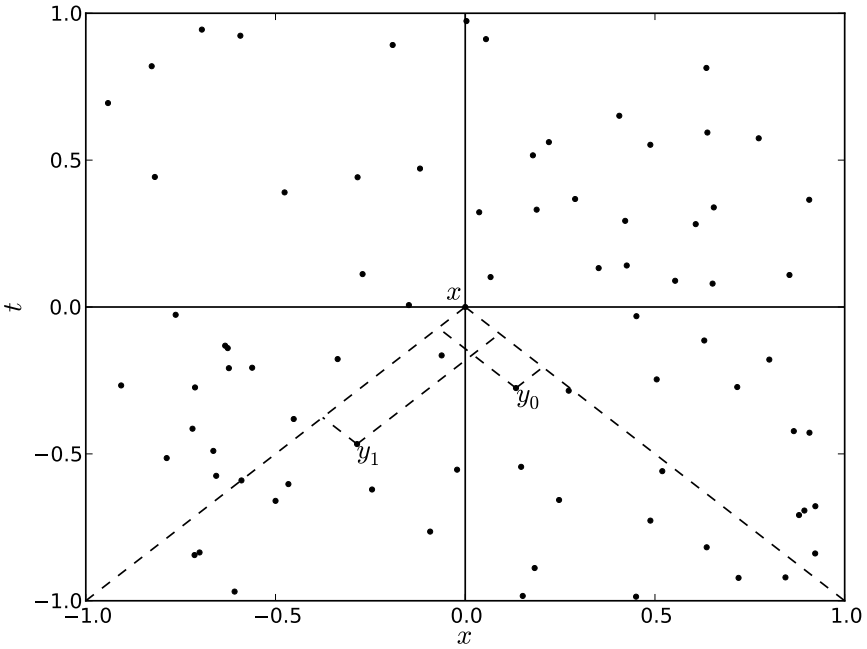


Figure 3.1: A Poisson sprinkling of $1+1$ Minkowski space at density $\rho = 80$. Here y_0 is a 0th neighbour of x because there are no elements which are both to the future of y_0 and the past of x. Similarly, y_1 is a first neighbour of x. The contributions of the points y_0 and y_1 to $\rho^{-1}(B_\rho^{(2)}\Phi)(x)$ 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 $(\Box_\rho^{(2)}\Phi)(x)$.

sprinklings at density ρ:

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

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,[3] any plane wave $e^{ip\cdot x}$ is an eigenfunction of $\Box^{(2)}_\rho$ (provided that the integrals in (3.3) converge, so that the left hand side is well defined):

$$\Box^{(2)}_\rho e^{ip\cdot x} = g^{(2)}_\rho(p) e^{ip\cdot x}, \tag{3.4}$$

where $p\cdot x \equiv \eta_{\mu\nu}p^\mu x^\nu$ and $\eta_{\mu\nu} = \mathrm{diag}(-1,1)$. Interestingly enough, $g^{(2)}_\rho(p)$ in this case can be expressed in closed form:[4]

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

where $\mathrm{E}_2(z)$ is a generalized exponential integral function (see e.g. 8.19 of [154]) and

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

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 [154]):

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

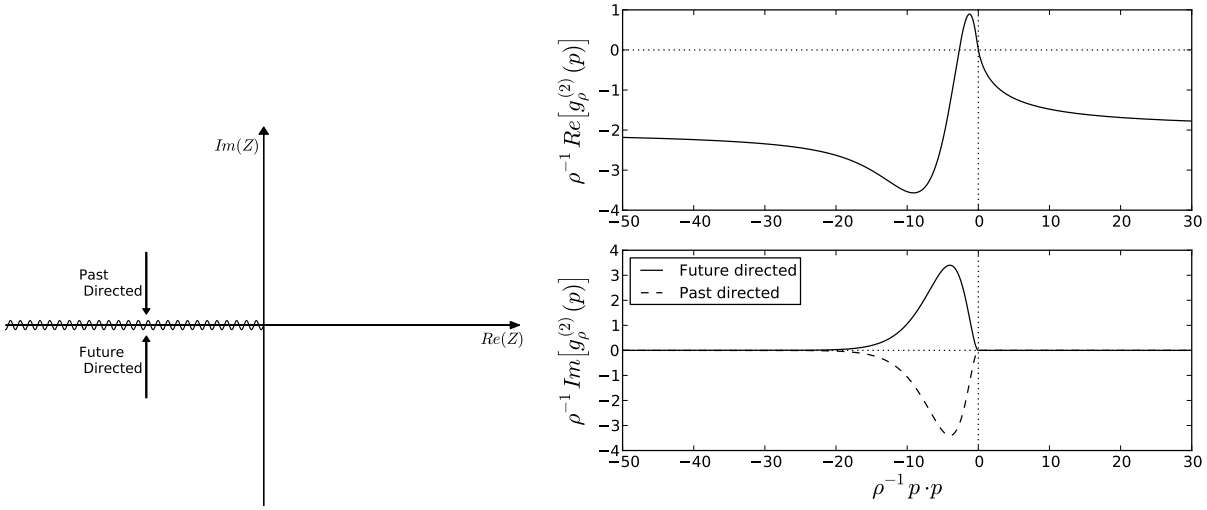
$$\rho^{-1} g^{(2)}_\rho(p) \xrightarrow{Z\to \infty} -2 + \frac{8}{Z} + \cdots . \tag{3.8}$$

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

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

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



(a) (b)

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.

$Z = \rho^{-1}p\cdot p = 0$. To demonstrate this, we note that $g_\rho^{(2)}(p)$ has the following representation (see e.g. 8.19.1 and 8.6.4 of [154]):

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

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 $\Box_\rho^{(2)}\Phi = 0$ 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 $\Box_\rho^{(2)}\Phi = 0$ 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 $\Box_\rho^{(2)}$, 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 $\Box_\rho^{(2)}\Phi(x) = 0$ 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 $p_I^0 > 0$. 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 and $p_R \cdot p_I = 0$. 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 $\Box_\rho^{(2)}\Phi$ 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 [24] :

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

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

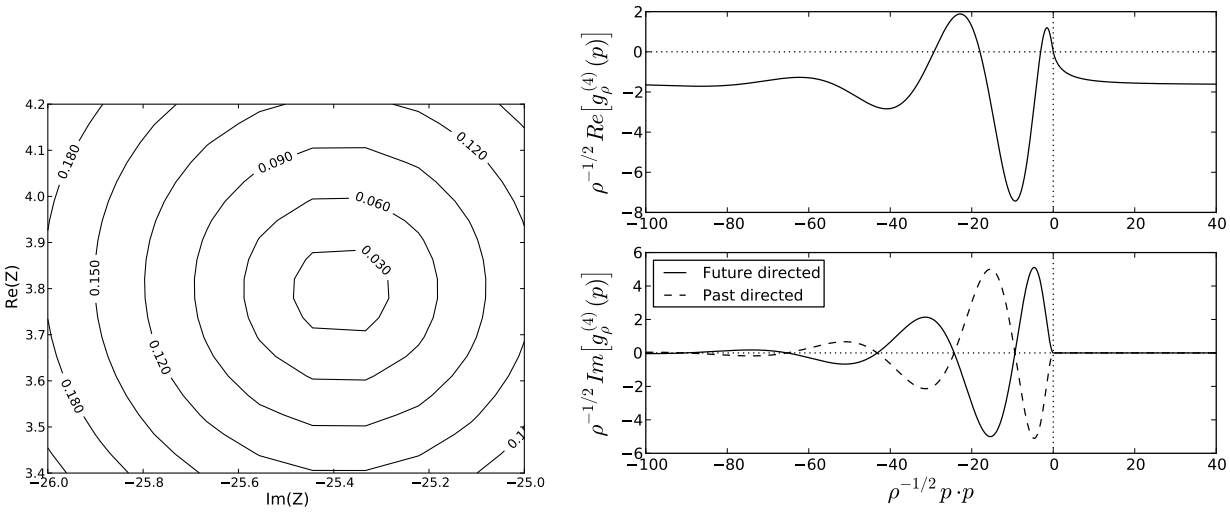
$$Z \equiv \rho^{-1/2} p\cdot p, \qquad C_4 = \frac{\pi}{24}. \tag{3.15}$$

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

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

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



(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.

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.

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:

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

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 [145], 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 Section B.4 of Appendix B; 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

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

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

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

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 [52]. 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 $\mathrm{ph}(z) = 0$ [154].

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 Section B.1 of Appendix B, 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 Section B.1 of Appendix B for the details.

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

need to be satisfied are

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

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

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

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 Section B.2 of Appendix B, 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 $G_R(x,y) = 0 \;\forall\; x \not\succeq y$.

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 $g_\rho^{(D)}(p)$:

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

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 $\int d^D p\; p^{-nD}, n = 1, 2, \cdots$, 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

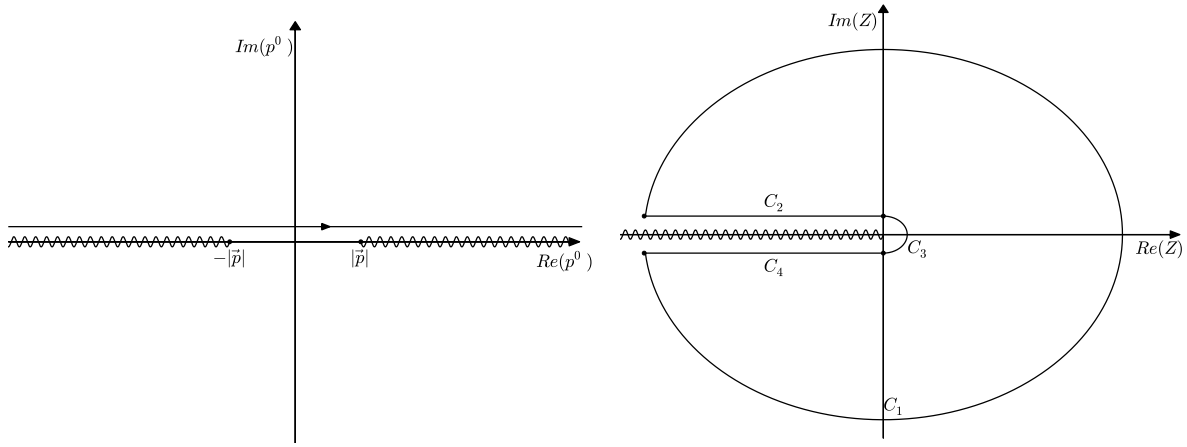
$$\rho^{2/D} G(x-y) \equiv \rho^{2/D} G_R(x-y) - \frac{1}{a}\delta^{(D)}(x-y). \tag{3.38}$$

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

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

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

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



(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.

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

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

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

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

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

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 [52] to compute $G(x-y)$:

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

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 [154]), (3.45) can be simplified into the following form for even D:

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

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 [154]) and the fact that $\mathrm{Im}\left[\tilde{g}(-\xi^2+i\epsilon)\right]$ is exponentially damped for large ξ (see Section B.2.1 of Appendix B), 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.[5] Strictly speaking, the discussion above only analyzes the behavior of $G(x-y)$ as τ_{xy} *approaches* 0, and consequently

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

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.

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 Chapter. 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 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 $p_R =< \pi^0_R, \vec{\pi}_R >$ 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

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

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 $\Box^{(D)}_\rho$ 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 $x < 0$:

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

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 [145]. 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.[6] 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 [24] 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 [67].) Would this feature persist for all of the GCB operators? This feature has also been used to define an action-functional for causal sets [24]. A final question then is whether the instability found above has any consequences for this causal set action?

[6] 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

A Distinguished Vacuum State for a Quantum Field in a Curved Spacetime

4.1 Introduction

The framework known as "quantum field theory in curved spacetime" concerns the interaction of quantum fields with gravity, but only in an asymmetrical sense. Non-gravitational, "matter" fields are treated in accord with quantum principles while their gravitational "back reaction" is either ignored entirely or described by a semiclassical form of the Einstein equations. Although not a fundamental theory of nature, this framework has provided us with profound insights into an eventual theory of quantum gravity. Important examples include Hawking radiation by black holes [72], the Unruh effect [31, 155], and the generation of Gaussian-distributed random perturbations in the theory of cosmic inflation [105]. In all these examples a choice of vacuum — or at least a reasonable reference state of the field — is of crucial importance. It therefore seems unsatisfactory that as it stands, quantum field theory lacks a general notion of "vacuum" which extends very far beyond flat spacetime.

Formulations of quantum field theory in Minkowski spacetime do provide a distinguished vacuum, but it rests heavily on a particle interpretation of the field that is closely tied to the properties of the Fourier transform and the availability of plane waves. More abstract treatments tend to trace the uniqueness of the vacuum to Poincaré-invariance, but that is tied even more closely to flat space. It is thus unclear how one might extend the

notion of vacuum beyond the case of spacetimes with a high degree of symmetry. Moreover, even a large symmetry-group does not always yield a unique vacuum without further input. In de Sitter space, for example, there exist one complex-parameter family of vacua, the so-called "α-vacua" [12], all of which are invariant under the full de Sitter group.To single one value of α out from the rest, one needs to impose the further condition that the two-point function take the so-called Hadamard form.

One might even question whether a quantum field theory is well-defined at all before a vacuum is specified. What is probably the best studied mathematical framework for quantum field theory in flat space, that of the Wightman axioms, incorporates assertions about the vacuum among its basic assumptions, and it relies on them in proving such central results as the CPT and spin-statistics theorems. It is therefore noteworthy that the so-called algebraic approach to quantum field theory has been able to proceed a great distance without relying on a notion of vacuum, or indeed any unique representation of the quantum fields at all. In place of a Poincaré-invariant vacuum, it has been proposed to rely on a distinguished *class* of states, the so-called Hadamard states (which are well-suited to renormalization of the stress-tensor by "point-splitting"), supplemented by an assumption about a short-distance asymptotic expansion for products of quantum fields, namely the operator product expansion or "OPE" (see for example [158], [159] and references therein). If such a "purely algebraic" approach were to establish itself more generally, it might diminish the interest in distinguished "vacua" for curved spacetimes. Conversely, if a reasonable definition of a preferred vacuum state could be obtained, it might remove some of the motivation for a purely algebraic formulation of quantum field theory.[1]

Let us remark also that histories-based formulations of quantum mechanics tend to fuse the concept of state with that of equation of motion. This shows up clearly in formulations that start from the "quantum measure" [57, 143, 146] or "decoherence functional" [71], neither of which can be defined without furnishing a suitable set of "initial conditions". In this sense, one has no dynamical law at all before a distinguished "initial state" is specified.

At a less formal level, the ability to think in terms of particles offers an obvious benefit to one's intuition. And, especially in relation to cosmology, great interest attaches to the question whether certain sorts of states can be regarded as "natural" to certain regions of spacetime, a question we return to briefly in Section 4.6.2. These, then, are two more reasons why the availability of a distinguished vacuum could be welcome, whether or not it is logically necessary to quantum field theory as such.

[1] We suspect that lasting enlightenment about the "best" formulation of quantum field theory will only arrive together with a solution of the problem of quantum gravity, by means of a greater theory within which that of quantum field theory in curved spacetime will have to be subsumed.

Moreover, what is logically necessary can change drastically if one passes from the spacetime continuum to some more fundamental structure, especially if that structure is discrete. As we will review later, the entire quantization process — as usually conceived — boils down to selecting an appropriate subspace of the solution space of the Klein-Gordon equation. But that way of organizing the problem seems to break down in the case of a causal set. There, the notion of "approximate solution" seems to be the best that is available, and one therefore requires a different starting point.

In [84], such a starting point was found in (the discrete analog of) the retarded Green's function. On that basis a complete counterpart of the quantum field theory of a free scalar field was built up, and a unique "vacuum state" was derived. Herein, we generalize that derivation to quantum fields on continuum spacetimes, showing thereby that there is a sensible way to uniquely define a vacuum state for a scalar field in any globally hyperbolic spacetime or region of spacetime. More precisely, we consider the case of a *free scalar field* in a *globally hyperbolic* spacetime or region of spacetime, and in that context we put forward a definition of *distinguished vacuum state* that applies to all compact regions and to a large class of noncompact regions. We will refer to this state as the *Sorkin-Johnston (SJ) vacuum*.[2]

It is thus possible to carry the concept of vacuum far beyond the confines of Minkowski space by means of definitions we expose in detail below. Although, for all of the reasons indicated above, this possibility is of interest in itself, one naturally wants to know to what extent, and in what sense, our proposal is "the right one"? To that question, only a sufficient number of particular instances of our vacuum would seem to be germane. The examples of Minkowski spacetime and of globally static spacetimes furnish important evidence, but they contain little that is new physically. To judge the ultimate fruitfulness of our prescription, one should, for example, test it against the behavior of the "matter fields" that one actually encounters in the early universe. In Sections 4.5 and 4.6 we make a start on this kind of test, beginning with de Sitter space, as well as the case of a spatially homegeneous and isotropic cosmology.

The rest of this Chapter is organized as follows. In Section 4.2, we review the quantization of a free real scalar field in a globally hyperbolic spacetime. The SJ prescription is then introduced in Section 4.3. We prove in Section 4.4 that the SJ vacuum agrees with the ground state of the Hamiltonian in the case of static spacetimes. In Section 4.5, we determine the SJ vacuum for general $d+1$ dimensional de Sitter space, in both the

[2] In [84], Steven Johnston defined a free scalar quantum field theory on a causal set, using only the discrete analogue of the retarded Green's function. Rafael Sorkin then noticed that Johnston's derivation can be generalized to quantum fields on continuum curved spacetimes.

Poincaré and global patches. Section 4.6 applies the SJ prescription to a massive and massless field in a spatially homegeneous and isotropic universe filled with radiation only. In Section 4.7, we illustrate the inherent non-locality of the SJ prescription with the example of a spacetime which sandwiches a region with curvature in-between flat initial and final regions. Finally, in Section 4.8, we compute the SJ vacuum on a causal set corresponding to a causal diamond in $1+1$ dimensional de Sitter space.

4.2 Quantum Field Theory in Curved Spacetime

In this Section, we briefly review the quantization, along traditional lines, of a free real scalar field $\phi(x)$ in a $D = d+1$ dimensional globally hyperbolic spacetime $(M, g_{\mu\nu})$.[3] Such a spacetime admits a foliation by Cauchy surfaces Σ_t labelled by a time coordinate t. The classical equation of motion of the field is given by the Klein-Gordon (KG) equation

$$(\Box - m^2)\phi(x) = 0, \tag{4.1}$$

where $\Box\phi = \frac{1}{\sqrt{-g}}\partial_\mu\left(\sqrt{-g}g^{\mu\nu}\partial_\nu\phi\right)$ is the d'Alembertian operator and g is the determinant of the metric. The retarded and advanced Green's functions $G_{R,A}(x,y)$ associated with (4.1) are solutions to

$$(\Box - m^2)G_{R,A}(x,y) = \frac{1}{\sqrt{-g}}\delta^{(D)}(x-y), \tag{4.2}$$

where by definition $G_R(x,y) = 0$ unless $x \succ y$ (meaning that x is inside or on the future light cone of y) and $G_A(x,y) = 0$ unless $y \succ x$. *These solutions are unique when* $(M, g_{\mu\nu})$ *is globally hyperbolic* [158]. Let us also define the Klein-Gordon "inner product" $(\cdot\,,\cdot)_{KG}$ on pairs of complex solutions to (4.1):

$$(f,g)_{KG} := i\int_{\Sigma_t}\left(\overline{f}n^\mu\nabla_\mu g - g n^\mu\nabla_\mu\overline{f}\right)d\Sigma_t, \tag{4.3}$$

where bar denotes complex conjugation, Σ_t is an arbitrary Cauchy surface in M, n^μ is the future-directed unit normal to Σ_t, and $d\Sigma_t$ is the induced volume element on Σ_t. This is a well-defined "inner product" because it is independent of t for solutions of the Klein-Gordon equation. (See Section C.1 of Appendix C for a proof.) The reason we have surrounded "inner product" in quotation marks is that $(\cdot\,,\cdot)_{KG}$ is not positive-definite. It may checked, for instance, that $(f,f)_{KG} = 1$ implies $(\overline{f},\overline{f})_{KG} = -1$. Nevertheless, we will continue to refer to $(\cdot\,,\cdot)_{KG}$ as an inner product.

[3] We use signature $(-+++\cdots)$ and set $\hbar = c = 1$.

To quantize the theory, we promote $\phi(x)$ to an operator (more appropriately an operator-valued distrbution) $\widehat{\phi}(x)$ which satisfies the KG equation

$$(\Box - m^2)\widehat{\phi}(x) = 0, \tag{4.4}$$

and the commutation relations

$$[\widehat{\phi}(x), \widehat{\phi}(y)] = i\Delta(x, y), \tag{4.5}$$

where $\Delta(x, y)$ is the Pauli-Jordan function, defined as the difference between the retarded and advanced Green's functions:

$$\Delta(x, y) := G_R(x, y) - G_A(x, y). \tag{4.6}$$

This is the so-called Peierls form of the commutation relations, which is entirely equivalent to (but more explicitly covariant than) the more commonly seen equal-time commutation relations. (See Section C.2 of Appendix C for a proof.) A simple but important observation is that $\Delta(x, y)$ satisfies the KG equation:

$$\begin{aligned}(\Box_x - m^2)\Delta(x, y) &= (\Box_x - m^2)G_R(x, y) - (\Box_x - m^2)G_A(x, y) \\ &= \frac{1}{\sqrt{-g}}\delta^{(D)}(x - y) - \frac{1}{\sqrt{-g}}\delta^{(D)}(x - y) \\ &= 0.\end{aligned} \tag{4.7}$$

It is because of this fact that the commutation relations (4.5) are consistent with the equation of motion (4.4).[4]

To construct a Hilbert space $\mathcal{H}$ and a representation of the field operators $\widehat{\phi}(x)$, consider a set $\{u_{\mathbf{k}}(x)\}$ of complex solutions of the KG equation:

$$(\Box - m^2)u_{\mathbf{k}}(x) = 0, \qquad \forall \qquad \mathbf{k}. \tag{4.8}$$

We assume that $\{u_{\mathbf{k}}(x)\}$, together with their complex conjugates $\{\overline{u}_{\mathbf{k}}(x)\}$, span the space of all complex solutions of the KG equation. We require these so-called modefunctions to be orthornormal with respect to the KG inner product in the following sense

$$(u_{\mathbf{k}}, u_{\mathbf{q}})_{KG} = -(\overline{u}_{\mathbf{k}}, \overline{u}_{\mathbf{q}})_{KG} = \delta_{\mathbf{kq}}, \qquad (u_{\mathbf{k}}, \overline{u}_{\mathbf{q}})_{KG} = 0. \tag{4.9}$$

Throughout the text, we will refer to any set $\{u_{\mathbf{k}}(x)\}$ of complex functions which satisfy (4.8) and (4.9) as an *orthornormal basis of the KG solution space*. The vector space

[4] $0 = [(\Box_x - m^2)\widehat{\phi}(x), \widehat{\phi}(y)] = (\Box_x - m^2)[\widehat{\phi}(x), \widehat{\phi}(y)] = i(\Box_x - m^2)\Delta(x, y) = 0.$

spanned by $\{u_{\mathbf{k}}(x)\}$ and equipped with the KG inner product forms a (genuine) Hilbert space $\mathbb{H}$. The Hilbert space $\mathcal{H}$ of physical states is then taken to be the symmetric Fock space associated with $\mathbb{H}$, with the operators $\widehat{\phi}(x)$ defined as

$$\widehat{\phi}(x) = \sum_{\mathbf{k}} u_{\mathbf{k}}(x)\widehat{a}_{\mathbf{k}} + \overline{u}_{\mathbf{k}}(x)\widehat{a}^{\dagger}_{\mathbf{k}}, \tag{4.10}$$

where $\widehat{a}_{\mathbf{k}}$ and $\widehat{a}^{\dagger}_{\mathbf{k}}$ are the annihilation and creation operators associated with $\{u_{\mathbf{k}}(x)\}$. They satisfy the usual commutation relations

$$[\widehat{a}_{\mathbf{k}}, \widehat{a}^{\dagger}_{\mathbf{k}'}] = \delta_{\mathbf{k}\mathbf{k}'}, \qquad [\widehat{a}_{\mathbf{k}}, \widehat{a}_{\mathbf{k}'}] = [\widehat{a}^{\dagger}_{\mathbf{k}}, \widehat{a}^{\dagger}_{\mathbf{k}'}] = 0. \tag{4.11}$$

The vacuum state $|0\rangle$ associated with this expansion is defined by the condition that $\widehat{a}_{\mathbf{k}}|0\rangle = 0 \; \forall \; \mathbf{k}$. We will refer to the state $|0\rangle$ defined in this manner as the "vacuum state associated with the modefunctions $\{u_{\mathbf{k}}(x)\}$".

To see that (4.10) defines a valid representation, first note that the right hand side satisfies the KG equation. (This follows trivially from (4.8).) It is less obvious to see why the commutation relations (4.5) are also satisfied. By substituting the right hand side of (4.10) in $[\widehat{\phi}(x), \widehat{\phi}(y)]$ and using (4.11), we find

$$[\widehat{\phi}(x), \widehat{\phi}(y)] = \sum_{\mathbf{k}} \left[u_{\mathbf{k}}(x)\overline{u}_{\mathbf{k}}(y) - \overline{u}_{\mathbf{k}}(x)u_{\mathbf{k}}(y)\right]. \tag{4.12}$$

Therefore, the commutation relations (4.5) are satisfied if and only if

$$i\Delta(x, y) = \sum_{\mathbf{k}} \left[u_{\mathbf{k}}(x)\overline{u}_{\mathbf{k}}(y) - \overline{u}_{\mathbf{k}}(x)u_{\mathbf{k}}(y)\right], \tag{4.13}$$

which is indeed the case (see Section C.3 of Appendix C for a proof).

As is well-known, the construction outlined above is not unique. A different set of modes $\{u'_{\mathbf{k}}(x)\}$ defined by a so-called Bogoliubov transformation of the modes $\{u_{\mathbf{k}}(x)\}$,

$$u'_{\mathbf{k}}(x) = \sum_{\mathbf{q}} A_{\mathbf{k}\mathbf{q}}u_{\mathbf{q}}(x) + B_{\mathbf{k}\mathbf{q}}\overline{u}_{\mathbf{q}}(x), \tag{4.14}$$

define a different representation

$$\widehat{\phi}(x) = \sum_{\mathbf{k}} u'_{\mathbf{k}}(x)\widehat{a'}_{\mathbf{k}} + \overline{u}'_{\mathbf{k}}(x)\widehat{a'}^{\dagger}_{\mathbf{k}}, \tag{4.15}$$

which is also consistent with the commutation relations (4.5) so long as

$$\begin{aligned}\sum_{\mathbf{k}} A_{\mathbf{ak}}B_{\mathbf{bk}} - B_{\mathbf{ak}}A_{\mathbf{bk}} &= 0 \\ \sum_{\mathbf{k}} A_{\mathbf{ak}}\overline{A}_{\mathbf{bk}} - B_{\mathbf{ak}}\overline{B}_{\mathbf{bk}} &= \delta_{\mathbf{ab}}.\end{aligned} \tag{4.16}$$

The vacuum state $|0'\rangle$ associated with these modes, i.e. the state defined by $\widehat{a}'_{\mathbf{k}}|0'\rangle = 0 \; \forall \; \mathbf{k}$, is different from $|0\rangle$ unless $B_{\mathbf{kq}} = 0 \; \forall \; \mathbf{k}, \mathbf{q}$, since otherwise $\widehat{a}'_{\mathbf{k}}|0\rangle \neq 0$.

The Wightman (two-point) function of the field in the state $|0\rangle$ is defined as

$$W_0(x, y) := \langle 0|\phi(x)\phi(y)|0\rangle. \tag{4.17}$$

When $|0\rangle$ is a Gaussian state, knowledge of this function fully specifies the quantum theory, since Wick's theorem then guarantees that all field correlators reduce to polynomials in $W_0(x, y)$. We will assume that $|0\rangle$ is Gaussian, since we are dealing with a non-interacting field. Using the definition of the commutation relations and the Wightman function, it follows that

$$W_0(x, y) = \frac{1}{2}H_0(x, y) + \frac{i}{2}\Delta(x, y), \tag{4.18}$$

where we have defined the Hadamard function or anticommutator

$$H_0(x, y) := 2\mathrm{Re}\left[W_0(x, y)\right] = \langle 0|\{\phi(x), \phi(y)\}|0\rangle. \tag{4.19}$$

We see that the choice of a ground state $|0\rangle$ specifies the function $H_0(x, y)$, which in turn fully encodes the state, since *any* state consistent with the canonical commutation relations will have the same Pauli-Jordan function.

4.3 The Sorkin–Johnston (SJ) Vacuum

4.3.1 The SJ Proposal

The SJ formalism defines a unique vacuum state on any bounded globally hyperbolic region $(M, g_{\mu\nu})$ of spacetime, by identifying the two-point function $W_{SJ}(x, y)$ with the "positive part" of $i\Delta(x, y) = iG_R(x, y) - iG_A(x, y)$. Let us explain what is meant by this. The kernel $i\Delta(x, y)$ has two basic properties:

- Antisymmetric — i.e. $i\Delta(y,x) = -i\Delta(x,y)$, because $G_A(x,y) = G_R(y,x)$,[5]
- Hermitian — i.e. $\overline{i\Delta(y,x)} = i\Delta(x,y)$.

Informally, if we think of $i\Delta(x,y)$ as a hermitian and antisymmetric matrix $[i\Delta]_{xy}$, its nonzero eigenvalues are all real and appear in pairs with equal magnitude but opposite signs. The SJ prescription then amounts to throwing away the negative eigenvalues and defining $[W_{SJ}]_{xy}$ as the positive part of $[i\Delta]_{xy}$. Let us look at a concrete example. Suppose, for instance, that spacetime has only two points and that

$$G_R = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \qquad \Delta = G_R - G_R^T = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{4.20}$$

This is a valid example because Δ is antisymmetric and real. The orthonormal eigenvectors of $i\Delta$ are

$$v_+ = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}, \qquad v_- = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}, \tag{4.21}$$

which have corresponding eigenvalues $+1$ and -1. In other words:

$$i\Delta v_+ = v_+, \qquad i\Delta v_- = -v_-, \qquad v_+^\dagger v_+ = v_-^\dagger v_- = 1, \qquad v_+^\dagger v_- = 0. \tag{4.22}$$

In this eigen basis, $i\Delta$ can be written as

$$i\Delta = 1 \times v_+ v_+^\dagger - 1 \times v_- v_-^\dagger. \tag{4.23}$$

The SJ prescription then amounts to defining the two-point function W_{SJ} to be the positive part of $i\Delta$, in the sense that only eigenvectors with positive eigenvalues are kept in the above expansion:

$$W_{SJ} = 1 \times v_1 v_1^\dagger = \frac{1}{2} \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix}. \tag{4.24}$$

With the general idea in mind, let us state the SJ prescription more carefully. Consider the space $L^2(M)$ of all square-integrable functions on M with the usual inner product $(\forall f, g \in L^2(M))$

$$\langle f, g \rangle := \int_M \overline{f}(x) g(x) dV_x, \tag{4.25}$$

[5] See Section C.4 of Appendix C for a proof.

where $dV_x = \sqrt{-g(x)}d^D x$ is the natural volume element on (M, g). We define the Pauli-Jordan operator as the integral operator whose kernel is $i\Delta(x, y)$:

$$(i\Delta f)(x) = \int_M i\Delta(x, y) f(y) dV_y. \tag{4.26}$$

Then, $i\Delta$ defines a self-adjoint (or hermitian) operator on $L^2(M)$, meaning that $\langle f, i\Delta g\rangle = \langle i\Delta f, g\rangle$:

$$\begin{aligned}
\langle f, i\Delta g\rangle &= \int_M \overline{f}(x)(i\Delta g)(x)\, \mathrm{d}V_x \\
&= \int_M \int_M i\Delta(x, y)\overline{f}(x) g(y)\, \mathrm{d}V_x\, \mathrm{d}V_y \\
&= -\int_M \int_M i\Delta(y, x)\overline{f}(x) g(y)\, \mathrm{d}V_x\, \mathrm{d}V_y \\
&= \int_M \overline{(i\Delta f)(y)} g(y)\, \mathrm{d}V_y \\
&= \langle i\Delta f, g\rangle,
\end{aligned} \tag{4.27}$$

where in the second line we have used the fact that $\Delta(x, y)$ is antisymmetric. There is an important technical issue which we have overlooked here: $i\Delta$ defines a *symmetric* operator on a dense subset of $L^2(M)$ (smooth functions of compact support) *only when* M *is bounded*.[6] In this case, as is shown in Section C.5 of Appendix C, $i\Delta$ is actually a bounded operator on $L^2(M)$, which then implies that it is self-adjoint. Therefore, for the SJ formalism to go through, it is necessary to consider only bounded globally hyperbolic spacetimes. The spectral theorem then guarantees that $i\Delta$ has a set of real eigenvalues $\{\lambda_{\mathbf{a}}\}$, as well as a complete orthonormal set of eigenvectors $\{\mathfrak{u}_{\mathbf{a}}(x)\}$ which satisfy [129]

$$i\Delta \mathfrak{u}_{\mathbf{a}} = \lambda_{\mathbf{a}} \mathfrak{u}_{\mathbf{a}}, \qquad \lambda_{\mathbf{a}} \in \mathbb{R}. \tag{4.28}$$

Since $\Delta(x, y)$ is a real function, it follows (by complex conjugating both sides of (4.28)) that

$$i\Delta \mathfrak{u}_{\mathbf{a}} = \lambda_{\mathbf{a}} \mathfrak{u}_{\mathbf{a}}(x) \implies i\Delta \overline{\mathfrak{u}}_{\mathbf{a}} = -\lambda_{\mathbf{a}} \overline{\mathfrak{u}}_{\mathbf{a}}, \tag{4.29}$$

which means that the non-zero eigenvectors of $i\Delta$ come in pairs:

$$i\Delta \mathfrak{u}^{\pm}_{\mathbf{a}} = \pm\lambda_{\mathbf{a}} \mathfrak{u}^{\pm}_{\mathbf{a}}, \tag{4.30}$$

[6] Bounded = having compact closure.

where by definition $\lambda_{\mathbf{a}} > 0$ and $\mathfrak{u}_{\mathbf{a}}^- = \overline{\mathfrak{u}}_{\mathbf{a}}^+$. Moreover, these functions are orthonormal in the $L^2(M)$ inner product:

$$\langle \mathfrak{u}_{\mathbf{a}}^{\pm}, \mathfrak{u}_{\mathbf{b}}^{\pm} \rangle = \delta_{\mathbf{ab}}, \qquad \langle \mathfrak{u}_{\mathbf{a}}^{+}, \mathfrak{u}_{\mathbf{b}}^{-} \rangle = 0. \tag{4.31}$$

We can now split $i\Delta(x,y)$ into a positive and negative part

$$i\Delta(x,y) = \sum_{\mathbf{a}} Q(x,y) - \overline{Q(x,y)}, \tag{4.32}$$

where

$$Q(x,y) = \sum_{\mathbf{a}} \lambda_{\mathbf{a}} \mathfrak{u}_{\mathbf{a}}^+(x) \overline{\mathfrak{u}}_{\mathbf{a}}^+(y). \tag{4.33}$$

The SJ vacuum $|SJ\rangle$ is then defined by

$$W_{SJ}(x,y) = \langle SJ | \widehat{\phi}(x) \widehat{\phi}(y) | SJ \rangle := Q(x,y). \tag{4.34}$$

In Section C.6 of Appendix C we show that this definition is valid also when the spectrum of $i\Delta$ is degenerate. It follows from (4.33) and (4.34) that the field operator $\widehat{\phi}(x)$ can be expanded as a mode sum

$$\widehat{\phi}(x) = \sum_{\mathbf{a}} u_{\mathbf{a}}^{SJ}(x) \widehat{a}_{\mathbf{a}} + \overline{u}_{\mathbf{a}}^{SJ}(x) \widehat{a}_{\mathbf{a}}^\dagger, \tag{4.35}$$

where the SJ modefunctions $\{u_{\mathbf{a}}^{SJ}\}$ are given by

$$u_{\mathbf{a}}^{SJ}(x) := \sqrt{\lambda_{\mathbf{a}}} \mathfrak{u}_{\mathbf{a}}^+(x), \tag{4.36}$$

and the corresponding creation and annihilation operators satisfy the usual commutation relations

$$[\widehat{a}_{\mathbf{a}}, \widehat{a}_{\mathbf{a}'}^\dagger] = \delta_{\mathbf{a}\mathbf{a}'}, \qquad [\widehat{a}_{\mathbf{a}}, \widehat{a}_{\mathbf{a}'}] = [\widehat{a}_{\mathbf{a}}^\dagger, \widehat{a}_{\mathbf{a}'}^\dagger] = 0. \tag{4.37}$$

The SJ vacuum is then defined by $\widehat{a}_{\mathbf{a}} |SJ\rangle = 0 \; \forall \; \mathbf{a}$.

Is $W_{SJ}(x,y)$ a valid definition for a two-point function? Certainly, not every function of two variables is. To our knowledge, any Wightman function $W(x,y) = \langle 0 | \widehat{\phi}(x) \widehat{\phi}(y) | 0 \rangle$ must have the following properties:

1. $\langle f, Wf \rangle \geq 0 \; \forall f$. To see why this should be true, notice that

$$\begin{aligned} \langle f, Wf \rangle &= \int_M \int_M W(x,y) \overline{f}(x) f(y) \, \mathrm{d}V_x \, \mathrm{d}V_y \\ &= \left(\int_M \langle 0 | \widehat{\phi}(x) \overline{f}(x) \, \mathrm{d}V_x \right) \left(\int_M f(y) \widehat{\phi}(y) | 0 \rangle \, \mathrm{d}V_y \right). \end{aligned} \tag{4.38}$$

Defining $|\psi\rangle = \int_M f(y)\widehat{\phi}(y)|0\rangle \, \mathrm{d}V_y$, the second line in (4.38) becomes $\langle\psi|\psi\rangle$, which is always greater of equal to zero.

2. $W(x,y) - W(y,x) = i\Delta(x,y)$. This should be true because

$$\begin{aligned} W(x,y) - W(y,x) &= \langle 0|\widehat{\phi}(x)\widehat{\phi}(y)|0\rangle - \langle 0|\widehat{\phi}(y)\widehat{\phi}(x)|0\rangle \\ &= \langle 0|[\widehat{\phi}(x), \widehat{\phi}(y)]|0\rangle \\ &= i\Delta(x,y), \end{aligned} \tag{4.39}$$

   where we have assumed that $|0\rangle$ is normalized.

3. $(\Box_x - m^2)W(x,y) = 0$. (This follows trivially from (4.4).)

The SJ two-point function $W_{SJ}(x,y)$ satisfies all these conditions because

1. it is positive semi-definite:

$$\begin{aligned} \langle f, W_{SJ} f\rangle &= \int_M \int_M \left\{ \sum_{\mathbf{a}} \lambda_{\mathbf{a}} \mathfrak{u}^+_{\mathbf{a}}(x) \overline{\mathfrak{u}}^+_{\mathbf{a}}(y) \right\} \mathrm{d}V_x \, \mathrm{d}V_y \\ &= \sum_{\mathbf{a}} \lambda_{\mathbf{a}} \left(\int_M \mathfrak{u}^+_{\mathbf{a}}(x) \, \mathrm{d}V_x \right) \left(\int_M \overline{\mathfrak{u}}^+_{\mathbf{a}}(y) \, \mathrm{d}V_y \right) \\ &= \sum_{\mathbf{a}} \lambda_{\mathbf{a}} \left| \int_M \mathfrak{u}^+_{\mathbf{a}}(x) \, \mathrm{d}V_x \right|^2 \geq 0. \end{aligned} \tag{4.40}$$

2. its anti-symmetrization produces $i\Delta(x,y)$, by definition.

3. it satisfies the KG equation: $(\Box_x - m^2)W_{SJ}(x,y) = \sum_{\mathbf{a}} \lambda_{\mathbf{a}} (\Box_x - m^2) u^+_{\mathbf{a}}(x) \overline{\mathfrak{u}}^+_{\mathbf{a}}(y) = 0$. To see why the last equality is true, recall that $(\Box_x - m^2)\Delta(x,y) = 0$. Therefore, $(\Box_x - m^2)(i\Delta f)(x) = 0$ for all f, which implies $(\Box_x - m^2)\mathfrak{u}^+_{\mathbf{a}}(x) = (\Box_x - m^2)(i\Delta \mathfrak{u}^+_{\mathbf{a}})(x)/\lambda_{\mathbf{a}} = 0$.

Before delving into the diagonlization of $i\Delta$, a few generic remarks about the SJ formalism are in order. Although selfadjointness of $i\Delta$ might seem to be merely a technical issue, it highlights the fact that the SJ vacuum depends on a choice of (globally hyperbolic) spacetime region. Indeed, as we have already argued, the SJ prescription is not guaranteed to be well defined unless one chooses a region that is bounded, *both spatially and temporally.*

Thus arise two questions: To what extent does the SJ prescription depend on boundary condtions, and to what extent does it remain well-defined in unbounded spacetimes?

In answering the first question, one must distinguish between spatial boundaries (also referred to as timelike boundaries) and temporal ones. Spatial boundaries are familiar to us from putting fields in a box, Casimir effect, etc; and they seem unproblematic. When they are present the SJ vacuum will be sensitive to one's choice of boundary conditions, because the retarded and advanced Green's functions depend on them. But this is as it should be since the physics genuinely depends on the boundary conditions. We will also consider below regions which are unbounded spatially, but no special difficulties will arise from that feature.

The case of a temporal boundary (spacelike or null) is less familiar. The first thing to notice is that *boundary conditions are neither needed nor possible* in this case, since the region is (by assumption) globally hyperbolic. Mathematically, this very satisfactory feature stems from the fact that Δ is not a differential operator but an integral one. Nevertheless, one must bear in mind that the "ground state" one ends up with, does depend on the region with which one begins. In itself, this dependence on the region merely expresses the nonlocal character of our definition. One might for example be interested in which "vacuum" would be appropriate to an early stage of expansion of the cosmos, and one would not want in that case to apply our definition to the full spacetime, including its whole future development. However, one might also want to apply the definition to unbounded spacetimes like Minkowski space, and in such cases one needs to worry about dependence on an eventual infrared cutoff. If the metric is static, for instance, why should time play any role in what the vacuum state looks like?

In dealing with such instances, it is always possible to work first in a truncated spacetime, and later send the temporal boundary to infinity. In Section C.7 of Appendix C, we apply this method to the simple harmonic oscillator and show that it succeeds in the sense that the resulting SJ vacuum is the minimum energy state of the Hamiltonian. As we will later demonstrate, this continues to be true for all static spacetimes. Another example of such a calculation can be found in [6], where the spectrum of $i\Delta$ is computed in a 1+1 dimensional causal-diamond, and it is found that (up to the usual infrared ambiguities that affect massless scalars in 2d) the resulting two-point function has the correct (i.e. Minkowski) limiting behaviour as the boundaries of the diamond tend to infinity. However, there are also cases where taking a temporal cutoff to infinity is an ill-defined procedure. We will later show that the so-called Poincaré patch of de Sitter space provides an example of such a case. (We suspect that this kind of ambiguity can be understood intuitively as the failure of $i\Delta$ to admit a selfadjoint extension which is unique. However we don't know how to pose such a question properly, because Δ is densely defined only as a quadratic

form on $L^2(M)$, not as an operator from $L^2(M)$ to itself.)

4.3.2 Diagonalizing $i\Delta$

Diagonalizing $i\Delta$, i.e. solving the eigenvalue problem (4.28), can be reduced to a set of algebraic equations as follows. Given any expansion of the field in terms of a set of modes $\{u_{\mathbf{k}}(x)\}$ which form an orthonormal basis of the KG solution space, the commutator function $i\Delta(x,y)$ can be expressed as the mode sum (4.13). It then follows from (4.26) and (4.25) that

$$i\Delta u^{SJ}_{\mathbf{a}}(x) = \sum_{\mathbf{q}} \langle u_{\mathbf{q}}, u^{SJ}_{\mathbf{a}}\rangle u_{\mathbf{q}}(x) - \langle \overline{u}_{\mathbf{q}}, u^{SJ}_{\mathbf{a}}\rangle \overline{u}_{\mathbf{q}}(x). \tag{4.41}$$

This implies that we can rewrite (4.28) for an eigenfunction $u^{SJ}_{\mathbf{a}}$ with positive eigenvalue $\lambda_{\mathbf{a}}$ as

$$u^{SJ}_{\mathbf{a}}(x) = \sum_{\mathbf{q}} A_{\mathbf{aq}} u_{\mathbf{q}}(x) + B_{\mathbf{aq}} \overline{u}_{\mathbf{q}}(x), \tag{4.42}$$

where we have defined

$$\begin{aligned} A_{\mathbf{aq}} &= \quad \lambda_{\mathbf{a}}^{-1} \langle u_{\mathbf{q}}, u^{SJ}_{\mathbf{a}}\rangle, \\ B_{\mathbf{aq}} &= -\lambda_{\mathbf{a}}^{-1} \langle \overline{u}_{\mathbf{q}}, u^{SJ}_{\mathbf{a}}\rangle. \end{aligned} \tag{4.43}$$

As the notation is meant to indicate, these coefficients define a Bogoliubov transformation. Acting on (4.42) with $\langle u_{\mathbf{k}}, \cdot\,\rangle$ and $\langle \overline{u}_{\mathbf{k}}, \cdot\,\rangle$ yields

$$A_{\mathbf{ak}} = \frac{1}{\lambda_{\mathbf{a}}} \sum_{\mathbf{q}} A_{\mathbf{aq}} \langle u_{\mathbf{k}}, u_{\mathbf{q}}\rangle + B_{\mathbf{aq}} \langle u_{\mathbf{k}}, \overline{u}_{\mathbf{q}}\rangle, \tag{4.44}$$

$$B_{\mathbf{ak}} = \frac{-1}{\lambda_{\mathbf{a}}} \sum_{\mathbf{q}} A_{\mathbf{aq}} \langle \overline{u}_{\mathbf{k}}, u_{\mathbf{q}}\rangle + B_{\mathbf{aq}} \langle u_{\mathbf{q}}, u_{\mathbf{k}}\rangle. \tag{4.45}$$

Complementing these equations with the L^2 orthonormality conditions (4.31) on the SJ modes, we find the Bogoliubov conditions

$$\sum_{\mathbf{k}} A_{\mathbf{ak}} B_{\mathbf{bk}} - B_{\mathbf{ak}} A_{\mathbf{bk}} = 0 \tag{4.46}$$

$$\sum_{\mathbf{k}} A_{\mathbf{ak}} \overline{A}_{\mathbf{bk}} - B_{\mathbf{ak}} \overline{B}_{\mathbf{bk}} = \delta_{\mathbf{ab}}. \tag{4.47}$$

Finding the SJ vacuum now reduces to solving the above system of equations for $A_{\mathbf{ak}}$ and $B_{\mathbf{ak}}$. Note again that this construction is only valid in a bounded region of spacetime, since otherwise the inner products might diverge.

Lemma 1. *If the algebraic relations (4.44)–(4.47) imply $B_{\mathbf{ak}} = 0$, then the SJ state $|SJ\rangle$ coincides with the vacuum state $|0_u\rangle$ corresponding to the modefunctions $\{u_{\mathbf{k}}(x)\}$.*

Proof. To see why this is, note that:

$$\begin{aligned}\langle SJ|\widehat{\phi}(x)\widehat{\phi}(y)|SJ\rangle &= \sum_{\mathbf{a}} u_{\mathbf{a}}^{SJ}(x)\overline{u}_{\mathbf{a}}^{SJ}(y) \\ &= \sum_{\mathbf{qq'}} u_{\mathbf{q}}(x)\overline{u}_{\mathbf{q'}}(y)\sum_{\mathbf{a}} A_{\mathbf{aq}}\overline{A}_{\mathbf{aq'}} \\ &= \sum_{\mathbf{q}} u_{\mathbf{q}}(x)\overline{u}_{\mathbf{q}}(y) \\ &= \langle 0_u|\widehat{\phi}(x)\widehat{\phi}(y)|0_u\rangle,\end{aligned} \tag{4.48}$$

where in the second equality we have substituted the right hand side of (4.42) for $u_{\mathbf{a}}^{SJ}(x)$ with $B_{\mathbf{ak}} = 0$, and the third equality follows from (4.47). Since the theory at hand is non-interacting by assumption, this is enough to show that $|SJ\rangle = |0_u\rangle$. □

Interestingly, it can be shown that $B_{\mathbf{ak}} = 0$ whenever $\langle u_{\mathbf{k}}, \overline{u}_{\mathbf{q}}\rangle = 0 \ \forall \ \mathbf{k}, \mathbf{q}$.

Theorem 2. *The SJ proposal picks the orthonormal basis $\{u_{\mathbf{k}}(x)\}$ of the Klein-Gordon solution space which enjoys the addition property $\langle u_{\mathbf{k}}, \overline{u}_{\mathbf{q}}\rangle = 0 \ \forall \ \mathbf{k}, \mathbf{q}$.*

Proof. Define the matrices $\mathbf{U}_{\mathbf{kq}} = \langle u_{\mathbf{k}}, u_{\mathbf{q}}\rangle$ and $\mathbf{\Lambda}_{\mathbf{aa'}} = \lambda_{\mathbf{a}}\delta_{\mathbf{aa'}}$. Note that both $\mathbf{U}$ and $\mathbf{\Lambda}$ are positive matrices:

$$f^{\dagger}\mathbf{\Lambda}f = \sum_{\mathbf{a},\mathbf{a'}} \overline{f}_{\mathbf{a}}\mathbf{\Lambda}_{\mathbf{aa'}}f_{\mathbf{a'}} = \sum_{\mathbf{a}} \lambda_{\mathbf{a}}|f_{\mathbf{a}}|^2 > 0. \tag{4.49}$$

and

$$\begin{aligned}f^{\dagger}\mathbf{U}f &= \sum_{\mathbf{k},\mathbf{q}} \overline{f}_{\mathbf{k}}\mathbf{U}_{\mathbf{kq}}f_{\mathbf{q}} = \sum_{\mathbf{k},\mathbf{q}} \overline{f}_{\mathbf{k}}\langle u_{\mathbf{k}}, u_{\mathbf{q}}\rangle f_{\mathbf{q}} \\ &= \sum_{\mathbf{k},\mathbf{q}} \overline{f}_{\mathbf{k}}\left(\int_M \overline{u}_{\mathbf{k}}(x)u_{\mathbf{q}}(x)dV_x\right)f_{\mathbf{q}} \\ &= \int_M \left(\overline{\sum_{\mathbf{k}} f_{\mathbf{k}}u_{\mathbf{k}}(x)}\right)\left(\sum_{\mathbf{q}} f_{\mathbf{q}}u_{\mathbf{q}}(x)\right)dV_x \\ &= \int_M \left|\sum_{\mathbf{k}} f_{\mathbf{k}}u_{\mathbf{k}}(x)\right|^2 dV_x > 0.\end{aligned} \tag{4.50}$$

When $\langle u_{\mathbf{k}}, \overline{u}_{\mathbf{q}} \rangle = 0$, (4.45) simplifies to

$$\mathbf{\Lambda B} = -\mathbf{BU}. \tag{4.51}$$

Multiplying both side of (4.51) with $\mathbf{B}^\dagger$ from the left:

$$\mathbf{B}^\dagger \mathbf{\Lambda B} = -\mathbf{B}^\dagger \mathbf{BU}. \tag{4.52}$$

The left hand side of (4.52) is a positive-definite operator because $\mathbf{\Lambda}$ is positive:

$$f^\dagger \mathbf{B}^\dagger \mathbf{\Lambda B} f = (\mathbf{B}f)^\dagger \mathbf{\Lambda} (\mathbf{B}f) \geq 0. \tag{4.53}$$

The right hand side of (4.52), however, need not be positive. For instance, if we let f be an eigenvector of $\mathbf{U}$ with eigenvalue $\gamma_f > 0$ (γ_f is positive because $\mathbf{U}$ is a positive matrix):

$$-f^\dagger \mathbf{B}^\dagger \mathbf{BU} f = -\gamma_f f^\dagger \mathbf{B}^\dagger \mathbf{B} f = -\gamma_f (\mathbf{B}f)^\dagger (\mathbf{B}f) \leq 0. \tag{4.54}$$

This can only happen if $\mathbf{B} = 0$, or equivalently $B_{\mathbf{ak}} = 0$. As was shown in Lemma (1), when this is the case, the SJ state coincides with the corresponding vacuum state of the modefunctions $\{u_{\mathbf{k}}(x)\}$.

Note that the matrices $\mathbf{U}$, $\mathbf{\Lambda}$, etc, are really infinite dimensional matrices. Therefore, it could happen that the sums involved in the matrix multiplications above do not converge, in which case, this proof does not go through. □

4.4 The SJ Vacuum on Static Spacetimes

In static spacetimes, i.e. spacetimes which admit an everywhere time-like and hypersurface-orthogonal Killing vector ξ^μ, a natural choice of vacuum modefunctions exists. These are the solutions of the Klein-Gordon equation which are purely positive frequency with respect to the Killing time. The corresponding vacuum state minimizes the Hamiltonian, which in this case is time independent. In this Section, we show that the SJ state coincides with this usual ground state in static spacetimes, for a free massive scalar field.

Let us first review the construction of the natural positive frequency modefunctions. In static spacetimes, it is always possible to choose a coordinate system in which the line element takes the form

$$ds^2 = -N^2(\mathbf{x})dt^2 + \gamma_{ij}(\mathbf{x})dx^i dx^j, \tag{4.55}$$

where $N(\mathbf{x})$ is any smooth and everywhere-positive lapse function, γ_{ij} is the induced metric on the spatial hypersurface Σ, and $i, j \in \{1, \cdots, d\}$. The Klein-Gordon equation for a massive scalar field $\phi(x)$ then reads

$$\left(\frac{\partial^2}{\partial t^2} + K\right)\phi(x) = 0, \tag{4.56}$$

where $K = N^2(\mathbf{x})\Big\{-\frac{1}{\sqrt{\gamma(\mathbf{x})}N(\mathbf{x})}\partial_i\big[\sqrt{\gamma(\mathbf{x})}N(\mathbf{x})\gamma^{ij}(\mathbf{x})\partial_j\big] + m^2\Big\}$ is a purely spatial differential operator. We keep $m > 0$ to avoid potential infrared difficulties with zero modes. Let $L^2(\Sigma)$ denote the Hilbert space of all L^2 functions on Σ with inner product

$$\langle f, g\rangle_S = \int_\Sigma \overline{f}(\mathbf{x})g(\mathbf{x})\frac{\sqrt{\gamma(\mathbf{x})}}{||\xi(\mathbf{x})||}\,\mathrm{d}^d\mathbf{x}, \tag{4.57}$$

where $||\xi|| = \sqrt{-\xi_\mu\xi^\mu} = N(\mathbf{x})$, and γ is the determinant of γ_{ij}. Let us assume that K is a self-adjoint and strictly positive operator on $L^2(\Sigma)$, so that it has a well-defined positive spectrum

$$(K\psi_{\mathbf{k}})(\mathbf{x}) = \omega(\mathbf{k})^2\psi_{\mathbf{k}}(\mathbf{x}), \qquad \omega(\mathbf{k}) > 0. \tag{4.58}$$

We also assume that $\{\psi_{\mathbf{k}}(\mathbf{x})\}$ form an orthonormal basis (which can always be arranged):

$$\langle\psi_{\mathbf{k}}, \psi_{\mathbf{q}}\rangle_S = \delta_{\mathbf{kq}}. \tag{4.59}$$

Then, it is always possible to find complex solutions of the Klein-Gordon equation of the form

$$u_{\mathbf{k}}(t, \mathbf{x}) = \frac{e^{-i\omega(\mathbf{k})t}}{\sqrt{2\omega(\mathbf{k})}}\psi_{\mathbf{k}}(\mathbf{x}). \tag{4.60}$$

These modefunctions also form an orthonormal basis of the Klein-Gordon solution space:[7]

$$(u_{\mathbf{k}}, u_{\mathbf{q}})_{KG} = \frac{\omega(\mathbf{k}) + \omega(\mathbf{q})}{\sqrt{2\omega(\mathbf{k})}\sqrt{2\omega(\mathbf{q})}}e^{it(\omega(\mathbf{k})-\omega(\mathbf{q}))}\langle\psi_{\mathbf{k}}, \psi_{\mathbf{q}}\rangle_S = \delta_{\mathbf{kq}} \tag{4.61}$$

$$(\bar{u}_{\mathbf{k}}, u_{\mathbf{q}})_{KG} = \frac{\omega(\mathbf{q}) - \omega(\mathbf{k})}{\sqrt{2\omega(\mathbf{k})}\sqrt{2\omega(\mathbf{q})}}e^{-it(\omega(\mathbf{k})+\omega(\mathbf{q}))}\langle\bar{\psi}_{\mathbf{k}}, \psi_{\mathbf{q}}\rangle_S = 0. \tag{4.62}$$

[7] The KG norm is defined in (4.3). The unit normal to Σ is $n^\mu = \frac{\xi^\mu}{||\xi||}$. Note that $(\bar{u}_{\mathbf{k}}, u_{\mathbf{q}})_{KG}$ also vanishes when $\omega(\mathbf{k}) \neq \omega(\mathbf{q})$ because in that case $\langle\bar{\psi}_{\mathbf{k}}, \psi_{\mathbf{q}}\rangle_S = 0$. This follows from the fact that $\bar{\psi}_{\mathbf{k}}$ is also an eigenfunction of K with eigenvalue $\omega(\mathbf{k})^2$. Therefore, when $\omega(\mathbf{k}) \neq \omega(\mathbf{q})$, $\bar{\psi}_{\mathbf{k}}$ and $\psi_{\mathbf{q}}$ are two eigenfunctions of K with different eigenvalues, which implies that their inner product must vanish.

The corresponding vacuum state of these positive-frequency modefunctions is indeed the ground state of the Hamiltonian.

We now turn to the SJ prescription. It Section 4.3.2, it was shown how the SJ modefunctions can be constructed from any orthonormal basis of the KG solution space $\{u_{\mathbf{k}}(x)\}$. Using the positive frequency modes described above for our calculation, it can be shown that

$$\langle u_{\mathbf{k}}, \overline{u}_{\mathbf{q}} \rangle = \frac{\pi \delta(\omega(\mathbf{k}) + \omega(\mathbf{q}))}{\sqrt{\omega(\mathbf{k})\omega(\mathbf{q})}} \int \overline{\psi}_{\mathbf{k}}(\mathbf{x}) \overline{\psi}_{\mathbf{q}}(\mathbf{x}) N(\mathbf{x}) \sqrt{\gamma} d^d \mathbf{x} = 0, \tag{4.63}$$

where we have taken $t \in (-\infty, \infty)$. The last equality follows because $\omega(\mathbf{k}) > 0$, and therefore $\delta(\omega(\mathbf{k}) + \omega(\mathbf{q})) = 0$. By theorem 2, this is sufficient to show that the SJ state coincides with the corresponding vacuum state of the positive-frequency modefunctions $\{u_{\mathbf{k}}(x)\}$.

It is worth noting that the foregoing analysis does *not* apply to stationary spacetimes that are not static, including cases where the Killing vector under consideration is not everywhere timelike. It would be particularly interesting to investigate the SJ vacuum in the spacetime of a rotating star with an ergo-region.

4.5 The SJ Vacuum on de Sitter Space

In this Section, we apply the SJ formalism to a free massive scalar field in $D = d + 1$ dimensional de Sitter space, which is a particularly interesting setting for various reasons. Firstly, as we saw in the previous Section, the SJ vacuum agrees with the ground state of the Hamiltonian in static spacetimes. Because de Sitter space and its half spaces are not static (or stationary), computing the SJ vacuum thereon is not merely another "consistency check". Secondly, the SJ formalism is sensitive to the global structure of spacetime. (Section 4.7 explores this feature in greater detail.) By evaluating it on the the full de Sitter hyperboloid as well as its Poincaré half space, we can investigate further its nonlocal nature. Thirdly, as was mentioned in Section 4.3.1, strictly speaking, the SJ vacuum is only defined on *bounded* regions of spacetime. One strategy to find the SJ vacuum on an unbounded region is to first compute it for a bounded globally hyperbolic subregion, and then take the appropriate limits to recover the entire spacetime. In the case of de Sitter space, we will see that this procedure gives meaningful answers in most circumstances, but that it also fails in some cases. Fourthly, it is worth investigating whether, or in which circumstances, the SJ vacuum obeys the so-called Hadamard condition. An explicit calculation in [62] shows that the SJ vacuum is not always Hadamard. We find that for certain

ranges of the scalar field mass and values of spacetime dimension, the SJ vacuum on de Sitter space is also not Hadamard. Finally, de Sitter space is appropriate for studying potential phenomenological applications of the SJ vacuum to cosmology.

In order to diagonalise $i\Delta$, as described in Section 4.3.2, we need to pick an arbitrary complete set of modes $\{u_{\mathbf{k}}\}$, in terms of which we can obtain the SJ modes. A convenient choice are the modes associated with the so-called *Euclidean* or *Bunch-Davies* (BD) state [41]. The modes that define this vacuum on the full space (denoted dS^D) and on the Poincaré half-space (denoted dS^D_P) will be referred to as the Euclidean modes $u^E_{Lj}(x_G)$, and BD modes $u^{BD}_{\mathbf{k}}(x_P)$, respectively, where x_G and x_P denote the coordinates on the two patches. These modes define the same state on dS^D_P, i.e. their two-point functions are identical. The Euclidean/BD state belongs to a two-real-parameter family of de Sitter-invariant vacuum states, known as the *Mottola-Allen* or α-vacua [104, 12]. We have listed some relevant facts about de Sitter geometry and α-vacua in Sections C.8 and C.9 of Appendix C, respectively.

4.5.1 The SJ vacuum on the Poincaré patch

In cosmological coordinates, the de Sitter metric reads (see Section C.8.2 of Appendix C)

$$ds^2 = \frac{\ell^2}{\eta^2}\left[-d\eta^2 + \sum_{i=1}^{d} dx_i^2\right], \tag{4.64}$$

where $\eta \in (-\infty, 0)$, and $x_i \in (-\infty, +\infty)$. The positive-frequency modes that define the BD vacuum $|BD\rangle$ on dS^D_P take the form (see Section C.9.1 of Appendix C)

$$u^{BD}_{\mathbf{k}}(\eta, \mathbf{x}) = \frac{e^{i\mathbf{k}\cdot\mathbf{x}}}{(2\pi)^{d/2}}\chi_k(\eta), \qquad \chi_k(\eta) = \sqrt{\frac{\pi\ell}{4}} e^{i\pi\left(\frac{\nu}{2} - \frac{d+2}{4}\right)} \left(\frac{-\eta}{\ell}\right)^{d/2} H^{(1)}_\nu(-k\eta), \tag{4.65}$$

where

$$\nu = \ell\sqrt{m_*^2 - m^2}, \qquad m_* = \frac{d}{2\ell}, \qquad k = |\mathbf{k}|. \tag{4.66}$$

We shall refer to m_* as the *critical mass*. As m increases from 0 to m_*, ν decreases along the real line from $\frac{d}{2\ell}$ to 0, and as m increases further across m_*, ν moves up the imaginary axis. (The critical mass plays a central role in the representation theory of the de Sitter group; see [153, 37] for some considerations on the physical significance of m_*.) As shown in Section C.9.1 of Appendix C, these modes satisfy the KG equation and are orthonormal

with respect to the KG inner product. The L^2 inner product (see (4.25)) of these modes is also of interest to us:

$$\langle u^{BD}_{\mathbf{k}}, u^{BD}_{\mathbf{q}} \rangle = \delta^{(d)}(\mathbf{k}-\mathbf{q})\langle \chi_k, \chi_k \rangle_\eta, \tag{4.67}$$

$$\langle u^{BD}_{\mathbf{k}}, \overline{u}^{BD}_{\mathbf{q}} \rangle = \delta^{(d)}(\mathbf{k}+\mathbf{q})\langle \chi_k, \overline{\chi}_k \rangle_\eta, \tag{4.68}$$

where we have defined the inner product $\langle \cdot, \cdot \rangle_\eta$ for functions of η only:

$$\langle f, g \rangle_\eta := \int_{\eta_{\min}}^{\eta_{max}} \overline{f(\eta)} g(\eta) \left(\frac{-\ell}{\eta}\right)^{d+1} d\eta. \tag{4.69}$$

We have introduced η_{min} and η_{max} as regulators which will be sent to $-\infty$ and 0 (respectively) after the SJ vacuum has been computed. The algebraic relations (4.44)–(4.47) can now be solved for:

$$\begin{aligned} A_{\mathbf{kq}} &= \delta^{(d)}(\mathbf{k}-\mathbf{q})\cosh(\alpha_k) \\ B_{\mathbf{kq}} &= \delta^{(d)}(\mathbf{k}+\mathbf{q})\sinh(\alpha_k)e^{i\beta_k} \\ \lambda_{\mathbf{k}} &= \sqrt{\langle \chi_k, \chi_k \rangle_\eta^2 - |\langle \overline{\chi}_k, \chi_k \rangle_\eta|^2}, \end{aligned} \tag{4.70}$$

where

$$\alpha_k = \frac{1}{2}\tanh^{-1}|r_k|, \qquad \beta_k = \arg(r_k) + \pi, \tag{4.71}$$

and

$$r_k := \frac{\langle \overline{\chi}_k, \chi_k \rangle_\eta}{\langle \chi_k, \chi_k \rangle_\eta}. \tag{4.72}$$

The SJ modefunctions (4.42) are then given by

$$u^{SJ}_{\mathbf{k}}(x) = \cosh(\alpha_k) u^{BD}_{\mathbf{k}}(x) + \sinh(\alpha_k) e^{i\beta_k} \overline{u}^{BD}_{-\mathbf{k}}(x). \tag{4.73}$$

The above expressions are valid only when $|r_k| \neq 1$. When $|r_k| = 1$, the Bogoliubov coefficients blow up and the SJ prescription is no longer valid. In Section C.10.1 of Appendix C, we have computed r_k in the limit $\eta_{min} \to -\infty$ and $\eta_{max} \to 0$:

$$r_k = \begin{cases} e^{i\pi\left(\nu - \frac{d}{2}\right)} & \text{if } m \leq m_*, \\ e^{-i\pi\frac{d}{2}}\operatorname{sech}(\pi|\nu|) & \text{if } m \geq m_*. \end{cases} \tag{4.74}$$

We see that for masses $m \leq m_*$, the SJ prescription is not well defined in the limit $\eta_{max} \to 0$, since in that case $|r_k| \to 1$. When $m > m_*$, we find that the Bogoliubov coefficients are

$$\alpha_k = \tanh^{-1} e^{-\pi|\nu|} \quad \text{and} \quad \beta_k = -\frac{D+1}{2}\pi. \tag{4.75}$$

This corresponds to the particular α-vacuum known as the *out*-vacuum (see Section C.9.3 of Appendix C). More specifically, when $m > m_*$, the two point function of the SJ vacuum in the Poincaré patch is equal to the restriction of the out-vacuum two-point function in this region.

4.5.2 The SJ vacuum on the global patch

In global coordinates, the de Sitter metric reads (see Section C.8.1 of Appendix C)

$$ds^2 = -dt^2 + \ell^2 \cosh^2(t/\ell)\, d\Omega_d^2, \tag{4.76}$$

where $d\Omega_d^2$ is the line element on the $d-$Sphere (S^d) and $t \in (-\infty, +\infty)$. Letting $z(t) = 1 + e^{2t/\ell}$, the positive-frequency modes that define the Euclidean vacuum on dS^D take the form (see Section C.9.2 of Appendix C)

$$u^E_{Lj}(t,\Omega) = y^E_L(t) Y_{Lj}(\Omega), \qquad y^E_L(t) = \mathcal{N}_L e^{(a+\nu)t/\ell} \cosh^L(t/\ell) F(a, a+\nu; 2a; z(t) - i\epsilon), \tag{4.77}$$

where

$$\mathcal{N}_L = \frac{e^{i\frac{\pi}{2}(a+\nu)}}{2^a \ell^{\frac{d-1}{2}}} \frac{\sqrt{\Gamma(a+\nu)\Gamma(a-\nu)}}{\Gamma(a+\frac{1}{2})}, \qquad a = L + d/2. \tag{4.78}$$

Here F denotes the hypergeometric function ${}_2F_1$ and $-i\epsilon$ determines the side of the branch cut (from 1 to ∞ along the real axis) where it should be evaluated. The functions $Y_{Lj}(\Omega)$ are spherical harmonics on S^d, whose relevant properties we have included in Section C.9.1 of Appendix C. Also, $L \in \{0, 1, 2, \dots\}$ and j is a collective index for $j_1, j_2, \dots, j_{d-1}$, which run over values $|j_{d-1}| \leq j_{d-2} \leq \cdots \leq j_1 \leq L$. These modes satisfy the Klein-Gordon equation and are orthonormal with respect to the Klein-Gordon inner product. The L^2 inner products of interest are

$$\langle u^E_{Lj}, u^E_{L'j'} \rangle = \langle y^E_L, y^E_L \rangle_t \delta_{LL'} \delta_{jj'}, \tag{4.79}$$

$$\langle \overline{u}^E_{Lj}, u^E_{L'j'} \rangle = \langle \overline{y}^E_L, y^E_L \rangle_t (-1)^L \delta_{LL'} \delta_{jj'}, \tag{4.80}$$

where we have defined an inner product $\langle \cdot, \cdot \rangle_t$ for functions of t only:

$$\langle f, g \rangle_t = \int_{-T}^{T} \overline{f(t)} g(t) \ell^d \cosh^d(t/\ell) dt. \tag{4.81}$$

We have introduced T as a regulator which will be sent to ∞ once the SJ vacuum is computed. This procedure clearly breaks de Sitter invariance, but we shall see that when

the limit is taken, we obtain a state that is de Sitter invariant. The algebraic relations (4.44)–(4.47) can now be solved for in complete analogy with the previous section:

$$\begin{aligned} A_{Lj,L'j'} &= \cosh(\alpha_L)\delta_{LL'}\delta_{jj'} \\ B_{Lj,L'j'} &= \sinh(\alpha_L)e^{i\beta_L}\delta_{LL'}\delta_{jj'} \\ \lambda_{Lj} &= \sqrt{\langle y_L^E, y_L^E\rangle_t^2 - \left|\langle \overline{y}_L^E, y_L^E\rangle_t\right|^2}, \end{aligned} \tag{4.82}$$

where

$$\alpha_L = \frac{1}{2}\tanh^{-1}|r_L|, \qquad \beta_L = \arg(r_L) + \pi, \tag{4.83}$$

and

$$r_L := (-1)^L \frac{\langle \overline{y}_L^E, y_L^E\rangle_t}{\langle y_L^E, y_L^E\rangle_t}. \tag{4.84}$$

The SJ modefunctions (4.42) are then given by

$$u_{Lj}^{SJ}(x) = \cosh(\alpha_L)u_{Lj}^E(x) + \sinh(\alpha_L)e^{i\beta_L}\overline{u}_{Lj}^E(x). \tag{4.85}$$

In Section C.10.2 of Appendix C, we have computed r_L in the limit $T \to \infty$:

$$r_L = \begin{cases} \sin\dfrac{D}{2}\pi \operatorname{sech}\pi|\nu| & \text{if } m \geq m_*, \\ \sin\left[\left(\dfrac{D}{2} - \nu\right)\pi\right] & \text{if } 0 < m \leq m_*. \end{cases} \tag{4.86}$$

Regardless of the spacetime dimension or mass of the field, the SJ vacuum is invariant under the full de Sitter group in the global patch (see Section C.9.3 of Appendix C). As a result, it is always an α-vacuum. The case of even and odd spacetime dimensions look quite different, so we consider them in turn. For even D, (4.86) reduces to

$$r_L = \begin{cases} 0 & \text{if } m \geq m_*, \\ (-i)^{D-2}\sin\pi\nu & \text{if } 0 < m \leq m_*, \end{cases} \tag{4.87}$$

and for odd D we have

$$r_L = \begin{cases} (-i)^{D-1}\operatorname{sech}\pi|\nu| & \text{if } m \geq m_*, \\ (-i)^{D-1}\cos\pi\nu & \text{if } 0 < m \leq m_*. \end{cases} \tag{4.88}$$

When $m \geq m_*$ and D is even, $\alpha_L = 0$ and the SJ vacuum is equal to the Euclidean state. In odd spacetime dimensions and above the critical mass we have

$$\alpha_L = \tanh^{-1} e^{-\pi|\nu|} \qquad \text{and} \qquad \beta_L = -\frac{D+1}{2}\pi, \tag{4.89}$$

which means that the SJ vacuum is the *in*/*out*-vacuum. (The *in* and *out*-vacua are the same in odd dimensions [36, 89].) Below the critical mass, the Bogoliubov coefficients for even D are:

$$\alpha_L = \frac{1}{2}\tanh^{-1}|\sin\pi\nu| \quad \text{and} \quad \beta_L = \left[\frac{D}{2} + \theta(-\sin(\pi\nu))\right]\pi \tag{4.90}$$

and for odd D:

$$\alpha_L = \frac{1}{2}\tanh^{-1}|\cos\pi\nu| \quad \text{and} \quad \beta_L = \left[\frac{D+1}{2} + \theta(-\cos(\pi\nu))\right]\pi, \tag{4.91}$$

where $\theta(x)$ is the Heaviside step function. In even dimensions, we obtain $\alpha = 0$ whenever $|\nu|$ is an integer, in which case the SJ vacuum then corresponds to the Euclidean state. Whenever $|\nu|$ is a half-integer, the Bogoliubov coefficients diverge. The same holds in odd dimensions but with integer $\leftrightarrow$ half-integer. It is worth noting that the conformally coupled massless field corresponds in every spacetime dimension to the value $\nu = \frac{1}{2}$ [8] Let us take a closer look at the case of macroscopic physical spacetime, $D = 3+1$. As we have shown above, the SJ vacuum is the Euclidean state when $m \geq m_* = 3/2\ell$. Below the critical mass, the SJ vacuum is a de Sitter invariant α-vacuum, except when $m = m_{cc} = \sqrt{2}/\ell$, in which case the SJ prescription is not well-defined because the Bogoliubov coefficients diverge. The magnitude of the second Bogoliubov coefficient as a function of m is shown in Figure 4.1.

4.6 SJ Vacuum of a Radiation-Filled Cosmos

Quantum field theory on time-dependent backgrounds is of particular importance because the universe we live in is not static. The choice of vacuum in such cases is not at all trivial.

[8] A free massless conformally coupled scalar field ϕ obeys the equation

$$(\Box - \xi(D)R)\phi = 0, \tag{4.92}$$

where R is the Ricci scalar and

$$\xi(D) = \frac{D-2}{4(D-1)}. \tag{4.93}$$

If ϕ is a solution to (4.92) in spacetime $(M, g_{\mu\nu})$, then $\tilde{\phi} = \Omega^{1-\frac{D}{2}}\phi$ is a solution to $(\tilde{\Box} - \xi(D)\tilde{R})\tilde{\phi} = 0$ in spacetime $(M, \tilde{g}_{\mu\nu} = \Omega^2 g_{\mu\nu})$. In de Sitter space $R = D(D-1)/\ell^2$ is constant and (4.92) can be written as

$$(\Box - m_{eff}^2)\phi = 0, \qquad m_{eff} = \frac{\sqrt{D(D-2)}}{2\ell}. \tag{4.94}$$

It then follows that in this case $\nu = \ell\sqrt{m_*^2 - m_{eff}^2} = 1/2$.

Patch	Spacetime Dimension	SJ state for $\mathbf{m} \geq \mathbf{m}_*$	SJ state for $\mathbf{m} < \mathbf{m}_*$
Global	even	Euclidean	α-vacuum (4.90)
	odd	in = out	α-vacuum (4.91)
Poincaré	even	out	not defined
	odd	in = out	not defined

Table 4.1: The Sorkin-Johnston vacuum in the global and Poincaré patches of de Sitter space. Depending on the mass m of the field, the SJ vacuum corresponds to different α-vacua (the Euclidean, in- and out- vacua are all special cases of α-vacua and in odd spacetime dimensions the in- and out-vacua coincide). The critical mass that marks these transitions is $m_* = \frac{D-1}{2\ell}$, where D is the spacetime dimension and ℓ is the de Sitter radius.

For example, in a Friedmann-Lemaître-Robertson-Walker (FLRW) spacetime, one choice of instantaneous vacuum is obtained by minimizing the Hamiltonian at the given instant in time. This might seem like a natural generalization from static spacetimes, but as is well-known by now, it suffers from severe physical problems like infinite particle production [66].

In this section, we will work out the SJ vacuum state in a 3+1 dimensional spatially-flat FLRW spacetime for some specific cases. The metric reads

$$ds^2 = a(\eta)^2[-d\eta^2 + d\mathbf{x}^2], \tag{4.95}$$

where $a(\eta)$ is the scale factor and $\eta \in (0, \infty)$ is the conformal time. An orthornomal basis[9] $\{u_{\mathbf{k}}(x)\}$ for the Klein-Gordon solution space may be constructed as

$$u_{\mathbf{k}}(\eta, \mathbf{x}) = \frac{e^{i\mathbf{k}\cdot\mathbf{x}}}{\sqrt{(2\pi)^3}} \frac{g_k(\eta)}{a(\eta)}, \tag{4.96}$$

where $\mathbf{k}$ is the comoving Fourier wavenumber and $g_k(\eta)$ satisfies

$$\left(\partial_\eta^2 + k^2 + m^2 a^2 - \frac{a''}{a}\right) g_k(\eta) = 0 \tag{4.97}$$

$$g_k(\eta)\overline{g}_k'(\eta) - \overline{g}_k(\eta)g_k'(\eta) = i. \tag{4.98}$$

[9] See below (4.9) for the definition of "basis".

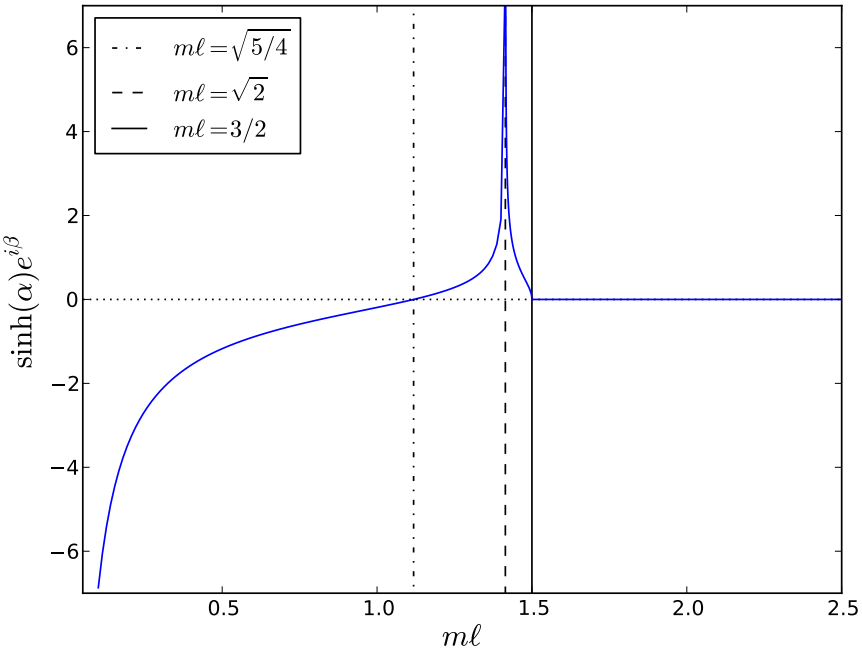


Figure 4.1: The Sorkin-Johnston (SJ) vacuum in the global patch of $3+1$ dimensional de Sitter space. The SJ modefunctions u^{SJ}_{Lj} are related to those of the Euclidean vacuum u^{E}_{Lj} by the Bogoliubov transformation $u^{SJ}_{Lj} = \cosh(\alpha)u^{E}_{Lj} + \sinh(\alpha)e^{i\beta}\overline{u}^{E}_{Lj}$, the second coefficient of which is plotted here. Depending on the product $m\ell$, where m is the mass of the field and ℓ is the de Sitter radius, the SJ vacuum corresponds to different α-vacua. For $m\ell \geq 3/2$ and $m\ell = \sqrt{5/4}$, it coincides with the Euclidean vacuum. The prescription fails for $m\ell = \sqrt{2}$.

Here $k = |\mathbf{k}|$ and $'$ denotes differentiation with respect to η. It may be verified that satisfying (4.97) and (4.98) is equivalent to satisfying the Klein-Gordan equation and the orthonromality relations (4.9), respectively. The L^2 inner products take the form

$$\langle u_{\mathbf{k}}, u_{\mathbf{k}'}\rangle = \langle g_k, g_k\rangle_\eta \delta^{(3)}(\mathbf{k}-\mathbf{k}') \tag{4.99}$$

$$\langle u_{\mathbf{k}}, \overline{u}_{\mathbf{k}'}\rangle = \langle g_k, \overline{g}_k\rangle_\eta \delta^{(3)}(\mathbf{k}+\mathbf{k}'), \tag{4.100}$$

where we have defined the inner product $\langle\cdot,\cdot\rangle_\eta$ for functions of η only:

$$\langle f, g\rangle_\eta \equiv \int_0^{\eta_{max}} \overline{f}(\eta)g(\eta)a^2(\eta)\,\mathrm{d}\eta. \tag{4.101}$$

We have introduced η_{max} as a regulator which will be sent to ∞ after the SJ vacuum has been computed. The algebraic relations (4.44)–(4.47) can now be solved for:

$$\begin{aligned} A_{\mathbf{kq}} &= \delta^{(3)}(\mathbf{k}-\mathbf{q})\cosh(\alpha_k) \\ B_{\mathbf{kq}} &= \delta^{(3)}(\mathbf{k}+\mathbf{q})\sinh(\alpha_k)e^{i\beta_k} \\ \lambda_{\mathbf{k}} &= \sqrt{\langle g_k, g_k\rangle_\eta^2 - |\langle \overline{g}_k, g_k\rangle_\eta|^2}, \end{aligned} \tag{4.102}$$

where

$$\alpha_k = \frac{1}{2}\tanh^{-1}|r_k|, \qquad \beta_k = \arg(r_k) + \pi, \tag{4.103}$$

and

$$r_k := \frac{\langle \overline{g}_k, g_k\rangle_\eta}{\langle g_k, g_k\rangle_\eta}. \tag{4.104}$$

The SJ modefunctions (4.42) are then given by

$$u_{\mathbf{k}}^{SJ}(x) = \cosh(\alpha_k)u_{\mathbf{k}}(x) + \sinh(\alpha_k)e^{i\beta_k}\overline{u}_{-\mathbf{k}}(x). \tag{4.105}$$

We shall now specialize to the case of a radiation-filled universe, where the scale factor takes the form

$$a(\eta) = \frac{\eta}{\eta_0}, \tag{4.106}$$

where η_0 is conformal time at which $a = 1$.

4.6.1 Massless Field

We begin by considering a massless field, i.e. we set $m = 0$. It may be verified that in this case

$$g_k(\eta) = \frac{1}{\sqrt{2k}}e^{-ik\eta}, \tag{4.107}$$

satisfies both (4.97) and (4.98). Also:

$$\lim_{\eta_{max}\to\infty}|r_k| = \lim_{\eta_{max}\to\infty}\frac{|\int_0^{\eta_{max}} e^{-2ik\eta}\eta^2\,\mathrm{d}\eta|}{\int_0^{\eta_{max}}\eta^2\,\mathrm{d}\eta} = 0. \tag{4.108}$$

It then follows from (4.103) and (4.105) that

$$u_{\mathbf{k}}^{SJ}(\eta,\mathbf{x}) = \frac{\eta_0/\eta}{\sqrt{(2\pi)^3 2k}}e^{-i(k\eta-\mathbf{k}\cdot\mathbf{x})}. \tag{4.109}$$

These are the so-called adiabatic-vacuum modefunctions, for which an exact solution exists in the case of a massless scalar field in a radiation dominated cosmos [30].

4.6.2 Massive Field

For a massive field in the radiation era, (4.97) can still be solved analytically. Let

$$z = i\tilde{m}^2\eta^2, \qquad \tilde{m} = \sqrt{m/\eta_0}, \tag{4.110}$$

and define the function G through the relation

$$g_k(\eta) \equiv \frac{G(i\tilde{m}^2\eta^2)}{\sqrt{2m\eta/\eta_0}}. \tag{4.111}$$

With these definitions, (4.97) is equivalent to

$$\frac{\partial^2 G}{\partial z^2} + \big(-\frac{1}{4} - \frac{i(k/\tilde{m})^2}{4z} + \frac{3}{16z^2}\big)G = 0. \tag{4.112}$$

This equation has two independent solutions $W_{\frac{-ik^2}{4\tilde{m}^2},\frac{1}{4}}(z)$ and $W_{\frac{ik^2}{4\tilde{m}^2},\frac{1}{4}}(-z)$, called Whittaker functions (see e.g. Chapter 13 of [154]). In our case, these two functions are complex conjugates of one another. Using the properties of Whittaker functions (see e.g. 13.14.30 of [154]), it can be shown that $g_k(\eta)$ satisfies the Wronskian condition (4.98) with the normalization:

$$g_k(\eta) = \frac{e^{\frac{-\pi k^2}{8\tilde{m}^2}}}{\sqrt{2m\eta/\eta_0}} W_{\frac{-ik^2}{4\tilde{m}^2},\frac{1}{4}}(i\tilde{m}^2\eta^2). \tag{4.113}$$

Let us now turn to the inner products $\langle \overline{g}_k, g_k\rangle_\eta$ and $\langle g_k, g_k\rangle_\eta$, which are important for computing the SJ vacuum. These inner products do not suffer any divergences in the $\eta \to 0$ limit, since $|g_k(\eta)|$ is constant for small η (see e.g. 13.14.18 of [154]). Divergences arise for large η, though. In this regime, $W_{\lambda,\mu}(z) \longrightarrow e^{-\frac{1}{2}z}z^\lambda$ (see e.g. 13.14.21 of [154]) and plugging this into (4.113) we find:

$$g_k(\eta) \xrightarrow{\eta\to\infty} \frac{1}{\sqrt{2m\eta/\eta_0}} e^{-\frac{i}{2}[\tilde{m}^2\eta^2 + \frac{k^2}{2\tilde{m}^2}\ln(\tilde{m}^2\eta^2)]}. \tag{4.114}$$

It can then be verified that $\langle g_k, g_k\rangle_\eta$ diverges quadratically in η_{max}, while $\langle \overline{g}_k, g_k\rangle_\eta$ oscillates$\sim e^{-i\tilde{m}^2\eta_{max}^2}$. Therefore, in the $\eta_{\max} \to \infty$ limit:

$$\lim_{\eta_{max}\to\infty} |r_k| = \lim_{\eta_{max}\to\infty} \frac{\langle \overline{g}_k, g_k\rangle_\eta}{\langle g_k, g_k\rangle_\eta} = 0. \tag{4.115}$$

It then follows from (4.103) and (4.105) that the SJ modefunctions are

$$u_{\mathbf{k}}^{SJ}(\eta, \mathbf{x}) = \frac{e^{i\mathbf{k}\cdot\mathbf{x}} e^{\frac{-\pi k^2}{8\tilde{m}^2}}}{\sqrt{2m(2\pi)^3}(\eta/\eta_0)^{3/2}} W_{\frac{-ik^2}{4\tilde{m}^2},\frac{1}{4}}(i\tilde{m}^2\eta^2). \tag{4.116}$$

It is reasonable to ask whether this vacuum state could potentially have observable effects. One way of approaching this problem is to calculate the response rate of a comoving detector, such as the Unruh-Dewitt detector, when the field is in the SJ vacuum state. Even more ambitiously, one could (in principle) derive the SJ vacuum for a general scale-factor $a(\eta)$ (with reasonable boundary conditions), and study its back-reaction on the underlying geometry via the renormalized stress-energy-momentum tensor. Unfortunately, these computations are fairly cumbersome and a detailed treatment of them is beyond the scope of this Chapter. In order to gain *some* intuition, however, we will compute $\langle SJ|\widehat{\rho}|SJ\rangle \equiv -\langle SJ|\widehat{T}^0{}_0|SJ\rangle = a^{-2}\langle SJ|\widehat{T}_{00}|SJ\rangle$, where the expectation value of the (unrenormalized) energy momentum tensor $\widehat{T}_{\mu\nu}$ takes the form [30]:

$$\langle SJ|\widehat{T}_{\mu\nu}|SJ\rangle = \int T_{\mu\nu}\left[u^{SJ}_{\mathbf{k}}, \overline{u}^{SJ}_{\mathbf{k}}\right] \mathrm{d}^3\mathbf{k}, \tag{4.117}$$

where

$$T_{\mu\nu}[\phi,\psi] = \nabla_\mu\phi\nabla_\nu\psi - \frac{1}{2}g_{\mu\nu}\left[\nabla^\alpha\phi\nabla_\alpha\psi + m^2\phi\psi\right]. \tag{4.118}$$

It can be checked that

$$T_{00}\left[u^{SJ}_{\mathbf{k}}, \overline{u}^{SJ}_{\mathbf{k}}\right] = \frac{1}{2}\left|\partial_\eta u^{SJ}_{\mathbf{k}}\right|^2 + \frac{1}{2}\delta^{ij}\partial_i u^{SJ}_{\mathbf{k}}\partial_j\overline{u}^{SJ}_{\mathbf{k}} + \frac{1}{2}m^2a^2\left|u^{SJ}_{\mathbf{k}}\right|^2. \tag{4.119}$$

Substituting (4.116) for $u^{SJ}_{\mathbf{k}}$ in (4.119) we find

$$\langle SJ|\widehat{\rho}(\eta)|SJ\rangle = \int \frac{\mathrm{d}^3\mathbf{p}(\eta)}{(2\pi)^3}\sqrt{p(\eta)^2+m^2}(n_{SJ}+\frac{1}{2}), \tag{4.120}$$

where

$$n_{SJ} = \frac{1}{4}\frac{e^{-i\pi\lambda}}{\sqrt{\frac{p^2}{m^2}+1}}\left\{\left(\frac{p^2}{m^2}+1\right)\left|W_{\lambda,\frac{1}{4}}(z)\right|^2 - \frac{1}{z^2}\left|2z\frac{dW_{\lambda,\frac{1}{4}}(z)}{dz} - \frac{3}{2}W_{\lambda,\frac{1}{4}}(z)\right|^2\right\} - \frac{1}{2}. \tag{4.121}$$

The variables used above are defined as follows:

$$\mathbf{p}(\eta) = \frac{\mathbf{k}}{a(\eta)} = \frac{\mathbf{k}}{\eta/\eta_0}, \quad H(\eta) = \frac{a'}{a^2} = \frac{\eta_0}{\eta^2}, \qquad z = i\tilde{m}^2\eta^2 = i\frac{m}{H}, \tag{4.122}$$

$$\lambda = \frac{-ik^2}{4\tilde{m}^2} = \frac{-ip^2}{4mH}, \qquad k = |\mathbf{k}|, \qquad p(\eta) = |\mathbf{p}(\eta)|. \tag{4.123}$$

Note that $H(\eta)$ is the Hubble parameter and $\mathbf{p}(\eta)$ is the physical momentum of the Fourier mode with comoving wavenumber $\mathbf{k}$ at conformal time η.

For a thermal bath of relativistic bosons at temperature T, the energy density takes the form $\rho = \int \frac{\mathrm{d}^3\mathbf{p}}{(2\pi)^3}\, E(p)\, n_{BE}(p)$, where $E(p) = \sqrt{p^2+m^2}$ and $n_{BE} = \frac{1}{e^{E/T}-1}$ is the Bose-Einstein distribution. This relation can be inverted to get $T = \frac{E}{\ln\left(1+\frac{1}{n_{BE}}\right)}$. In order to see how "close to thermal" our state is, we similarly define the "effective temperature" of a mode as

$$T_{SJ}(p) = \frac{\sqrt{p^2+m^2}}{\ln\left(1+\frac{1}{n_{SJ}}\right)}. \tag{4.124}$$

The "more constant" T_{SJ} is as a function of p, the closer the distribution n_{SJ} is to being thermal. Here we define n_{SJ} to include only excitations *above* the state $|GS\rangle$ that minimizes the Hamiltonian at a particular instant of time, and for which $\langle GS|\widehat{\rho}|GS\rangle = \int \frac{\mathrm{d}^3\vec{p}}{(2\pi)^3}\frac{1}{2}\sqrt{p^2+m^2}$. Then $\langle SJ|\widehat{\rho}|SJ\rangle - \langle GS|\widehat{\rho}|GS\rangle = \int \frac{\mathrm{d}^3\mathbf{p}}{(2\pi)^3}\sqrt{p^2+m^2}\, n_{SJ}$.

Figure 4.2 shows the behaviour of T_{SJ} for different ratios of $\frac{m}{H}$ and $\frac{p}{m}$. It is evident that the long wavelength modes are in fact at a constant effective temperature. For example, in the regime where $m \ll H$ and $p \ll \sqrt{mH}$, the Whittaker function has a simple asymptotic expansion

$$W_{\lambda,\frac{1}{4}}(z) \longrightarrow \frac{\sqrt{\pi}}{\Gamma(\frac{3}{4}-\lambda)} z^{\frac{1}{4}} - \frac{2\sqrt{\pi}}{\Gamma(\frac{1}{4}-\lambda)} z^{3/2} + \mathcal{O}(z^{5/4}), \tag{4.125}$$

using which it can be shown that

$$T_{SJ}(p) \sim \frac{\pi}{4\,|\Gamma(3/4)|^2}\frac{H^{3/2}}{\sqrt{m}}. \tag{4.126}$$

This result suggests that there are correlations on super-horizon scales. It is noteworthy that these correlations have appeared without the help of any previous epoch of accelerated expansion. Potentially, they could therefore open up a new perspective on the question of primordial fluctuations and on the related puzzle sometimes called "horizon problem".

4.7 Causality and the SJ Vacuum

Like other vacuua, the SJ vacuum is defined globally, and it depends on both the causal past *and* future of the spacetime. Consider for example a spacetime which is first static, then expands for a short time, then goes back to being static again. [10] In light of the inherent time-reversal symmetry of the conditions defining the SJ vacuum-state, it is clear that this

[10] Of course, such a spacetime is not necessarily a solution to the Einstein equations.

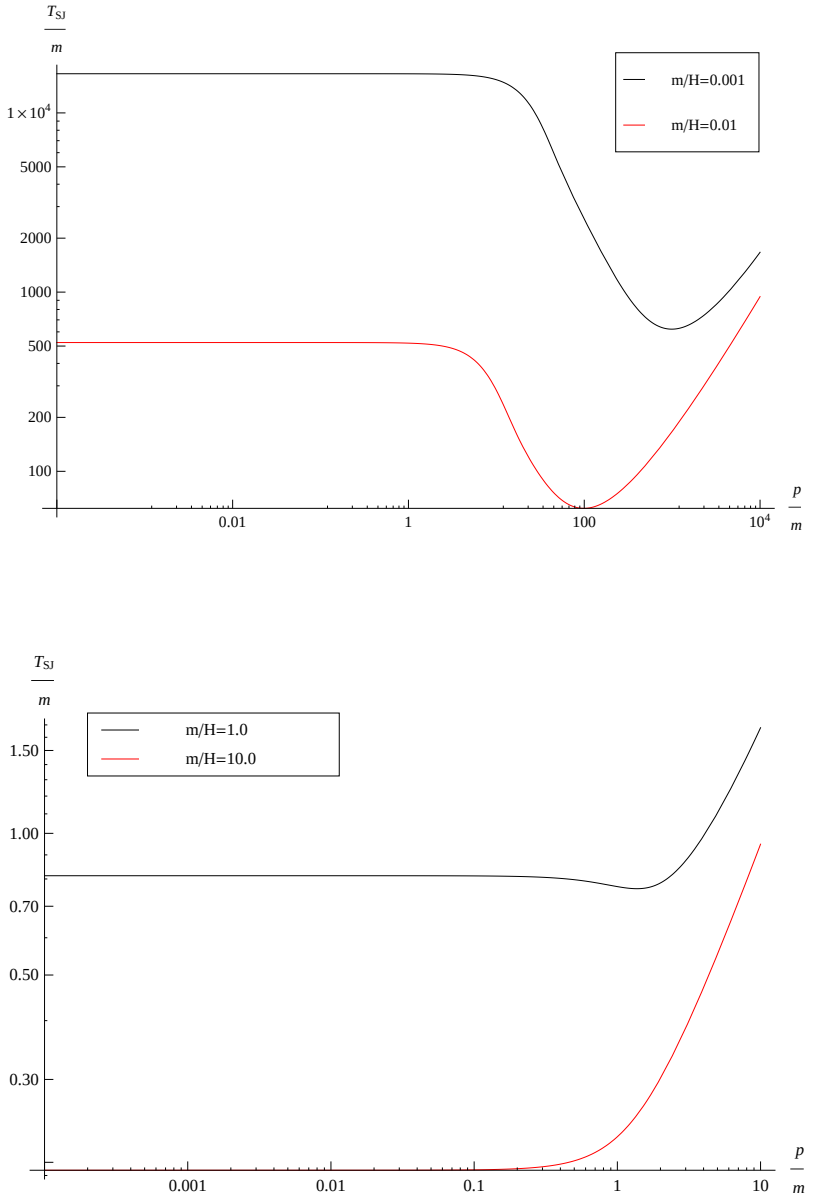


Figure 4.2: Behaviour of $T_{SJ}(p)$ (see (4.124)) for different ratios of $\frac{m}{H}$ and $\frac{p}{H}$.

state can agree neither with the early-time vacuum (the state of minimum energy at that time), nor with the late-time vacuum. Rather, it must strike some sort of "compromise" between them.

In the present section we will illustrate this behavior with a simple example, but before doing so, we would like to dwell for a bit on the question of whether one should interpret this type of dependence on the future as a failure of causality. By construction, our definition of the vacuum depends on the full spacetime geometry. That it thereby fails to be what John Bell called "locally causal" is no surprise because, as is well understood by now, any reasonable quantum state must incorporate nonlocal correlations and entanglement. Certainly the Minkowski vacuum does so. But does this type of nonlocality also imply genuine acausality?

The prior question that begs for an answer here is what is meant by acausality in the context of quantum field theory, considering also that quantum field theory must ultimately find its place within a theory of full quantum gravity. If we remain within the "operationalist" framework of external agents, "measurements" and state-vector collapse, then causality (in the sense of relativistic causality) reduces to the impossibility of superluminal signalling. In this sense, there is no question of acausality as long as the twin conditions of spacelike commutativity and hyperbolicity of the field equations are respected, which by construction they are in the field theory we are working with in this paper.[11] On the other hand, if we try to adopt a more "objective" framework which dispenses with external agents, then we seem to be left without any clear definition of relativistic causality at all. That is, we lack an *intrinsic* criterion which could decide whether or not physical influences are propagating outside the light cone or "into the past". But without such a criterion, the meaning of relativistic causality in general is called into question.

A further observation also seems relevant here, even if it does not turn out to be decisive. Namely, the assumption we have made of a fixed, non-dynamical spacetime is already "anticausal" in a certain sense. In a full quantum gravity theory the future geometry must evolve together with, and in mutual dependence on the future matter-field. Hence, any attempt to specify the geometry in advance amounts to imposing a future boundary condition on the combined system of metric plus scalar field. Given this, it would not be surprising if a correct semiclassical treatment of the scalar were also to involve some degree of "dependence on the future".

The specific model we will consider is a 1+1 dimensional FLRW universe with metric $ds^2 = C(\eta)(-d\eta^2 + dx^2)$, where $C(\eta) = A + B\tanh(\rho\eta)$. In the infinite past $C(\eta) \to A - B$

[11] The theory of [84] retains spacelike commutativity, but hyperbolicity becomes, together with the notion of field-equation itself, approximate at best.

and in the infinite future $C(\eta) \to A + B$. It is known that there are modefunctions $u_k^{in}(\eta, x)$ (which form an orthonormal basis of the KG solution space) that behave like positive frequency Minkowski-space modes in the remote past $\eta \to -\infty$: (see Section 3.4 of [30])

$$\begin{aligned} u_k^{in}(\eta, x) &= \frac{1}{\sqrt{4\pi\omega_k^{in}}} e^{ikx - i\omega_k^+\eta - (i\omega_k^-/\rho)\ln[2\cosh(\rho\eta)]} \\ &\times F\left(1 + i\omega_k^-/\rho, i\omega_k^-/\rho; 1 - i\omega_k^{in}/\rho; \frac{1+\tanh(\rho\eta)}{2}\right) \\ &\xrightarrow{\eta\to-\infty} \frac{1}{\sqrt{4\pi\omega_k^{in}}} e^{ikx - i\omega_k^{in}\eta}, \end{aligned} \tag{4.127}$$

where F denotes the hypergeometric function ${}_2F_1$ and

$$\begin{aligned} \omega_k^{in} &= \sqrt{k^2 + m^2(A - B)} \\ \omega_k^{out} &= \sqrt{k^2 + m^2(A + B)} \\ \omega_k^{\pm} &= \frac{1}{2}(\omega_k^{out} \pm \omega_k^{in}). \end{aligned} \tag{4.128}$$

Similarly, there are orthonormal modes $u_k^{out}(\eta, x)$ that behave like the positive frequency Minkowski-space modes in the remote future $\eta \to \infty$:

$$\begin{aligned} u_k^{out}(\eta, x) &= \frac{1}{\sqrt{4\pi\omega_k^{out}}} e^{ikx - i\omega_k^+\eta - (i\omega_k^-/\rho)\ln[2\cosh(\rho\eta)]} \\ &\times F\left(1 + i\omega_k^-/\rho, i\omega_k^-/\rho; 1 + i\omega_k^{out}/\rho; \frac{1-\tanh(\rho\eta)}{2}\right) \\ &\xrightarrow{\eta\to\infty} \frac{1}{\sqrt{4\pi\omega_k^{out}}} e^{ikx - i\omega_k^{out}\eta}. \end{aligned} \tag{4.129}$$

The *in* and *out* modes are related to eachother by the following Bogolubov transformation

$$u_k^{in}(\eta, x) = \gamma_k u_k^{out}(\eta, x) + \mu_k \bar{u}_{-k}^{out}, \tag{4.130}$$

where

$$\gamma_k = \sqrt{\frac{\omega_k^{out}}{\omega_k^{in}}} \frac{\Gamma(1 - i\omega_k^{in}/\rho)\Gamma(-i\omega_k^{out}/\rho)}{\Gamma(1 - i\omega_k^+/\rho)\Gamma(-i\omega_k^+/\rho)} \tag{4.131}$$

$$\mu_k = \sqrt{\frac{\omega_k^{out}}{\omega_k^{in}}}) \frac{\Gamma(1 - i\omega_k^{in}/\rho)\Gamma(-i\omega_k^{out}/\rho)}{\Gamma(1 + i\omega_k^-/\rho)\Gamma(i\omega_k^-/\rho)}. \tag{4.132}$$

The modes u_k^{in} and u_k^{out} define vacuum states at early and late times, respectively. If the system is at first ($\eta \to -\infty$) in the in-vacuum state, i.e. the no particle state, it will have $|\mu_k|^2$ particles of momentum k with respect to the out-vacuum after the expansion ($\eta \to \infty$). The SJ vacuum has a different nature, simply because the vacuum state in the $\eta \to -\infty$ region depends on what happens in the infinite future (and vice-versa).

Let us now turn to computing the SJ modefunctions. Define $g_k^{in}(\eta)$ through

$$u_k^{in}(\eta, x) = \frac{e^{ikx}}{\sqrt{2\pi}} g_k^{in}(\eta). \tag{4.133}$$

The L^2 inner products take the form

$$\langle u_k^{in}, u_{k'}^{in} \rangle = \langle g_k^{in}, g_k^{in} \rangle_\eta \delta(k - k') \tag{4.134}$$

$$\langle u_k^{in}, \overline{u}_{k'}^{in} \rangle = \langle g_k^{in}, \overline{g}_k^{in} \rangle_\eta \delta(k + k'), \tag{4.135}$$

where we have defined the inner product $\langle \cdot, \cdot \rangle_\eta$ for functions of η only:

$$\langle f, g \rangle_\eta \equiv \int_{-\eta_{max}}^{\eta_{max}} \overline{f}(\eta) g(\eta) C(\eta) \, \mathrm{d}\eta. \tag{4.136}$$

We have introduced η_{max} as a regulator which will be sent to ∞ after the SJ vacuum has been computed. The algebraic relations (4.44)–(4.47) can now be solved for:

$$\begin{aligned} A_{kq} &= \delta(k - q) \cosh(\alpha_k) \\ B_{kq} &= \delta(k + q) \sinh(\alpha_k) e^{i\beta_k} \\ \lambda_k &= \sqrt{\langle g_k^{in}, g_k^{in} \rangle_\eta^2 - \left| \langle \overline{g}_k^{in}, g_k^{in} \rangle_\eta \right|^2}, \end{aligned} \tag{4.137}$$

where

$$\alpha_k = \frac{1}{2} \tanh^{-1} |r_k|, \qquad \beta_k = \arg(r_k) + \pi, \tag{4.138}$$

and

$$r_k := \frac{\langle \overline{g}_k^{in}, g_k^{in} \rangle_\eta}{\langle g_k^{in}, g_k^{in} \rangle_\eta}. \tag{4.139}$$

The SJ modefunctions (4.42) are then given by

$$u_k^{SJ}(\eta, x) = \cosh(\alpha_k) u_k^{in}(\eta, x) + \sinh(\alpha_k) e^{i\beta_k} \overline{u}_{-k}^{in}(\eta, x). \tag{4.140}$$

The asymptotic behaviour of $g_k^{in}(\eta)$ is given by:

$$g_k^{in}(\eta) \xrightarrow{\eta\to-\infty} \frac{1}{\sqrt{2\omega_k^{in}}} e^{-i\omega_k^{in}\eta} \tag{4.141}$$

$$g_k^{in}(\eta) \xrightarrow{\eta\to\infty} \frac{1}{\sqrt{2\omega_k^{out}}} (\gamma_k e^{-i\omega_k^{out}\eta} + \mu_k e^{i\omega_k^{out}\eta}), \tag{4.142}$$

using which it can be shown that

$$\lim_{\eta_{max}\to\infty} r_k = \frac{2\gamma_k\mu_k}{|\gamma_k|^2 + |\mu_k|^2 + \frac{\omega_{out}}{\omega_{in}}\frac{A-B}{A+B}}. \tag{4.143}$$

Fig. 4.3 shows the difference between the SJ and "in" vacuua for a specific set of parameters and frequencies. As one would expect, this deviation is only significant for low-frequency modes, which are more sensitive to the rate of expansion ρ.

4.8 The SJ vacuum on a causal set

While the methods of canonical quantisation are not available on a causal set, the SJ formalism admits a natural discrete formulation [84, 85]. In fact, on a causal set, the formalism is free of many of the technicalities that arise in the continuum and accordingly simpler to outline. For the massive scalar field in $D = 1 + 1$ dimensional flat space, it has been shown numerically that the mean of the discrete SJ two-point function approximates that of the continuum Minkowski vacuum state [84] in the appropriate "continuum limit". In this section, we will carry out a similar analysis in the case of two-dimensional de Sitter space.

4.8.1 Causal sets and the discrete SJ vacuum

Let us briefly review the necessary background on causal sets. [12] A *causal set* $(\mathcal{C}, \preceq)$ is a set $\mathcal{C}$ with a partial order relation $\preceq$ which is

(i) reflexive : $x \preceq x$

(ii) acyclic : $x \preceq y \preceq x \implies x = y$

(iii) transitive : $x \preceq y \preceq z \implies x \preceq z$

(iv) locally finite : $|I(x, y)| < \infty$

[12] Although causal sets are introduced in Chapter 2, we summarize the necessary background here for the convenience of the reader.

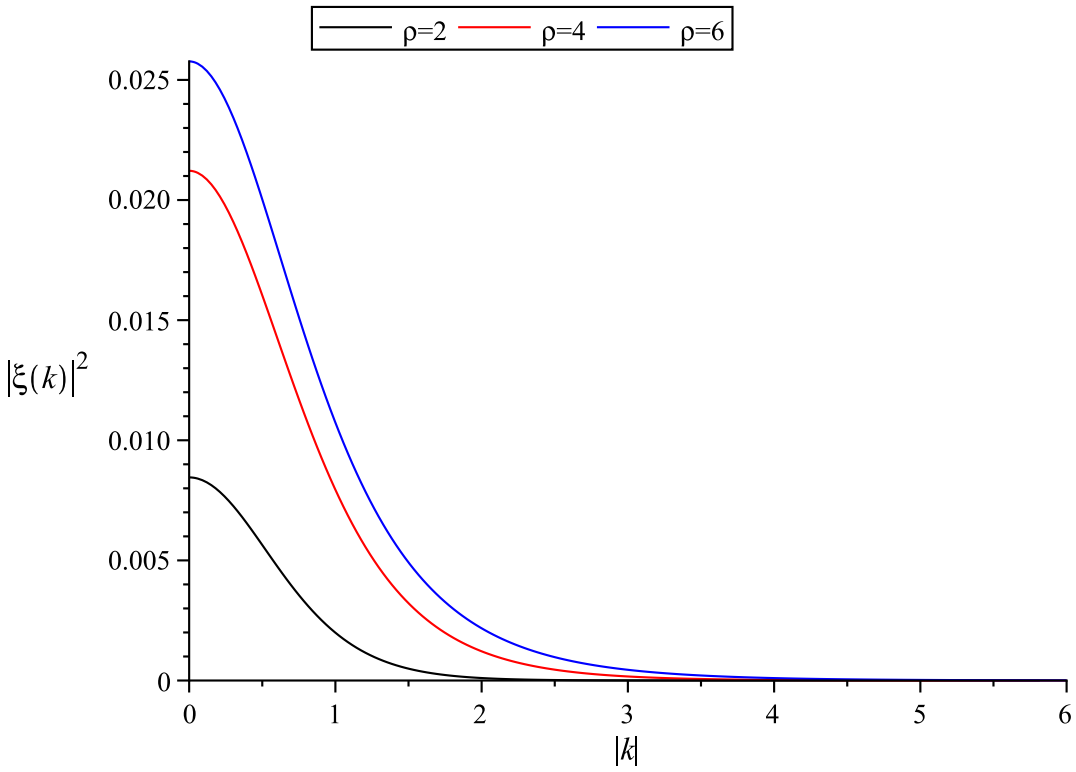


Figure 4.3: The deviation between the SJ and "in" vacuua, as measured by $\sinh(\alpha_k)^2$ (see (4.140)). The parameters used here are $A = 2.0$, $B = 1.0$, and $m = 1.0$.

for all $x, y, z \in \mathcal{C}$, where $I(x, y) := \{z \in \mathcal{C} \,|\, x \preceq z \preceq y\}$ is the (inclusive) order interval between two elements $x, y \in \mathcal{C}$ and $|\cdot|$ denotes cardinality. We write $x \prec y$ when $x \preceq y$ and $x \neq y$.

A causal set is fully encoded in its adjacency or *causal matrix* $\mathbf{C}$, defined as the $|\mathcal{C}| \times |\mathcal{C}|$-matrix with entries

$$C_{ij} := \begin{cases} 1 & \text{if } \nu_i \prec \nu_j \\ 0 & \text{otherwise,} \end{cases} \tag{4.144}$$

for $\nu_i, \nu_j \in \mathcal{C}$, where $i, j \in \{1, 2, \ldots, |\mathcal{C}|\}$ are indices that label the elements in $\mathcal{C}$.

A *sprinkling* is a procedure for generating a causal set $(\mathcal{C}_M, \preceq)$ given a continuum spacetime region $(M, g_{\mu\nu})$. Points are placed at random in M using a Poisson process with "density" $\rho := |\mathcal{C}_M|/V_M$, where V_M denotes the spacetime volume of M, in such a way that the expected number of points in any region of spacetime volume V is ρV. This

generates a causal set whose elements are the sprinkled points, and whose partial order relation can be "read off" from that of the underlying spacetime. Such a causal set provides a discretisation of $(M, g_{\mu\nu})$ which, unlike a regular lattice, is statistically Lorentz invariant (see Chapter 2). In order to reduce the computational cost of the simulations described below, we keep the geodesic distance information of $(M, g_{\mu\nu})$ for all pairs of causally related elements in $\mathcal{C}_M$, meaning that for all such pairs $\nu_i, \nu_j \in \mathcal{C}_M$ with coordinates x_i, x_j in M, we record the values $d_{ij} := d(x_i, x_j)$, where $d(x_i, x_j)$ denotes geodesic distance in $(M, g_{\mu\nu})$. While this information is not explicitly contained in $(\mathcal{C}_M, \preceq)$, it can be recovered by known algorithms [79].

Let $(\mathcal{C}_M, \preceq)$ be an N-element causal set generated by a sprinkling into a $1+1$ dimensional spacetime region $(M, g_{\mu\nu})$. To define the SJ vacuum on the causal set, we start with the "discrete retarded propagator", which in two dimensions can be defined for a scalar field of mass m on $\mathcal{C}_M$ as [85]

$$\mathbf{R} = \frac{1}{2}\mathbf{C}\left(\mathbf{1} + \frac{m^2}{2\rho}\mathbf{C}\right)^{-1}, \tag{4.145}$$

where $\mathbf{C}$ denotes the causal matrix defined in (4.144). It has been shown that if $(M, g_{\mu\nu})$ is a causal diamond[13] in two-dimensional Minkowski space, the mean of R_{ij} as a function of the geodesic distance d_{ij} is in agreement with the known continuum retarded propagator $G_R(x, y)$ for high sprinkling density and mass range $0 < m \ll \sqrt{\rho}$ [85]. We have obtained similar evidence for the case where $(M, g_{\mu\nu})$ is a causal diamond in de Sitter space (see below). Given a retarded propagator, we define the *discrete Pauli-Jordan function* $\mathbf{\Delta}$ on $\mathcal{C}_M$ in analogy with its continuum counterpart:

$$\mathbf{\Delta} := \mathbf{R} - \mathbf{R}^T, \tag{4.146}$$

where T denotes the matrix transpose. It is then natural to define the *discrete SJ two-point function* as the positive spectral projection of $i\mathbf{\Delta}$:

$$\mathbf{W}_{SJ} := \mathrm{Pos}(i\mathbf{\Delta}). \tag{4.147}$$

Since $i\mathbf{\Delta}$ is now a finite Hermitian matrix (at least for causal sets of finite cardinality), its positive part is completely well-defined and specifies $\mathbf{W}_{SJ}$ uniquely. We also define the discrete analogue of the Hadamard function

$$\mathbf{H}_{SJ} := 2\mathrm{Re}\mathbf{W}_{SJ}, \tag{4.148}$$

[13] A causal diamond is the intersection of the interior of the past lightcone of a point q with the interior of the future light cone of a point p that lies to the causal past of q.

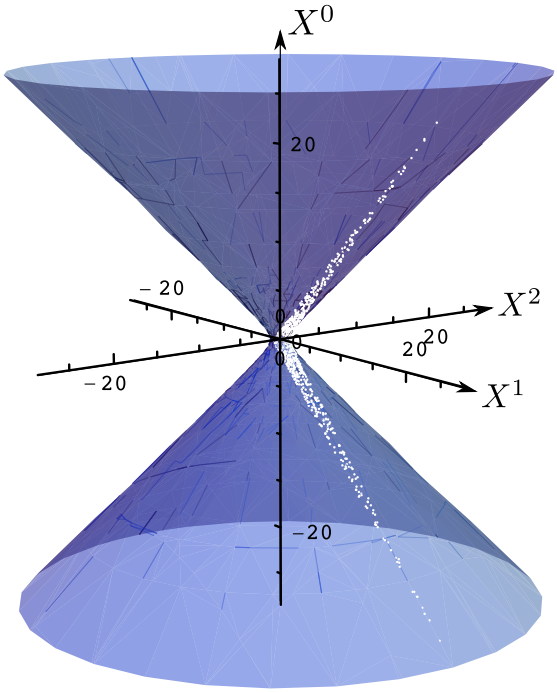


Figure 4.4: An $N = 1010$ element sprinkling with density $\rho = 76\ell^{-2}$ into a causal diamond of length $\tau = 8\ell$ in two-dimensional de Sitter space, visualised in the embedding three-dimensional Minkowski space (see Section C.8 of Appendix C). The de Sitter radius has been set to $\ell = 1$.

such that

$$\mathbf{W}_{SJ} = \frac{1}{2}\mathbf{H}_{SJ} + \frac{i}{2}\mathbf{\Delta}. \tag{4.149}$$

To compare the discrete SJ two-point function with the known propagators in continuum de Sitter space, we evaluate it on a causal set that is obtained by a sprinkling into a causal interval (diamond) in $1+1$ dimensional continuum de Sitter space. For any two points $x \prec y$, the causal interval between them is the intersection of the future of x with the past of y. In de Sitter space, the spacetime volume V of the causal interval between two timelike points depends only on their Lorentzian distance τ: $V = 4\ell^2 \ln(\cosh(\tau\ell^{-1}/2))$. We shall refer to a causal diamond of *length* τ as one whose volume is given by the formula above.

4.8.2 Simulation results

In order to compare causal set results with those of the continuum, we have computed the retarded propagator $\mathbf{R}$, and subsequently the discrete Hadamard function $\mathbf{H}_{SJ}$, on an $N = 1010$ element sprinkling into a causal diamond of length $\tau = 8\ell$ in $1+1$ dimensional de Sitter space (implying $\rho \simeq 76\ell^{-2}$). The sprinkling is shown in Figure 4.4, where we have set $\ell = 1$.

Figure 4.5 shows values of the retarded propagator $\mathbf{R}_{ij}$ for all pairs of related events $(\nu_i, \nu_j) \in \mathcal{C}_M$, plotted as a function their geodesic distance d_{ij}. There is good agreement between the mean of $\mathbf{R}$ and the continuum retarded Green function, which further validates the proposal (4.145). At large $\tau \gg \ell$, we see a slight deviation between the mean of the causal set data and the continuum retarded Green function. This discrepancy can be associated with edge-effects due to the finite size of the causal diamond: pairs of points separated by a geodesic distance comparable to the size of the diamond will feel the boundaries of the spacetime region (the effect of spacetime boundaries has been addressed in more detail in [6]). Figure 4.6 shows the discrete SJ Hamadard function $\mathbf{H}_{SJ}$, computed for both timelike and spacelike pairs of events. Since we have no expression for the continuum SJ vacuum in the causal diamond itself, we cannot compare $\mathbf{H}_{SJ}$ with its *exact* continuum counterpart. However, the expectation would be that the discrete SJ two-point function approximates that of a de Sitter invariant vacuum in the centre of the diamond (where the boundaries of the diamond are felt the least). Indeed, Figure 4.6 shows a very good agreement between the mean of $\mathbf{H}_{SJ}$ and the Hadamard function associated with the Euclidean vacuum ($\alpha = 0$). At large $\tau \gg \ell$, the boundary effects become noticeable again. To highlight the particular agreement with the Euclidean ($\alpha = \beta = 0$) Hadamard function, we have also plotted in Figure 4.6 the Hadamard function of two other α-vacua with $(\alpha, \beta) = (1, 0)$ and $(\alpha, \beta) = (0.1, 0)$. Note that $H_{\alpha,\beta}(x, y)$ is more sensitive to variations in α for spacelike separated arguments because of the extra antipodal singularity at $d(x, y) = \pi\ell$, i.e. $Z(x, y) = -1$, present in every α-vacuum except the Euclidean one (see Section C.9.3 of Appendix C). For instance, for the range of parameters we have probed in our simulations, including those of Figure 4.6, the function $H_{0.1,0}$ as a function of the geodesic distance can be distinguished from the Euclidean Hadamard function for spacelike separated arguments, whereas it lies on top of the Euclidean Hadamard function for timelike separated arguments (and has thus been omitted from the timelike plot). With the parameters probed in our simulations, we cannot discriminate between the in/out and the Euclidean vacua, since they are very "close" unless $m \sim m_*$. Indeed, for the values presented here we have $\alpha_{in} = \alpha_{out} = \mathcal{O}(10^{-4})$. Discriminating between the in/out and Euclidean vacua is more demanding computationally. A full treatment of this matter will

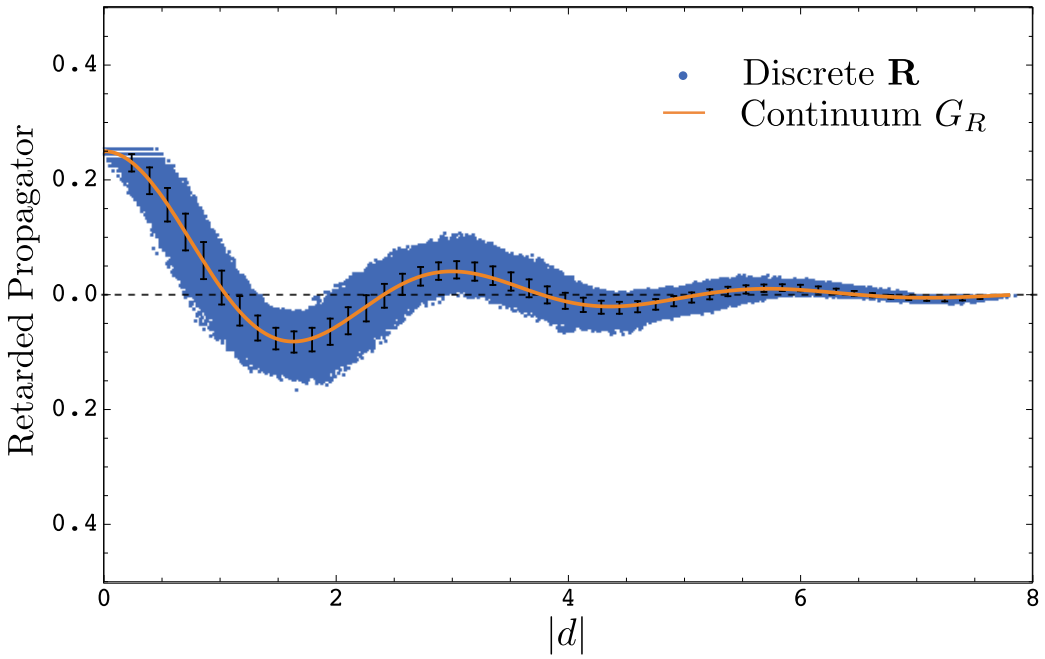


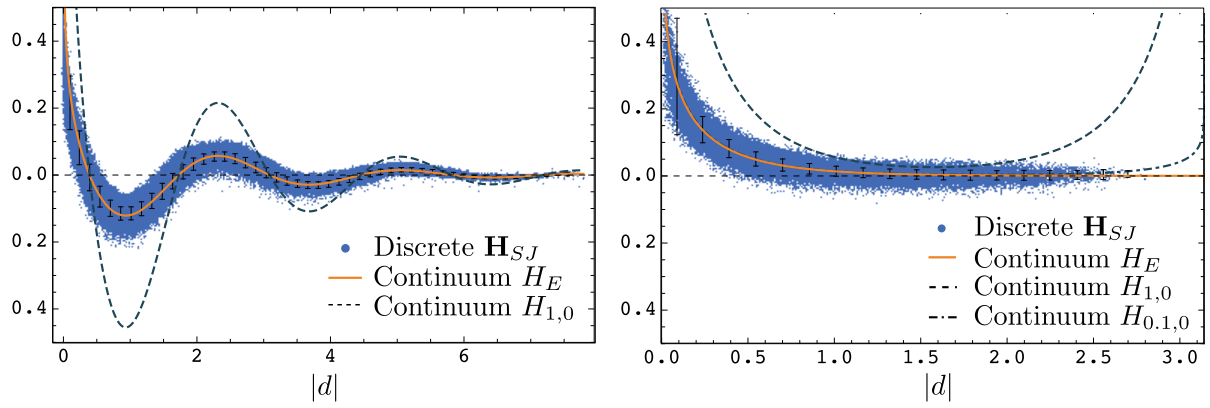
Figure 4.5: The retarded propagator $\mathbf{R}$, computed on a causal set obtained via a $N = 1010$ sprinkling into a causal diamond of length $\tau = 8\ell$ in $1+1$ dimensional de Sitter space. The mass of the field is $m = 2.36\ell^{-1}$ and the de Sitter radius ℓ is set to unity. The geodesic distance $|d|$ between the two arguments of the function is plotted on the horizontal axis. The error bars show the standard deviation about the mean of $\mathbf{R}$ for binned values of $|d|$. The continuum propagator G_R is shown with the thick orange line.

require more extensive simulations.

4.9 Conclusions and Discussions

We have defined a distinguished vacuum for a free quantum field in a globally hyperbolic region of an arbitrarily curved spacetime. This Sorkin–Johnston (SJ) state is well-defined for all compact regions and for a large class of noncompact ones. We have shown that for static spacetimes, it coincides with the usual ground state.

We have determined the SJ vacuum for massive free scalar field in $D = d + 1$ dimensional de Sitter space. In those cases where the prescription is well-defined, we find that the SJ vacuum always corresponds to one of the de Sitter-invariant α-vacua. This is re-



(a) Timelike (b) Spacelike

Figure 4.6: The Hadamard function $\mathbf{H}_{SJ}$ on a causal set obtained through an $N = 1010$ sprinkling of a causal diamond of length $\tau = 8\ell$ in $1+1$ dimensional de Sitter space. The mass of the field is taken to be $m = 2.36\ell^{-1}$, and the de Sitter radius ℓ is set to unity. The geodesic distance $|d|$ between the two arguments of the function is plotted on the horizontal axis for (a) timelike and (b) spacelike separated points. The error bars show the standard deviation about the mean of $\mathbf{H}_{SJ}$ for binned values of $|d|$. $H_{\alpha,\beta}(x,y)$ refers to the Hadamard function of the α-vacua (see Section C.9.3 of Appendix C). The function $H_{0.1,0}$ has been omitted in (a), since it is indistinguishable from the Euclidean function H_E.

assuring, because a covariant approach should give rise to a vacuum state that respects the symmetries of the underlying spacetime. We find that the SJ vacuum depends on *(i)* whether the mass of the field is in the complementary or principal series (i.e. below or above the critical value $(D-1)/(2\ell)$, where ℓ is the de Sitter radius), *(ii)* whether it is evaluated on the complete de Sitter manifold or its Poincaré half-space, and *(iii)* whether the spacetime dimension is even or odd. For a field of mass $m < (D-1)/(2\ell)$ on the Poincaré patch, the SJ prescription cannot be applied to the entire spacetime, but only a bounded globally hyperbolic subregion of it.It would be interesting to investigate whether a physical account can be given for the failure of the procedure in this particular case (an example of another vacuum prescription which fails for light masses is the instantaneous ground state of the Hamiltonian, particularly in the global patch [65]). Here it is worth noting that the complementary and principal series also exhibit different behaviours in the case of interacting theories [95, 82]. For instance, quantum-corrected fields whose bare mass belong to the principal series, unlike the complementary series, decay faster than the free KG field in past/future infinity. This has important consequences for objects such as the S-matrix for QFTs on global de Sitter space [96, 38, 37].

We find that the SJ vacuum in de Sitter space does not in general correspond to the Bunch-Davies or Euclidean state, and as a result is not always Hadamard [12]. (See [62] for another instance where the SJ state is not Hadamard.) The main advantage of Hadamard states is that for such states it is known how to construct physically relevant expectation values, such as those of the stress-energy tensor, on arbitrarily curved spacetimes [39, 76, 77]. Although it has not been proven that this cannot be done for α-vacua, it is known that standard prescriptions such as point-splitting and normal ordering fail [40]. Exploring the consequences of these facts for the SJ formalism would certainly be of interest.

Using the discrete SJ formalism on a causal set, we have determined the SJ state on a sprinkling of a causal diamond in $1+1$ dimensional de Sitter space. As part of our analysis, we have found evidence that the "discrete retarded propagator" proposed in [83] agrees well with the continuum retarded propagator in de Sitter space. Our simulation also shows that the mean of the discrete SJ two-point function is consistent with that of an α-vacuum and in particular with that of the Euclidean vacuum in the centre of the diamond (away from the edges) for a field of mass $m \ll \sqrt{\rho}$. This is encouraging, since the QFT defined on causal sets by the SJ formalism seems to reproduce what one would expect: a state that respects the spacetime isometries in the appropriate "continuum limit". It would be interesting to carry out further simulations to determine, with more statistical significance, which continuum state is best approximated by the discrete SJ state. This might be particularly illuminating when $m < 1/(2\ell)$, since the procedure in the continuum becomes pathological in the Poincaré patch in that case.

It is natural to wonder whether the SJ formalism could have phenomenological implications in relation to cosmology. We would like to raise two potential difficulties in this direction. Firstly, because of its non-local nature, it is not clear what portion of spacetime one should use to compute the SJ vacuum. For instance, should one consider the behaviour of late-time cosmology to determine the SJ vacuum for the early universe? Secondly, if we ultimately aim to make a prediction for the primary anisotropy spectrum of the Cosmic Microwave Background, how are we to interpret the scalar field whose vacuum state we compute using the SJ formalism? Does it also involve scalar metric perturbations? If so, one is likely to run into trouble with gauge-invariance, because the SJ formalism is not invariant under field re-definitions.

To gain *some* intuition in this regard, however, we have computed the SJ state for a scalar field of mass $m \geq 0$ in a radiation-filled, spatially flat, homogeneous and isotropic cosmos. In that connection we also computed an "effective temperature" that can be defined for the super-horizon modes of the massive field. The correlations found thereby could open up a new perspective on the question of primordial fluctuations and the so-called "horizon problem".

A peculiar aspect of the SJ prescription is its temporal non-locality. We demonstrated this feature by the example of a spacetime which sandwiches a region with curvature in-between flat initial and final regions, but we did not explore its phenomenological implications any further.

Chapter 5

Phenomenology of *Gravitational Aether* as a Solution to the *Old* Cosmological Constant Problem

5.1 Introduction

The discovery of recent acceleration of cosmic expansion was one of the most surprising findings in modern cosmology [130, 112]. The standard cosmological model (also known as the concordance model) drives this expansion with a cosmological constant Λ, which appears in the Einstein field equations

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G_N T_{\mu\nu}. \tag{5.1}$$

The value of Λ has been measured with remarkable percision:

$$\Lambda = 3.57 \times 10^{-52} \Omega_\Lambda h^2 m^{-2} = 9.32 \times 10^{-122} \Omega_\Lambda h^2 {l_p}^{-2}, \tag{5.2}$$

where $l_p = \sqrt{\hbar G_N / c^3}$ is the Planck length and recent measurements from the *Planck* satellite give [118]

$$\Omega_\Lambda = 0.692 \pm 0.01, \qquad h = 0.6780 \pm 0.0077.^{1} \tag{5.3}$$

The cosmological constant is consistent with (nearly) all current cosmological observations. There is, however, something unsettling about its extremely small (in Planck units) and yet nonzero value, especially if we are to accept it as a fundamental constant.

[1] See Table 5 of [118]. We have used the $Planck + WP + highL + BAO$ values here.

This becomes even more puzzling when we consider the vacuum energy of quantum fields. We expect from general covariance that the vacuum expectation value of the energy momentum tensor should take the form $\langle T_{\mu\nu}\rangle = -\langle\rho_V\rangle g_{\mu\nu} + \cdots$. Therefore, the zero point energy of quantum fields acts just like a cosmological constant in linear order. Of course, $\langle\rho_V\rangle$ is formally divergent. In practical calculations, one needs to use a regulator which preserves general covariance, such as dimensional regularization, to even arrive at $\langle T_{\mu\nu}\rangle = -\langle\rho_V\rangle g_{\mu\nu} + \cdots$. Naively speaking though, if we believe general relativity is the correct description of gravity up to an energy scale E_Λ, we would *expect* $\langle\rho_V\rangle \sim E_\Lambda^4$. Taking the cut-off E_Λ to be on the the order of Planck energy, we find the famous discrepancy of some 120 orders of magnitude between $\langle\rho_V\rangle$ and $\rho_\Lambda \equiv \Lambda/(8\pi G_N)$ (see e.g. [44] and [161]). Even if we cut off the theory at energy scales well below the Planck scale, say ~ 1 GeV, a huge discrepancy remains. This is the so-called cosmological constant "problem": how is it that the vacuum energy density of quantum fields, each of which is expected to contribute $\sim E_\Lambda^4$, cancel out to such remarkable precision (e.g. 120 decimal places), to produce the observed value ρ_Λ? Often, this puzzle is split into two parts (largely for historical reasons): (i) finding a mechanism which drives $\langle\rho_V\rangle$ to zero or decouples it from gravity (the *old* cosmological constant "problem") and (ii) explaining why the cosmological constant has the specific nonzero value that it does (the *new* cosmological constant "problem").

We prefer using the word *puzzle* as opposed to *problem* because there is no logical inconsistency here. Strictly speaking, quantum field theory does not predict anything since $\langle\rho_V\rangle$ is formally divergent. The argument which is typically put forth for the "realty" of zero-point vacuum energy is the Casimir effect (see e.g. [27] and references therein). Casimir showed that quantum fluctuations in the space between two conducting plates which are at a distance d from one another produce a force per unit area of $\hbar c\pi^2/240d^4$ (see e.g. [30]). To obtain this finite value value, however, it is necessary to subtract the zero-point energy of empty space from that between the plates (after regularizing both values, of course). If anything, this calculation suggests that only the *change* in zero-point vacuum energy is physical, and not the infinite zero-point energy of empty space itself. Furthermore, the context in which this puzzle is being posed is that of semi-classical gravity, where the dynamics are governed by a semiclassical form of the Einstein equations: $G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G_N\langle T_{\mu\nu}\rangle$. It may very well be that this is the wrong framework to use, and that a full theory of quantum gravity, where both matter *and* gravitational degrees of freedom are "quantized", is needed.

In this chapter, we will take the viewpoint that the QFT expectation of the nature and magnitude of zero-point quantum fluctuations is reasonable. Once this is assumed, there is no choice but to abandon the idea that vacuum energy should gravitate. This, however, requires modifying Einstein's theory of gravity, in which *all* sources of energy gravitate.

Attempts in this direction have been proposed in the context of massive gravity [59], or braneworld models of extra dimensions such as cascading gravity [49, 51], or supersymmetric large extra dimensions (*e.g.,* [42]). However, efforts to find explicit cosmological solutions that de-gravitate vacuum have proven difficult (*e.g.,* [11, 50]). In [3], Afshordi proposed a novel approach in which quantum vacuum fluctuations (of linear order in the metric) are decoupled from gravity through the introduction of a fluid called the *Gravitational Aether* (GA). In this Chapter, we will explore in detail the predictions of this theory in the context of cosmology and solar-system tests of gravity.

5.2 Gravitational Aether

In [3], Afshordi proposed a modification to the right hand side of the Einstein field equations as follows:[2]

$$(8\pi\tilde{G})^{-1}G_{\mu\nu} = T_{\mu\nu} - \frac{1}{4}T^{\alpha}{}_{\alpha}g_{\mu\nu} + \tilde{T}_{\mu\nu}, \tag{5.4}$$

where $T_{\mu\nu}$ and $\tilde{T}_{\mu\nu}$ denote the energy-momentum tensor of ordinary matter and the aether fluid, respectively. The only constant of the theory is $\tilde{G}$, which we will later show is related to Newton's gravitational constant G_N. The point of subtracting the trace part of $T_{\mu\nu}$ is that any contribution of the form $T^{vac}_{\mu\nu} = \Lambda g_{\mu\nu}$, where Λ is constant, drops out of the right hand side of (5.4):

$$T^{vac}_{\mu\nu} - \frac{1}{4}g^{\alpha\beta}T^{vac}_{\alpha\beta}g_{\mu\nu} = 0. \tag{5.5}$$

As with ordinary general relativity, it is assumed that the energy-momentum tensor of ordinary matter is conserved

$$\nabla^{\mu}T_{\mu\nu} = 0. \tag{5.6}$$

Choosing to "de-gravitate" quantum vacuum fluctuations in this fashion, and further assuming the conservation of matter, makes it *necessary* to add $\tilde{T}_{\mu\nu}$ to the right hand side of (5.4). To see why this should be, let us apply ∇^{μ} to both sides of (5.4). Noting that $\nabla^{\mu}G_{\mu\nu} = 0$ (by the Bianchi identities) and $\nabla^{\mu}T_{\mu\nu} = 0$ (by definition), we find

$$\nabla^{\mu}\tilde{T}_{\mu\nu} = \frac{1}{4}\nabla_{\nu}T, \tag{5.7}$$

where $T = T^{\alpha}{}_{\alpha}$. Had we taken $\tilde{T}_{\mu\nu} = 0$, we would have ended up with the constraint $T = const$, which is of course nonsense. In principle, $\tilde{T}_{\mu\nu}$ has 10 independent components,

[2] Our metric signature is $(-,+,+,+)$. Throughout the paper we set the speed of light c=1.

since it is a symmetric rank-2 tensor. However, only 4 of these would be constrained through (5.7). In order not to add any new degrees of freedom,[3] Afshordi assumed $\tilde{T}_{\mu\nu}$ to be an incompressible perfect fluid (which only has 4 independent components):

$$\tilde{T}_{\mu\nu} = \tilde{p}(\tilde{u}_\mu \tilde{u}_\nu + g_{\mu\nu}), \qquad \tilde{u}^\mu \tilde{u}_\mu = -1, \tag{5.8}$$

where $\tilde{p}$ and $\tilde{u}^\mu$ are the pressure and four-velocity of aether, respectively. Given this form of the aether fluid, (5.7) can be cast as a relativistic version of the continuity and Euler equations: (see Section D.1 of Appendix D)

$$\tilde{p}\nabla_\mu \tilde{u}^\mu = -\frac{1}{4}\tilde{u}^\mu \nabla_\mu T, \tag{5.9}$$

$$\tilde{p}\tilde{u}^\mu \nabla_\mu \tilde{u}_\nu = -\tilde{\nabla}^\perp_\nu (\tilde{p} - T/4), \tag{5.10}$$

where

$$\tilde{\nabla}^\perp_\nu = \nabla_\nu + \tilde{u}_\nu \tilde{u}^\sigma \nabla_\sigma. \tag{5.11}$$

Are $\tilde{p}$ and $\tilde{u}^\mu$ uniquely determined by (5.9) and (5.10) once T is specified? This is an important question, especially in relation to solving the initial value problem of the field equations (5.4). Let us imagine we have specified initial data for the metric and the matter sector on a spacelike hypersurface. Would this set of initial data be sufficient to obtain a unique evolution of the metric using the gravitational aether field equations (5.4)? This depends on whether or not $\tilde{p}$ and $\tilde{u}^\mu$ are uniquely determined from (5.9)–(5.10). As we will see throughout this Chapter, there are many situations in which the aether degrees of freedom are fully constrained by the matter sector. However, there are situations in which this is not the case, as the following example will demonstrate. Let us choose a coordinate system which is co-moving with the aether fluid, i.e. $\tilde{u}^i = 0$, and consider only vacuum solutions (for which $T = 0$). Also, let the initial data for the metic be $g_{\mu\nu} = \eta_{\mu\nu}$ and $\partial_\alpha g_{\mu\nu} = 0$. It then follows from $\tilde{u}_\mu \tilde{u}^\mu = -1$ that $\tilde{u}^0 = 1$ and $\partial_0 \tilde{u}^0 = 0$ on the initial hypersurface. Therefore, $\nabla_\mu \tilde{u}^\mu = 0$ and (5.9) is satisfied trivially. Furthermore, the only nonreduntant information contained in (5.10) is $\nabla_i \tilde{p} = 0$, which implies $\tilde{p} = \tilde{p}_0$ on the initial hypersurface ($\tilde{p}_0$ being constant). Additional input is required to fix $\tilde{p}_0$ as well as how it evolves from one hypersurface to the next. Such peculiarities, although certainly worth investigating, will not concern us for the rest of this Chapter.

Let us now turn our attention to the only constant of the theory, $\tilde{G}$, whose value can be fixed by looking at the Newtonian limit of (5.4). In Section D.1 of Appendix

[3] This is one motivation for choosing the aether fluid to be an incompressible fluid. Another is the behaviour of a particular class of scalar field theories which have an infinite speed of sound but are causal [8, 7], the reason being they only couple to observables as constraints.

D, we show that when matter is a perfect fluid with a constant equation of state, i.e. $T_{\mu\nu} = (\rho + p)u_\mu u_\nu + p g_{\mu\nu}$ where $p = w\rho$ and w is constant, the gravitational aether field equations reduce to

$$(8\pi\tilde{G})^{-1} G_{\mu\nu} = \frac{3}{4}(1+w) T_{\mu\nu}. \tag{5.12}$$

In the Newtonian limit, $T_{\mu\nu}$ is well approximated by a perfect fluid with $w = 0$. Therefore, in order to recover Einstein's field equations in this limit, we must have

$$\tilde{G} = \frac{4}{3} G_N. \tag{5.13}$$

Before delving into the detailed predictions of the gravitational aether theory, a few generic remarks are in order. First, let us emphasize that GA is designed to solve the *old* cosmological constant problem, by way of degravitating quantum fluctuations. The present-day acceleration of cosmic expansion, or the so-called "new" cosmological constant problem, is not addressed by this proposal. Within the context of the same theory, however, it is argued in [128, 4] that quantum gravity effects in the presence of astrophysical black holes could lead to accelerated expansion of the universe. This is an intriguing proposal, but we will not pursue it any further in this Chapter. Whenever needed, we will use a standard cosmological constant to model the late-time acceleration of cosmic expansion. Secondly, it is not known whether the gravitational aether field equations can be derived from an action principle. This is likely to create a setback in any quantum-mechanical setting, where having an action principle is (almost always) necessary. At the classical level, however, having an equation of motion is all that is needed, as is the case with the Navier-Stokes fluid equations.

5.3 Cosmological Constraints on Gravitational Aether

As mentioned in the previous Section, when matter can be approximated by a perfect fluid with a constant equation of state w, the solutions to the gravitational aether theory become identical to those of general relativity, except with a renormalized gravitational constant:[4]

$$G_N \to G_{\text{eff}} = (1+w) G_N. \tag{5.14}$$

In other words, the gravitational coupling is not a constant anymore, and can change significantly for fluids with relativistic pressure. Not surprisingly, for the vacuum equation of state $w = -1$, $G_{\text{eff}} = 0$, which implies that vacuum does not gravitate.

[4] This follows from combining (5.13) and (5.12).

In particular, in the case of Friedmann-Lemaître-Robertson-Walker (FLRW) cosmology where the perfect fluid approximation is valid, this theory predicts that the effective gravitational constant is different in the matter and radiation eras:

$$\frac{G_N}{G_R} \equiv \frac{G_{\rm eff}(w=0)}{G_{\rm eff}(w=1/3)} = \frac{3}{4}. \tag{5.15}$$

This is the first cosmological prediction of this theory: *radiation energy gravitates more strongly than non-relativistic matter*. We will examine the consequences of this prediction for big bang nucleosynthesis in Section 5.3.1, and the evolution of cosmological perturbations in Section 5.3.2.

Before doing so, let us note that this result can also be thought of as a prediction for the effective number of neutrinos, which is often what is quoted in the literature. The expansion history in the radiation era depends on the product $G\rho_{\rm rad}$, and is constrained through different observational probes. The constraints are often described as the bound on the effective number of neutrinos $N_\nu^{\rm eff}$, which quantifies the total radiation density $\rho_{\rm rad}$. We can translate the constraints to those on $G_{\rm eff}$ by requiring

$$G_{\rm eff}\rho_{\rm rad}(N_\nu^{\rm eff}=3) = G_N\rho_{\rm rad}(N_\nu^{\rm eff}=3+\Delta N_\nu). \tag{5.16}$$

For example, at about $T \simeq 1$ MeV, the relativistic species that contribute most significantly to $\rho_{\rm rad}$ are photons and neutrinos, the latter being decoupled from thermal equilibrium by this time. The effective "temperature" of neutrinos is given by $T_\nu = (4/11)^{1/3}T_\gamma$, where T_γ is the temperature of photons (see e.g. [162]). The energy density of a gas of relativistic particles in thermal equilibrium at temperature T is given by $n \times \frac{\pi^2}{30}T^4$ for bosons, and $n \times \frac{7}{8}\frac{\pi^2}{30}T^4$ for fermions, where n is the number of degrees of freedom (e.g. number of helicity states, spin states, anti particles, etc). It then follows that

$$\begin{aligned}\rho_{\rm rad} &= n_\gamma \frac{\pi^2}{30}T_\gamma^4 + n_\nu \frac{7}{8}\frac{\pi^2}{30}T_\nu^4 \\ &= \frac{\pi^2}{30}T_\gamma^4\left[2 + N_\nu^{\rm eff}\times 2\times\frac{7}{8}\left(\frac{4}{11}\right)^{4/3}\right],\end{aligned} \tag{5.17}$$

where we have used $n_\gamma = 2$ and $n_\nu = N_\nu^{\rm eff}\times 2$. Combining (5.17), (5.16) and $G_{\rm eff} = \frac{4}{3}G_N$ we find:

$$\Delta N_\nu = 2.5. \tag{5.18}$$

It is important to emphasize that this correspondence is only meaningful for a homogeneous and radiation-dominated universe.

5.3.1 Big Bang Nucleosynthesis

It has been known that the increase of the gravitational constant at around $T = \mathcal{O}(1)$ MeV epoch induces earlier freezeout of the neutron to proton ratio because of a speed-up effect of the increased cosmic expansion. This raises the abundance of ^{4}He sensitively and deuterium (D) mildly, and can lower the abundance of ^{7}Be through $^7\mathrm{Be}\ (n,p)^7\mathrm{Li}(p,\alpha)^4\mathrm{He}$ (Note that the second p is thermal proton). For a relatively large baryon to photon ratio $\eta \gtrsim 3\times10^{-10}$, the dominant mode to produce ^{7}Li is the electron capture of ^{7}Be at a later epoch through $^7\mathrm{Be} + e^- \to {}^7\mathrm{Li} + \nu_e$. Therefore, the decrease of ^{7}Be makes the fitting better because so far any observational ^{7}Li abundances have been so low that they could not have agreed with theoretical prediction in Standard BBN (SBBN) at better than 3 σ [48].

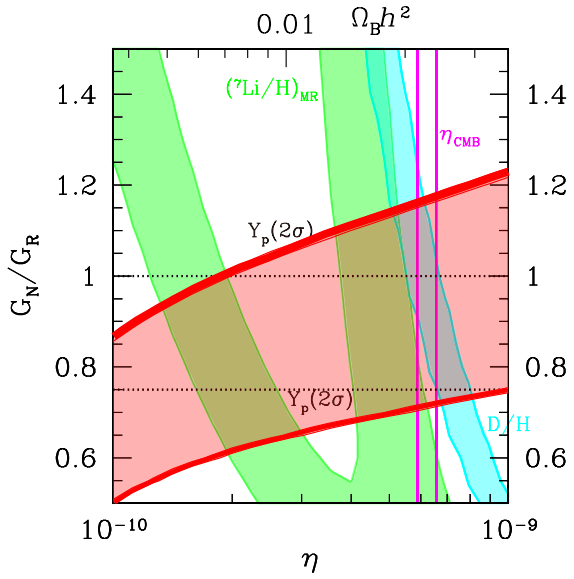


Figure 5.1: Allowed regions with 2 σ lines for D/H, Y_p and ^{7}Li/H are shown. The upper and lower horizontal dashed lines indicate GR and gravitational aether predictions, respectively. The thickness of Y_p means the uncertainty in measurements of neutron lifetime [108, 139]. We can translate the vertical axis into ΔN_ν by using a relation $G_N/G_R \simeq 1/(1{+}0.135\Delta N_\nu)$.

In this study, we adopt the following observational light element abundances as primordial values: the mass fraction of ^{4}He, $\mathrm{Y}_p = 0.2561 \pm 0.0108$ (stat) [22], the deuterium to hydrogen ratio, $\mathrm{D/H}=(2.80\pm0.20)\times10^{-5}$ [113], and the ^{7}Li to hydrogen ratio

$\mathrm{Log}_{10}(^7\mathrm{Li/H}) = -9.63 \pm 0.06$ [101] [5]. Theoretical errors come from experimental uncertainties in cross sections [141, 47, 48] and neutron lifetime [108, 139].

Comparing theoretical prediction with the observational light element aubndances provides a constraint on G_N/G_R. Fig. 6.1 shows the results of a comprehensive analysis for ^{4}He, D, and ^{7}Li. We also plotted a band for baryon to photon ratio, η which was reported from CMB observations by WMAP 7-year, $\eta = (6.225 \pm 0.170) \times 10^{-10}$ in case of $G_N/G_R = 1$ [87]. Then we can see that every light element agrees with the Gravitational Aether theory within 2 σ. It is notable that ^{7}Li in this theory fits the data better than that in SBBN. Performing χ^2 fitting for three elements with three degree of freedom, however, the model is allowed only at 99.7% (3 σ) in total.

However, notice that the main discrepancy is with deuterium abundance observed in quasar absorption lines, which suffer from an unexplained scatter. Moreover, deuterium could be depleted by absorption onto dust grains that would make its primordial value closer to our prediction (see [125] for a discussion).

5.3.2 Cosmological Fluctuations

In this Section we derive the equations which govern the behaviour of small cosmological fluctuations in the gravitational aether theory. These are then used to test the theory against cosmological observations, specifically the observed power spectrum of the cosmic microwave background.

We start by noting that the gravitational aether field equations can be derived as a special case of a more general theory, which we call the Generalized Gravitational Aether (GGA) theory. The GGA field equations are

$$(8\pi)^{-1} G_{\mu\nu} = G_R T_{\mu\nu} - (G_R - G_N) T^{\alpha}{}_{\alpha} g_{\mu\nu} + 4(G_R - G_N)\tilde{T}_{\mu\nu}, \tag{5.19}$$

where G_R is a free constant and all other quantities are defined as before. When $G_R = G_N$, the GGA field equations reduce to those of GR. When $G_R = \frac{4}{3}G_N$, they reduce to those of the gravitational aether theory. Therefore, this theory interpolates nicely between GR and GA through the parameter G_R, which can be constrained using cosmological observations. As was the case for gravitational aether, the Bianchi identities and the conservation of ordinary matter imply (5.7).

In analysing the evolution of cosmological perturbations, it is necessary to adequately model the late time expansion of the universe at the homogeneous level. However, the

[5] See also $\mathrm{Log}_{10}(^7\mathrm{Li/H}) = -9.90 \pm 0.09$ [35] for the lower value which makes fitting worse.

gravitational aether theory (at least as presented here) does not explain the late time acceleration of the cosmic expansion.[6] Therefore, to carry out our analysis of cosmological perturbations, we will add a cosmological constant Λ to the left hand side of the GGA field equations:

$$(8\pi)^{-1}(G_{\mu\nu} + \Lambda g_{\mu\nu}) = G_R T_{\mu\nu} - (G_R - G_N)T^{\alpha}{}_{\alpha} g_{\mu\nu} + 4(G_R - G_N)\tilde{T}_{\mu\nu}. \tag{5.20}$$

We are now in a position to tackle cosmological perturbations. Let us start by describing the dynamics at the background level. We will assume a spatially-flat, homogeneous, and isotropic FLRW universe. The metric reads

$$ds^2 = a(\tau)^2(-d\tau^2 + d\mathbf{x}^2), \tag{5.21}$$

where τ is the conformal time. (We have listed all relevant geometric objects derivable from this metric in Section D.2 of Appendix D.) As is shown in Section D.1 of Appendix D, when the matter energy-momentum tensor $T_{\mu\nu} = \sum_i T^{(i)}_{\mu\nu}$ consists of a collection of co-moving perfect fluids $T^{(i)}_{\mu\nu} = (\rho_i + p_i)u_\mu u_\nu + p_i g_{\mu\nu}$, each of which is separately conserved, i.e. $\nabla^\mu T^{(i)}_{\mu\nu} = 0 \ \forall \ i$, and has a constant equation of state, i.e. $p_i = w_i \rho_i$ where w_i is constant, then

$$\tilde{u}^\mu = u^\mu, \qquad \tilde{p} = \frac{1}{4}\sum_i (1 + w_i)(3w_i - 1)\rho_i \tag{5.22}$$

solves equation (5.7), or equivalently (5.9) and (5.10). Substituting $\tilde{p}$ and $\tilde{u}^\mu$ in (5.20), it may be verified that the GGA field equations simplify to

$$(8\pi)^{-1}(G_{\mu\nu} + \Lambda g_{\mu\nu}) = G_R \sum_i \left\{1 + (1 - G_N/G_R)(3w_i - 1)\right\} T^{(i)}_{\mu\nu}. \tag{5.23}$$

In the case of homogeneous cosmology, the main constituents of the universe are matter $\overline{T}^{(m)}_{\mu\nu}$ and radiation $\overline{T}^{(r)}_{\mu\nu}$, which are well-approximated by perfect fluids with equations of state $w = 0$ and $w = 1/3$, respectively. Furthermore, at the background level, matter and radiation are co-moving and separately conserved. It then follows from (5.23) that the GGA field equations reduce to

$$(8\pi)^{-1}(G_{\mu\nu} + \Lambda g_{\mu\nu}) = G_R \overline{T}^{(r)}_{\mu\nu} + G_N \overline{T}^{(m)}_{\mu\nu}. \tag{5.24}$$

If we let

$$\overline{T}^{(m)}_{\mu\nu} = \bar{\rho}_m \bar{u}_\mu \bar{u}_\nu, \qquad \overline{T}^{(r)}_{\mu\nu} = \frac{4}{3}\bar{\rho}_r \bar{u}_\mu \bar{u}_\nu + \frac{1}{3}\bar{\rho}_r \bar{g}_{\mu\nu}, \qquad \bar{u}_\mu = (-a(\tau), 0, 0, 0), \tag{5.25}$$

[6] See remarks at the end of Section 5.2.

where $\overline{g}_{\mu\nu} = a(\tau)^2\eta_{\mu\nu}$ is the FLRW metric, the two independent GGA field equations become

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G_N a^2}{3}\left\{\frac{G_R}{G_N}\bar{\rho}_r + \bar{\rho}_m + \rho_\Lambda\right\}, \tag{5.26}$$

$$\left(\frac{\dot{a}}{a}\right)^2 - 2\frac{\ddot{a}}{a} = \frac{8\pi G_N a^2}{3}\left\{\frac{G_R}{G_N}\bar{\rho}_r - 3\rho_\Lambda\right\}, \tag{5.27}$$

where dot denotes differentiation with respect to conformal time and

$$\rho_\Lambda = \frac{\Lambda}{8\pi G_N}. \tag{5.28}$$

Also, the conservation equations $\nabla^\mu \overline{T}^{(m)}_{\mu\nu} = 0$ and $\nabla^\mu \overline{T}^{(r)}_{\mu\nu} = 0$ are equivalent to

$$\dot{\bar{\rho}}_m + 3\frac{\dot{a}}{a}\bar{\rho}_m = 0, \qquad \dot{\bar{\rho}}_r + 4\frac{\dot{a}}{a}\bar{\rho}_r = 0. \tag{5.29}$$

Equations (5.26)–(5.29) then completely specify the dynamics at the background level.

We now turn to the perturbations. Our main focus will be solving the equations that couple aether to matter, i.e. (5.9) and (5.10), to first order in perturbation theory. Interestingly, we will be able to do this without having to choose any particular gauge. In what follows, b, dm, m, and r stand for baryon, dark matter, matter, and radiation, respectively. Also, all barred quantities are unperturbed. We define the perturbed metric as

$$g_{\mu\nu} = \overline{g}_{\mu\nu} + h_{\mu\nu}, \tag{5.30}$$

where $h_{\mu\nu}$ is the small perturbation about the background metric $\overline{g}_{\mu\nu}$. We will take the content of the universe to consist of baryons, cold dark matter, and radiation:

$$T_{\mu\nu} = T^b_{\mu\nu} + T^{dm}_{\mu\nu} + T^r_{\mu\nu}, \tag{5.31}$$

where

$$T^b_{\mu\nu} = \overline{T}^b_{\mu\nu} + \delta T^b_{\mu\nu}, \qquad T^{dm}_{\mu\nu} = \overline{T}^{dm}_{\mu\nu} + \delta T^{dm}_{\mu\nu}, \qquad T^r_{\mu\nu} = \overline{T}^r_{\mu\nu} + \delta T^r_{\mu\nu}. \tag{5.32}$$

At the homogeneous level, baryons and dark matter act like dust

$$\overline{T}^b_{\mu\nu} = \bar{\rho}_b \bar{u}_\mu \bar{u}_\nu, \qquad \overline{T}^{dm}_{\mu\nu} = \bar{\rho}_{dm} \bar{u}_\mu \bar{u}_\nu, \tag{5.33}$$

and constitute the non-relativistic matter component of the universe, i.e.

$$\overline{T}^m_{\mu\nu} = \overline{T}^b_{\mu\nu} + \overline{T}^{dm}_{\mu\nu}, \qquad \bar{\rho}_m = \bar{\rho}_b + \bar{\rho}_{dm}. \tag{5.34}$$

For each of these fluids, will use the following definitions for the perturbed part of the energy-momentum tensor:

$$\delta T^0{}_0 = -\delta\rho, \tag{5.35}$$

$$\delta T^0{}_i = (\bar{\rho} + \bar{p})\frac{\delta u_i}{a}, \tag{5.36}$$

$$\delta T^i{}_j = \delta p \delta_{ij} + \Sigma^i{}_j, \tag{5.37}$$

where $\Sigma^i{}_j$ is the traceless anisotropic shear stress perturbation. The aether pressure and four-velocity perturbations are defined as follows:

$$\tilde{p} = -\frac{\rho_m}{4} + \delta\tilde{p}, \tag{5.38}$$

$$\tilde{u}_\mu = u^{dm}_\mu + \delta\tilde{u}_\mu, \tag{5.39}$$

where

$$\rho_m = \rho_b + \rho_{dm} = \bar{\rho}_m + (\delta\rho_b + \delta\rho_{dm}), \qquad u^{dm}_\mu = \bar{u}_\mu + \delta u^{dm}_\mu. \tag{5.40}$$

Dark matter only interacts gravitationally and can be treated as a pressureless perfect fluid

$$T^{dm}_{\mu\nu} = \rho_{dm} u^{dm}_\mu u^{dm}_\nu, \qquad g^{\mu\nu} u^{dm}_\mu u^{dm}_\nu = -1, \tag{5.41}$$

which is separately conserved: (the following equations are equivalent to $\nabla^\mu T^{dm}_{\mu\nu} = 0$)

$$\nabla^\mu(\rho_{dm} u^{dm}_\mu) = 0 \tag{5.42}$$

$$u^{dm}_\mu \nabla^\mu u^{dm}_\nu = 0. \tag{5.43}$$

We assume negligible energy transfer between baryons and relativistic particles,[7] i.e.

$$\nabla^\mu(\rho_b u^b_\mu) = 0, \tag{5.44}$$

where u^b_μ is the four-velocity of baryons:

$$u^b_\mu = \bar{u}_\mu + \delta u^b_\mu, \qquad g^{\mu\nu} u^b_\mu u^b_\nu = -1. \tag{5.45}$$

We also assume

$$\delta p_r = \delta\rho_r/3, \qquad \delta p_b = 0, \tag{5.46}$$

[7] Baryons and photons are tightly coupled through matter-radiation equality and can be considered as one fluid because they exchange *momentum*.

so that

$$T^{\alpha}_{\ \alpha} = -\rho_b - \rho_{dm} = -\rho_m. \tag{5.47}$$

Then, to first order in perturbation theory (5.9) and (5.10) give: (see Section D.3 of Appendix D for a detailed derivation)

$$3\frac{\dot{a}}{a^2}\delta\tilde{p} = \frac{\bar{\rho}_m}{4}\partial_i(\delta\tilde{u}^i + \frac{\bar{\rho}_b}{\bar{\rho}_m}\delta w^i) \tag{5.48}$$

$$\partial_i\delta\tilde{p} = \frac{a\bar{\rho}_m}{4}(\delta\dot{\tilde{u}}^i + 2\frac{\dot{a}}{a}\delta\tilde{u}^i), \tag{5.49}$$

where $\delta w^i = a^{-2}(\delta u_i^{dm} - \delta u_i^b)$ and $\delta\tilde{u}^i = a^{-2}\delta\tilde{u}_i$. Taking the comoving divergence of (5.49) and applying the comoving Laplacian to (5.48), we can eliminate $\delta\tilde{p}$:

$$3\frac{\dot{a}}{a^3}\partial_\tau(a^2\Omega) - \nabla^2\Omega = \frac{\bar{\rho}_b}{a\bar{\rho}_m}\nabla^2(\dot{\delta}_b - \dot{\delta}_{dm}), \tag{5.50}$$

where

$$\Omega \equiv \partial_i\delta\tilde{u}^i, \qquad \delta_{dm} = \frac{\delta\rho_{dm}}{\bar{\rho}_{dm}}, \qquad \delta_b = \frac{\delta\rho_b}{\bar{\rho}_b}, \tag{5.51}$$

and we have used the fact that $\partial_i\delta w^i = \frac{1}{a}(\dot{\delta}_b - \dot{\delta}_{dm})$.[8] In Fourier space, this equation can be numerically integrated for modes of different wavelength, given the equations that govern δ_{dm} and δ_b. Once Ω is known, (5.48) can be used to find $\delta\tilde{p}$:

$$\delta\tilde{p} = \frac{\bar{\rho}_m a^2}{12\dot{a}}\left[\Omega + \frac{\bar{\rho}_b}{a\bar{\rho}_m}(\dot{\delta}_b - \dot{\delta}_{dm})\right]. \tag{5.52}$$

The rotational part of δu^i can be ignored because it decays with the expansion of the universe. To see this let $\delta u^i = \partial_i u_S + \partial\delta u_V^i$ where $\partial_i\delta u_V^i = 0$. Taking the curl of (5.49), it follows that $\nabla \times \delta\vec{u}_V \propto \frac{1}{a^2}$. As a result, the rotational part of the aether fluid decays and does not play a major role in cosmology. As a result, given Ω we can find δu^i in Fourier space ($\partial_j \to ik_j$):

$$\delta u^j = -i\frac{k_j}{k^2}\Omega, \tag{5.53}$$

where $k^2 = \delta^{ij}k_ik_j$.

So far, we did not need to choose a particular gauge. In order to write down Einstein's field equations, however, we will use the Conformal Newtonian Gauge [92]:

$$ds^2 = a^2(\tau)\{-[1 + 2\psi(\tau, \vec{x})]d\tau^2 + [1 - 2\phi(\tau, \vec{x})]d\mathbf{x}^2\}. \tag{5.54}$$

[8] See Section D.3 of Appendix D for a detailed proof.

We have listed all relevant (perturbed and unperturbed) geometric objects derivable from this metric in Section D.2 of Appendix D. Here we have confined ourselves to scalar perturbations of the metric only. The perturbed GGA equations are

$$(8\pi G_N)^{-1}\delta G^{\mu}{}_{\nu} = \delta\mathcal{T}^{\mu}{}_{\nu}, \tag{5.55}$$

where

$$\delta\mathcal{T}^{\mu}{}_{\nu} = \frac{G_R}{G_N}\delta T^{\mu}{}_{\nu} - \left(\frac{G_R}{G_N} - 1\right)\delta T^{\alpha}{}_{\alpha}\delta^{\mu}{}_{\nu} + 4\left(\frac{G_R}{G_N} - 1\right)\delta\tilde{T}^{\mu}{}_{\nu}. \tag{5.56}$$

Let us consider the different components. It can be checked that $(8\pi G_N)^{-1}\delta G^{0}{}_{0} = \delta\mathcal{T}^{0}_{0}$ leads, in Fourier space, to

$$k^2\phi + 3\frac{\dot{a}}{a}\left(\dot{\phi} + \frac{\dot{a}}{a}\psi\right) = 4\pi G_N a^2\left[\overline{\rho}_{dm}\delta_{dm} + \overline{\rho}_b\delta_b + \frac{G_R}{G_N}\overline{\rho}_r\delta_r\right], \tag{5.57}$$

where as before $\delta = \delta\rho/\overline{\rho}$ for every component (see (5.51)). Also, it follows from

$$(8\pi G_N)^{-1}\delta^{ij}\partial_j\delta G^{0}{}_{i} = \delta^{ij}\partial_j\delta\mathcal{T}^{0}_{i}$$

that

$$k^2\left(\dot{\phi} + \frac{\dot{a}}{a}\psi\right) = 4\pi G_N a^2\times \\ \left\{\overline{\rho}_{dm}\theta_{dm} + \overline{\rho}_b\theta_b + \frac{4}{3}\frac{G_R}{G_N}\overline{\rho}_r\theta_r + \left(1 - \frac{G_R}{G_N}\right)\left[a\overline{\rho}_m\Omega + \overline{\rho}_b(\dot{\delta}_b - \dot{\delta}_{dm})\right]\right\}, \tag{5.58}$$

where for every component $\theta = i\delta^{ij}k_i\delta u_j/a$. Taking the trace of the spatial part

$$(8\pi G_N)^{-1}\delta G^{i}{}_{i} = \delta\mathcal{T}^{i}_{i}$$

we find

$$\frac{k^2}{3}(\phi - \psi) + \ddot{\phi} + \left[2\frac{\ddot{a}}{a} - \left(\frac{\dot{a}}{a}\right)^2\right]\psi + \left(\frac{\dot{a}}{a}\right)(2\dot{\phi} + \dot{\psi}) = \frac{4\pi G_R}{3}a^2\times \\ \left\{\frac{1}{3}\bar{\rho}_r\delta_r + \left(1 - \frac{G_N}{G_R}\right)\frac{1}{aH}\left[a\overline{\rho}_m\Omega + \overline{\rho}_b(\dot{\delta}_b - \dot{\delta}_{dm})\right]\right\}, \tag{5.59}$$

where $H = \frac{\dot{a}}{a^2}$ is the Hubble parameter. Finally

$$(8\pi G_N)^{-1}\delta^{jk}(\partial_i\partial_k - \frac{1}{3}\delta_{ik})\delta G^{i}{}_{j} = \delta^{jk}(\partial_i\partial_k - \frac{1}{3}\delta_{ik})\delta\mathcal{T}^{i}_{j}$$

	G_N/G_R
WMAP+ACT	$0.73^{+0.31}_{-0.21}$
WMAP+ACT+SPT	$0.88^{+0.17}_{-0.13}$
WMAP+ACT+Hubble+BAO+Sne	$0.89^{+0.13}_{-0.11}$
WMAP+ACT+SPT+Hubble+BAO+Sne	$0.94^{+0.10}_{-0.09}$
WMAP+ACT+Sne+Ly-α (free Y_p)	$0.68^{+0.32}_{-0.25}$
WMAP+ACT+SPT+Sne+Ly-α (free Y_p)	$0.90^{+0.27}_{-0.23}$

Table 5.1: Summary of the constraints on G_N/G_R and the associated 95% confidence intervals for different combinations of observational data.

leads to

$$k^2(\phi-\psi) = 12\pi G_N a^2 \left[\overline{\rho}_{dm}\sigma_{dm} + \overline{\rho}_b\sigma_b + \frac{4}{3}\frac{G_R}{G_N}\overline{\rho}_r\sigma_r\right]. \tag{5.60}$$

where σ is defined through $(\bar{\rho}+\bar{p})\sigma \equiv (-\frac{k_i k_j}{k^2} + \frac{\delta_{ij}}{3})\Sigma^i{}_j$. Equations (5.57)–(5.60), once accompanied with (5.50) and the conservation equations (5.42)–(5.44), constitute the full set of (scalar) equations at first order in perturbation theory.

In [19], we modified the cosmological code Cmbeasy [53, 54] (using the GGA field equations derived in this Section) to constrain G_N/G_R using 7-year cosmic microwave background (CMB) data from WMAP [87]. The analysis assumed three massless neutrino species. Together with small-scale observations from the Atacama Cosmology Telescope (ACT) [58], the analysis yields $G_N/G_R = 0.73^{+0.31}_{-0.21}$ at 95%-confidence. Just like any additional relativistic component can be compensated by a higher fraction of dark matter in order to keep the time of matter-radiation equality constant, there is a high amount of degeneracy between G_N/G_R and $\Omega_m h^2$ and h (see Figure 5.2). Data from the South Pole Telescope (SPT), which measured the CMB power spectrum in the multipole range $650 < \ell < 3000$, significantly tightens the constraint and yields $0.88^{+0.17}_{-0.13}$. (For the combination of ACT and SPT data we adopted the SPT treatment of foreground nuisance parameters). A similar effect can be seen when adding Baryonic Acoustic Oscillations (BAO) [111] and constraints on the Hubble rate. Here we adopted the value of $H_0 = 73.8 \pm 2.4$ km^{-1} Mpc^{-1} [131]. Then, by breaking the degeneracy between the matter content and h, the combination WMAP+ACT+BAO+Sne+Hubble results in $G_N/G_R = 0.89^{+0.13}_{-0.11}$. The supernovae data of the Union catalog [14] do not significantly contribute to this constraint. Note that for both cases, i.e. adding either SPT data or adding the Hubble constraints to the basic WMAP+ACT set, move the gravitational Aether value of $G_N/G_R = 0.75$ to the border or just outside of the 95% confidence interval, while the

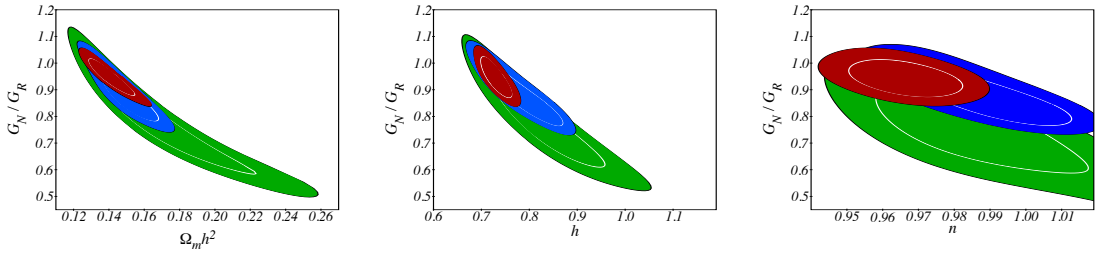


Figure 5.2: Constraints at the 95% confidence level for G_N/G_R from WMAP 7-year (background, green), WMAP+ACT+SPT (middle, blue) and WMAP+ACT+SPT+Sne+BAO+Hubble data (front, red). The white lines show the 68% confidence levels. Note that the Gravitational Aether prediction is $G_N/G_R = 0.75$, while in General Relativity $G_R = G_N$.

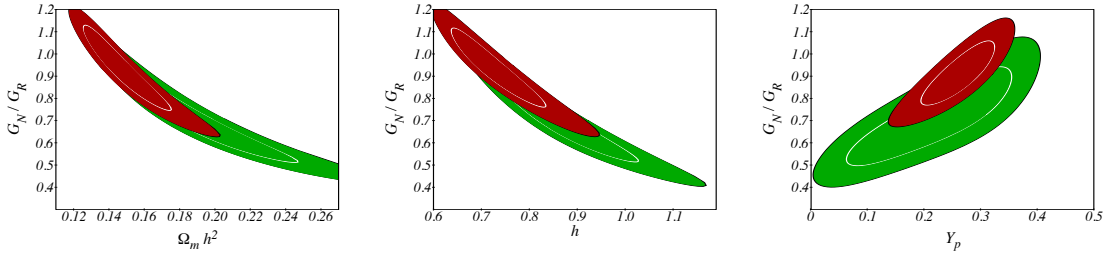


Figure 5.3: Constraints at the 95% confidence level for G_N/G_R from WMAP+ACT+Sne+Ly-α (background, green) and WMAP+ACT+SPT+Sne+Ly-α (front, red). The white lines show the 68% confidence levels. Note that the Gravitational Aether prediction is $G_N/G_R = 0.75$, while in General Relativity $G_R = G_N$.

GR value of $G_N/G_R = 1.0$ is well compatible with all combinations of data. Consequently, the full combination of data, i.e. WMAP+ACT+SPT+Hubble+BAO+Sne, constrains G_N/G_R to $0.94^{+0.10}_{-0.09}$.

In contrast, observational constraints at lower redshifts, in particular data of the Ly-α forest [138] prefer the aether prediction. Furthermore, additional degeneracies with e.g. the Helium mass fraction Y_p might shift the preferred values. Combining WMAP+ACT+Sne with the Ly-α forest constraints yields, $G_N/G_R = 0.68^{+0.32}_{-0.25}$ at 95% level, with Y_p as a free parameter. However, we should note that this result is more prone to systematic uncertainties due to the quasilinear nature of the Ly-α forest. Also, including the SPT data in this combination changes this result to the higher value of $0.90^{+0.27}_{-0.23}$. A summary of the constraints with different combinations of data is provided in Table 5.1.

Our analysis has recently been repeated using CMB data from the *Planck* satellite in [109]. They include also tensorial perturbations. Figure 5.4 shows the results of their measurements. It is quite remarkable how much smaller the error bars have gotten (almost by a factor of 5). It can be seen that the gravitational aether prediction $G_N/G_R = 0.75$ is ruled out at the 2.5–5σ level, depending on the combination of data sets used. (For certain combinations, GR is also ruled out at the 2.5σ level.) Although this is disappointing for the gravitational aether theory, it should be noted that the GGA formalism provides a way of measuring how pressure sources gravity in cosmology. Constraints on G_N/G_R can be directly translated to the anomalous pressure coupling to gravity ζ_4, which is introduced in the next Section.

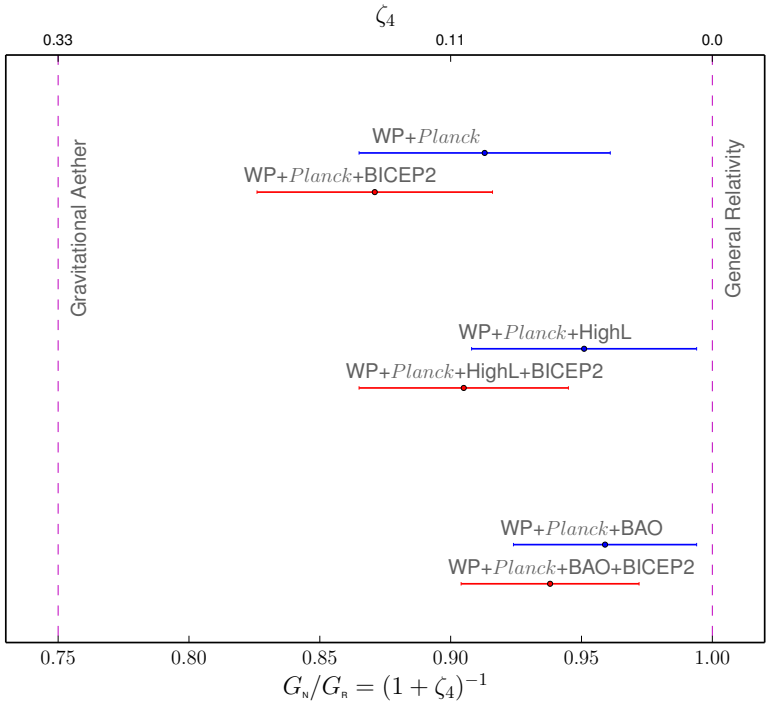


Figure 5.4: Measurement of G_N/G_R, by [109], using different data sets. The error bars show 68% confidence levels. WP and HighL refer to WMAP-9 polarization data and the higher multipole data sets ACT and SPT, respectively. (This Figure is taken from [109].)

5.4 Precision Tests of Gravity

Gravity on millimeter to solar system scales is well described by general relativity, which has passed many precision tests on these scales with flying colors (see *e.g.,* [163] for a review). That is why it is hard to imagine how an order unity change in the theory such as that of (5.4) can be consistent with these tests, without introducing any fine-tuned parameter. In this section, we argue that nearly all these tests are with gravitational sources that have negligible *pressure* or *vorticity*, which source deviations from GR predictions in gravitational aether theory.

5.4.1 Parametrized Post-Newtonian (PPN) formalism

In Section 5.3, we argued that for any perfect fluid with constant equation of state, w, the solutions of gravitational aether theory are identical to those of GR with a renormalized gravitational constant $\propto (1+w)$. However, for generic astrophysical applications, w is not constant except for pure radiation, or in the pressureless limit of $w = 0$. Focusing on the latter case, and given that pressure is 1st order in post-Newtonian expansion, we can quantify the gravitational aether theory through the Parametrized Post-Newtonian (PPN) formalism.

The Parametrized Post-Newtonian (PPN) formalism is defined in a weak field, slow motion limit, and describes the next-to-Newtonian order gravitational effects in terms of a standardized set of potentials and ten parameters. These PPN parameters will be determined by solving the field equations (5.4) order-by-order with a perfect fluid source in a standard coordinate gauge. The conventional introductory details of the formalism will be skipped over (see [64] for a more detailed explanation of the procedure and the general PPN formalism).

To be clear, though, we will assume a nearly globally Minkowskian coordinate system and basis with respect to which, at zeroth order, the metric is the Minkowski metric ($g_{\mu\nu} = \eta_{\mu\nu}$) and the fluid four-velocity u^μ is purely timelike ($u^0 = 1, u^i = 0$). The stress-energy tensor is taken to have the form $T_{\mu\nu} = (\rho + \rho\Pi + p)u_\mu u_\nu + p g_{\mu\nu}$ where u_μ, ρ, Π and p are the the unit four-velocity, rest-mass-energy density, internal energy density, and isotropic pressure of the fluid source, respectively. The fluid variables are assigned orders of $\rho \sim \Pi \sim \frac{p}{\rho} \sim u_i^2 \sim 1\text{PN}$. In the weak field limit, the metric can be written as a purturbation of the Minkowski metric: $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$. The components of the metric perturbations $h_{\mu\nu}$ with respect to this basis will be assumed to be of orders: $h_{00} \sim 1\text{PN} + 2\text{PN}$, $h_{ij} \sim 1\text{PN}$, and $h_{0i} \sim 1.5\text{PN}$. This choice preserve the Newtonian limit while allowing one to determine

just the first post-Newtonian corrections. Furthermore, the aether four-velocity $\tilde{u}_\mu$ will be assumed to be of the same order as that of the matter fluid.

Solving (5.9)–(5.10) to 1PN gives $\tilde{p} = -\rho/4$, which can be used in (5.4) to solve for g_{00} and g_{ij} to 1PN:

$$h_{00} = 2U \tag{5.61}$$

$$h_{ij} = 2U\delta_{ij}, \tag{5.62}$$

where U is the Newtonian potential and the following gauge condition is imposed: $\partial_j h_{ij} = \frac{1}{2}(\partial_i h_{jj} - \partial_i h_{00})$. Comparing the continuity equations for matter and aether (i.e. (5.9) to 1.5 PN), it can be shown that

$$\tilde{u}^i - u^i = t^i, \tag{5.63}$$

where t^i satisfies $\nabla^i(t_i\rho) = 0$. This implies that the rotational component of aether is not fixed by matter within the PN expansion formalism. Here we will make the assumption that $t^i = 0$ so that aether is completely dragged by matter. We will discuss this choice further in Section (5.4.2).

Previously we mentioned that in this case, an exact solution for $\tilde{u}_\mu$ and $\tilde{p}$ exists when matter has a constant equation of state. (It is worth noting that in the $t^i = 0$ case, higher PN equations appear to imply a nonstandard condition on the pressure $\nabla_a(u^a p) = 0$, which is satisfied for a constant equation of state.) Using this solution and an additional gauge condition $\partial_i h_{0i} = 3\partial_0 U$, the field equations can be solved for g_{0i} and g_{00} to 1.5PN and 2PN, respectively:

$$h_{0i} = -\frac{7}{2}V_i - \frac{1}{2}W_i, \tag{5.64}$$

$$h_{00} = 2U - 2U^2 + 4\phi_1 + 4\phi_2 + 2\phi_3 + 6(1 + \frac{1}{3})\phi_4, \tag{5.65}$$

where Appendix D.4 includes the definition for all potentials. Collecting all the results (5.61), (5.62), (5.64), and (5.65) indicates that all metric components are as in standard GR, except for the term in g_{00} with the pressure-dependent potential ζ_4. Consulting the parametrization rubric indicates that all PPN parameters have the standard values except ζ_4, which equals

$$\zeta_4 = \frac{1}{3}, \tag{5.66}$$

which was already pointed out in [3]. The most precise measurement of ζ_4 to date has been done in the context of cosmology, using the Generalized Gravitional Aether formalism introduced in Section 5.3.2 [109]. It is generically hard to measure ζ_4 because one needs to probe the relationship between gravity and pressure of an object with near-relativistic pressures. A notable exception is observation of neutron stars (or their mergers, via gravitational

wave observations), which can potentially measure ζ_4, assuming that the uncertainties in nuclear equation of state are under control [86].

5.4.2 Gravitomagnetic Effect

In the previous Section, we showed that rotation of aether is not fixed by matter in the non-relativistic regime. We further assumed that aether rotates with matter. Here we will argue that observational bounds on the gravitomagnetic effect provide a mild prefernce for this assumption.

Spacetime around a rotating object with a weak gravitational field, like earth, can be described in terms of a set of potentials. With appropriate definitions, these potentials satisfy equations analogous to Maxwell's equations [133]. Consider the gravitational field equations

$$G_{\mu\nu} = 8\pi\tilde{G}T^{\text{eff}}_{\mu\nu}. \tag{5.67}$$

For GR, we have $\tilde{G} = G_N$ and $T^{\text{eff}}_{\mu\nu} = T_{\mu\nu}$. For gravitational aether, $\tilde{G} = \frac{4}{3}G_N$ and $T^{\mu\nu}_{\text{eff}}$ is given by the RHS of (5.4). The gravitomagnetic effect describes the dragging of spacetime around a rotating object and can be quantified by a gravitomagnetic field $\mathbf{B}$ defined as:

$$\mathbf{B} = -4\frac{3(\mathbf{r}\cdot\mathbf{S})\mathbf{r} - r^2\mathbf{S}}{2r^5}, \tag{5.68}$$

$$S^i = 2\tilde{G}\int \epsilon^i{}_{jk}x^j T^{0k}_{\text{eff}}\mathrm{d}^3x. \tag{5.69}$$

where $\mathbf{r}$ is the position vector measured from the center of the object and $\epsilon^i{}_{jk}$ is the three-dimensional Levi-Civita tensor. The gravitomagnetic field causes the precession of the orbital angular momentum of a free falling test particle. The angular velocity of this precession is [133]

$$\mathbf{\Omega} = -\frac{\mathbf{B}}{2}. \tag{5.70}$$

If the aether fluid is irrotational, it can be checked that $T^{0k}_{\text{eff}} = T^{0k}$ to within the accuracy of linearized theory. It then follows from $\tilde{G} = \frac{4}{3}G_N$ that

$$\mathbf{\Omega}_{\text{aether}} = \frac{4}{3}\mathbf{\Omega}_{\text{GR}}. \tag{5.71}$$

Gravity Probe B (GP-B) is an experiment that measures the precession rate $\langle\mathbf{\Omega}\rangle$ of four gyroscopes orbiting the earth. In [61], GP-B reported a frame-dragging drift

rate of -37.2 ± 7.2 mas/yr, to be compared with the GR prediction of -39.2 mas/yr ('mas'=milliarc-second). Laser ranging to the LAGEOS and LAGEOS II satellites also provides a measurement of the frame-dragging effect. The total uncertainty in this case is still being debated; with optimistic estimates of $10\% - 15\%$ (*e.g.,* [46]), and more conservative estimates as large as $20\% - 30\%$ (*e.g.,* [80]).

Therefore, we conclude that even though perfect co-rotation of aether by matter is preferred by current tests of intrinsic gravitomagnetic effect, an irrotational aether is still consistent with present constraints at 2σ level.

5.5 Quantum Correction of Matter Fields and Gravitational Aether

In this Section we point out a curious fact about the behaviour of quantum corrections of matter fields in the gravitational aether theory. It is well known that the vacuum expectation value of the energy-momentum tensor $\langle \widehat{T}_{\mu\nu} \rangle_{vac}$ of a quantum field (e.g. a free scalar field) is divergent. The singularity structure of $\langle \widehat{T}_{\mu\nu} \rangle_{vac}$ is usually purely geometrical, because the divergences are ultraviolet (or short-distance) in nature.[9] In $3+1$ dimensions, a generally covariant regularization scheme would leads to

$$\langle \widehat{T}_{\mu\nu} \rangle_{vac} = c_g g_{\mu\nu} + c_G G_{\mu\nu} + c_{R^2} \mathcal{O}(R^2) + \text{finite terms}, \tag{5.72}$$

where c_g, c_G, and c_{R^2} are regularized constants which go to infinity as the regulator is removed. For instance, for a free minimally coupled scalar field with mass m, dimensional regularization for Hadamard states leads to [30]

$$c_g = -\frac{m^4}{32\pi^2} \left\{ \frac{1}{n-4} + \frac{1}{2} \left[\gamma + 2 \ln(m/\mu) \right] \right\} \tag{5.73}$$

$$c_G = \frac{m^2}{384\pi^2} \left\{ \frac{1}{n-4} + \frac{1}{2} \left[\gamma + 2 \ln(m/\mu) \right] \right\}, \tag{5.74}$$

where μ is the renormalization mass scale.

In general relativity, c_g and c_G merely renormalize the cosmological constant and the gravitational constant, respectively. Generically, the $\mathcal{O}(R^2)$ terms can only be renormalized

[9] This is not always the case, but it is true for the so-called Hadamard states (see e.g. [30]).

away if higher order terms are included in the gravitational action.[10] For the purpose of our current discussion, we shall ignore these terms. In the case of the gravitational aether theory, this story changes because the field equations are modified as

$$(8\pi\tilde{G}_b)^{-1}G_{\mu\nu} = T_{\mu\nu} - \frac{1}{4}T^{\alpha}{}_{\alpha}g_{\mu\nu} + \{\tilde{p}(1+1/\tilde{w}_b)\tilde{u}_\mu\tilde{u}_\nu + \tilde{p}g_{\mu\nu}\}\,, \tag{5.75}$$

where we have given the aether fluid an equation of state $\tilde{w}_b$, for reasons which will be discussed shortly. The subscript b reminds us that the constants $\tilde{G}_b$ and $\tilde{w}_b$ are *bare* quantities. Let

$$T_{\mu\nu} = \langle\widehat{T}_{\mu\nu}\rangle_{div} + T^{cl}_{\mu\nu}, \tag{5.76}$$

where

$$\langle\widehat{T}_{\mu\nu}\rangle_{div} = c_g g_{\mu\nu} + c_G G_{\mu\nu}, \tag{5.77}$$

and $T^{cl}_{\mu\nu}$ stands for the energy momentum tensor of "classical matter", as well as (possibly non-local) the finite-term corrections in (5.72). Note that

$$\begin{aligned}\langle\widehat{T}_{\mu\nu}\rangle_{div} - \frac{1}{4}\langle\widehat{T}^{\alpha}{}_{\alpha}\rangle_{div}g_{\mu\nu} &= c_G G_{\mu\nu} - \frac{c_g}{4}G^{\alpha}{}_{\alpha}g_{\mu\nu} \\ &= c_G G_{\mu\nu} - 2\pi\tilde{G}_b(3-1/\tilde{w}_b)c_G\tilde{p}g_{\mu\nu},\end{aligned} \tag{5.78}$$

where in the last line we have used the fact that $G^{\alpha}{}_{\alpha} = 8\pi\tilde{G}_b(3-1/\tilde{w}_b)\tilde{p}$, which can be obtained by taking the trace of (5.72). By plugging (5.76) into (5.75), it may be verified that the gravitational aether field equations can be rewritten as

$$\left[(8\pi\tilde{G}_b)^{-1} - c_G\right]G_{\mu\nu} = T^{cl}_{\mu\nu} - \frac{1}{4}T^{cl}g_{\mu\nu} + \{\tilde{p}_{\rm eff}(1/\tilde{w}_{\rm eff}+1)\tilde{u}_\mu\tilde{u}_\nu + \tilde{p}_{\rm eff}g_{\mu\nu}\}\,, \tag{5.79}$$

where

$$\tilde{p}_{\rm eff} = \left[1 - 2\pi\tilde{G}_b(3-1/\tilde{w}_b)c_G\right]\tilde{p}, \qquad \tilde{w}_{\rm eff} = \frac{1-2\pi\tilde{G}_b(3-1/\tilde{w}_b)c_G}{1/\tilde{w}_b + 2\pi\tilde{G}_b(3-1/\tilde{w}_b)c_G}. \tag{5.80}$$

Therefore, both the bare gravitational constant $\tilde{G}_b$ and aether equation of state $\tilde{w}_b$ get renormalized. Of course, we ought to require

$$(8\pi\tilde{G}_b)^{-1} - c_G = (8\pi\tilde{G})^{-1}, \tag{5.81}$$

[10] A notable exception is a massless conformally coupled scalar field, in which case c_{R^2} is rendered finite as the regulator is removed, and one finds the trace anomaly (see e.g. [30]).

where $\tilde{G} = \frac{4}{3}G_N$. Solving for $\tilde{G}_b = \tilde{G}/(1 + 8\pi\tilde{G}c_G)$ and plugging it back to (5.80), we find after removing the regulator, i.e. $c_G \to \infty$ (e.g. $n \to 4$ in (5.74)):

$$\tilde{w}_{\text{eff}} = \frac{1}{3}, \tag{5.82}$$

regardless of what the bare equation of state $\tilde{w}_b$ is. This seems to suggest that the gravitational aether equation of state is not stable against higher order quantum corrections in the matter sector. I would be interesting further pursue the consequences of this fact.

5.6 Conclusions and Discussions

In this chapter, we studied the phenomenological implications of the gravitational aether theory, a modification of general relativity which decouples quantum vacuum fluctuations from gravity through the introduction of an incompressible perfect fluid called the gravitational aether. In doing so, this proposal provides a solution to the so-called "old" cosmological constant problem. We showed that deviations from general relativity can only be significant in situations with relativistic pressure, or (potentially) relativistic vorticity. The most prominent prediction of this theory is that gravity should be 33% stronger in the cosmological radiation era than GR predictions. In other words, the effective gravitational coupling of relativistic matter, which we called G_R, is given in the radiation era by $G_R = \frac{4}{3}G_N$.

Comparing theoretical prediction of big bang nucleosynthesis with the observational light element abundances provides a constraint on G_N/G_R. We showed that every light element abundance agrees with the gravitational aether theory within 2σ. It is worth noting that the ^{7}Li primordial abundance prefers the prediction of gravitational aether to that of general relativity, which is of course $G_R = G_N$.

Furthermore, we studied the evolution of cosmological perturbations in the gravitational aether framework. In doing so, we introduced the generalized gravitational aether theory, which interpolates between general relativity and gravitational aether through an extra gravitational coupling constant G_R. This allowed us to constrain G_N/G_R using cosmic microwave background (CMB) measurements from WMAP, as well as other data sets. A summary of the constraints with different combinations of data is provided in Table 5.1. The gravitational aether theory is consistent at the $\sim 2\sigma$ level with all different combinations of data sets. In some cases, the gravitational aether prediction is preferred to that of general relativity. Our analysis has recently been repeated using CMB data from the *Planck* satellite [109]. Based on their measurements, the gravitational aether prediction is

ruled out at the 2.5–5σ level, depending on the combination of data sets used. (For certain combinations, GR is also ruled out at the 2σ level.) In contrast, observational constraints at lower redshifts, in particular data of the Ly-α forest [138], prefer the aether prediction. Furthermore, additional degeneracies with e.g. the Helium mass fraction Y_p might shift the preferred values. However, we should note that this result is more prone to systematic uncertainties due to the quasilinear nature of the Ly-α forest.

We then examined the implications for precision tests of gravity using the PPN formalism, and showed that the only PPN parameter that deviates from its GR value is ζ_4, the anomalous coupling to pressure, that has only recently been measured in [109], thanks to the generalized gravitational aether formalism. Moreover, we argued that current tests of Earth's gravitomagnetic effect mildly prefer a co-rotation of aether with matter, although they are consistent with an irrotational aether at 2σ level.

Finally, we pointed out a curious fact about the behaviour of quantum corrections of matter fields in the gravitational aether theory. We showed that the gravitational aether theory is not stable against higher order quantum corrections in the matter sector, in the sense that any bare equation of state for aether gets renormalized to the value 1/3.

Another interesting implication of this theory is for the cosmic baryon fraction. As we increase the gravity due to radiation, we need to increase the dark matter density to keep the redshift of equality constant, since it is well constrained by CMB power spectrum (see *e.g.,* [87]). This implies that the total matter density should be bigger by a factor of 4/3 (Fig. 5.2). Given that baryon density is insensitive to this change, the cosmic baryon fraction will decrease by a factor of 3/4, i.e. from 17% [87] to 13%. This leads us to the final Chapter of this thesis, where we introduce a model-independent and statistically optimal method for measuring the pressure profile of galaxy clusters, which has immediate implications for the cosmic baryonic budget.

Chapter 6

An Optimal and Model-Independent Measurement of the Intracluster Pressure Profile

6.1 Introduction

Clusters of galaxies have long been recognized as remarkable laboratories to test cosmological theories. They are the largest gravitationally bound objects in the universe, thought to have formed from the eventual collapse of initially overdense dark matter perturbations. Their abundance and large scale properties are sensitive to the expansion and initial conditions of the universe, making them excellent tools to constrain cosmological models. On smaller scales, the physics of clusters is dominated by complex baryonic processes such as gas cooling, star formation, and feedback from supernovae and active galactic nuclei. In order to use clusters as standard probes of the geometry and dynamics of the universe, it is necessary to reliably model these processes and distinguish amongst different feedback mechanisms (for recent reviews see [88, 13]).

The main baryonic budget of clusters is a hot plasma of ionized hydrogen and helium in the intracluster medium (ICM), making it the natural target for studying the complex astrophysical processes at play. This virialized plasma emits bremsstrahlung radiation in X-ray, making it possible to probe the dense regions of the ICM.[1] Historically, most of our observational understanding of the ICM has come from X-ray observations, leading to a

[1]This is because X-ray luminosity is proportional to the *square* of gas density (e.g. [162]).

fairly consistent picture of the scaling and structural properties of low-redshift clusters (see e.g. [32, 157, 16, 126, 149, 127]).

The thermal Sunyaev-Zeldovich (tSZ) effect [151] is another important probe of the ICM: as the Cosmic Microwave Background (CMB) photons inverse-Compton scatter off of the hot electrons in the ICM, their blackbody spectrum is distorted. The tSZ effect has the unique property that its signal is independent of redshift, making it a powerful observational tool for detecting clusters at cosmological redshifts, and hence a promising cosmological probe of dark energy (e.g. [29, 43]). Within the last few years, cluster surveys exploiting the SZ effect have started delivering cluster samples (e.g. [148, 97, 164, 120]) as well as constraints on cosmological parameters [156, 137]. Another important feature of the tSZ effect is that it is directly proportional to the integrated pressure of free electrons along the line of sight, which makes it a powerful probe of the ICM in the outskirts ($r \gtrsim R_{200}$), where X-ray emission is extremely faint.

Resolving the tSZ signal for individual clusters requires high resolution CMB measurements, which have become available only in the recent years. In fact, the only all-sky CMB survey with high enough resolution *and* sensitivity to detect individual SZ clusters is *Planck*. Even with *Planck*'s sensitivity, it is necessary to combine the tSZ signal from many clusters to meaningfully constrain physical quantities of interest, such as baryonic mass fraction [121]. Luckily, there is fairly strong evidence from X-ray observations and numerical simulations that clusters are self-similar to a good approximation (see e.g. [107, 16]). This fact justifies combining SZ signatures of many clusters to obtain constraints on the *mean* ICM properties. There have been quite a few efforts in this direction over the past few years. By analyzing WMAP-1 (WMAP-3 respectively) CMB data for 116 (193 respectively) X-ray detected clusters, [10, 9] (respectively) provided constraints on the ICM pressure profile out to $\sim R_{200}$. Other similar works include WMAP-3 stacking of over 700 clusters by [21], WMAP-5 analysis of about 900 ROSAT NORAS/REFLEX clusters [102], WMAP-7 analysis of 175 *Planck* ESZ clusters [93], and SZ measurements of 15 massive X-ray selected clusters obtained with the South Pole Telescope [115]. Most notably, [121] have studied the tSZ signal of 62 low-redshift massive clusters by using CMB data from the *Planck* satelite, constraining the mean pressure profile of the ICM out to $3R_{500}$ with unprecedented precision.

The practice of averaging signals from many clusters goes under the title "stacking". The basic idea is the following: the main sources of uncertainty in extracting the tSZ signal are the primary CMB anisotropies and instrumental noise. Since these sources of noise are random in nature, they "drop out" if the temperature profile around many clusters is averaged over. This procedure, in its various forms, is not statistically optimal for multiple reasons. Firstly, large-angle correlations of primary CMB fluctuations are ignored

when stacking. Secondly, it is not clear how contributions from different clusters should be optimally weighed in the averaging process. Typically, different weighing methods are adopted to see whether the effect on the extracted tSZ signal is significant or not (e.g. [21]). Thirdly, when using multiple frequency channels, the final resolution of the reconstructed tSZ map is determined by the *lowest* resolution of the combined frequency maps (e.g. [121]). Finally, the 3D pressure profiles are usually obtained a posteriori by deprojecting the tSZ signal, which may lead to noise amplification. For these reasons, stacking procedures either result in an underestimation of error or loss of statistical information.

We believe all the aforementioned shortcomings of stacking procedures can be overcome with the methodology we have formulated in this Chapter, which is more in line with optimized template fitting procedure of [87]. Our analysis includes an all-sky multi-channel fit to the mean pressure profile of the ICM which appropriately takes into account primary CMB and noise correlations on *all* scales, while using the maximum resolution and sensitivity of all channels to their full potential. Furthermore, following [9], our method is completely model-independent, thus eliminating any systematic uncertainty associated with theoretical modelling of the ICM. As a first application of our methodology, we will apply our formalism to WMAP-9 CMB data using the Meta-Catalogue of X-ray detected Clusters of galaxies [114].

The rest of this Chapter is organized as follows. In Sections 6.2.1 and 6.2.2, we review the tSZ effect and the concept of a universal pressure profile, describing how model-independence can be achieved. Sections 6.2.3 and 6.2.4 contain our main statistical and numerical methodology, outlining in detail how the mean pressure profile of ICM can be optimally constrained. Section 6.3 describes the CMB data and cluster sample we use to test our methodology, and is followed by a discussion of the resulting pressure profiles in Section 6.4.1. Section 6.4.2 presents gas mass fraction measurements of various subsamples of our cluster catalogue. We discuss future work and how we anticipate our results to improve by using *Planck* CMB data in Section 6.5, before concluding our findings in Section 6.6.

Throughout this Chapter we assume a ΛCDM cosmology with present matter density $\Omega_m = 0.3$, dark energy density $\Omega_\Lambda = 0.7$, and Hubble parameter $H_0 = 100\,h$ km s^{-1}Mpc^{-1}with $h = \frac{7}{10}h_{70} = 0.7$. We also denote the normalized Hubble parameter at redshift z by $E(z) \equiv \frac{H(z)}{H_0} = \sqrt{\Omega_m(1+z)^3 + \Omega_\Lambda}$.

6.2 Extracting the Pressure Profile

This Section contains the statistical and numerical methodology we use to extract the mean ICM pressure profile from a full-sky CMB experiment. Section 6.2.1 reviews the tSZ effect and how it is related to the electron pressure profile. In Section 6.2.2, we reduce the problem of finding the exact profile of each cluster to a single, universal up to normalization, pressure profile. In Section 6.2.3, we derive the maximum likelihood estimator of the profile and its covariance matrix. Finally, in Section 6.2.4, we describe how the components of the estimator are in practice computed numerically.

6.2.1 tSZ Effect Model

The contribution of the tSZ effect to the CMB temperature anisotropy at frequency ν and location $\widehat{\boldsymbol{n}}$ on the sky is proportional to the integral of the electron pressure along the line of sight: (see §2.5 of [162] for a derivation)

$$\begin{aligned} \delta T^{\mathrm{SZ}}(\widehat{\boldsymbol{n}};\nu) &= \frac{\sigma_{\mathrm{T}} T_{\mathrm{CMB}}}{m_{\mathrm{e}} c^2} F\left(\frac{h\nu}{k_{\mathrm{B}} T_{\mathrm{CMB}}}\right) \int dl_{\widehat{\boldsymbol{n}}} P_{\mathrm{e}}(l_{\widehat{\boldsymbol{n}}}), \\ F(x) &\equiv x \coth(x/2) - 4, \end{aligned} \tag{6.1}$$

where σ_{T} is the Thomson scattering cross-section, m_e is the mass of the electron, c is the speed of light, k_B is the Boltzmann constant, h is the Planck constant, [2] $T_{CMB} = 2.725$ K is the mean CMB temperature [99], and $P_{\mathrm{e}}(l_{\widehat{\boldsymbol{n}}})$ is the pressure of free electrons along the line of sight direction $\widehat{\boldsymbol{n}}$.

Our task is to constrain P_{e} through the tSZ effect. We will assume that P_{e} is spherically symmetric to a good approximation and denote the pressure profile of the a^{th} cluster by $P_{\mathrm{e}}^{(a)}(r)$. Furthermore, since it is not possible to constrain a continuous function without introducing model-dependence, we consider spherical bins around the centre of clusters, in each of which the pressure is assumed to be constant:

$$P_{\mathrm{e}}^{(a)}(r) = \begin{cases} P_1^{(a)} & \text{if } 0 < r < r_1^{(a)} \\ P_2^{(a)} & \text{if } r_1^{(a)} < r < r_2^{(a)} \\ \vdots & \\ P_{N_b}^{(a)} & \text{if } r_{N_b-1}^{(a)} < r < r_{N_b}^{(a)}. \end{cases} \tag{6.2}$$

[2] We've used h to denote Planck's constant only in Equations (6.1) and (6.4). Throughout the rest of our paper, h is the reduced Hubble constant.

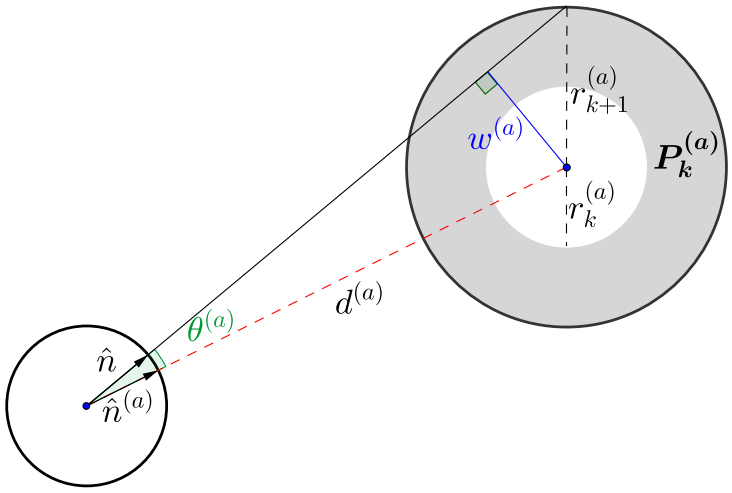


Figure 6.1: The k^{th} bin of the a^{th} cluster. The contribution of this bin to the temperature anisotropy of the CMB at frequency ν and location $\widehat{\boldsymbol{n}}$ on the sky is given by $P_k^{(a)} t_k^{(a)}(\widehat{\boldsymbol{n}}; \nu)$, where $t_k^{(a)}(\widehat{\boldsymbol{n}}; \nu)$ is defined in Equation (6.4).

Here $P_1^{(a)}, \ldots, P_{N_b}^{(a)}$ are all constants with units of pressure, r is the radius away from the centre of cluster, and N_b is the total number of bins. The value of the pressure in each bin may be better understood as the volume-weighed average of the pressure in that bin. With these simplifications, Equation (6.1) may be written as

$$\delta T^{\text{SZ}}(\widehat{\boldsymbol{n}}; \nu) = \sum_{k=1}^{N_b} \sum_{a=1}^{N_c} P_k^{(a)} t_k^{(a)}(\widehat{\boldsymbol{n}}; \nu), \tag{6.3}$$

where N_c is the total number of clusters and $t_k^{(a)}(\widehat{\boldsymbol{n}}; \nu)$ is given by:

$$t_k^{(a)}(\widehat{\boldsymbol{n}}; \nu) = \frac{\sigma_{\text{T}} T_{\text{CMB}}}{m_{\text{e}} c^2} F\left(\frac{h\nu}{k_{\text{B}} T_{\text{CMB}}}\right) \times 2 \begin{cases} l_{k+1}^{(a)}(\widehat{\boldsymbol{n}}) - l_k^{(a)}(\widehat{\boldsymbol{n}}) & \text{if } w^{(a)}(\widehat{\boldsymbol{n}}) \leq r_k^{(a)} \\ l_{k+1}^{(a)}(\widehat{\boldsymbol{n}}) & \text{if } r_k^{(a)} \leq w^{(a)}(\widehat{\boldsymbol{n}}) \leq r_{k+1}^{(a)} \\ 0 & \text{if } w^{(a)}(\widehat{\boldsymbol{n}}) \geq r_{k+1}^{(a)}. \end{cases} \tag{6.4}$$

The functions used in Equation (6.4) are defined as follows:

$$l_k^{(a)}(\widehat{\boldsymbol{n}}) = \sqrt{[r_k^{(a)}]^2 - [w^{(a)}(\widehat{\boldsymbol{n}})]^2}, \tag{6.5a}$$

$$w^{(a)}(\widehat{\boldsymbol{n}}) \equiv d^{(a)} \sin(\theta^{(a)}(\widehat{\boldsymbol{n}})), \tag{6.5b}$$

$$\cos(\theta^{(a)}(\widehat{\boldsymbol{n}})) = \widehat{\boldsymbol{n}} \cdot \widehat{\boldsymbol{n}}^{(a)}, \tag{6.5c}$$

where $d^{(a)}$ is the angular diameter distance to the a^{th} cluster and $\widehat{\boldsymbol{n}}^{(a)}$ is the unit vector pointing to its centre.[3] Figure 6.1 shows the basic geometry that underlies Equations (6.3-6.5).

6.2.2 Universal Pressure Profile

In principle, the analysis that will follow can be used to optimally estimate all parameters $P_1^{(a)}, \ldots, P_{N_b}^{(a)}$. Unfortunately, this is too computationally-intensive for a large sample of clusters, given the large angle correlations of primary CMB anisotropies and the current resolution of CMB experiments. However, there is fairly concrete evidence that the pressure profile of the hot gas in clusters is self-similar (see e.g. [107, 16]). This means that for a given cluster, there is a self-similarity scale $r_c^{(a)}$ such that the pressure profile takes the form

$$P_{\text{e}}^{(a)}(r) = \mathbb{P}(r/r_{\text{c}}^{(a)}) P_{\text{c}}^{(a)}, \tag{6.6}$$

where $P_{\text{c}}^{(a)}$ is a constant characteristic pressure of the a^{th} cluster, and $\mathbb{P}(x)$ is the so-called *universal pressure profile*. Within this context, it is natural to construct the radial bins so that $r_k^{(a)} = n_k r_c^{(a)}$, where $\{n_k\}_{k=1-N_b}$ are positive numbers satisfying $n_1 < n_2 < \cdots < n_{N_b}$. With these assumptions, discretization of the electron pressure profile amounts to $P_k^{(a)} = \mathbb{P}_k P_c^{(a)}$, and the tSZ contribution (6.3) to the CMB anisotropy takes the form

$$\delta T^{\text{SZ}}(\widehat{\boldsymbol{n}}; \nu) = \sum_{k=1}^{N_b} \mathbb{P}_k t_k^{(\nu)}(\widehat{\boldsymbol{n}}), \tag{6.7a}$$

$$t_k^{(\nu)}(\widehat{\boldsymbol{n}}) = \sum_{a=1}^{N_c} P_c^{(a)} t_k^{(a)}(\widehat{\boldsymbol{n}}; \nu). \tag{6.7b}$$

[3] Angular diameter distance to the a^{th} cluster with redshift z is given by $d^{(a)}(z) = \frac{c/H_0}{1+z} \int_0^z \frac{dy}{\sqrt{\Omega_\Lambda + \Omega_m (1+y)^3}}$.

If the characteristic scales $r_c^{(a)}$ and $P_c^{(a)}$ are fixed by external observations (such as X-ray's) for each cluster, our task is simplified to finding best estimate values (and their associated uncertainties) for N_b parameters: $\mathbb{P}_1, \ldots, \mathbb{P}_{N_b}$.

We use $R_{500}^{(a)}$ as the self-similarity length scale of the a^{th} cluster (i.e. we set $r_c^{(a)} = R_{500}^{(a)}$). The quantity $R_\Delta^{(a)}$ is defined as the radius up to which the matter density is Δ times the critical mass-density of the universe:

$$\begin{aligned} M_\Delta^{(a)} &\equiv \int_0^{R_\Delta^{(a)}} \rho_m^{(a)}(r) 4\pi r^2 dr \\ &= \Delta \times \frac{4}{3}\pi [R_\Delta^{(a)}]^3 \times \rho_{crit}(z_a), \end{aligned} \tag{6.8}$$

where z_a is the redshift of the a^{th} cluster, and $\rho_{crit}(z) = \frac{3H(z)^2}{8\pi G}$. We consider 8 bins with radii $r_k^{(a)} = 0.5k \times R_{500}^{(a)}$, where $k \in \{1, 2, \ldots, 8\}$. This is equivalent to setting $n_k = 0.5k$. We use two proposals for the characteristic cluster-dependent pressure $P_c^{(a)}$:

$$P_c^{(a)} = 1.65 \times 10^{-3} E(z_a)^{8/3} h_{70}^2 \left[\frac{M_{500}^{(a)}}{3 \times 10^{14} h_{70}^{-1} M_\odot} \right]^{2/3+\delta} \text{keV cm}^{-3}, \tag{6.9}$$

with $\delta = 0$ and $\delta = 0.12$. The former corresponds to the mass variation expected in the standard self-similar model based purely on gravitation (see [107, 16]). The $\delta = 0.12$ case is a modification to the standard self-similar model proposed by [16], which is a first approximation to quantifying how the mass scaling of $P_c^{(a)}$ changes with radius in the REXCESS sample [32]. Using the $\delta = 0.12$ scaling makes for a meaningful comparison of our results with those of [121], since this is what is used in their analysis.

6.2.3 Statistical Methods

We use the principle of maximum likelihood to find best-estimate values for the parameters $\mathbb{P}_1, \ldots, \mathbb{P}_{N_b}$. (For template fitting on CMB sky, see e.g. [68, 81, 87]). We assume that the only contributions to the temperature anisotropies of the CMB are the primordial anisotropies δT^{PA}, the tSZ effect δT^{SZ} (given by Equation (6.7)), possible residual monopole and dipole components, and the instrumental noise. Furthermore, we assume that primordial anisotropies follow the statistics of an isotropic Gaussian random field, for which we know the angular power spectrum C_ℓ. [4]

[4] We use the CAMB code to generate the expected values of C_ℓ's for the WMAP concordance ΛCDM cosmology [91, 78].

Let L_p (L_ν) denote the set of all pixels (frequency channels) we wish to use in our analysis. In Section E.1.1 of Appendix E we show that the log-likelihood of the observed temperature fluctuations $\{\delta T_{i\nu}\}_{i\in L_p}^{\nu\in L_\nu}$, given the binned pressure profile $\{\mathbb{P}_k\}$, is analytic and equal to

$$-\frac{1}{2}\chi^2(\{\delta T_{i\nu}\}_{i\in L_p}^{\nu\in L_\nu}|\{\mathbb{P}_k\}) = -\frac{1}{2}(\boldsymbol{\delta T} - \boldsymbol{\delta T}^{\rm SZ})^T\mathbf{C}^{-1}(\boldsymbol{\delta T} - \boldsymbol{\delta T}^{\rm SZ}), \tag{6.10}$$

where

$$\mathbf{C} = \mathbf{C_S} + \mathbf{C_N}. \tag{6.11}$$

Here $\mathbf{C_S}$ is the covariance matrix of the primary CMB fluctuations and $\mathbf{C_N}$ is the covariance of the instrumental noise. The former is related to the angular power spectrum C_ℓ through

$$[\mathbf{C_S}]_{i\nu,i'\nu'} \equiv \langle \delta T_{i\nu}^{\rm PA}\delta T_{i'\nu'}^{\rm PA}\rangle = \sum_{\ell=0}^{\ell_{max}}\left(\frac{2\ell+1}{4\pi}\right)C_\ell B_{\ell\nu}B_{\ell\nu'}W_\ell^2 P_\ell(\widehat{\boldsymbol{n}}_i\cdot\widehat{\boldsymbol{n}}_{i'}), \tag{6.12}$$

where $B_{\ell\nu}$ is the spherically averaged beam transfer function for the mode ℓ and frequency channel ν, W_ℓ is the spherically averaged pixel transfer function, $\widehat{\boldsymbol{n}}_i$ is the sky direction of the $i^{\rm th}$ pixel, and P_ℓ is the ℓ-th degree Legendre polynomial. We use $\ell_{max} = 2\times N_{side}$ throughout our analysis, where $N_{side} = 512$ is set by the HEALPix [69] resolution of the WMAP sky maps.

In the case of WMAP, the instrumental noise is largely uncorrelated both between pixels and between different frequency channels. Its properties are adequately modelled by a Gaussian distribution with covariance matrix

$$[\mathbf{C_N}]_{i\nu,i'\nu'} = n_{i\nu}^2\delta_{ii'}\delta_{\nu\nu'}. \tag{6.13}$$

In Equation (6.10), we have used $\delta T_{i\nu}^{\rm SZ}$ to stand for a pixelized version of Equation (6.7):

$$\delta T_{i\nu}^{\rm SZ} = \sum_{k=1}^{N_b}\mathbb{P}_k t_k^{(\nu)}(i), \tag{6.14a}$$

$$t_k^{(\nu)}(i) = \sum_{l=0}^{l_{max}}\sum_{m=-l}^{l}(t_k^{(\nu)})_{lm}B_{l\nu}W_l Y_{lm}(\widehat{\boldsymbol{n}}_i), \tag{6.14b}$$

where $(t_k^{(\nu)})_{lm}$ are the spherical harmonic coefficients of $t_k^{(\nu)}(\widehat{\boldsymbol{n}})$. In practice, we generate the templates at HEALPix resolution 12 ($N_{\rm side} = 4096$), and then downgrade to resolution

9 ($N_{\rm side} = 512$).[5] We then convolve all templates with instrumental beams to obtain the quantities $t_k^{(\nu)}(i)$.

Let us note again that $\mathbb{P}_k$ is the universal pressure of the $k^{\rm th}$ bin and $t_k^{(\nu)}(i)$ is the coefficient that multiplies it at pixel i and frequency channel ν. Assuming a uniform prior on $(\widehat{\mathbb{P}}_1, \ldots, \widehat{\mathbb{P}}_{N_b})$, the posterior probability function of these variables is also a Gaussian distribution which peaks at the maximum of the log-likelihood function given in Equation (6.10), which is

$$\widehat{\mathbb{P}}_k = \sum_{k'=1}^{N_b} [\boldsymbol{\alpha}^{-1}]_{k,k'} \beta_{k'}, \tag{6.15a}$$

$$\alpha_{k,k'} = \sum_{\nu \in L_\nu} \sum_{i \in L_p} t_k^{(\nu)}(i) X_k^{(\nu)}(i), \tag{6.15b}$$

$$\beta_k = \sum_{\nu \in L_\nu} \sum_{i \in L_p} \delta T_{i\nu} X_k^{(\nu)}(i), \tag{6.15c}$$

where we have introduced the inverse covariance weighed template

$$X_k^{(\nu)}(i) = \sum_{\nu' \in L_\nu} \sum_{i' \in L_p} [\mathbf{C}^{-1}]_{i\nu,i'\nu'} t_k^{(\nu')}(i'). \tag{6.16}$$

Finally, the covariance matrix $\mathbf{C}_{\mathbb{P}}$ of $\{\mathbb{P}_k\}$ is determined by the Hessian of the log-likelihood (6.10). Its matrix elements are

$$[\mathbf{C}_{\mathbb{P}}]_{k,k'} = \left\langle \left(\widehat{\mathbb{P}}_k - \mathbb{P}_k\right) \left(\widehat{\mathbb{P}}_{k'} - \mathbb{P}_{k'}\right) \right\rangle = [\boldsymbol{\alpha}^{-1}]_{k,k'}. \tag{6.17}$$

This matrix does not involve further computations as it is already required to obtain the best estimates $\{\widehat{\mathbb{P}}_k\}$. All the measurements on the pressure profiles and their attached error bars are obtained using only the expressions indicated in this Section.

In Section E.1.4 of Appendix E, we have shown how any residual monopole or dipole contribution can be conveniently accounted for in this formalism.

[5] To be more specific, following the notation of Section 6.2.2, we give the value $t_k^{(\nu)}(\widehat{\boldsymbol{n}}_i)$ to the i-th pixel of the k-th template in frequency channel ν, where $\widehat{\boldsymbol{n}}_i$ is the centre of the i-th pixel at a resolution 12. (At this resolution, the first radial bin of all MCXC clusters occupies at least one pixel.) We then downgrade the templates to resolution 9 and finally convolve all templates with the instrumental beams.

6.2.4 Numerical Methods

It is clear from Equation (6.15) that all quantities of interest can be calculated once the weighed templates $\boldsymbol{X}_{\boldsymbol{k}}^{(\boldsymbol{\nu})}$ are known. This is impossible to achieve by direct computation, which would involve inversion of the full covariance matrix $\mathbf{C}$. In this section, we will describe how we compute $\boldsymbol{X}_{\boldsymbol{k}}^{(\boldsymbol{\nu})}$ numerically.

We start by establishing some notation. The set of all masked (unmasked) pixels is denoted by $L_{\bar{p}}$ (L_p), so that $L = L_{\bar{p}} \cup L_p$ contains all pixels in the sky. We denote the total number of pixels (i.e. the size of L) by N_T.

In the statistical modelling described in Section 6.2.3, the covariance matrices $\mathbf{C_S}$ and $\mathbf{C_N}$ are only computed on the observed pixels L_p. However, numerical manipulation of these matrices is more efficient in harmonic space, using the spherical harmonic transform, which itself requires the knowledge of all pixels. Therefore, it is advantageous to compute the quantities of interest by extending the domain of $\mathbf{C_S}$ and $\mathbf{C_N}$ to the entire sky. We refer the reader to Section E.1.2 of Appendix E for details on how this can be achieved and state the final result here (see also [160, 90]).

We encode information about masking of pixels into a diagonal $N_T \times N_T$ matrix $\mathbf{M}$, with elements $M_{ii} = 0$ if $i \in L_{\bar{p}}$, and $M_{ii} = 1$ if $i \in L_p$. Furthermore, we let $\mathbf{S}$ be the full pixel to pixel covariance matrix due to primary CMB fluctuations: $(i, j \in L)$

$$[\mathbf{S}]_{ij} = \sum_{\ell=0}^{\ell_{max}} C_\ell \left(\frac{2\ell+1}{4\pi} \right) P_\ell(\widehat{\boldsymbol{n}}_i \cdot \widehat{\boldsymbol{n}}_j), \tag{6.18}$$

$\mathbf{N}_\nu$ the pixel-to-pixel covariance matrix of the instrumental noise:

$$[\mathbf{N}_\nu]_{ij} = n_{i\nu}^2 \delta_{ij}, \tag{6.19}$$

and $\mathbf{B}_\nu$ a model of the complete beam (pixelization and instrumental):

$$[\mathbf{B}_\nu]_{ij} = A_{\text{pix}} \sum_{\ell=0}^{\ell_{max}} \left(\frac{2\ell+1}{4\pi} \right) B_{\ell\nu} W_\ell P_\ell(\widehat{\boldsymbol{n}}_i \cdot \widehat{\boldsymbol{n}}_j), \tag{6.20}$$

where A_{pix} is the area of one pixel (which is equal to $4\pi/N_T$ for all pixels in the HEALPix scheme). We further define the square root of $\mathbf{S}$:

$$[\mathbf{S}^{1/2}]_{ij} \equiv \sqrt{A_{pix}} \sum_{\ell=0}^{\ell_{max}} \sqrt{C_\ell} \left(\frac{2\ell+1}{4\pi} \right) P_\ell(\widehat{\boldsymbol{n}}_i \cdot \widehat{\boldsymbol{n}}_j). \tag{6.21}$$

Equation (6.15) may now be rewritten using the full covariance matrices, temperature data, and templates:

$$\widehat{\mathbb{P}}_k = \sum_{k'=1}^{M} (\boldsymbol{\alpha}^{-1})_{k,k'} \beta_{k'}, \tag{6.22a}$$

$$\alpha_{k,k'} = \sum_{\nu \in L_\nu} \sum_{i \in L} t_k^{(\nu)}(i) X_{k'}^{(\nu)}(i), \tag{6.22b}$$

$$\beta_k = \sum_{\nu \in L_\nu} \sum_{i \in L} \delta T_{i\nu} X_k^{(\nu)}(i). \tag{6.22c}$$

where

$$X_k^{(\nu)}(i) = \sum_{\nu' \in L_\nu} \sum_{i' \in L} [G_{\nu,\nu'}]_{ii'} t_k^{(\nu')}(i'), \tag{6.23a}$$

$$\mathbf{G}_{\nu,\nu'} = \mathbf{M}\mathbf{N}_\nu^{-1}\mathbf{M}\delta_{\nu,\nu'} - \mathbf{M}\mathbf{N}_\nu^{-1}\mathbf{M}\mathbf{B}_\nu \mathbf{S}^{1/2}\mathbf{D}^{-1}\mathbf{S}^{1/2}\mathbf{B}_{\nu'}\mathbf{M}\mathbf{N}_{\nu'}^{-1}\mathbf{M}, \tag{6.23b}$$

$$\mathbf{D} = \mathbf{1} + \mathbf{S}^{1/2}\left(\sum_{\nu \in L_\nu} \mathbf{B}_\nu \mathbf{M}\mathbf{N}_\nu^{-1}\mathbf{M}\mathbf{B}_\nu\right)\mathbf{S}^{1/2}. \tag{6.23c}$$

The computation of $\boldsymbol{X}_k^{(\nu)}$ is now reduced to solving for the quantities $\boldsymbol{g}_k^{(\nu)} = \mathbf{D}^{-1}\tilde{\boldsymbol{t}}_k^{(\nu)}$, where $\tilde{\boldsymbol{t}}_k^{(\nu)} \equiv \mathbf{S}^{1/2}\mathbf{B}_\nu \mathbf{M}\mathbf{N}_\nu^{-1}\mathbf{M}\boldsymbol{t}_k^{(\nu)}$. The other operations may be done trivially as all involved operators are either diagonal in pixel space or in harmonic space. Computing $\boldsymbol{g}_k^{(\nu)}$ is equivalent to solving the equation $\mathbf{D}\boldsymbol{g}_k^{(\nu)} = \tilde{\boldsymbol{t}}_k^{(\nu)}$, for which a number of numerical techniques are available. We use the algorithm of the conjugate gradient method with preconditioning (see e.g. [140]), which is an iterative prescription for solving large linear systems equations of the type

$$\mathbf{A}\boldsymbol{x} = \boldsymbol{y}. \tag{6.24}$$

To speed up the convergence, we construct a preconditioner matrix $\mathbf{D}_0^{-1}$ (essentially an approximation of $\mathbf{D}^{-1}$) as follows: the block corresponding to all harmonic modes with $\ell \leq 60$ is taken to be the exact inversion of the same block in $\mathbf{D}$, which is computed using a Cholesky decomposition. The rest of $\mathbf{D}_0^{-1}$ is taken to be diagonal, the elements of which are reciprecals of the corresponding diagonal element of $\mathbf{D}$. This preconditioner has already been used in other works (see e.g. [60]). We stop the conjugate gradient algorithm whenever the relative error

$$\epsilon_n = \frac{||\mathbf{A}\boldsymbol{x}_n - \boldsymbol{y}||_2}{||\boldsymbol{y}||_2} \tag{6.25}$$

is less than a specified threshold. In the case of this work, we take $\epsilon_n < 10^{-6}$. We have checked that changing this threshold to 10^{-5} does not change the results, indicating that the solution has indeed converged (see Section E.2 of Appendix E for detailed convergence tests).

Finally, we note that all $\{\boldsymbol{g}_{\boldsymbol{k}}^{(\boldsymbol{\nu})}\}$ are independent and thus may be computed in parallel. We fully employ this property. Our software, ABYSS (the sphericAl BaYesian Statistical Sampler), runs in 25 hours and 52 minutes on seven nodes (16 cores) to solve for the 12 templates on an Intel Xeon E5620. We note that the monopole and dipole take significant more time to reach the same level of precision as the other maps.

6.3 Data

6.3.1 CMB Data

We use co-added inverse-noise weighted data from nine-year maps observed by WMAP at 41 GHZ (Q-band), 62 GHZ (V-band), and 94 GHz (W-band).[6] These maps are foreground cleaned [25] and are at HEALPix resolution 9 ($N_{side} = 512$). The standard deviation of the pixel noise in each map is given by (using notation of Section 6.2.4)

$$n_{i\nu} = \frac{\sigma_0^{(\nu)}}{\sqrt{N_i^{obs}}}, \tag{6.26}$$

where $\nu \in L_\nu = \{Q, V, W\}$, $i \in L$, and $\sigma_0^{(Q)} = 2.188$ mK, $\sigma_0^{(V)} = 3.131$ mK, $\sigma_0^{(W)} = 6.544$ mK. The number of observations N_i^{obs} at pixel i is included in the maps available from the LAMBDA website. In all of our analysis, we use the 'extended temperature data analysis mask' to exclude foreground-contaminated regions of the sky from the analysis. The beam transfer function for every differencing assembly is also provided on the LAMBDA website. For a single value of ℓ, we average beam transfer function values for all differencing assemblies belonging to the same frequency channel. This is how we obtain the quantities $B_{\ell\nu}$ introduced in Section 6.2.3.

[6] http://lambda.gsfc.nasa.gov

6.3.2 Cluster Sample

We use the Meta-Catalogue of X-ray detected Clusters of galaxies (MCXC) to extract the universal pressure profile of the ICM [114].[7] The MCXC provides (amongst other quantities) sky coordinates, redshift, and M_{500} data for all clusters. With a few exceptions, luminosity is used as a mass proxy for all clusters (see equation (2) of [114]).

We perform our analysis on all 1743 MCXC clusters, as well as a subsample of 162 clusters whose first radial bin ($= 0.5 \times R_{500}$, as discussed in the next Section) is resolved by the W frequency channel of WMAP. More specifically, we obtain this subsample by requiring $d(z)\theta^{(W)} < 0.5R_{500}$, where $\theta^{(W)} = 0.12^{\circ}$ is the effective angular radius of the (averaged) W-channel detector beam. We will refer to clusters in this subsample as *resolved MCXC clusters*. Figure 6.2 shows the redshift-R_{500} distribution of all MCXC clusters, differentiating between the resolved and unresolved ones. The redshift of all (resolved) MCXC clusters range from $0.0031 - 1.26$ ($0.0031 - 0.077$) with a median of 0.14 (0.028), and their masses range from $\frac{M_{500}}{10^{14}M_{\odot}} = 0.0096 - 22.1$ ($0.0096 - 7.27$) with a median of 1.77 (0.86).

Since most MCXC clusters cannot be resolved, one expects numerical uncertainties to become important. This is why we have chosen to study a subsample in which all clusters are resolved. However, even unresolved clusters contribute to the tSZ signal, especially in the outer bins. Therefore, the price one pays for ignoring unresolved clusters is statistical information. We have analyzed both samples to see how this trade-off manifests itself in practice.

We also analyze subsamples of MCXC clusters binned according to mass. This allows us to study the dependence of various quantities, such as pressure and gas mass fraction, on the mass of clusters. Table 6.1 shows the mass range and number of clusters in every bin. We have subdivided the resolved MCXC clusters into three mass bins, and the entire MCXC sample into four bins. These bins have been chosen so that they lead to roughly similar signal-to-noise properties, characterized by the null chi-squared of pressure measurements. We have excluded the 45 most massive clusters because none of them are resolved, resulting in a measurement with extremely low significance and a nearly degenerate covariance matrix.

[7] All information about MCXC clusters may be found here: http://vizier.cfa.harvard.edu/viz-bin/VizieR?-source=J/A+A/534/A109

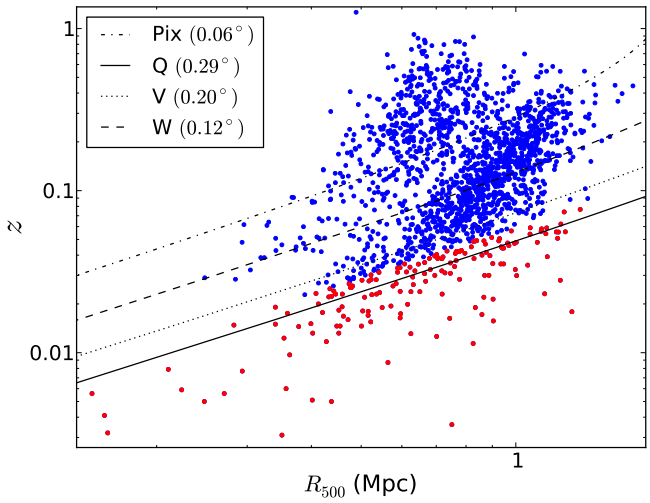


Figure 6.2: The redshift-R_{500} distribution of MCXC clusters. The red points represent clusters whose first radial bin is resolved by the W frequency channel of WMAP. The three lines plot $d(z)\theta^{(\nu)}$ for different WMAP frequency bands, where $\theta^{(\nu)} = \sqrt{\Omega^{(\nu)}/\pi}$ is the angular radius of the disk with the same effective area as the detector beam in frequency channel ν. Here $\Omega^{(\nu)}$ is the beam solid angle of frequency channel ν, which is provided on the LAMBDA website: $\Omega^{(Q)} = 0.51^2$, $\Omega^{(V)} = 0.35^2$, and $\Omega^{(W)} = 0.22^2$ (deg^2). The curve labeled 'Pix' is constructed similarly and reflects the resolution associated with pixelization.

6.4 Results

In this section, we describe our two main results. Section 6.4.1 discusses WMAP constraints on the universal pressure profile $\mathbb{P}$, and Section 6.4.2 includes our gas mass fraction measurements.

6.4.1 WMAP Constraints on the Universal Pressure Profile of the ICM

Figure 6.3 shows the result of our pressure measurements as applied to all MCXC clusters, as well as the resolved subsample defined in Section 6.3.2. As was mentioned in Section 6.2.2, we perform our analysis using the standard self-similar model $P_c^{(a)} \propto M_{500}^{2/3}$ ($\delta = 0$

Table 6.1: Binning MCXC clusters according to their mass.

Bin number	M_{500} range ($10^{14} M_\odot$)	Number of clusters
1	0.0096-2.41	1140
2	2.41-4.17	364
3	4.18-5.31	124
4	5.32-7.27	70

(a) All but the 45 most massive MCXC clusters.

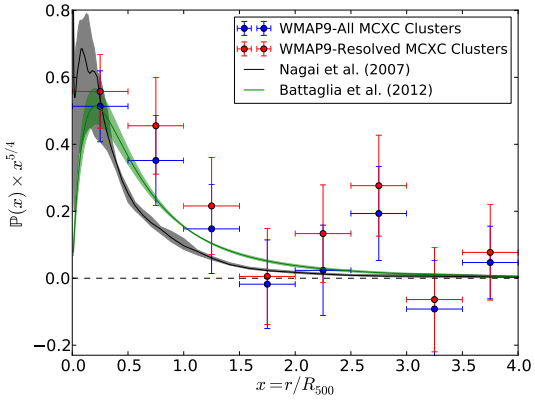
Bin number	M_{500} range ($10^{14} M_\odot$)	Number of clusters
1	0.0096-2.71	138
2	2.83-4.56	15
3	5.17-7.27	9

(b) Resolved MCXC clusters.

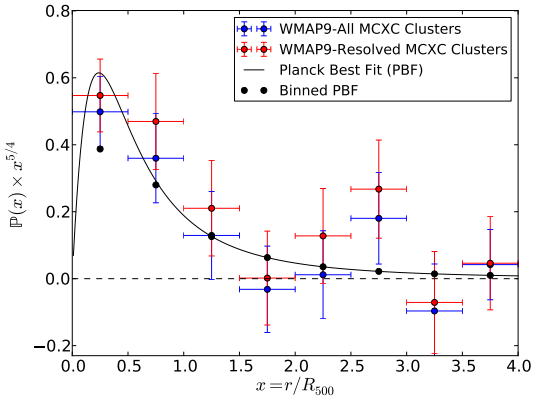
in Equation (6.9)), as well as the modified scaling $P_c^{(a)} \propto M_{500}^{2/3+0.12}$ ($\delta = 0.12$ in Equation (6.9)). These results are shown in Figure 6.3a and Figure 6.3b, respectively. There is essentially no signal beyond $1.5R_{500}$. The best fit pressure values even become negative for some bins in this regime. We have decided not to impose positivity of pressure as a prior in order to keep the statistics Gaussian and not spoil the analytic results (6.15) and (6.17). Repeating these measurements with *Planck* CMB data is expected to provide a *significantly* tighter constraint on the universal pressure profile (see Section 6.5 below).

In Figure 6.3a, we have compared our pressure measurements with two sets of simulations [107, 23], which include treatment of radiative cooling, star formation and energy feedback from supernova explosions. It is worth noting that [23] also account for feedback from active galactic nuclei, while [107] consider the effect that electrons and ions are not kept in thermal equilibrium in the outskirts [132]. Comparison of our measurements with the simulated profiles of [107] is straightforward because they compute the exact same quantity. This is not the case for [23], however, because they use R_{200} as the self-similarity scale and also consider the variation of $\mathbb{P}$ with mass and redshift. In this case, we use the $c_{200} - M_{200}$ relation of [26] to compute R_{200}, and use the fitting formula of [23] (equations (11-12) and table 1) to compute $\mathbb{P}$ for all MCXC clusters. In Figure 6.3a we have plotted in green the average of these profiles, as well as the standard deviation about their mean. Where there is signal, our pressure measurements are slightly more consistent with those of [23]. Due to the large statistical uncertainties, however, we cannot meaningfully discriminate between the two.

In Figure 6.3b, we have compared our pressure measurements with those of [121]. Given



(a) Standard self-similar scaling of pressure with mass ($\delta = 0$ in Equation (6.9)). The shaded areas mark the dispersion about the mean profiles of simulated clusters from [107] (gray), and [23] (green).



(b) Modified self-similar scaling of pressure with mass ($\delta = 0.12$ in Equation (6.9)). The black curve is the best fit GNFW profile to pressure measurements of *Planck*. The black points are obtained by a volume weighed average of *Planck*'s best fit profile over our radial bins.

Figure 6.3: WMAP-9 constraints on the universal pressure profile $\mathbb{P}$ of the ICM. The blue (red) data points are the resulting pressure profiles for all (only resolved) MCXC clusters. A cluster is considered resolved if its first radial bin subtends a solid angle larger than the effective beam area of the W frequency channel (see Section 6.3.2). Defined in Equation (6.9), δ characterizes deviation from the standard self-similar model. $\delta = 0$ corresponds to the mass variation expected in the standard self-similar model (see [107, 16]), and $\delta = 0.12$ is a modification which better captures the variation of mass scaling with radius in the REXCESS sample [16, 32].

Table 6.2: Level of detection for various pressure measurements. $\delta = 0$ (0.12) corresponds to measurements presented in Figure 6.3a (6.3b), respectively. The null chi-squared is given by $\chi_0^2 = \widehat{\mathbb{P}}^T \mathbf{C}_{\mathbb{P}}^{-1} \widehat{\mathbb{P}}$, where $\widehat{\mathbb{P}}$ are the best fit pressure measurements and $\mathbf{C}_{\mathbb{P}}$ is their associated covariance matrix. The level of detection is calculated for 8 degrees of freedom, i.e. number of radial bins.

Measurement	χ_0^2	Detection (σ)
All MCXC clusters, $\delta = 0$	259.3	15.1
All MCXC clusters, $\delta = 0.12$	262.2	15.2
Resolved MCXC clusters, $\delta = 0$	115.6	9.5
Resolved MCXC clusters, $\delta = 0.12$	118.6	9.6

that we use different radial bins, and more importantly that *Planck*'s measurements are a lot more precise, it suffices to compare our measurements with their best fit generalized Navarro-Frenk White (GNFW) profile [110, 107]. We discretize this profile by a volume-weighed average over our radial bins, which makes for a more meaningful comparison with the discretized universal pressure profile we have defined, i.e. $\{\mathbb{P}_k\}$. Our measurements are in good agreement.

Table 6.2 shows the level of confidence for our various pressure measurements. The difference between the standard and modified scalings of $P_c^{(a)}$ with mass is very small. However, the significance of detection reduces from 15.1σ to 9.5σ if we limit our sample to the resolved clusters. This may seem surprising because, looking at Figure 6.3, the uncertainties are similar in both cases and the best fit pressure values are even consistently higher in the case of resolved clusters. Note, however, that Figure 6.3 does not compare the off-diagonal elements of the covariance matrices − i.e. correlation between different bins. *In fact, the extra statistical information coming from unresolved clusters is encoded almost entirely in the off-diagonal correlations.* We refer the reader to Section E.5 of Appendix E for the full covariance matrix and numerical values of $\{\widehat{\mathbb{P}}_k\}$. To get some sense for the nature of correlations, however, we have shown a normalized version of $\mathbf{C}_{\mathbb{P}}$ for both samples in Table 6.3. Nearby bins are anti-correlated in both cases, but more so for the sample containing all clusters. The extra information contained in these anti-correlations can be quantitatively described by examining the eigenvalues and eigenvectors of the covariance matrix $\mathbf{C}_{\mathbb{P}}$. We refer the reader to Section E.4 of Appendix E for a detailed discussion of this point and state the results here. In the case of the resolved clusters, the eigenvectors with the three largest eigenvalues are responsible for most of the contribution to χ_0^2. For the whole MCXC sample, however, all eigenvectors contribute more or less equally. Moreover, eigenvectors corresponding to larger eigenvalues carry most of their weight from the inner

Table 6.3: The normalized covariance matrix of the universal pressure profile for all MCXC clusters (blue or top-right), as well as only the resolved ones (red or bottom-left). The modified self-similar model is used for these measurements (i.e. $\delta = 0.12$ in Equation (6.9)). To construct these matrices, let $\mathbf{C}_{\mathbb{P}}^{\text{all}}$ be the covariance matrix for analysis done on all MCXC clusters. Construct a diagonal matrix $\mathbf{\Delta}^{\text{all}}$ such that $\mathbf{\Delta}_{k,k}^{\text{all}} = \sqrt{[\mathbf{C}_{\mathbb{P}}^{\text{all}}]_{k,k}}$, where k runs over the different radial bins. We define the normalized covariance matrix via $\mathbf{D}^{\text{all}} = [\mathbf{\Delta}^{\text{all}}]^{-1}\mathbf{C}_{\mathbb{P}}[\mathbf{\Delta}^{\text{all}}]^{-1}$, which normalizes all diagonal elements of $\mathbf{C}_{\mathbb{P}}^{\text{all}}$ to one. By the same construction, let $\mathbf{D}^{\text{res}}$ be the resulting normalized covariance matrix for analysis done on resolved MCXC clusters. The blue or top-right (red or bottom-left) numbers in this table denote the off-diagonal elements of $\mathbf{D}^{\text{all}}$ ($\mathbf{D}^{\text{res}}$), respectively.

blue	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7	Bin 8
Bin 1	1.000	-0.727	0.282	-0.082	0.029	-0.022	0.011	-0.001
Bin 2	-0.612	1.000	-0.671	0.214	-0.051	0.005	-0.003	0.006
Bin 3	0.277	-0.497	1.000	-0.633	0.144	0.000	-0.023	0.018
Bin 4	-0.069	0.232	-0.445	1.000	-0.598	0.109	0.009	-0.014
Bin 5	0.033	-0.039	0.147	-0.434	1.000	-0.594	0.117	-0.004
Bin 6	-0.024	-0.007	-0.013	0.091	-0.451	1.000	-0.597	0.156
Bin 7	-0.001	-0.023	-0.041	-0.015	0.108	-0.442	1.000	-0.697
Bin 8	-0.011	-0.043	-0.044	-0.075	-0.053	0.085	-0.476	1.000

bins. Therefore, in the case of resolved clusters, mostly the inner bins are contributing to the signal, whereas for the whole MCXC sample, there is also contribution from outer bins. This analysis reassures us that even unresolved clusters contribute to the tSZ signal in the outskirts of the ICM.

Although the unresolved clusters add to the tSZ signal in the outer bins, one expects numerical uncertainties associated with them. This is especially worrisome for those on sub-pixel scales, where certain approximations, such as a spherically averaged pixel transfer function, break down. In order to get an estimate for how large such uncertainties are, we performed our analysis on all MCXC clusters using higher resolution WMAP sky maps ($N_{side} = 1024$). The result is shown in Section E.2.2 of Appendix E. For all radial bins, this discrepancy is at most at the 1σ level and is random in nature.

Our pressure measurements are also affected by the uncertainty present in determining masses of clusters. In Section E.3 of Appendix E, we have investigated this issue by considering 62 MCXC clusters which are also in the Early Release SZ (ESZ) sample [120].

The ESZ mass estimates are systematically higher on average by about 12 percent. This results in systematically lower pressure measurements (where there is actual signal), but it is only at the 1σ level (see Section E.3 of Appendix E for details). We will return to this issue in Section 6.4.2, because this effect is no longer small when determining gas mass fraction.

The results of our analysis as applied to the cluster subsamples introduced in Table 6.1 are included in Section E.5 of Appendix E. Because of the large statistical uncertainties, comparing the pressure profile of different mass bins is not terribly illuminating. We will, however, discuss the implications for gas mass fraction in the next Section.

6.4.2 Gas Mass Fraction

The density of gas $\rho_g^{(a)}(r)$ in the a^{th} cluster with temperature profile $T^{(a)}(r)$ takes the form

$$\rho_g^{(a)}(r) = \frac{\mu_e m_p P_e^{(a)}(r)}{k_B T^{(a)}(r)}, \tag{6.27}$$

where m_p is the proton mass and $\mu_e = \frac{2}{X+1} \simeq 1.14$ is the mean molecular weight per free electron for a cosmic hydrogen abundance of $X \simeq 0.76$.[8] As it has been the case for the electron pressure profile (see Equation (6.6)), we assume a universal temperature profile

$$T^{(a)}(r) = T_c^{(a)} \mathbb{T}(r/r_c^{(a)}), \tag{6.28}$$

which in turn implies

$$\rho_g^{(a)}(r) = \rho_c^{(a)} \frac{\mathbb{P}(r/r_c^{(a)})}{\mathbb{T}(r/r_c^{(a)})}, \tag{6.29a}$$

$$\rho_c^{(a)} \equiv \frac{\mu_e m_p P_c^{(a)}}{k_B T_c^{(a)}}. \tag{6.29b}$$

The volume-averaged gas density at radius r takes the form

$$\begin{aligned} \bar{\rho}_g^{(a)}(<r) &= \frac{1}{\frac{4}{3}\pi r^3} \int_0^r \rho_g^{(a)}(r') 4\pi r'^2 dr' \\ &= 3\rho_c^{(a)} (r/r_c^{(a)})^{-3} \int_0^{r/r_c^{(a)}} \frac{\mathbb{P}(x)}{\mathbb{T}(x)} x^2 dx. \end{aligned} \tag{6.30}$$

[8] The free electron number density is $n_e = n_H + 2n_{He}$, where n_H and n_{He} are the Hydrogen and Helium number density. The cosmic hydrogen abundance is $X = n_H/(n_H + 4n_{He})$. It then follows that $\rho_b = m_p(n_H + 4n_{He}) = \frac{2m_p}{X+1} n_e = \frac{2m_p}{X+1} \frac{P_e}{k_B T}$, where m_p is the proton mass.

Given that we have considered radial bins throughout which $\mathbb{P}$ is constant, we shall also bin the temperature profile:

$$\mathbb{T}_k = \frac{1}{\frac{4\pi}{3}(n_k^3 - n_{k-1}^3)} \int_{n_{k-1}}^{n_k} \mathbb{T}(x) x^2 dx, \tag{6.31}$$

where n_k is the value of the k^{th} radial bin in units of $r_c^{(a)}$, with $n_0 \equiv 0$ (see Section 6.2.2).[9] It then follows that

$$\bar{\rho}_g^{(a)}(< r) = \rho_c^{(a)} \sum_{k=1}^{N_b} V_k(r/r_c^{(a)}) \mathbb{P}_k, \tag{6.32}$$

where N_b is the total number of radial bins and

$$V_k(x) = \begin{cases} \frac{1}{\mathbb{T}_k x^3}\left[n_k^3 - n_{k-1}^3\right] & \text{if } k \leq k_* \\ \frac{1}{\mathbb{T}_k x^3}\left[x^3 - n_{k_*}^3\right] & \text{if } k = k_* + 1 \\ 0 & \text{if } k > k_* + 1. \end{cases} \tag{6.33}$$

Here k_* is an integer defined via $n_{k_*} \leq x < n_{k_*+1}$. We assume the total matter density of the a^{th} cluster to be of the NFW form:

$$\rho_m^{(a)}(r) = \frac{\rho_s^{(a)}}{r/r_s^{(a)}\left(1 + r/r_s^{(a)}\right)^2}. \tag{6.34}$$

Defining $c_\Delta^{(a)} \equiv R_\Delta^{(a)}/r_s^{(a)}$ and $\delta^{(a)} = \rho_s^{(a)}/\rho_{crit}(z_a)$, it may be checked that the total mass enclosed within a radius r is equal to [10]

$$M_m^{(a)}(< r) = \frac{m(c_\Delta^{(a)} r/R_\Delta^{(a)})}{m(c_\Delta^{(a)})} M_\Delta^{(a)}, \tag{6.35a}$$

$$m(x) \equiv \ln(1+x) - \frac{x}{1+x}, \tag{6.35b}$$

where $R_\Delta^{(a)}$ and $M_\Delta^{(a)}$ were defined in Equation (6.8). Also, it follows from Equation (6.8) that

$$\frac{m(c_\Delta^{(a)})}{[c_\Delta^{(a)}]^3} = \frac{\Delta}{3\delta^{(a)}}. \tag{6.36}$$

[9] We will discretize all continuous profiles over our radial bins, because our pressure measurements are discrete by construction. In the case of temperature, it might seem more natural from Equation (6.30) to discretize $1/\mathbb{T}$ instead of $\mathbb{T}$. We have checked that the difference between these discretization schemes is insignificant.

[10] Here we have adopted the notation of [26].

Therefore, knowing $c_\Delta^{(a)}$ (for any Δ) determines $\delta^{(a)}$, or equivalently $\rho_s^{(a)}$. We estimate the concentration parameter from the $c_{200} - M_{200}$ relation of [26]:

$$c_{200} = 5.9 D(z)^{0.54} \nu(M_{200}, z)^{-0.35}, \tag{6.37a}$$

$$\nu(M, z) \simeq \frac{1}{D(z)} \left[1.12 \left(\frac{M}{5 \times 10^{13} h^{-1} M_\odot} \right)^{0.3} + 0.53 \right], \tag{6.37b}$$

where $D(z)$ is the linear growth factor normalized to 1 at $z = 0$.[11] As was the case with the temperature of baryons, we similarly bin $\rho_m^{(a)}(r)$

$$\rho_{m,k}^{(a)} = \frac{M_m^{(a)}(< r_k^{(a)}) - M_m^{(a)}(< r_{k-1}^{(a)})}{\frac{4\pi}{3} \left\{ [r_k^{(a)}]^3 - [r_{k-1}^{(a)}]^3 \right\}}, \tag{6.39}$$

where as before $r_k^{(a)} = n_k r_c^{(a)}$. Finally, the volume-averaged matter density $\bar{\rho}_m^{(a)}$ up to radius r is

$$\bar{\rho}_m^{(a)}(< r) = \sum_{k=1}^{N_b} \tilde{V}_k(r/r_c^{(a)}) \rho_{m,k}^{(a)}, \tag{6.40}$$

where (k_* being defined as above)

$$\tilde{V}_k(x) = \begin{cases} \frac{1}{x^3} \left[n_k^3 - n_{k-1}^3 \right] & \text{if } k \leq k_* \\ \frac{1}{x^3} \left[x^3 - n_{k_*}^3 \right] & \text{if } k = k_* + 1 \\ 0 & \text{if } k > k_* + 1. \end{cases} \tag{6.41}$$

The average gas mass-fraction up to radius r in the a^{th} cluster takes the form:

$$\begin{aligned} f_{gas}^{(a)}(< r) &\equiv \frac{\bar{\rho}_g^{(a)}(< r)}{\bar{\rho}_m^{(a)}(< r)} \\ &= \frac{\rho_c^{(a)}}{\bar{\rho}_m^{(a)}(< r)} \sum_{k=1}^{N_b} V_k(r/r_c^{(a)}) \mathbb{P}_k. \end{aligned} \tag{6.42}$$

[11] A good approximation to $D(z)$ is given by $D(z) = \frac{D_1(z)}{D_1(0)}$, where (see [45])

$$D_1(z) \simeq \frac{5\Omega_m(z)}{2(1+z)} \left\{ \Omega_m(z)^{4/7} - \Omega_\Lambda(z) + [1 + \Omega_m(z)/2] [1 + \Omega_\Lambda(z)/70] \right\}^{-1}. \tag{6.38}$$

Here $\Omega_m(z) = \Omega_m (1+z)^3 / E(z)^2$ and $\Omega_\Lambda(z) = \Omega_\Lambda / E(z)^2$.

In order to make a meaningful comparison with the universal gas mass-fraction, we average this quantity over all clusters

$$f_{gas}(< x) \equiv \frac{1}{N_c}\sum_{a=1}^{N_c} f_{gas}^{(a)}(< xr_c^{(a)}),$$

$$= \mathcal{A}(x)\sum_{k=1}^{N_b} V_k(x)\mathbb{P}_k, \tag{6.43a}$$

$$\mathcal{A}(x) \equiv \frac{1}{N_c}\sum_{a=1}^{N_c}\frac{\rho_c^{(a)}}{\bar{\rho}_m^{(a)}(< xr_c^{(a)})}. \tag{6.43b}$$

Since $f_{gas}(< x)$ is a linear combination of $\{\mathbb{P}_k\}$, it is a Gaussian random variable with mean and variance

$$\langle f_{gas}(< x)\rangle = \mathcal{A}(x)\sum_{k=1}^{N_b} V_k(x)\widehat{\mathbb{P}}_k \tag{6.44}$$

$$\sigma^2_{f_{gas}(<x)} \equiv \left\langle [f_{gas}(< x) - \langle f_{gas}(< x)\rangle]^2 \right\rangle$$

$$= \mathcal{A}(x)^2\sum_{k,k'=1}^{N_b} V_k(x)V_{k'}(x)[\mathbf{C}_{\mathbb{P}}]_{kk'}. \tag{6.45}$$

The averaging scheme we have adopted in Equation (6.43) may seem arbitrary. One could, for instance, assign different weights to different clusters. If $P_c^{(a)}$ and $T_c^{(a)}$ scale similarly with mass, different averaging schemes would differ by a negligible amount. This is because the only variation in f_{gas} amongst different clusters would be due to the scaling of c_{500} with mass, which is fairly mild. As a result, given the temperature profile we have adopted (see Equation (6.46)), we use the standard self-similar scaling of $P_c^{(a)}$ with mass ($\delta = 0$ in Equation (6.9) below) to compute f_{gas}.

We use the average temperature profile of [157]:

$$\mathbb{T}(x) = 1.35\frac{(x/0.045)^{1.9} + 0.45}{(x/0.045)^{1.9} + 1}\frac{1}{[1 + (x/0.6)^2]^{0.45}}, \tag{6.46a}$$

$$\frac{T_c^{(a)}}{5\text{ keV}} = \left[\frac{M_{500}^{(a)}E(z)}{3.41\times 10^{14}h^{-1}M_\odot}\right]^{1/1.51}. \tag{6.46b}$$

This is an approximation to the averaged profile of about a dozen low-redshift X-ray clusters, with measurements obtained for $r < R_{500}$ [157]. The scatter about the mean profile

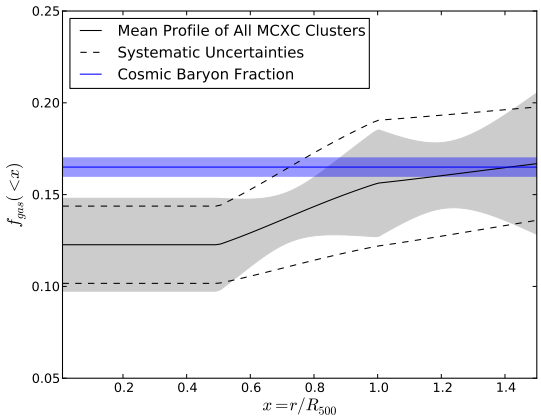
is about 15 percent. The assumption of universality (i.e. Equation (6.28)) may be easily relaxed if temperature measurements for individual clusters are available. In the case of our present work, however, this option is not viable since a large cluster sample is required to compensate for WMAP's insufficient sensitivity. To get an estimate for how this assumption affects our f_{gas} measurements, consider an average 100ϵ percent scatter about the universal profile. For a relatively large sample of clusters, as is the case with our measurements, changing the individual temperature profiles by $\sim 100\epsilon$ percent would introduce a systematic uncertainty of order ϵ^2 in f_{gas}.[12] Given the large statistical uncertainties in our pressure measurements, this effect is small.

Systematic uncertainties associated with estimating masses of clusters have a more significant effect on gas mass fraction. As mentioned in the previous section, we have investigated this issue in Appendix E.3 by considering 62 MCXC clusters which are also in the ESZ sample [120]. The ESZ mass estimates are systematically higher by about 12 percent, which causes lower pressure measurements at the 1σ level. Because of the scaling of temperature with mass (Equation (6.46)), this decreases f_{gas} by about **20** percent (see Figure E.3). Repeating our analysis with *Planck* CMB data is expected to reduce statistical errors significantly (see Section 6.5 and [121]). In that case, f_{gas} measurements would be solely dominated by systematic uncertainties associated with determining masses of clusters.

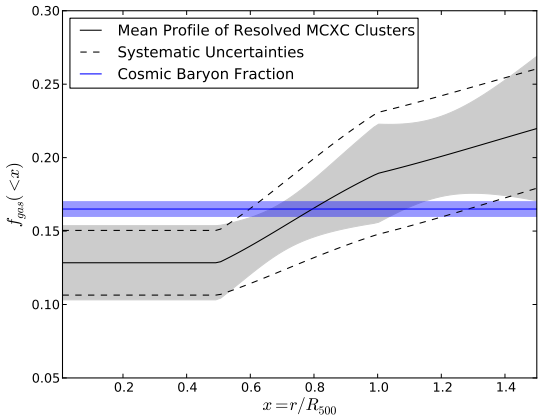
Figure 6.4 shows the result of our analysis, as applied to all MCXC clusters, as well as just the resolved ones. Given the large statistical and systematic uncertainties, we extrapolated the temperature profile of [157] out to R_{200}. The black solid curves show the average gas mass fraction, computed using Equation (6.44). The shaded areas represent the standard deviation in the measurement of f_{gas} as given by Equation (6.45). The dashed black lines show the expected systematic uncertainty about the mean gas mass profile ($\sim$ 20 percent), mostly due to cluster mass estimates. Considering both statistical and systematic errors, our results are fully consistent with the cosmic baryonic fraction up to R_{200}. Given the large error bars, accounting for all baryons in stars does not change this conclusion.

Figure 6.5 shows the dependence of gas mass fraction (up to R_{500}) on the cluster subsamples which we have binned according to mass (see Table 6.1). A general trend of increasing f_{gas} with M_{500} can be seen. Due to the large error bars, however, we cannot

[12] This is because $f_{gas} \propto \sum_a 1/T^{(a)}$, where $T^{(a)}$ is the temperature of the a^{th} cluster. Let $\{\epsilon_a\}$ be realizations of a gaussian random variable with zero mean and standard deviation ϵ. Changing the temperature $T^{(a)} \to T^{(a)}(1+\epsilon_a)$ is equivalent to $\sum_a 1/T^{(a)} \to \sum_a 1/T^{(a)}(1-\epsilon_a+\epsilon_a^2+\cdots)$. When $T^{(a)}$ doesn't change drastically from cluster to cluster, the contribution of the term linear in ϵ_a is on the order of $\sqrt{N}\epsilon$, while the second order term contributes about $N\epsilon^2$. When $N > 1/\epsilon^2$, the ϵ^2 term dominates.

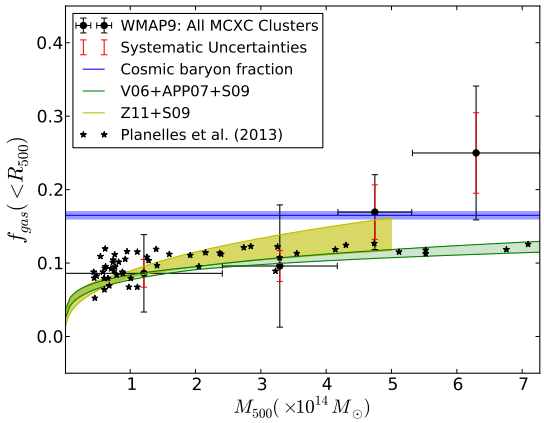


(a) Gas mass fraction for all MCXC clusters.

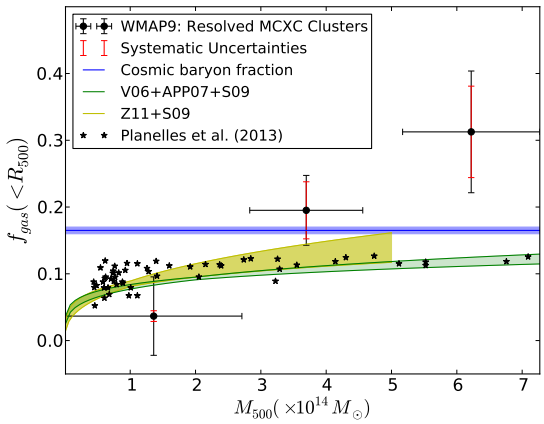


(b) Gas mass fraction for resolved MCXC clusters.

Figure 6.4: Gas mass fraction of all MCXC, as well as the resolved subsample. A cluster is considered resolved if its first radial bin subtends a solid angle larger than the effective beam area of the W frequency channel (see Section 3.2). The standard self-similar model of pressure is assumed for both measurements ($\delta = 0$ in Equation (6.9)). The black solid curves show the average gas mass fraction, computed using Equation (6.44). The shaded areas represent the standard deviation in the measurement of f_{gas} as given by Equation (6.45). The dashed black lines show the expected systematic uncertainty about the mean gas mass profile, mostly due to cluster mass estimates. The cosmic gas mass fraction is obtained by fitting ΛCDM to WMAP9+SPT+ACT data and is equal to $\Omega_b/\Omega_m = 0.165 \pm 0.005$ [75].



(a) Binning the entire MCXC sample according to mass, as shown in Table 6.1a.



(b) Binning the resolved MCXC sample according to mass, as shown in Table 6.1b.

Figure 6.5: Dependence of f_{gas} on the mass of clusters. All mass bins are defined in Table 6.1. The black dots show f_{gas} measurements up to R_{500}, with black vertical bars denoting statistical errors. The red vertical bars show systematic uncertainties expected due to cluster mass estimates. Following [124], we compare our measurements with two different observational samples: a combined sample of 41 clusters and groups from [157], [15] and [150] (V06+APP07+S09), shown as the green region, and the sample obtained from the combination of the data by [165] and [150] (Z11+S09), shown as the yellow area (see Table 1 of [124]). The black stars show f_{gas} obtained from hydrodynamical simulations carried out by [124], which include radiative cooling, star formation and feedback from supernovae and active galactic nuclei.

make any statistically significant statement about this dependence. Following [124], we compare our measurements with two different observational samples: a combined sample of 41 clusters and groups from [157], [15] and [150] (V06+APP07+S09), shown as the green region, and the sample obtained from the combination of the data by [165] and [150] (Z11+S09), shown as the yellow area (see Table 1 of [124]). The black stars show f_{gas} obtained from hydrodynamical simulations carried out by [124], which include radiative cooling, star formation and feedback from supernovae and active galactic nuclei. Considering both statistical and systematic errors, our measurements are consistent with both the observational and numerical results.

6.5 Discussion and Future Work

To get an idea for how much our measurements could improve using *Planck* CMB data, we *estimate* here the pressure covariance matrix $\mathbf{C}_{\mathbb{P}}$ expected from *Planck*. We consider the six *Planck*-HFI channels, which have central frequencies $100, 143, 217, 353, 545$, and 857 GHz, at HEALPix resolution $N_{side} = 2048$. We assume a homogeneous detector noise constructed by averaging the noise variance of all pixels for a given frequency channel:

$$[\mathbf{C_N}]_{i\nu,i'\nu'} = n_\nu^2 \delta_{\nu\nu'}\delta_{ii'}, \tag{6.47}$$

where $n_{100} = 50.6$ μK, $n_{143} = 20.1$ μK, $n_{217} = 27.1$ μK, $n_{353} = 0.1$ mK, $n_{545} = 28.1$ mK, $n_{857} = 27.9$ mK. We assume Gaussian instrumental beams with Full-Width-Half-Maximum (FWHM) of $9.5, 7.1, 5.0, 5.0, 5.0, 5.0$ arcmins, for the six HFI channels in order of increasing frequency [123]. Finally, we assume no masking (i.e. $\mathbf{M} = \mathbf{1}$) but remove all clusters that are masked out from our templates.

Because we have assumed a homogeneous detector noise and no masking, the matrix $\mathbf{D}$ introduced in Equation (6.23) can be inverted analytically, resulting in the pressure covariance matrix:

$$[\mathbf{C}_{\mathbb{P}}^{-1}]_{kk'} = \sum_{\ell=0}^{\ell_{max}} \sum_{m=-\ell}^{\ell} \sum_{\nu} (\bar{t}_k^{(\nu)})_{lm} (t_{k'}^{(\nu)})_{lm} \xi_\ell^{(\nu)} - \frac{C_\ell}{1 + C_\ell \xi_\ell} \left[\sum_{\nu} (\bar{t}_k^{(\nu)})_{lm} \xi_\ell^{(\nu)} \right] \left[\sum_{\nu} (t_{k'}^{(\nu)})_{lm} \xi_\ell^{(\nu)} \right], \tag{6.48}$$

where

$$\xi_\ell^{(\nu)} = \frac{(B_{\ell\nu} W_\ell)^2}{N_\ell^{(\nu)}} \tag{6.49a}$$

$$\xi_\ell = \sum_\nu \xi_\ell^{(\nu)} \tag{6.49b}$$

$$N_\ell^{(\nu)} = A_{pix} n_\nu^2. \tag{6.49c}$$

Here A_{pix} denotes the pixel area and all other quantities are defined in Sections 6.2.3 and 6.2.4.

We generate our templates $\boldsymbol{t_k^{(\nu)}}$ using the resolved MCXC clusters which are not masked (total of 122 clusters), with a standard self-similar pressure-dependent scaling ($\delta = 0$ in Equation (6.9)). In order to make sure this estimate is reasonable, we computed the same quantity with WMAP9 data, using the Q, V, and W channels. Assuming the best-fit pressure values $\widehat{\mathbb{P}}$ remain the same, this leads to a null chi-squared of $\chi_0^2 = 124.551$, which is reasonably close to the actual value $\chi_0^2 = 115.626$. (We use the same 8 radial bins as for our pressure measurements, i.e. there are 8 degrees of freedom here.) Estimating the covariance matrix for *Planck* and using the same best-fit pressure values, we obtain $\chi_0^2 = 66154.8$. Therefore, assuming that the signal does not change, we expect the statistical uncertainties to reduce by a factor of $\sim \sqrt{66154.8/124.551} = \mathbf{23.9}$. This is a *significant* improvement, which will allow us to consider finer bins and possibly probe the ICM pressure to larger radii. Figure 6.6 compares the expected error for different bins with those of WMAP9.

Our analysis does not account for the uncertainty present in modelling of beams. In the case of *Planck*, the beam uncertainty is modelled by

$$B_{\ell\nu} = B_{\ell\nu}^{\text{mean}} \exp\left[\sum_{k=1}^{n_{modes}} g^k E_{\ell\nu}^k\right], \tag{6.50}$$

where $\{g^k\}$ are independent Gaussian random variables with unit variance, and $E_{\ell\nu}^k$ is the k^{th} eigenmode of the beam covariance matrix [117].[13] In order to see how this uncertainty affects our results, we have computed the best-fit pressure profile $\{\widehat{\mathbb{P}}_k\}$ for 100 different realizations of the beams. To do so, we created mock CMB skies which contain the SZ signal, primary CMB and noise, and repeated the analysis outlined above for our *Planck* forecast.[14] We find $\delta\widehat{\mathbb{P}}_k \sim 0.01\sqrt{[\mathbf{C}_{\mathbb{P}}]_{kk}}$, where $\delta\widehat{\mathbb{P}}_k$ denotes the standard deviation of the

[13] The coefficients $E_{\ell\nu}^k$ are contained in the RIMO beam files of *Planck*. Also, we use $n_{modes} = 5$.

[14] The input pressure profile needed to create the SZ signal is taken to be the best-fit pressure values used in our *Planck* forecast analysis.

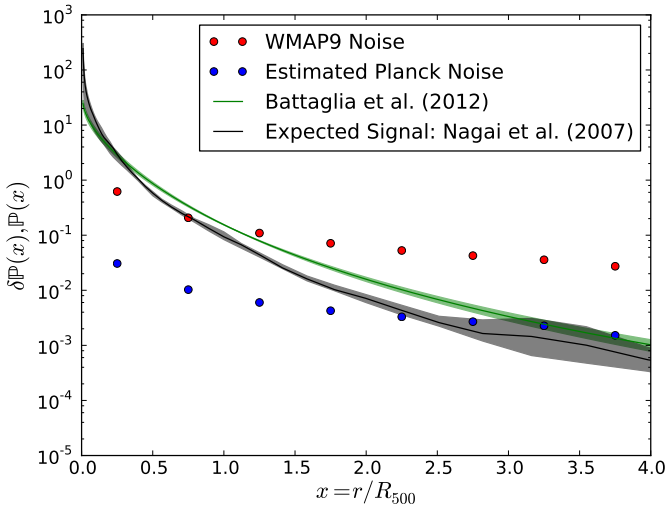


Figure 6.6: Comparison of the statistical noise expected from analysis with *Planck* CMB data vs. WMAP9. The blue data points are the estimated noise expected from repeating our analysis with *Planck* CMB data (see Equation (6.48)). The red points correspond to the same quantity for WMAP9 measurements, applied to the resolved MCXC clusters with standard self-similar pressure-dependent scaling ($\delta = 0$ in Equation (6.9)). The green and black curves show the expected pressure signal from simulated clusters of [23] and [107], respectively.

100 values of $\widehat{\mathbb{P}}_k$ obtained through our simulations. Therefore, effects of beam modeling are quite small relative to the statistical uncertainty due to primary CMB and instrumental noise.

We have also ignored the impact of foreground residuals in our formalism. Our current model is sufficient for WMAP foreground cleaned maps but not for *Planck*, because dust emission dominates at high frequencies and there are other emissions (e.g. CIB, radio and infrared point sources) which are not negligible and can not be modelled easily. To get an estimate for this effect, we consider foreground templates created by taking the difference of low and high frequency sky maps. More specifically, we created four templates by taking the difference between $030 - 044$, $044 - 070$, $353 - 545$, and $545 - 857$ *Planck* sky maps. Because different frequency channels have different beam and noise properties, we smoothed these maps using a Gaussian window function with FWHM$= 0.006$ radians $= 21'$. Considered as a template, each difference-map contributes a different coefficient to

the total temperature anisotropy, depending on the frequency band. We then estimated the expected pressure covariance matrix as above, this time using only $100, 143, 217$ GHz frequency channels. Assuming the same best-fit pressure values, accounting for foreground residuals decreases the null chi-squared by about 0.5 percent.

We have not addressed the issue of point source contamination so far. In our framework, it is not feasible to fit locally for contribution of point sources, given the large number of clusters. We did try to account for them by assigning a constant absolute luminosity per frequency channels to all clusters. The results, however, change only by a negligible amount.

6.6 Conclusions

We have introduced a statistically-optimal and model-independent framework for extracting the universal pressure profile of the hot gas in the intracluster medium. The thermal Sunyaev-Zeldovich effect makes this possible because it is linearly proportional to the integral of the electron pressure along the line of sight. We use the principle of maximum likelihood to find best-estimate values for the radially binned values of the pressure profile, as well as the full covariance matrix governing their uncertainties. Once reformulated in the proper mathematical framework, the main technical challenge is solving a very large system of linear equations, which we do numerically by employing the conjugate gradient method.

We applied our methodology to WMAP9 data and various subsamples of the MCXC catalogue. In the case of all MCXC clusters, we extract the pressure profile with a high accuracy at $\sim 15\sigma$ confidence, with possible systematics uncertainties dragging our detection down to $\sim 14\sigma$. We also considered a subsample of the MCXC clusters completely resolved by the W frequency channel of WMAP, resulting in a $\sim 9\sigma$ detection. An estimation of the pressure covariance matrix expected from *Planck* suggests that the current signal-to-noise will improve by a factor of ~ 24.

Assuming a temperature profile motivated by X-ray observations, we computed the average gas mass fraction as a function of radius. We argued that systematic uncertainties associated with estimating mass of clusters could have a drastic effect (~ 20 percent) on gas mass fraction. Considering both statistical and systematic errors, our results are fully consistent with the cosmic baryonic fraction and the expected gas mass fraction in halos, up to R_{200}.

We also made a first attempt at studying the dependence of gas mass fraction on the

mass of clusters. Due to the large error bars, we cannot make any statistically significant statements about this dependence. Nevertheless, our measurements suggest that gas mass fraction increases with the mass of clusters, which is consistent with findings from X-ray measurements [157, 15, 150, 165] and numerical simulations [124] .

APPENDICES

Appendix A

Supplementary material for Chapter 2

A.1 Definitions regarding causal structure of spacetime

Let (M, g) be a connected, four-dimensional smooth manifold M without boundary, together with a smooth pseudo-Riemannian metric of Lorentzian signature g. Further, assume that (M, g) is temporally oriented.

Definition 1. *The chronological future (past) $I^+(p)$ ($I^-(p)$) of a point $p \in M$ is the set of points $q \in M$ such that there exists a future (past)-directed smooth timelike curve $\gamma : I \to M$ (i.e. a smooth curve whose tangent vector is everywhere non-vanishing, timelike, and future (past)-pointing) and points $t_1, t_2 \in I$ such that $t_1 \leq t_2$, $\gamma(t_1) = p$ and $\gamma(t_2) = q$. Here I is a connected subset of the real line. If $q \in I^+(p)$, we write $p \prec q$, and simililary if $q \in I^-(p)$ we write $p \succ q$.*

Definition 2. *The causal future (past) $J^+(p)$ ($J^-(p)$) of a point $p \in M$ is the set of points $q \in M$ such that there exists a future (past)-directed smooth non-spacelike (i.e. timelike or null) curve $\gamma : I \to M$ and points $t_1, t_2 \in I$ such that $t_1 \leq t_2$, $\gamma(t_1) = p$ and $\gamma(t_2) = q$. Here I is a connected subset of the real line. If $q \in J^+(p)$, we write $p \preceq q$, and simililary if $q \in J^-(p)$ we write $p \succeq q$.*

Definition 3. *A spacetime (M, g) is future (past) distinguishing if and only if for all $p, q \in M$: $I^+(p) = I^+(q) \Rightarrow p = q$ ($I^-(p) = I^-(q) \Rightarrow p = q$).*

Definition 4. *Given two spacetimes (M, g) and (M', g'), a bijection $f : M \to M'$ is a smooth conformal isometry if and only if f and f^{-1} are smooth, and there is a smooth non vanishing map $\Omega : M' \to \mathbb{R}$ such that $f_*(g) = \Omega^2 g'$, where f_* is the push-forward map.*

Definition 5. *Given two spacetimes (M, g) and (M', g'), a map $f : M \to M'$ is a causal isomorphism if it is a bijection and for all points $x, y \in M$, $x', y' \in M' : x \preceq y \Leftrightarrow f(x) \preceq f(y)$ and similarly $x' \preceq y' \Leftrightarrow f^{-1}(x') \preceq f^{-1}(y')$.*

A.2 Proof of Inequality (2.7)

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.3 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) $k_1 k_4 - k_2 k_3 = 1$. Then, if we let $\cosh(\phi) = \frac{k_1+k_4}{2}$, $\gamma = \sqrt{k_3/k_2}$, and $\sinh(\chi) = \frac{k_1-k_4}{2\sqrt{k_2 k_3}}$, 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 2.2b 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

[154]), 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 [154]) 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 [154])

$$\int_0^\infty s^\mu e^{-as^D}\ln(s)\,\mathrm{d}s = \frac{\Gamma(\frac{\mu+1}{D})}{D^2a^{\frac{\mu+1}{D}}}\left[\psi(\frac{\mu+1}{D})-\ln(a)\right], \tag{B.8}$$

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 [154]), it follows that

$$(Z^{\frac{1}{2}} s)^{-N+\frac{1}{2}} K_{N-\frac{1}{2}}(Z^{\frac{1}{2}} s) = (-1)^{N-1} 2^{N-\frac{3}{2}} \pi (Z^{\frac{1}{2}} s)^{-2N+1} \sum_{k=0}^{N} \frac{(Z s^2/4)^k}{k! \Gamma(k - N + \frac{3}{2})} \tag{B.13a}$$

$$+ \frac{(-1)^N 2^{-N-\frac{1}{2}} \pi}{\Gamma(N + \frac{1}{2})} \tag{B.13b}$$

$$+ \frac{(-1)^N 2^{-N-\frac{5}{2}} \pi}{\Gamma(N + \frac{3}{2})} s^2 Z \tag{B.13c}$$

$$+ \mathcal{O}(Z^{3/2}).$$

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 [52], 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 [154] 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 [154],

$$K_1(Z^{\frac{1}{2}} s) = \frac{-1}{Z^{\frac{1}{2}}} \frac{d}{ds} K_0(Z^{\frac{1}{2}} s). \tag{B.27}$$

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 [154], 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 [154]),

$$e^Z E_1(Z) = 2\int_0^\infty e^{-t}K_0(\sqrt{2zt})\,\mathrm{d}t, \tag{B.41}$$

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 [154]),

$$\frac{d}{dz}\left[e^z E_p(z)\right] = e^z E_p(z)\left(1+\frac{p-1}{z}\right) - \frac{1}{z}, \tag{B.43}$$

$$pE_{p+1}(z) + zE_p(z) = e^{-z}, \tag{B.44}$$

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 [145] 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 [145] and [24]. 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 Klein-Gordon Inner-Product

Here we show that the KG inner product (4.3) is independent of the Cauchy surface on which it is evaluated, for solutions of the Klein-Gordon equation. Let $\mathcal{M}$ be the region bounded by a pair of Cauchy surfaces Σ_{t_1} and Σ_{t_2}, such that Σ_{t_1} is to the causal past of Σ_{t_2}. Let $f(x)$ and $g(x)$ be two complex solutions of the KG equation

$$(\Box - m^2)f(x) = 0, \qquad (\Box - m^2)g(x) = 0, \tag{C.1}$$

which induce initial data of compact support on all Σ_t. Then:

$$0 = i\int_{\mathcal{M}} \left\{\overline{f}(x)(\Box - m^2)g(x) - g(\Box - m^2)\overline{f}(x)\right\} \mathrm{d}V_x \tag{C.2}$$

$$= i\int_{\mathcal{M}} \left\{\overline{f}(x)\nabla_\mu\nabla^\mu g(x) - g(x)\nabla_\mu\nabla^\mu\overline{f}(x)\right\} \mathrm{d}V_x \tag{C.3}$$

$$= i\int_{\mathcal{M}} \nabla_\mu\left(\overline{f}(x)\nabla^\mu g(x) - g(x)\nabla^\mu\overline{f}(x)\right) \mathrm{d}V_x \tag{C.4}$$

$$= i\int_{\partial\mathcal{M}} \left(\overline{f}(x)k^\mu\nabla_\mu g(x) - g(x)k^\mu\nabla_\mu\overline{f}(x)\right) \mathrm{d}S, \tag{C.5}$$

where $dV_x = \sqrt{-g}d^Dx$ and in arriving at the last equality we have used Stoke's theorem. Here k^μ is the unit normal to the boundary $\partial\mathcal{M}$ and $\mathrm{d}S$ is the induced volume element on

$\partial\mathcal{M}$. Because f and g induce initial data of compact support on all equal-time spatial slices Σ_t (by definition), there is no contribution to (C.5) from timelike boundaries. Therefore, the only non-vanishing contributions to (C.5) are the integrals over Σ_{t_1} and Σ_{t_2}. Recall that n^μ appearing in (4.3) is the *future-pointing* unit normal to Σ_t, therefore $n^\mu|_{\Sigma_{t_1}} = k^\mu|_{\Sigma_{t_1}}$ and $n^\mu|_{\Sigma_{t_2}} = -k^\mu|_{\Sigma_{t_2}}$. It then follows that

$$0 = i\int_{\partial\mathcal{M}} \left(\overline{f}k^\mu\nabla_\mu g - gk^\mu\nabla_\mu\overline{f}\right)\,\mathrm{d}S \tag{C.6}$$

$$= i\int_{\Sigma_{t_1}} \left(\overline{f}n^\mu\nabla_\mu g - gn^\mu\nabla_\mu\overline{f}\right)\,\mathrm{d}\Sigma_{t_1} - i\int_{\Sigma_{t_2}} \left(\overline{f}n^\mu\nabla_\mu g - gn^\mu\nabla_\mu\overline{f}\right)\,\mathrm{d}\Sigma_{t_2}, \tag{C.7}$$

or equivalently

$$i\int_{\Sigma_{t_1}} \left(\overline{f}n^\mu\nabla_\mu g - gn^\mu\nabla_\mu\overline{f}\right)\,\mathrm{d}\Sigma_{t_1} = i\int_{\Sigma_{t_2}} \left(\overline{f}n^\mu\nabla_\mu g - gn^\mu\nabla_\mu\overline{f}\right)\,\mathrm{d}\Sigma_{t_2}. \tag{C.8}$$

C.2 Commutation Relations

Here we show that the Peierls form of the commutation relations is equivalent to the more usually quoted canonical equal-time commutation relations. We consider a $D = d + 1$-dimensional globally hyperbolic spacetime $(M, g_{\mu\nu})$, use the metric signature $-+++\cdots$, and set $\hbar = c = 1$.

The retarded and advanced Green's functions $G_{R,A}(x, y)$ are solutions to

$$(\Box_x - m^2)G_{R,A}(x, y) = \frac{1}{\sqrt{-g}}\delta^{(D)}(x - y), \tag{C.9}$$

where by definition $G_R(x, y) = 0$ unless $x \succ y$ (meaning that x is inside or on the future light cone of y) and $G_A(x, y) = 0$ unless $y \succ x$. The Peierls form of the commutation relations is then

$$[\widehat{\phi}(x), \widehat{\phi}(y)] = i\Delta(x, y), \tag{C.10}$$

where $\Delta(x, y)$ is the Pauli-Jordan function, defined as the difference between the retarded and advanced Green's functions:

$$\Delta(x, y) := G_R(x, y) - G_A(x, y). \tag{C.11}$$

To state the canonical commutation relations, we start with the classical action functional for the real scalar field $\phi(x)$:

$$S = \int \mathcal{L}(\phi, \partial_\mu \phi) d^D x, \qquad \mathcal{L} = -\frac{1}{2}\sqrt{-g}\left(g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi + m^2\phi^2\right). \tag{C.12}$$

The conjugate momentum $\Pi(x)$ is given by

$$\Pi(x) = \frac{\partial \mathcal{L}}{\partial\left(\partial_0 \phi\right)} = -\sqrt{-g}g^{0\mu}\partial_\mu\phi. \tag{C.13}$$

The equal-time commutation relations then read

$$[\widehat{\phi}(x^0, \mathbf{x}), \widehat{\Pi}(x^0, \mathbf{y})] = i\delta^{(d)}(\mathbf{x} - \mathbf{y}), \tag{C.14}$$

$$[\widehat{\phi}(x^0, \mathbf{x}), \widehat{\phi}(x^0, \mathbf{y})] = 0. \tag{C.15}$$

It is easy to see that (C.15) is a direct consequence of (C.10): $G_R(x, y)$ and $G_A(x, y)$ vanish, by definition, when x and y are spacelike. It then follows that $\Delta(x, y)$, and consequently $[\widehat{\phi}(x), \widehat{\phi}(y)]$, vanish when x and y are spacelike. Therefore, all that is left to prove is that (C.14) also follows from (C.10).

To do so, let us start by noting that

$$G_R(x, y) = \theta(x^0 - y^0)\Delta(x, y), \tag{C.16}$$

where θ is the Heaviside function. Then:

$$(\Box_x - m^2)\left(\theta(x^0 - y^0)[\widehat{\phi}(x), \widehat{\phi}(y)]\right) = i(\Box_x - m^2)\left(\theta(x^0 - y^0)\Delta(x, y)\right) \tag{C.17}$$

$$= i(\Box_x - m^2)G_R(x, y) \tag{C.18}$$

$$= i\frac{\delta^{(D)}(x - y)}{\sqrt{-g(y)}}, \tag{C.19}$$

where the first, second, and third equalities follow from (C.10), (C.16), and (C.9), respectively. Smearing the above equality with a suitable test function f over M:

$$\int_M (\Box_x - m^2)\left(\theta(x^0 - y^0)[\widehat{\phi}(x), \widehat{\phi}(y)]\right) f(y)\sqrt{-g(y)}d^D y = if(x). \tag{C.20}$$

It will be shown in what follows that

$$\int_M (\Box_x - m^2)\left(\theta(x^0 - y^0)[\widehat{\phi}(x), \widehat{\phi}(y)]\right) f(y)\sqrt{-g(y)}d^D y$$

$$= \int [\widehat{\phi}(x^0, \mathbf{x}), \widehat{\Pi}(x^0, \mathbf{y})] f(x^0, \mathbf{y})\sqrt{\frac{-g(x^0, \mathbf{y})}{-g(x^0, \mathbf{x})}}d^d\mathbf{y}. \tag{C.21}$$

which when combined with (C.20), implies the commutation relation (C.14) as desired.

It then remains to prove (C.21). Using the definition of the d'Alembertian $\Box_x = \frac{1}{\sqrt{-g(x)}}\frac{\partial}{\partial x^\mu}\left(\sqrt{-g(x)}g^{\mu\nu}(x)\frac{\partial}{\partial x^\nu}\right)$, it may be verified that

$$(\Box_x - m^2)\left(\theta(x^0-y^0)[\widehat{\phi}(x),\widehat{\phi}(y)]\right) = \theta(x^0-y^0)[(\Box_x - m^2)\widehat{\phi}(x),\widehat{\phi}(y)] \tag{C.22}$$

$$+ g^{\mu 0}(x)[\frac{\partial}{\partial x^\mu}\widehat{\phi}(x),\widehat{\phi}(y)]\delta(x^0-y^0) \tag{C.23}$$

$$+ \frac{1}{\sqrt{-g(x)}}\frac{\partial}{\partial x^\mu}\left(\sqrt{-g(x)}g^{\mu 0}(x)\delta(x^0-y^0)[\widehat{\phi}(x),\widehat{\phi}(y)]\right). \tag{C.24}$$

The term (C.22) vanishes identically because the quantum field satisfies the Klein-Gordon equation: $(\Box_x - m^2)\widehat{\phi}(x) = 0$. Smearing (C.23) with a test function f:

$$\int_M g^{\mu 0}(x)[\frac{\partial}{\partial x^\nu}\widehat{\phi}(x),\widehat{\phi}(y)]\delta(x^0-y^0)f(y)\sqrt{-g(y)}d^Dy \tag{C.25}$$

$$= \int g^{\mu 0}(x)[\frac{\partial}{\partial x^\nu}\widehat{\phi}(x),\widehat{\phi}(x^0,\mathbf{y})]f(x^0,\mathbf{y})\sqrt{-g(x^0,\mathbf{y})}d^d\mathbf{y} \tag{C.26}$$

$$= \int [\widehat{\phi}(x^0,\mathbf{y}),\widehat{\Pi}(x^0,\mathbf{x})]f(x^0,\mathbf{y})\frac{\sqrt{-g(x^0,\mathbf{y})}}{\sqrt{-g(x^0,\mathbf{x})}}d^d\mathbf{y}, \tag{C.27}$$

where the first equality follows from performing the integration over y^0, and the second equality from (C.13). Smearing (C.24) with a test function f:

$$\int_M \frac{1}{\sqrt{-g(x)}}\frac{\partial}{\partial x^\mu}\left(\sqrt{-g(x)}g^{\mu 0}(x)\delta(x^0-y^0)[\widehat{\phi}(x),\widehat{\phi}(y)]\right)f(y)\sqrt{-g(y)}d^Dy \tag{C.28}$$

$$= \frac{1}{\sqrt{-g(x)}}\frac{\partial}{\partial x^\mu}\left(\sqrt{-g(x)}g^{\mu 0}(x)\int \delta(x^0-y^0)[\widehat{\phi}(x),\widehat{\phi}(y)]f(y)\sqrt{-g(y)}d^Dy\right) \tag{C.29}$$

$$= \frac{1}{\sqrt{-g(x)}}\frac{\partial}{\partial x^\mu}\left(\sqrt{-g(x)}g^{\mu 0}(x)\int [\widehat{\phi}(x^0,\mathbf{x}),\widehat{\phi}(x^0,\mathbf{y})]f(x^0,\mathbf{y})\sqrt{-g(x^0,\mathbf{y})}d^d\mathbf{y}\right) \tag{C.30}$$

$$= 0, \tag{C.31}$$

where the third equality follows from performing the integration over y^0, and the last equality from (C.15). At last, (C.21) follows from putting (C.27) and (C.31) together, which, as explained above, is sufficient to prove the equivalence of the Peierls form of the commutation relations with the canonical equal-time commutation relations.

C.3 Equation (4.13) as an equality between distributions

Here we show that the right and left hand sides of (4.13) are equal if they are integrated against a smooth test function of compact support. To do so, we start by proving a useful theorem. Let $(M, g_{\mu\nu})$ be a globally hyperbolic spacetime which admits a foliation by Cauchy surfaces Σ_t labelled by a time coordinate t. Also, let $\mathrm{C}_0^\infty(M)$ be the space of all real and smooth functions of compact support on M.

Theorem 4. [1] *For all $f \in C_0^\infty(M)$ and solutions ϕ of the Klein-Gordon equation (i.e. all ϕ which satisfy $(\Box - m^2)\phi = 0$):*[2]

$$(\phi, i\Delta f)_{KG} = \langle \phi, f \rangle. \tag{C.32}$$

Proof. Let $\mathcal{M}$ be the region bounded by a pair of Cauchy surfaces Σ_{t_1} and Σ_{t_2}, such that Σ_{t_1} is to the causal past of Σ_{t_2} and $f(x) = 0$ for all $x \notin \mathcal{M}$. Also, let $(G_{R,A} f)(x)$ denote the retarded and advanced solutions of the Klein-Gordon equation with source f:

$$(G_{R,A} f)(x) = \int_M G_{R,A}(x, y) f(y) dV_y, \tag{C.33}$$

where $dV_y = \sqrt{-g(y)} d^D y$. It follows from (4.2) that

$$(\Box - m^2)(G_{R,A} f)(x) = f(x). \tag{C.34}$$

The corresponding integral operator to $\Delta(x, y)$ is defined similarly:

$$(\Delta f)(x) = \int_M \Delta(x, y) f(y) dV_y \tag{C.35}$$

$$= (G_R f)(x) - (G_A f)(x), \tag{C.36}$$

[1] This is a generalization of Lemma 3.2.1 of [158].

[2] See (4.3) and (4.25) for the definition of the KG inner product $(\cdot, \cdot)_{KG}$ and the L^2 inner product $\langle \cdot, \cdot \rangle$.

where the last line follows from the definition of $\Delta(x,y)$ (see (4.6)). Then:

$$\begin{aligned}
\langle \phi, f\rangle &= \int_M \overline{\phi}(x) f(x)\, \mathrm{d}V_x = \int_{\mathcal{M}} \overline{\phi}(x) f(x)\, \mathrm{d}V_x \\
&= \int_{\mathcal{M}} \overline{\phi}(x)(\nabla_\mu \nabla^\mu - m^2)(G_A f)(x)\, \mathrm{d}V \\
&= \int_{\mathcal{M}} \left\{ \nabla_\mu \left[\overline{\phi}(x) \nabla^\mu (G_A f)(x) \right] - \nabla_\mu (G_A f)(x) \nabla^\mu \overline{\phi}(x) - m^2 \overline{\phi}(x)(G_A f)(x) \right\} \mathrm{d}V_x \\
&= \int_{\mathcal{M}} \left\{ \nabla_\mu \left[\overline{\phi}(x) \nabla^\mu (G_A f)(x) - (G_A f)(x) \nabla^\mu \overline{\phi}(x) \right] + (G_A f)(x)(\Box - m^2)\overline{\phi}(x) \right\} \mathrm{d}V_x \\
&= \int_{\partial \mathcal{M}} \{ \overline{\phi}(x) n^\mu \nabla_\mu (G_A f)(x) - (G_A f)(x) n^\mu \nabla_\mu \overline{\phi}(x) \}\, \mathrm{d}S, \qquad \text{(C.37)}
\end{aligned}$$

where in arriving at the last equality we have used Stoke's theorem, as well as the fact that ϕ satisfies the KG equation. Here n^μ is the unit normal to the boundary $\partial\mathcal{M}$ and $\mathrm{d}S$ is the induced volume element on $\partial\mathcal{M}$. Because $G_A f$ induces initial data of compact support on all equal-time spatial slices Σ_t, there is no contribution to (C.37) from timelike boundaries. Also, since $G_A f$ vanishes outside of the causal past of the support of f: $G_A f|_{\Sigma_{t_2}} = 0$. Therefore, the only non-vanishing contribution to (C.37) is the integral over Σ_{t_1}. Because $G_R f$ vanishes outside of the causal future of the support of f, it follows that $\Delta f|_{\Sigma_{t_1}} = G_R f|_{\Sigma_{t_1}} - G_A f|_{\Sigma_{t_1}} = -G_A f|_{\Sigma_{t_1}}$. Putting these facts together:

$$\begin{aligned}
\langle \phi, f\rangle &= \int_{\partial \mathcal{M}} \{ \overline{\phi}(x) n^\mu \nabla_\mu (G_A f)(x) - (G_A f)(x) n^\mu \nabla_\mu \overline{\phi}(x) \}\, \mathrm{d}S \\
&= \int_{\Sigma_{t_1}} \{ \overline{\phi}(x) n^\mu \nabla_\mu (G_A f)(x) - (G_A f)(x) n^\mu \nabla_\mu \overline{\phi}(x) \}\, \mathrm{d}\Sigma_{t_1} \\
&= i \int_{\Sigma_{t_1}} \{ \overline{\phi}(x) n^\mu \nabla_\mu (-iG_A f)(x) - (-iG_A f)(x) n^\mu \nabla_\mu \overline{\phi}(x) \}\, \mathrm{d}\Sigma_{t_1} \qquad \text{(C.38)} \\
&= i \int_{\Sigma_{t_1}} \{ \overline{\phi}(x) n^\mu \nabla_\mu (i\Delta f)(x) - (i\Delta f)(x) n^\mu \nabla_\mu \overline{\phi}(x) \}\, \mathrm{d}\Sigma_{t_1} \\
&= (\phi, i\Delta f)_{KG},
\end{aligned}$$

which concludes the proof. □

We are now in a position to show that (4.13) is an equality between distributions. It follows from (4.7) that for any $f \in \mathrm{C}_0^\infty(M)$

$$(\Box - m^2)(\Delta f)(x) = 0. \qquad \text{(C.39)}$$

Since Δf is a solution of the KG equation, we can expand it out in terms of $\{u_{\mathbf{k}}\}$ and $\{\overline{u}_{\mathbf{k}}\}$:

$$i\Delta f(x) = \sum_{\mathbf{k}} \alpha_k u_{\mathbf{k}}(x) + \beta_k \overline{u}_{\mathbf{k}}(x), \tag{C.40}$$

where α_k's and β_k's are constants. It may be verified using the orthonormality conditions (4.9) that

$$\alpha_{\mathbf{k}} = (u_{\mathbf{k}}, i\Delta f)_{KG}, \qquad \beta_{\mathbf{k}} = -(\overline{u}_{\mathbf{k}}, i\Delta f)_{KG}. \tag{C.41}$$

Then,

$$\begin{aligned} i\Delta f(x) &= \sum_k (u_k, i\Delta f)_{KG} u_k(x) - (\overline{u}_k, i\Delta f)_{KG} \overline{u}_k(x) \\ &= \sum_k \langle u_k, f\rangle u_k(x) - \langle \overline{u}_k, f\rangle \overline{u}_k(x) \\ &= \int_M \left[\sum_k u_k(x)\overline{u}_k(y) - \overline{u}_k(x)u_k(y)\right] f(y)\, \mathrm{d}V_y, \end{aligned} \tag{C.42}$$

where in arriving at the second line we have used (C.32).

C.4 Proof of $G_R(x,y) = G_A(y,x)$

Let M be the region bounded by a pair of Cauchy surfaces, one to the future of both x and z, the other to their past. Then

$$\begin{aligned} G_R(x,y) - G_A(y,x) &= \int_M \big\{G_R(z,y)\delta^4(z-x) - G_A(z,x)\delta^4(z-y)\big\}\, \mathrm{d}^4 z \\ &= \int_M \big\{G_R(z,y)(\Box_z - m^2)G_A(z,x) - G_A(z,x)(\Box_z - m^2)G_R(z,y)\big\}\, \mathrm{d}V_z \\ &= \int_M \big\{G_R(z,y)\Box_z G_A(z,x) - G_A(z,x)\Box_z G_R(z,y)\big\}\, \mathrm{d}V_z \\ &= \int_M \nabla^z_\mu \big\{G_R(z,y)\nabla^\mu_z G_A(z,x) - G_A(z,x)\nabla^\mu_z G_R(z,y)\big\}\, \mathrm{d}V_z \\ &= \int_{\partial M} \big\{G_R(z,y)n^\mu(z)\nabla^z_\mu G_A(z,x) - G_A(z,x)n^\mu(z)\nabla^z_\mu G_R(z,y)\big\}\, \mathrm{d}S_z = 0, \end{aligned}$$

where $\mathrm{d}V_z = \sqrt{-g(z)}\,\mathrm{d}^4 z$, and a super- or subscript of z on $\Box$ or ∇ reminds us that the derivatives are being taken with respect to the variable z. In the last line, we have

used Stoke's theorem to recast the spacetime volume of the total divergence in terms of an integral over the boundary ∂M, whose unit normal and induced volume element are denoted by n^μ and dS, respectively. Let us now explain why the last expression is identically zero. By definition, it is only when z is in the causal future of y and the causal past of x where this expression could be nonzero. However, because z is being evaluated at the boundary this is never possible in a globally hyperbolic spacetime.

C.5 When is $i\Delta$ self-adjoint?

Here, we will show that $i\Delta$ is self-adjoint for a bounded region of $3+1$ dimensional Minkowski ($\mathbb{M}^4$) space, and will argue that this conclusion should continue to hold in all curved spacetimes. The retarded Green's function in $\mathbb{M}^4$ takes the form

$$\begin{aligned} G_R(t,\mathbf{x};t',\mathbf{x}') = &-\frac{1}{2\pi}\theta(t-t')\delta((t-t')^2-|\mathbf{x}-\mathbf{x}'|^2) \\ &+\frac{m}{4\pi}\theta(t-t')\theta((t-t')^2-|\mathbf{x}-\mathbf{x}'|^2)\frac{J_1(m\sqrt{(t-t')^2-|\mathbf{x}-\mathbf{x}'|^2})}{\sqrt{(t-t')^2-|\mathbf{x}-\mathbf{x}'|^2}}, \end{aligned} \tag{C.43}$$

where J_1 is the Bessel function of the first kind and m is the mass of the scalar field. By definition, $G_R(t,\mathbf{x};t',\mathbf{x}')$ satisfies

$$(-\partial_t^2+\nabla_{\mathbf{x}}^2-m^2)G_R(t,\mathbf{x};t',\mathbf{x}') = \delta(t-t')\delta^{(3)}(\mathbf{x}-\mathbf{x}'). \tag{C.44}$$

Consider a spacetime box $t,\mathbf{x}\in[-L,L]$ about the origin in $\mathbb{M}^4$. We shall denote this region by $\mathbb{L}^4$. The action of $i\Delta$ on a function $f(x)$ in $\mathbb{L}^4$ is given by

$$(i\Delta f)(t,\mathbf{x}) = i(G_R f)(t,\mathbf{x}) - i(G_A f)(t,\mathbf{x}), \tag{C.45}$$

where

$$(G_R f)(t,\mathbf{x}) = \int_{\mathbb{L}^4} G_R(t,\mathbf{x};t',\mathbf{x}')f(t',\mathbf{x}')\,\mathrm{d}t'\,\mathrm{d}^3\mathbf{x}', \tag{C.46}$$

$$(G_A f)(t,\mathbf{x}) = \int_{\mathbb{L}^4} G_R(t',\mathbf{x}';t,\mathbf{x})f(t',\mathbf{x}')\,\mathrm{d}t'\,\mathrm{d}^3\mathbf{x}'. \tag{C.47}$$

We will now show that $i\Delta$ is a bounded operator on the space $L^2(\mathbb{L}^4)$ of all square integrable functions in $\mathbb{L}^4$. In other words, we will prove there exists $N>0$ such that

$$||i\Delta f|| \le N||f|| \qquad \forall \qquad f\in L^2(\mathbb{L}^4), \tag{C.48}$$

where $||\cdot||$ denotes the $L^2(\mathbb{L}^4)$ norm.[3] If $i\Delta$ is bounded on $L^2(\mathbb{L}^3)$, it would mean that it is Hermitian (or symmetric) on all of $L^2(\mathbb{L}^3)$. This then implies, using standard results from functional analysis, that $i\Delta$ is self-adjoint [129].

Before delving into the details of the proof, let us make the following definitions:

$$G_R^{(0)}(t,\mathbf{x};t',\mathbf{x}') = -\frac{1}{2\pi}\theta(t-t')\delta((t-t')^2-|\mathbf{x}-\mathbf{x}'|^2), \tag{C.50}$$

$$G_R^{(m)}(t,\mathbf{x};t',\mathbf{x}') = \frac{m}{4\pi}\theta(t-t')\theta((t-t')^2-|\mathbf{x}-\mathbf{x}'|^2)\frac{J_1(m\sqrt{(t-t')^2-|\mathbf{x}-\mathbf{x}'|^2})}{\sqrt{(t-t')^2-|\mathbf{x}-\mathbf{x}'|^2}}, \tag{C.51}$$

so that

$$G_R(t,\mathbf{x};t',\mathbf{x}') = G_R^{(0)}(t,\mathbf{x};t',\mathbf{x}') + G_R^{(m)}(t,\mathbf{x};t',\mathbf{x}'), \tag{C.52}$$

or equivalently (for all $f \in L^2(\mathbb{L}^4)$)

$$(G_R f)(t,\mathbf{x}) = (G_R^{(0)} f)(t,\mathbf{x}) + (G_R^{(m)} f)(t,\mathbf{x}). \tag{C.53}$$

We start by noting that it suffices to show G_R and G_A are bounded on $L^2(\mathbb{L}^4)$, because by the triangular inequality:

$$||i\Delta f|| = ||iG_R f - iG_A f|| \leq ||G_R f|| + ||G_A f||. \tag{C.54}$$

It follows from $|J_1(x)|/x \leq 1/2$ (see e.g. 10.14.4 of [154]) that

$$|G_R^{(m)}(t,\mathbf{x};t',\mathbf{x}')| \leq \frac{m^2}{8\pi}, \tag{C.55}$$

which in turn implies

$$\begin{aligned}
|(G_R^{(m)} f)(t,\mathbf{x})|^2 &= \left|\int_{\mathbb{L}^4} G_R^{(m)}(t,\mathbf{x};t',\mathbf{x}') f(t',\mathbf{x}')\,\mathrm{d}t'\,\mathrm{d}^3\mathbf{x}'\right|^2 \\
&\leq \left\{\int_{\mathbb{L}^4} |G_R^{(m)}(t,\mathbf{x};t',\mathbf{x}')|^2\,\mathrm{d}t'\,\mathrm{d}^3\mathbf{x}'\right\} \times \left\{\int_{\mathbb{L}^4} |f(t',\mathbf{x}')|^2\,\mathrm{d}t'\,\mathrm{d}^3\mathbf{x}'\right\} \\
&\leq \left(\frac{m^2}{8\pi}\right)^2 (2L)^4 ||f||^2 = \left(\frac{m^2L^2}{2\pi}\right)^2 ||f||^2,
\end{aligned} \tag{C.56}$$

[3] The L^2 inner product $\langle\cdot,\cdot\rangle$ and norm $||\cdot||$ on $\mathbb{L}^4$ are defined as usual:

$$\langle f,g\rangle = \int_{\mathbb{L}^4} \bar{f}(t,\mathbf{x}) g(t,\mathbf{x})\,\mathrm{d}t\,\mathrm{d}^3\mathbf{x}, \qquad ||f|| = \sqrt{\langle f,f\rangle}. \tag{C.49}$$

where in the second line we have used the Cauchy-Schwarz inequality and the last line follows from (C.55). Furthermore:

$$||G_R^{(m)} f||^2 = \int_{\mathbb{L}^4} |(G_R^{(m)} f)(t, \mathbf{x})|^2 \, \mathrm{d}t \, \mathrm{d}^3\mathbf{x} \tag{C.57}$$

$$\leq (2L)^4 \left(\frac{m^2 L^2}{2\pi} \right)^2 ||f||^2 = \left(\frac{2m^2 L^4}{\pi} \right)^2 ||f||^2, \tag{C.58}$$

or equivalently

$$||G_R^{(m)} f|| \leq \frac{2m^2 L^4}{\pi} ||f||. \tag{C.59}$$

Let us now turn our attention to $G_R^{(0)} f$. It can be checked that:

$$(G_R^{(0)} f)(t, \mathbf{x}) = -\frac{1}{4\pi} \int_{\mathbb{L}^3} \frac{f(t - |\mathbf{x} - \mathbf{x}'|, \mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} \theta(t - |\mathbf{x} - \mathbf{x}'| + L) \, \mathrm{d}^3\mathbf{x}'. \tag{C.60}$$

Using the Cauchy-Schwarz inequality:

$$|(G_R^{(0)} f)(t, \mathbf{x})|^2 \leq \frac{1}{16\pi^2} \left\{ \int_{\mathbb{L}^3} \frac{\mathrm{d}^3\mathbf{x}'}{|\mathbf{x} - \mathbf{x}'|^2} \right\} \times \tag{C.61}$$

$$\left\{ \int_{\mathbb{L}^3} |f(t - |\mathbf{x} - \mathbf{x}'|, \mathbf{x}')|^2 \theta(t - |\mathbf{x} - \mathbf{x}'| + L) \, \mathrm{d}^3\mathbf{x}' \right\}. \tag{C.62}$$

Also

$$\begin{aligned} \int_{\mathbb{L}^3} \frac{\mathrm{d}^3\mathbf{x}'}{|\mathbf{x} - \mathbf{x}'|^2} &= \int_{-L}^{L} \mathrm{d}x_1' \int_{-L}^{L} \mathrm{d}x_2' \int_{-L}^{L} \mathrm{d}x_3' \frac{1}{\sum_{i=1}^3 (x_i' - x_i)^2} \\ &= \int_{-L-x_1}^{L-x_1} \mathrm{d}y_1 \int_{-L-x_2}^{L-x_2} \mathrm{d}y_2 \int_{-L-x_3}^{L-x_3} \mathrm{d}y_3 \frac{1}{\sum_{i=1}^3 y_i^2} \\ &\leq \int_{(2\mathbb{L})^3} \frac{\mathrm{d}^3\mathbf{y}}{|\mathbf{y}|^2} \\ &\leq \int_{\mathbb{S}^3(2\sqrt{3}L)} \frac{\mathrm{d}^3\mathbf{y}}{|\mathbf{y}|^2} = \int_0^{2\sqrt{3}L} \frac{4\pi|\mathbf{y}|^2}{|\mathbf{y}|^2} d|\mathbf{y}| = 8\sqrt{3}\pi L, \end{aligned} \tag{C.63}$$

where $(2\mathbb{L})^3$ is the spatial box around the origin whose sides are $4L$ long, and $\mathbb{S}^3(2\sqrt{3}L)$ is a solid sphere with radius $2\sqrt{3}L$. In the second line we have changed integration variables to $\mathbf{y} = \mathbf{x}' - \mathbf{x}$, and the last inequality is true because $\mathbb{S}^3(2\sqrt{3}L)$ contains $(2\mathbb{L})^3$. Therefore:

$$\left| (G_R^{(0)} f)(t, \mathbf{x}) \right|^2 \leq \frac{\sqrt{3}L}{2\pi} \int_{\mathbb{L}^3} |f(t - |\mathbf{x} - \mathbf{x}'|, \mathbf{x}')|^2 \, \theta(t - |\mathbf{x} - \mathbf{x}'| + L) \, \mathrm{d}^3\mathbf{x}'. \tag{C.64}$$

It then follows that

$$\begin{aligned}
||G_R^{(0)} f||^2 &= \int_{\mathbb{L}^3} \mathrm{d}^3\mathbf{x} \int_{-L}^{L} \mathrm{d}t \left|(G_R^{(0)} f)(t,\mathbf{x})\right|^2 \\
&\leq \frac{\sqrt{3}L}{2\pi} \int_{\mathbb{L}^3} \mathrm{d}^3\mathbf{x} \int_{-L}^{L} \mathrm{d}t \int_{\mathbb{L}^3} \mathrm{d}^3\mathbf{x}' \left|f(t-|\mathbf{x}-\mathbf{x}'|,\mathbf{x}')\right|^2 \theta(t-|\mathbf{x}-\mathbf{x}'|+L) \\
&= \frac{\sqrt{3}L}{2\pi} \int_{\mathbb{L}^3} \mathrm{d}^3\mathbf{x} \int_{\mathbb{L}^3} \mathrm{d}^3\mathbf{x}' \int_{-L}^{L} \mathrm{d}t \left|f(t-|\mathbf{x}-\mathbf{x}'|,\mathbf{x}')\right|^2 \theta(t-|\mathbf{x}-\mathbf{x}'|+L) \\
&= \frac{\sqrt{3}L}{2\pi} \int_{\mathbb{L}^3} \mathrm{d}^3\mathbf{x} \int_{\mathbb{L}^3} \mathrm{d}^3\mathbf{x}' \int_{-L}^{L-|\mathbf{x}-\mathbf{x}'|} \mathrm{d}t' \left|f(t',\mathbf{x}')\right|^2 \\
&\leq \frac{\sqrt{3}L}{2\pi} \int_{xf} \mathrm{d}^3\mathbf{x} \int_{\mathbb{L}^3} \mathrm{d}^3\mathbf{x}' \int_{-L}^{L} \mathrm{d}t' \left|f(t',\mathbf{x}')\right|^2 \\
&= \frac{\sqrt{3}L}{2\pi}(2L)^3 ||f||^2 = \frac{4\sqrt{3}L^4}{\pi}||f||^2,
\end{aligned} \tag{C.65}$$

where in the fourth line we have changed integration variables to $t' = t - |\mathbf{x}-\mathbf{x}'|$. In summary:

$$||G_R^{(0)} f|| \leq \left(\frac{48}{\pi^2}\right)^{1/4} L^2 ||f||. \tag{C.66}$$

Putting together (C.66), (C.66), and using the triangular inequality:

$$\begin{aligned}
||G_R f|| &= ||G_R^{(0)} f + G_R^{(m)} f|| \leq ||G_R^{(0)} f|| + ||G_R^{(m)} f|| \\
&\leq \frac{2m^2L^4}{\pi}||f|| + \left(\frac{48}{\pi^2}\right)^{1/4} L^2||f|| = \left[\frac{2}{\pi}(mL)^2 + \left(\frac{48}{\pi^2}\right)^{1/4}\right] L^2||f||.
\end{aligned} \tag{C.67}$$

A similar analysis for the advanced solution $G_A f$ results in the same bound. Finally, using (C.54), we arrive at

$$||i\Delta f|| \leq 2\left[\frac{2}{\pi}(mL)^2 + \left(\frac{48}{\pi^2}\right)^{1/4}\right] L^2||f||. \tag{C.68}$$

It should be clear that the boundedness of $i\Delta$ has everything to do with the singularity structure of $i\Delta(x,y)$ (once we restrict ourselves to bounded spacetimes). It is not terribly unrealistic to assume that this singularity structure remains (more or less) the same in curved spacetimes. This is certainly true in the coincidence limit, if the equivalence principle is respected. Based on these arguments, we assume in this chapter that $i\Delta$ is a self-adjoint operator on $L^2(M)$, for all bounded globally-hyperbolic spacetimes M.

C.6 Degeneracy of the Spectrum of $i\Delta$

There are often (infinitely) many linearly independent eigenfunctions of $i\Delta$ which have the same eigenvalue. In the subspace spanned by all eigenfunctions which share the same eigenvalue, there is no unique orthonormal basis. We had glossed over this issue when introducing the SJ formalism. Here we show that the positive part of $i\Delta(x, y)$, as defined in Section 4.3.1, is insensitive to this choice.

Let $\{\lambda_{\mathbf{a}}\}$ denote the set of all positive eigenvalues of $i\Delta$, with the corresponding set of orthonormal eigenvectors $\{\mathfrak{u}^+_{\mathbf{a},\mathbf{b}}(x)\}$ (now accounting also for any degeneracy, with the subscript $\mathbf{b}$):

$$i\Delta\mathfrak{u}^+_{\mathbf{a},\mathbf{b}} = \lambda_{\mathbf{a}}\mathfrak{u}^+_{\mathbf{a},\mathbf{b}}, \qquad \langle \mathfrak{u}^+_{\mathbf{a},\mathbf{b}}, \mathfrak{u}^+_{\mathbf{a}',\mathbf{b}'}\rangle = \delta_{\mathbf{a}\mathbf{a}'}\delta_{\mathbf{b}\mathbf{b}'}. \tag{C.69}$$

The SJ formalism then defines the two-point function in the SJ state to be the positive part of $i\Delta(x, y)$ as follows:

$$W_{SJ}(x, y) := \sum_{\mathbf{a}} \lambda_{\mathbf{a}} \sum_{\mathbf{b}} \mathfrak{u}^+_{\mathbf{a},\mathbf{b}}(x)\overline{\mathfrak{u}}^+_{\mathbf{a},\mathbf{b}}(y). \tag{C.70}$$

Consider now the set of functions $\{\mathfrak{v}^+_{\mathbf{a},\mathbf{b}}(x)\}$ defined by:

$$\mathfrak{v}^+_{\mathbf{a},\mathbf{b}}(x) = \sum_{\mathbf{b}'} \alpha_{\mathbf{b}\mathbf{b}'}\mathfrak{u}^+_{\mathbf{a},\mathbf{b}'}(x). \tag{C.71}$$

It can be easily shown that $\{\mathfrak{v}^+_{\mathbf{a},\mathbf{b}}(x)\}$ also form an orthonormal set of eigenfunctions of $i\Delta$:

$$i\Delta\mathfrak{v}^+_{\mathbf{a},\mathbf{b}} = \lambda_{\mathbf{a}}\mathfrak{v}^+_{\mathbf{a},\mathbf{b}}, \qquad \langle \mathfrak{v}^+_{\mathbf{a},\mathbf{b}}, \mathfrak{v}^+_{\mathbf{a}',\mathbf{b}'}\rangle = \delta_{\mathbf{a}\mathbf{a}'}\delta_{\mathbf{b}\mathbf{b}'}, \tag{C.72}$$

if and only if

$$\sum_{\mathbf{c}} \overline{\alpha}_{\mathbf{b}\mathbf{c}}\alpha_{\mathbf{b}'\mathbf{c}} = \delta_{\mathbf{b}\mathbf{b}'}, \tag{C.73}$$

or in matrix language

$$\boldsymbol{\alpha}\boldsymbol{\alpha}^\dagger = \mathbf{1}. \tag{C.74}$$

Defining the SJ two-point function using $\{\mathfrak{v}^+_{\mathbf{a},\mathbf{b}}(x)\}$, as opposed to $\{\mathfrak{u}^+_{\mathbf{a},\mathbf{b}}(x)\}$, leads to the exact the same answer, because:

$$\sum_{\mathbf{b}} \mathfrak{v}^+_{\mathbf{a},\mathbf{b}}(x)\overline{\mathfrak{v}}^+_{\mathbf{a},\mathbf{b}}(y) = \sum_{\mathbf{c},\mathbf{c}'} \mathfrak{u}^+_{\mathbf{a},\mathbf{c}}(x)\overline{\mathfrak{u}}^+_{\mathbf{a},\mathbf{c}'}(y) \sum_{\mathbf{b}} \overline{\alpha}_{\mathbf{b}\mathbf{c}'}\alpha_{\mathbf{b}\mathbf{c}} \tag{C.75}$$

$$= \sum_{\mathbf{c},\mathbf{c}'} \mathfrak{u}^+_{\mathbf{a},\mathbf{c}}(x)\overline{\mathfrak{u}}^+_{\mathbf{a},\mathbf{c}'}(y)[\boldsymbol{\alpha}^\dagger\boldsymbol{\alpha}]_{\mathbf{c}'\mathbf{c}} \tag{C.76}$$

$$= \sum_{\mathbf{c}} \mathfrak{u}^+_{\mathbf{a},\mathbf{c}}(x)\overline{\mathfrak{u}}^+_{\mathbf{a},\mathbf{c}}(y). \tag{C.77}$$

C.7 The SJ Vacuum and the Simple Harmonic Oscillator

This simple example illustrates the technical difficulties one faces when diagonalizing $i\Delta$, and how they can be resolved. Consider a simple harmonic oscillator with unit mass and frequency ω, whose position $q(t)$ satisfies $(\frac{d^2}{dt^2} + \omega^2)q(t) = 0$. The associated retarded Green's function satisfies $(\frac{d^2}{dt^2} + \omega^2)G_R(t,t') = -\delta(t-t')$, with $G_R(t,t') = 0$ for $t < t'$. The solution to this equation is $G_R(t,t') = -\theta(t-t')\frac{\sin[\omega(t-t')]}{\omega}$, which in turn gives

$$\Delta(t,t') = G_R(t,t') - G_R(t',t) = -\frac{\sin[\omega(t-t')]}{\omega}. \tag{C.78}$$

Equivalently:

$$i\Delta(t,t') = \frac{1}{2\omega}\big[e^{-i\omega(t-t')} - e^{i\omega(t-t')}\big]. \tag{C.79}$$

Taking $t \in (-\infty, \infty)$, it may be verified that, formally,

$$i\Delta e^{\pm i\omega t} = \int_{-\infty}^{\infty} i\Delta(t,t')e^{\pm i\omega t'}\,\mathrm{d}t' = \frac{\mp\pi\delta(0)}{\omega}e^{\pm i\omega t}. \tag{C.80}$$

Keeping the $\delta(0)$'s around, we see that $\mathfrak{u}^{\mp}(t) = \frac{e^{\pm i\omega t}}{\sqrt{2\pi\delta(0)}}$ are orthonormal eigenfunctions of $i\Delta$ with eigenvalues $\lambda_{\pm} = \pm\frac{\pi\delta(0)}{\omega}$. According to our prescription, the resulting SJ mode-function is $u^{SJ}(t) = \sqrt{\lambda_+}\mathfrak{u}^+(t) = \frac{e^{-i\omega t}}{\sqrt{2\omega}}$, which is completely well defined and agrees with the ground state of the harmonic oscillator, i.e. the state $|SJ>$ annihilated by $\widehat{a}$ operator expansion $\widehat{q}(t) = u^{SJ}(t)\widehat{a} + \bar{u}^{SJ}(t)\widehat{a}^\dagger$ is ithe minimum energy state of the Hamiltonian. Therefore, the infinities appearing in the spectrum of $i\Delta$ end up being harmless. In other words, this example seems to suggest that the SJ vacuum state should not depend on how $\delta(0)$ is regularized.

One regularization scheme, for example, is to first restrict the time interval to $t \in I = [-T, T]$. The $L^2(I)$ inner product and the action of $i\Delta$ are now defined (for all functions $f(t)$ and $g(t)$) as

$$(i\Delta f)(t) = \int_{-T}^{T} i\Delta(t,t')f(t')\,\mathrm{d}t', \qquad \langle f, g\rangle = \int_{-T}^{T} \bar{f}(t)g(t)\,\mathrm{d}t. \tag{C.81}$$

We can now diagonalize $i\Delta$ and then take the limit $T \to \infty$ once the spectrum of $i\Delta$ has been computed. To do so, let $u(t) = \frac{e^{-i\omega t}}{\sqrt{2\omega}}$ so that $i\Delta(t,t') = u(t)\bar{u}(t') - u(t)\bar{u}(t')$. It

can then be confirmed that $i\Delta$ has two orthonormal eigenfunctions $\mathfrak{u}^+(t)$ and $\bar{\mathfrak{u}}^+(t)$, with corresponding eigenvalues $\lambda > 0$ and $-\lambda$:

$$u^{SJ}(t) = \sqrt{\lambda}\mathfrak{u}^+(t) = Au(t) + B\bar{u}(t), \tag{C.82}$$

where

$$A = \cosh(\alpha), \qquad B = \sinh(\alpha)e^{i\beta}, \qquad \alpha = \frac{1}{2}\tanh^{-1}|r|, \tag{C.83}$$

$$\beta = \arg(r) + \pi, \qquad r = \frac{\langle \bar{u}, u\rangle}{\langle u, u\rangle}, \qquad \lambda = \sqrt{\langle u, u\rangle^2 - |\langle u, \bar{u}\rangle|^2}. \tag{C.84}$$

All quantities above are now completely well-defined, since the inner products are finite. In the limit $T \to \infty$, the ratio $\frac{|<u,\bar{u}>|^2}{<u,u>^2} \to 0$ and we recover $u^{SJ}(t) = \frac{e^{-i\omega t}}{\sqrt{2\omega}}$.

C.8 Geometry of de Sitter Space

De Sitter space is the maximally symmetric spacetime of constant positive curvature (a comprehensive review of de Sitter geometry can be found in [136]). We denote de Sitter space in $D = d + 1$ dimensions by dS^D. It can be viewed as the hyperboloid

$$X \cdot X = +\ell^2 \tag{C.85}$$

in an embedding $D+1$ dimensional Minkowski space $\mathbb{M}^{D+1}$ with Cartesian coordinates X^a $(a = 0, 1, \ldots, D)$ and a Lorentzian metric $\eta_{ab} = \mathrm{diag}(-1, 1, \ldots, 1)$ that defines the product $X \cdot Y = \eta_{ab}X^aY^b$. The de Sitter metric $g_{\mu\nu}$ $(\mu = 0, \ldots, D-1)$ is induced by the restriction of η_{ab} onto the hyperboloid.

The geodesic distance between two points $p, q \in dS^D$ takes a particularly simple form in terms of the product between the embedding coordinates, which we denote by

$$Z(p, q) := \ell^{-2}X(p) \cdot X(q). \tag{C.86}$$

In terms of Z, the geodesic distance is

$$d(p, q) := \int_{\lambda_i}^{\lambda_f} \sqrt{g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}}d\lambda = \ell\cos^{-1}Z(p, q), \tag{C.87}$$

where λ parametrises the geodesic $x^\mu(\lambda)$, and p and q have coordinates $x^\mu(\lambda_i)$ and $x^\mu(\lambda_f)$, respectively. For points that can be joined by a geodesic, the range of Z is $-1 \leq Z < \infty$,

where $Z > 1$, $Z = 1$ and $-1 \leq Z < 1$ correspond to timelike, null, and spacelike separations, respectively.

One of the symmetries of de Sitter space that will be relevant below is the antipodal map $A : p \to p^A$, which sends a point $p \in dS^D$ to its "antipode", denoted by p^A. In embedding coordinates, A takes the simple form of a reflection about the origin of $\mathbb{M}^{D+1}$:

$$X^a(p^A) = -X^a(p). \tag{C.88}$$

It is clear from the invariance of (C.85) under A that $p \in dS^D \iff p^A \in dS^D$. Note also that $Z(p, q)$ and $d(p, q)$ are invariant under the action of A.

We will consider two coordinate charts on de Sitter space: *closed global coordinates*, which cover the entire de Sitter manifold defined by (C.85), and *cosmological coordinates*, which cover only the half space $X^0 + X^1 > 0$, known as the (expanding) Poincaré patch (the contracting Poincaré patch corresponds to the other half $X^0 + X^1 < 0$). We will denote the Poincaré patch by dS^D_P. It is highlighted in Figure C.1 and corresponds to the causal future of an observer at the north pole of the d-Sphere (S^d) at past timelike infinity (the bottom left corner of the Penrose diagram). De Sitter space, as well as its upper and lower half spaces, constitute globally hyperbolic manifolds in their own right, but neither admits a global timelike Killing vector field [147] that would serve to define a unique "minimum energy" state.

C.8.1 Global patch of de Sitter (dS^D)

The global chart is given by the coordinates $x^\mu_G = (t, \theta^1, \ldots, \theta^d)$ defined by

$$\begin{aligned} X^0 &= \ell \sinh(t/\ell) \\ X^i &= \ell \cosh(t/\ell)\, \omega^i \qquad\qquad 1 \leq i \leq D, \end{aligned} \tag{C.89}$$

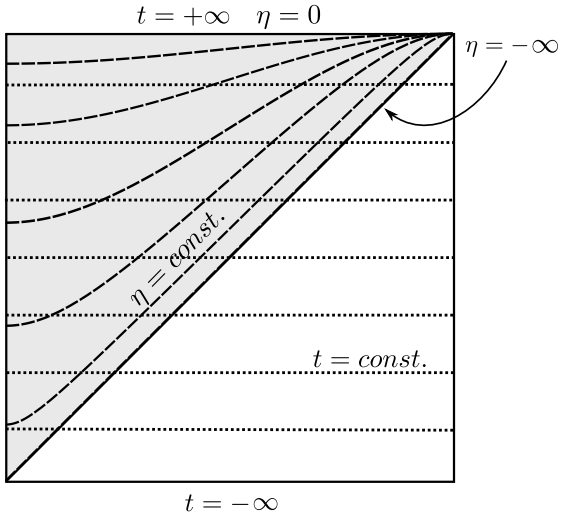


Figure C.1: The Penrose diagram of de Sitter space. The shaded area represents the (expanding) Poincaré patch. Dotted lines are surfaces of constant t (d-spheres), dashed lines are surfaces of constant η (d-planes).

where θ^i are the standard hyperspherical coordinates on S^d and ω^i are given by

$$\begin{aligned}
\omega^1 &= \cos(\theta^1) \\
\omega^2 &= \sin(\theta^1)\cos(\theta^2) \\
\omega^3 &= \sin(\theta^1)\sin(\theta^2)\cos(\theta^3) \\
&\vdots \\
\omega^{D-1} &= \sin(\theta^1)\ldots\sin(\theta^{d-1})\cos(\theta^d) \\
\omega^D &= \sin(\theta^1)\ldots\sin(\theta^{d-1})\sin(\theta^d).
\end{aligned} \tag{C.90}$$

These coordinates range over the values

$$t \in (-\infty, \infty), \qquad \theta^1, \ldots, \theta^{d-1} \in [0, \pi], \qquad \theta^d \in [0, 2\pi). \tag{C.91}$$

The metric in global coordinates takes the form

$$ds^2 = -dt^2 + \ell^2 \cosh^2(t/\ell)\, d\Omega_d^2, \tag{C.92}$$

where $d\Omega_d^2$ is the line element on S^d. The antipode of a point p with coordinates $x_G^\mu(p) = (t, \theta^1, \theta^2, \ldots, \theta^d)$ has coordinates $x_G^\mu(p^A) = (-t, \pi - \theta^1, \pi - \theta^2, \ldots, \pi - \theta^{d-1}, \theta^d \pm \pi)$, where the $+$ and $-$ are for $0 \le \theta^d < \pi$ and $\pi \le \theta^d < 2\pi$, respectively.

C.8.2 Cosmological/Poincaré patch of de Sitter (dS_P^D)

The cosmological chart is defined by the coordinates $x_P^\mu = (\eta, \mathbf{x})$ where $\mathbf{x} \in \mathbb{R}^d$ and

$$\begin{aligned} X^0 &= \frac{-1}{2\eta}\left(\ell^2 - \eta^2 + \mathbf{x}^2\right) \\ X^1 &= \frac{-1}{2\eta}\left(\ell^2 + \eta^2 - \mathbf{x}^2\right) \\ X^i &= \frac{-1}{\eta} x^{i-1} \qquad\qquad 2 \leq i \leq D, \end{aligned} \tag{C.93}$$

with $\mathbf{x}^2 = \sum_{i=1}^d (x^i)^2$. The range of the (conformal) time coordinate is $\eta \in (-\infty, 0)$, i.e. we work in the convention where time flows in the positive η-direction. The spatial coordinates range over the whole real line. The line element is then given by

$$ds^2 = \frac{\ell^2}{\eta^2}\left[-d\eta^2 + \sum_{i=1}^d dx_i^2\right], \tag{C.94}$$

which corresponds to an exponentially expanding Friedmann-Lemaître-Robertson-Walker universe with flat spatial sections.

The antipodal map A is not defined on dS_P^D: if q is a point on the Poincaré patch, its antipode q^A is *not* a point on the Poincaré patch, since the antipodal map in cosmological coordinates takes the form $x_P^\mu(q) = (\eta, \mathbf{x}) \implies x_P^\mu(q^A) = (-\eta, \mathbf{x})$, and η is only defined on the negative real line. Bearing this in mind, we shall still use the notation x_P^A on cosmological coordinates to mean "switch the sign of η" in x_P.

C.9 Vacuum states on de Sitter space

Here we review the so-called Euclidean or Bunch-Davies (BD) vacuum state for a massive free scalar field on de Sitter space. The Euclidean/BD state belongs to a two-real-parameter family of de Sitter-invariant vacuum states, known as the *Mottola-Allen* or α-vacua. We review below how these states are constructed and how they are related to each other.

C.9.1 Bunch-Davies modes on dS_P^D

In cosmological coordinates, the de Sitter metric is given by (C.94). Consider the mode functions

$$u_{\mathbf{k}}(\eta, \mathbf{x}) = \frac{e^{i\mathbf{k}\cdot\mathbf{x}}}{(2\pi)^{d/2}}\chi_k(\eta), \qquad \chi_k(\eta) = \mathcal{N}_{\mathbf{k}}(-\eta)^{d/2}\psi_k(\eta), \tag{C.95}$$

where $\mathcal{N}_{\mathbf{k}}$ is a normalisation constant and $k := |\mathbf{k}|$. These modes satisfy the Klein-Gordon equation if $\psi_k(\eta)$ satisfies Bessel's differential equation:

$$z^2\frac{d^2\psi_k}{dz^2} + z\frac{d\psi_k}{dz} + (z^2 - \nu^2)\psi_k = 0, \tag{C.96}$$

where

$$z = -k\eta, \qquad \nu^2 = \frac{d^2}{4} - m^2\ell^2. \tag{C.97}$$

The BD positive-frequency modes are taken to be $\psi_k^{BD}(\eta) = H_\nu^{(1)}(-k\eta)$, where $H_\nu^{(1)}$ is the Hankel function of the first kind. In order to fix the normalisation $\mathcal{N}_{\mathbf{k}}$, we use the fact that these modes should be orthonormal with respect to the Klein-Gordon inner-product:

$$(u_{\mathbf{k}}^{BD}, u_{\mathbf{q}}^{BD}) = -(\overline{u}_{\mathbf{k}}^{BD}, \overline{u}_{\mathbf{q}}^{BD}) = \delta^{(d)}(\mathbf{k}-\mathbf{q}), \qquad (u_{\mathbf{k}}^{BD}, \overline{u}_{\mathbf{q}}^{BD}) = 0. \tag{C.98}$$

These conditions require the norm of $\mathcal{N}_{\mathbf{k}}$ to be $|\mathcal{N}_{\mathbf{k}}| = \sqrt{\frac{\pi}{4}}\ell^{\frac{-d+1}{2}}e^{-\pi\mathrm{Im}(\nu)/2}$, while leaving its phase unconstrained.[4] We choose the phase of $\mathcal{N}_{\mathbf{k}}$ such that the mode functions satisfy the property $\overline{u}_{\mathbf{k}}(x_P) = u_{-\mathbf{k}}(x_P^A)$, where x_P^A is the antipode of x_P. The function $\chi_k(\eta)$ has

[4] To derive this, note that in this foliation $n^0 = \frac{-\eta}{l}$, $n^i = 0$, and $d\Sigma = \left(\frac{-\ell}{\eta}\right)^d d^dx$. (See (4.3) for the definition of these quantities). Then

$$\begin{aligned}(u_{\mathbf{k}}^{BD}, u_{\mathbf{q}}^{BD}) &= i\int \frac{e^{i(\mathbf{q}-\mathbf{k})\cdot\mathbf{x}}}{(2\pi)^d}\left(\frac{-\ell}{\eta}\right)^{d-1}\left[\overline{\chi}_k\partial_\eta\chi_q - \overline{\chi}_q\partial_\eta\chi_k\right] \\ &= i|\mathcal{N}_{\mathbf{k}}|^2\ell^{d-1}\int\frac{e^{i(\mathbf{q}-\mathbf{k})\cdot\mathbf{x}}}{(2\pi)^d}(-\eta)\left[\overline{H}_\nu^{(1)}(-k\eta)\partial_\eta H_\nu^{(1)}(-q\eta) - H_\nu^{(1)}(-q\eta)\partial_\eta\overline{H}_\nu^{(1)}(-k\eta)\right].\end{aligned}$$

Since this inner product is conserved with time, it suffices to evaluate it for $\eta \to -\infty$, where the Hankel function has the simple asymptotic form $H_\nu(-k\eta) \to \sqrt{\frac{-2}{\pi k\eta}}e^{-i\left(k\eta + \frac{\pi\nu}{2} + \frac{\pi}{4}\right)}$ (see 10.2.5 of [154]). Plugging this back into the above expression, we find

$$(u_{\mathbf{k}}^{BD}, u_{\mathbf{q}}^{BD}) = \ell^{d-1}\frac{4}{\pi}e^{\pi\mathrm{Im}(\nu)}|\mathcal{N}_{\mathbf{k}}|^2\delta^{(d)}(\mathbf{k}-\mathbf{q}). \tag{C.99}$$

The desired result now follows by requiring (C.98).

a branch cut that can be taken to be the negative real axis, so the more precise statement is that we require

$$\overline{u}_{\mathbf{k}}(\eta, \mathbf{x}) = u_{-\mathbf{k}}(-\eta - i\epsilon, \mathbf{x}). \tag{C.100}$$

When ν is either purely real or imaginary, $\overline{H}^{(1)}_{\nu}(x) = -e^{i\pi \mathrm{Re}(\nu)} H^{(1)}_{\nu}(-x + i\epsilon)$ for real $x > 0$ and small positive ϵ. [5] Using this fact, we find that (C.100) will be satisfied if the phase of $\mathcal{N}_{\mathbf{k}}$ is $e^{i\pi\left(\frac{\mathrm{Re}(\nu)}{2} - \frac{d+2}{4}\right)}$ and so

$$\mathcal{N}_{\mathbf{k}} = |\mathcal{N}_{\mathbf{k}}| e^{i\pi\left(\frac{\mathrm{Re}(\nu)}{2} + \frac{d}{4}\right)} = \sqrt{\frac{\pi}{4}} \ell^{\frac{-d+1}{2}} e^{i\pi\left(\frac{\nu}{2} - \frac{d+2}{4}\right)}. \tag{C.101}$$

Collecting our results, the positive-frequency modes that define the BD vacuum $|BD\rangle$ take the form

$$u^{BD}_{\mathbf{k}}(\eta, \mathbf{x}) = \frac{e^{i\mathbf{k}\cdot\mathbf{x}}}{(2\pi)^{d/2}} \chi_k(\eta), \qquad \chi_k(\eta) = \sqrt{\frac{\pi\ell}{4}} e^{i\pi\left(\frac{\nu}{2} - \frac{d+2}{4}\right)} \left(\frac{-\eta}{\ell}\right)^{d/2} H^{(1)}_{\nu}(-k\eta). \tag{C.102}$$

It may also be verified that these modes minimise the Hamiltonian on the spatial slice at $\eta \to -\infty$.

C.9.2 Euclidean modes on dS^D

Our introduction of the Euclidean modes will follow that of [36], with some relevant additional details spelt out. In global coordinates, the de Sitter metric is given by (C.92). Since the spatial sections are d-spheres, it is natural to introduce spherical harmonics $Y_{Lj}(\Omega)$, which are a complete and orthonormal eigenbasis of the Laplacian $\nabla^2_{S^d}$ on S^d:

$$\nabla^2_{S^d} Y_{Lj} = -L(L + d - 1) Y_{Lj}, \tag{C.103}$$

$$\sum_{Lj} Y_{Lj}(\Omega) \overline{Y}_{Lj}(\Omega') = \delta^{(d)}(\Omega, \Omega'), \qquad \int Y_{Lj}(\Omega) \overline{Y}_{Lj}(\Omega) d\Omega = \delta_{LL'} \delta_{jj'}. \tag{C.104}$$

Here $L \in \{0, 1, 2, \dots\}$ and j is a collective index for $j_1, j_2, \dots, j_{d-1}$, which run over values $|j_{d-1}| \leq j_{d-2} \leq \cdots \leq j_1 \leq L$. We work with a particular choice of harmonics $Y_{Lj}(\Omega)$ (see [36]), which enjoy the useful property

$$\overline{Y}_{Lj}(\Omega) = (-1)^L Y_{Lj}(\Omega) = Y_{Lj}(\Omega^A), \tag{C.105}$$

[5] It follows from 10.11.9 and 10.11.5 of [154] that $H^{(1)}_{\nu}(-z) = -e^{-i\pi\nu} \overline{H}^{(1)}_{\overline{\nu}}(\overline{z})$. Letting $z = x - i\epsilon$, we find $\overline{H}^{(1)}_{\nu}(x) = -e^{-i\pi\overline{\nu}} H^{(1)}_{\overline{\nu}}(-x + i\epsilon)$. For real ν, the desired relation follows. For purely imaginary ν, we get the same result by using $H^{(1)}_{-\nu}(z) = e^{i\pi\nu} H^{(1)}_{\nu}(z)$ (10.4.6 of [154]).

where Ω^A is the antipodal point to Ω on S^d. Consider the modefunctions

$$u_{Lj}(t,\Omega) = y_L(t)Y_{Lj}(\Omega), \qquad y_L(t) = e^{(a+\nu)t/\ell}\cosh^L(t/\ell)v_L(t), \tag{C.106}$$

where ν is given by (C.97) and

$$a = L + d/2. \tag{C.107}$$

These modes satisfy the Klein-Gordon equation if $v_L(t)$ is a solution to the hypergeometric differential equation

$$z(1-z)\frac{d^2v_L}{dz^2} + [c - (a+b+1)\,z]\,\frac{dv_L}{dz} - abv_L = 0, \tag{C.108}$$

where $c = 2a$, $b = a + \nu$ and

$$z = z(t) = 1 + e^{2t/\ell}. \tag{C.109}$$

The Euclidean mode functions are defined by

$$v_L(t) = \mathcal{N}_L F(a, a+\nu; 2a; z(t) - i\epsilon), \tag{C.110}$$

where F is the hypergeometric function ${}_2F_1$ and $\mathcal{N}_L$ is a normalisation constant. More precisely, F stands for the hypergeometric function obtained by introducing a cut from 1 to ∞ on the real axis. This is exactly the range of interest to us and $-i\epsilon$ determines the side of the branch cut on which the function should be evaluated. The normalisation constant $\mathcal{N}_L$ is determined by requiring the modes to be orthonormal in the KG norm:

$$(u_{Lj}, u_{L'j'}) = -(\overline{u}_{Lj}, \overline{u}_{L'j'}) = \delta_{LL'}\delta_{jj'}, \qquad (\overline{u}_{Lj}, u_{L'j'}) = 0, \tag{C.111}$$

which is equivalent to

$$\begin{aligned} i &= \ell^d \cosh^d(t/\ell)\left[y_L\frac{d\overline{y}_L}{dt} - \frac{dy_L}{dt}\overline{y}_L\right] \\ &= \frac{\ell^{d-1}}{2^{2a-1}} z^{2a}(z-1)^{\mathrm{Re}(\nu)}\left\{(z-1)\left[v_L\frac{d\overline{v}_L}{dz} - \overline{v}_L\frac{dv_L}{dz}\right] - i\mathrm{Im}(\nu)v_L\overline{v}_L\right\}. \end{aligned} \tag{C.112}$$

Since the above expression is conserved in time, it suffices to look at the $z \to \infty$ (i.e. $t \to \infty$) limit. In that limit:

$$F(a, a+\nu; 2a; z(t) - i\epsilon) \xrightarrow{z\to\infty} z^{-a}e^{-i\pi a}\left[\gamma + \xi e^{-\nu\ln z}e^{-i\pi\nu}\right] \tag{C.113}$$

$$\frac{d}{dz}F(a, a+\nu; 2a; z(t) - i\epsilon) \xrightarrow{z\to\infty} z^{-a-1}e^{-i\pi(a+1)}\left[a\gamma + (a+\nu)\xi e^{-\nu\ln z}e^{-i\pi\nu}\right], \tag{C.114}$$

where all functions assume their principal values [6] and

$$\gamma = \frac{\Gamma(\nu)\Gamma(2a)}{\Gamma(a+\nu)\Gamma(a)}, \qquad \xi = \frac{\Gamma(-\nu)\Gamma(2a)}{\Gamma(a-\nu)\Gamma(a)}. \tag{C.115}$$

This expression is valid when $\nu \neq 0, \pm 1, \pm 2, \ldots$, $a \neq \nu$. [7] Note that because $\Gamma(\overline{z}) = \overline{\Gamma}(z)$, both γ and ξ are real when ν is real, and $\overline{\gamma} = \xi$ when ν is purely imaginary. Using these facts, evaluating (C.112) in the limit $z \to \infty$ constrains the norm of $\mathcal{N}_L$ to: [8]

$$|\mathcal{N}_L|^2 = \frac{e^{-\pi \mathrm{Im}(\nu)}}{2^{2a}\ell^{d-1}} \frac{\Gamma(a+\nu)\Gamma(a-\nu)}{\Gamma(a+\frac{1}{2})^2}. \tag{C.119}$$

Although the derivation of this result uses relations which are only valid for $\nu \neq 0, 1, 2, \ldots$, the final result is completely well-defined for such values. Therefore, we could imagine a limiting procedure in which we add a tiny amount ϵ to an integer value of ν, go through the same derivation, and then let ϵ go to zero.

We use the freedom in the phase of $\mathcal{N}_L$ to choose mode functions with the useful property

$$u_{Lj}(x_G^A) = \overline{u}_{Lj}(x_G). \tag{C.120}$$

[6] If z and c are two complex numbers, then $z^c = e^{c\mathrm{Log}(z)}$, where $\mathrm{Log}(z) = \ln(|z|) + i\Theta$, with $z = |z|e^{i\Theta}$ and $-\pi < \Theta \leq \pi$.

[7] To arrive at these expressions, we have used 15.1.1, 15.1.2, and 15.8.2 of [154] to obtain

$$\begin{aligned} \frac{\sin(\pi(b-a))}{\pi\Gamma(c)} F(a,b;c;z) = & \quad \frac{1}{\Gamma(b)\Gamma(c-a)\Gamma(a-b+1)} (-z)^{-a} F(a, a-c+1; a-b+1; 1/z) \\ + & \quad \frac{1}{\Gamma(a)\Gamma(c-b)\Gamma(b-a+1)} (-z)^{-b} F(b, b-c+1; b-a+1; 1/z). \end{aligned} \tag{C.116}$$

Here all functions assume their principal values, $|\mathrm{ph}(-z)| < \pi$, and $(b-a) \neq 0, \pm 1, \ldots$ Then using (C.148) to rewrite $\sin(\pi(b-a))$ in terms of Gamma functions and (C.147) to get $\Gamma(\pm(a-b)+1) = \pm(a-b)\Gamma(\pm(a-b))$, we find

$$\begin{aligned} F(a,b;c;z) = & \quad \frac{\Gamma(b-a)\Gamma(c)}{\Gamma(b)\Gamma(c-a)} (-z)^{-a} F(a, a-c+1; a-b+1; \frac{1}{z}) \\ + & \quad \frac{\Gamma(a-b)\Gamma(c)}{\Gamma(a)\Gamma(c-b)} (-z)^{-b} F(b, b-c+1; b-a+1; \frac{1}{z}). \end{aligned} \tag{C.117}$$

We can also relate the derivative of F to another hypergeometric function using 15.5.1 of [154]:

$$\frac{d}{dz} F(a,b;c;z) = \frac{ab}{c} F(a+1, b+1; c+1; z). \tag{C.118}$$

Noting that for any complex number c and $1 < z < \infty$ we have $(z - i\epsilon)^c = e^{c \ln z} e^{ic\pi}$, and also using the fact that $F(a,b;c;0) = 1$, the desired expressions follow.

[8] Here we have used 15.5.5 of [154] to rewrite $\Gamma(2a) = \pi^{-1/2} 2^{2a-1} \Gamma(a)\Gamma(a+1/2)$.

Given that we have chosen spherical harmonics with the property $\overline{Y}_{Lj}(\Omega) = Y_{Lj}(\Omega^A)$, this condition reduces to

$$y_L(-t) = \overline{y}_L(t), \tag{C.121}$$

which can be achieved by setting [9]

$$\mathcal{N}_L = |\mathcal{N}_L| e^{i\frac{\pi}{2}[a+\mathrm{Re}(\nu)]}. \tag{C.127}$$

Collecting our results, the Euclidean modes are

$$u^E_{Lj}(t,\Omega) = y^E_L(t) Y_{Lj}(\Omega), \qquad y^E_L(t) = \mathcal{N}_L e^{(a+\nu)t/\ell} \cosh^L(t/\ell) F(a, a+\nu; 2a; z(t) - i\epsilon), \tag{C.128}$$

where $z(t) = 1 + e^{2t/\ell}$, $a = L + d/2$ and

$$\mathcal{N}_L = \frac{e^{i\frac{\pi}{2}(a+\nu)}}{2^a \ell^{\frac{d-1}{2}}} \frac{\sqrt{\Gamma(a+\nu)\Gamma(a-\nu)}}{\Gamma(a+\frac{1}{2})}. \tag{C.129}$$

C.9.3 Two-point functions and α-vacua

That the Euclidean and the BD modes define the same physical state is made apparent by the fact that the two-point function W_E associated with the Euclidean modes (C.128), when restricted to the Poincaré patch, coincides with the two-point function W_{BD} associated with the Bunch-Davies modes (C.102). (They are functions of the geodesic distance

[9] To see this, let $\mathcal{N}_L = |\mathcal{N}_L| e^{i\Theta}$. It follows from the definition of F (see e.g. 15.2.1 of [154]) and $\Gamma(\overline{z}) = \overline{\Gamma}(z)$:

$$\overline{F}(a, a+\nu; 2a; z - i\epsilon) = \begin{cases} F(a, a+\nu; 2a; z + i\epsilon) & \nu \text{ real} \\ F(a, a-\nu; 2a; z + i\epsilon) & \nu \text{ imaginary.} \end{cases} \tag{C.122}$$

Using 15.8.1 of [154], it may be checked that

$$F(a, a+\nu; 2a; z(t) + i\epsilon) = (1 - z(t) - i\epsilon)^{-a-\nu} F(a, a+\nu; 2a; z(t)/\left(z(t) - 1\right) - i\epsilon) \tag{C.123}$$

$$= e^{-2(a+\nu)t/l} e^{i\pi(a+\nu)} F(a, a+\nu; 2a; z(-t) - i\epsilon). \tag{C.124}$$

Using the relations above when ν is real, it follows from the definition of $y_L(t)$ that $y_L(-t) = e^{2i\Theta} e^{-i\pi(a+\nu)} \overline{y}_L(t)$. The same formula in [154] also guarantees

$$F(a, a-\nu; 2a; z(t) + i\epsilon) = (1 - z(t) - i\epsilon)^{-a} F(a, a+\nu; 2a; z(t)/\left(z(t) - 1\right) - i\epsilon) \tag{C.125}$$

$$= e^{-2at/l} e^{i\pi a} F(a, a+\nu; 2a; z(-t) - i\epsilon). \tag{C.126}$$

Using this expression and (C.122) when ν is purely imaginary, it follows that $y_L(-t) = e^{2i\Theta} e^{-i\pi a} \overline{y}_L(t)$. Combining these results we find $\Theta = \frac{\pi}{2}[a + \mathrm{Re}(\nu)]$.

and the causal relation between their arguments, which are both coordinate independent quantities [41, 153].) For the Euclidean state, the two-point function is given by [153]

$$W_E(x,y) = \frac{\Gamma[h_+]\Gamma[h_-]}{4\pi\ell^2\Gamma[\frac{D}{2}]}{}_2F_1\left(h_+, h_-, \frac{D}{2}; \frac{1+Z(x,y)+i\epsilon\,\mathrm{sign}(x^0-y^0)}{2}\right), \tag{C.130}$$

where $h_\pm = \frac{d}{2} \pm \nu$ and ${}_2F_1(a,b,c;z)$ is the hypergeometric function (see (C.86) and (C.97) for the definitions of Z and ν). The $i\epsilon$ prescription selects the side of the branch cut from $Z = 1$ to $Z = +\infty$ on which the function should be evaluated when x and y are causally related (when x and y are spacelike, then $Z < 1$ and the values of the function below and above the real line coincide). The Hadamard function is equal to the real part $H_E(x,y) = \mathrm{Re}\,[W_E(x,y)]$, which depends only on the coordinate-independent quantity $Z(x,y)$. The Pauli-Jordan function and the retarded Green function can be written in terms of $W_E(x,y)$, since $i\Delta(x,y) = 2\mathrm{Im}\,[W_E(x,y)]$ and $G_R(x,y) = \theta(x^0-y^0)\Delta(x,y)$.

We denote the two-real-parameter family of dS-invariant α-vacua by $|\alpha,\beta\rangle$. Their modefunctions can be obtained through a Bogoliubov transformation [12]

$$u_{\mathbf{k}}^{(\alpha,\beta)} = \cosh(\alpha)u_{\mathbf{k}}^{BD} + \sinh(\alpha)e^{i\beta}\overline{u}_{-\mathbf{k}}^{BD}, \tag{C.131}$$

for the BD modes, and

$$u_{Lj}^{(\alpha,\beta)} = \cosh(\alpha)u_{Lj}^{E} + \sinh(\alpha)e^{i\beta}\overline{u}_{Lj}^{E}, \tag{C.132}$$

for the Euclidean modes. Here $\alpha \in \mathbb{R}^+$ and $\beta \in \mathbb{R}$ is defined modulo 2π. Recall the relations between negative frequency modes and positive frequency modes taking antipodal arguments, which can be obtained for both the Euclidean modes (C.120) [12], and the Poincaré modes (C.100) by appropriate choices of the arbitrary complex phases in the normalisation factors. Because of these relations, it is possible to express the two-point function $W_{\alpha,\beta}(x,y)$ associated to an arbitrary α-vacuum in terms of the Euclidean/BD two-point function $W_E(x,y)$ (C.130). The imaginary part of $W_{\alpha,\beta}(x,y)$ is always equal to $i\Delta(x,y)$ and hence identical for all α-vacua. The real part, i.e. the Hadamard function, depends on α and β. By computing the mode sums using the α-modes, the family of de Sitter invariant Hadamard functions $H_{\alpha,\beta}(x,y)$ can be obtained and reads [12]:

$$H_{\alpha,\beta}(x,y) = \cosh 2\alpha\, H_E(x,y) + \sinh 2\alpha\left[\cos\beta\, H_E(x^A,y) - \sin\beta\Delta(x^A,y)\right]. \tag{C.133}$$

The two-point function for an α-vacuum is thus given by $W_{\alpha,\beta}(x,y) = \frac{1}{2}H_{\alpha,\beta}(x,y) + \frac{i}{2}\Delta(x,y)$. In this particular parametrisation of the α-vacua [12], the Euclidean state cor-

responds to $\alpha = 0$.[10] The derivation of (C.133) for modes on the Poincaré patch requires evaluating the BD Hadamard function outside its domain of validity. Specifically, one uses the property that

$$H^{BD}(\eta_x, \mathbf{x}; -\eta_y - i\epsilon, \mathbf{y}) := \int d^d\mathbf{k} \left[u^{BD}_{\mathbf{k}}(\eta_x, \mathbf{x})\overline{u}^{BD}_{\mathbf{k}}(-\eta_y - i\epsilon, \mathbf{y}) + \overline{u}^{BD}_{\mathbf{k}}(\eta_x, \mathbf{x})u^{BD}_{\mathbf{k}}(-\eta_y - i\epsilon, \mathbf{y})\right]$$
$$= H^E(x, y^A), \tag{C.134}$$

where $H^E(x, y^A)$ is the Hadamard function of the Euclidean vacuum, which is of course defined on all of de Sitter space. This implies that for a given choice of α and β, the two-point function associated with the modes (C.131) is the restriction of the global α-vacua two-point function, defined via the modes (C.132), to the Poincaré patch.

Two α-vacua which will be of special interest to us are the *in*- and *out*-vacua [104, 36] :

$$\alpha_{in} = \alpha_{out} = \tanh^{-1} e^{-\pi|\nu|}, \qquad \beta_{in} = -\beta_{out} = \frac{D+1}{2}\pi, \tag{C.135}$$

which have no incoming/outgoing particles at past/future infinity, respectively. [11] In other words, they minimise the Hamiltonian on spatial slices at $t \to \pm\infty$ in global coordinates, as shown in Figure C.1. Notice that in odd spacetime dimensions, the *in* and *out*-vacua are the same, i.e. they are related by a trivial Bogoliubov transformation, since then $\exp(i\beta_{in}) = \exp(i\beta_{out})$ ("odd-dimensional de Sitter space is transparent" [36, 89]). It is also worth pointing out that for masses much larger than the Hubble radius, $m \gg m_* = (D-1)/2\ell$, the *in*/*out* states are "exponentially close" to the Euclidean state, since then $|\nu| = \frac{1}{2}\ell\sqrt{m^2 - m_*^2} \gg 1$ and $\sinh(\alpha) \sim e^{-\pi|\nu|}$.

[10] The relation between the parametrisation used here and that of [104, 36], which uses a single complex parameter $\tilde{\alpha}$, is $\mathrm{Re}(\tilde{\alpha}) = \ln\tanh\alpha$ and $\mathrm{Im}(\tilde{\alpha}) = \beta$. The notation used here will be more convenient in the analysis of the SJ vacuum on a causal set, because the Euclidean state then corresponds to a finite value $\alpha = 0$ instead of $\tilde{\alpha} = -\infty$.

[11] The modefunctions associated with these choices of α and β correspond to $\tilde{\phi}^{in}_{Lj}$ and $\tilde{\phi}^{out}_{Lj}$ defined in [36], which differ from the usually defined in/out modes by a constant phase. Of course, these two choices define the same vacuum state because the two-point function is insensitive to any constant-phase rescaling of modefunctions.

C.10 Calculation of Inner Products

C.10.1 Poincaré chart

Here we shall evaluate (4.72):

$$r_k = \frac{\langle \overline{\chi}_k, \chi_k \rangle_\eta}{\langle \chi_k, \chi_k \rangle_\eta}. \tag{C.136}$$

It follows from the definition of χ_k^{BD} that

$$\langle \overline{\chi}_k, \chi_k \rangle_\eta = \frac{\pi \ell}{4} e^{i\pi\left(\nu - \frac{d+2}{2}\right)} \int_{\eta_{min}}^{\eta_{max}} \left[H_\nu^{(1)}(-k\eta)\right]^2 \left(\frac{-\ell}{\eta}\right) d\eta \tag{C.137}$$

$$\langle \chi_k, \chi_k \rangle_\eta = \frac{\pi \ell}{4} e^{-\pi \mathrm{Im}(\nu)} \int_{\eta_{min}}^{\eta_{max}} \left|H_\nu^{(1)}(-k\eta)\right|^2 \left(\frac{-\ell}{\eta}\right) d\eta. \tag{C.138}$$

Changing integration variables to $x = -k\eta$, and defining $\epsilon = -k\eta_{min}$, $x_m = -k\eta_{max}$, we find:

$$r_k = e^{i\pi\left(\mathrm{Re}(\nu) - \frac{d+2}{2}\right)} F(\epsilon, x_m) \qquad \text{where} \qquad F(\epsilon, x_m) = \frac{\int_\epsilon^{x_m} \frac{dx}{x} \left[H_\nu^{(1)}(x)\right]^2}{\int_\epsilon^{x_m} \frac{dx}{x} \left|H_\nu^{(1)}(x)\right|^2}. \tag{C.139}$$

Let us list a few useful properties of the Hankel function $H_\nu^{(1)}(z)$. It satisfies the Bessel equation $[z^2 \frac{d^2}{dz^2} + z\frac{d}{dz} + (z^2 - \nu^2)]H_\nu^{(1)}(z) = 0$ and has the defining property (see 10.2.5 of [154])

$$H_\nu^{(1)}(z) \to \sqrt{\frac{2}{\pi z}} e^{i\left(z - \frac{\pi\nu}{2} - \frac{\pi}{4}\right)}, \tag{C.140}$$

as $z \to \infty$ in $-\pi + \delta \leq \mathrm{ph} z \leq 2\pi - \delta$, where δ is an arbitrary small positive number. It has a branch point at $z = 0$ and its principal branch corresponds to the principal value of the square root in (C.140), with a branch cut along $(-\infty, 0]$. [12] From here on out $H_\nu^{(1)}(z)$ will denote the principal value of this function. The asymptotic behaviour of $H_\nu^{(1)}(z)$ as $z \to 0$

[12] $\mathrm{PV}(z^{-\frac{1}{2}}) = e^{-\frac{1}{2}\mathrm{Log}(z)}$, where $\mathrm{Log}(z) = \ln(r) + i\Theta$ with $z = re^{i\Theta}$ and $-\pi < \Theta \leq \pi$.

is also of interest to us:[13]

$$H_0^{(1)}(z) \to \left(\frac{2i}{\pi}\right) \mathrm{Log}(z) \tag{C.141}$$

$$H_\nu^{(1)}(z) \to -\left(\frac{i}{\pi}\right) \Gamma(\nu) e^{-\nu \mathrm{Log}(z/2)}, \qquad \mathrm{Re}(\nu) > 0 \tag{C.142}$$

$$H_{i\nu}^{(1)}(z) \to A_\nu e^{i\nu \mathrm{Log}(z/2)} + B_\nu e^{-i\nu \mathrm{Log}(z/2)}, \qquad \nu \in \mathbb{R}, \nu \neq 0. \tag{C.143}$$

where

$$A_\nu = \frac{1 + \coth(\pi\nu)}{\Gamma(1 + i\nu)}, \qquad B_\nu = -\frac{\mathrm{csch}(\pi\nu)}{\Gamma(1 - i\nu)}. \tag{C.144}$$

Since our goal is to evaluate (C.139), we are only interested in positive values of z, for which $\mathrm{Log}(z) = \ln(x)$. For finite ϵ, as can be seen from (C.140), both integrals in the numerator and denominator of $F(\epsilon, x_m)$ converge as $x_m \to \infty$. Moreover, (C.141)−(C.143) show that both integrals diverge in the limit $\epsilon \to 0$, which means we can let $x_m = \infty$ and only concern ourselves with the behaviour of the integrands close to zero. Doing so, (C.141) and (C.142) imply

$$\lim_{\substack{\epsilon \to 0 \\ x_m \to \infty}} F(\epsilon, x_m) = -1 \qquad \text{for} \qquad \nu \geq 0. \tag{C.145}$$

Similarly, (C.142) implies

$$\lim_{\substack{\epsilon \to 0 \\ x_m \to \infty}} F(\epsilon, x_m) = \frac{2A_\nu B_\nu}{|A_\nu|^2 + |B_\nu|^2} = -\mathrm{sech}(\pi|\nu|) \qquad \text{for} \qquad \nu = i|\nu|, \nu \neq 0. \tag{C.146}$$

To derive this last equality, we have used the following properties of the Gamma function (see 5.5.1, 5.5.3 and 5.4.3 of [154]):

$$\Gamma(z+1) = z\Gamma(z) \tag{C.147}$$

$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)} \qquad z \neq 0, \pm 1, \pm 2, \ldots \tag{C.148}$$

$$|\Gamma(iy)| = \sqrt{\frac{\pi}{y \sinh(\pi y)}} \qquad y \in \mathbb{R}. \tag{C.149}$$

It then follows that

$$\Gamma(1 + i|\nu|)\Gamma(1 - i|\nu|) = (i|\nu|)\Gamma(i|\nu|)\Gamma(1 - i|\nu|) = i|\nu|\pi / \sin(i\pi|\nu|) = \pi|\nu| / \sinh(\pi|\nu|) \tag{C.150}$$

[13] In [154] see 10.7.2 for (C.141), 10.7.7 for (C.142), and a combination of 10.4.3, 10.7.3, and 10.7.6 for (C.141).

and

$$|\Gamma(1 \pm i|\nu|)| = |\pm i|\nu|\Gamma(\pm i|\nu|)| = \sqrt{\pi|\nu|/\sinh(\pi|\nu|)}. \tag{C.151}$$

Using these expressions we obtain

$$\frac{2A_\nu B_\nu}{|A_\nu|^2 + |B_\nu|^2} = \frac{-2[1+\coth(\pi|\nu|)]\mathrm{csch}(\pi|\nu|)}{[1+\coth(\pi|\nu|)]^2 + \mathrm{csch}^2(\pi|\nu|)} = -\mathrm{sech}(\pi|\nu|).$$

Figure C.2 provides numerical evidence for these calculations. We have computed $|F(\epsilon, x_m)|$ numerically and plotted its behaviour as a function of ϵ. These results are consistent with the analytical arguments provided above.

Summarising our results:

$$r_k = \begin{cases} e^{i\pi\left(\nu - \frac{d}{2}\right)} & \text{if } m \leq m_*, \\ e^{-i\pi\frac{d}{2}}\mathrm{sech}(\pi|\nu|) & \text{if } m \geq m_*. \end{cases} \tag{C.152}$$

C.10.2 Global chart

Here we shall evaluate (4.84):

$$r_L := (-1)^L \frac{\langle \overline{y}_L^E, y_L^E \rangle_t}{\langle y_L^E, y_L^E \rangle_t}. \tag{C.153}$$

It follows from (C.121) that

$$\langle y_L^E, y_L^E \rangle_t = 2\int_0^T |y_L^E(t)|^2 l^d \cosh^d(t/l) dt \tag{C.154}$$

$$\langle \overline{y}_L^E, y_L^E \rangle_t = 2\int_0^T \mathrm{Re}\left[y_L^E(t)^2\right] l^d \cosh^d(t/l) dt. \tag{C.155}$$

Changing integration variables to $z(t) = 1 + e^{2t/l}$, these integrals become

$$\langle \overline{y}_L^E, y_L^E \rangle_t = \frac{l^{d+1}}{2^{2a}} \mathrm{Re}\left\{\int_2^{z_T} \mathcal{N}_L^2 I_1(z) dz\right\}, \qquad \langle y_L^E, y_L^E \rangle_t = \frac{l^{d+1}}{2^{2a}} \int_2^{z_T} |\mathcal{N}_L|^2 I_2(z) dz, \tag{C.156}$$

where

$$I_1(z) := \frac{1}{\mathcal{N}_L^2} v_L(z)^2 (z-1)^{\nu-1} z^{2a}, \qquad I_2(z) := \frac{1}{|\mathcal{N}_L|^2} |v_L(z)|^2 (z-1)^{\mathrm{Re}(\nu)-1} z^{2a}. \tag{C.157}$$

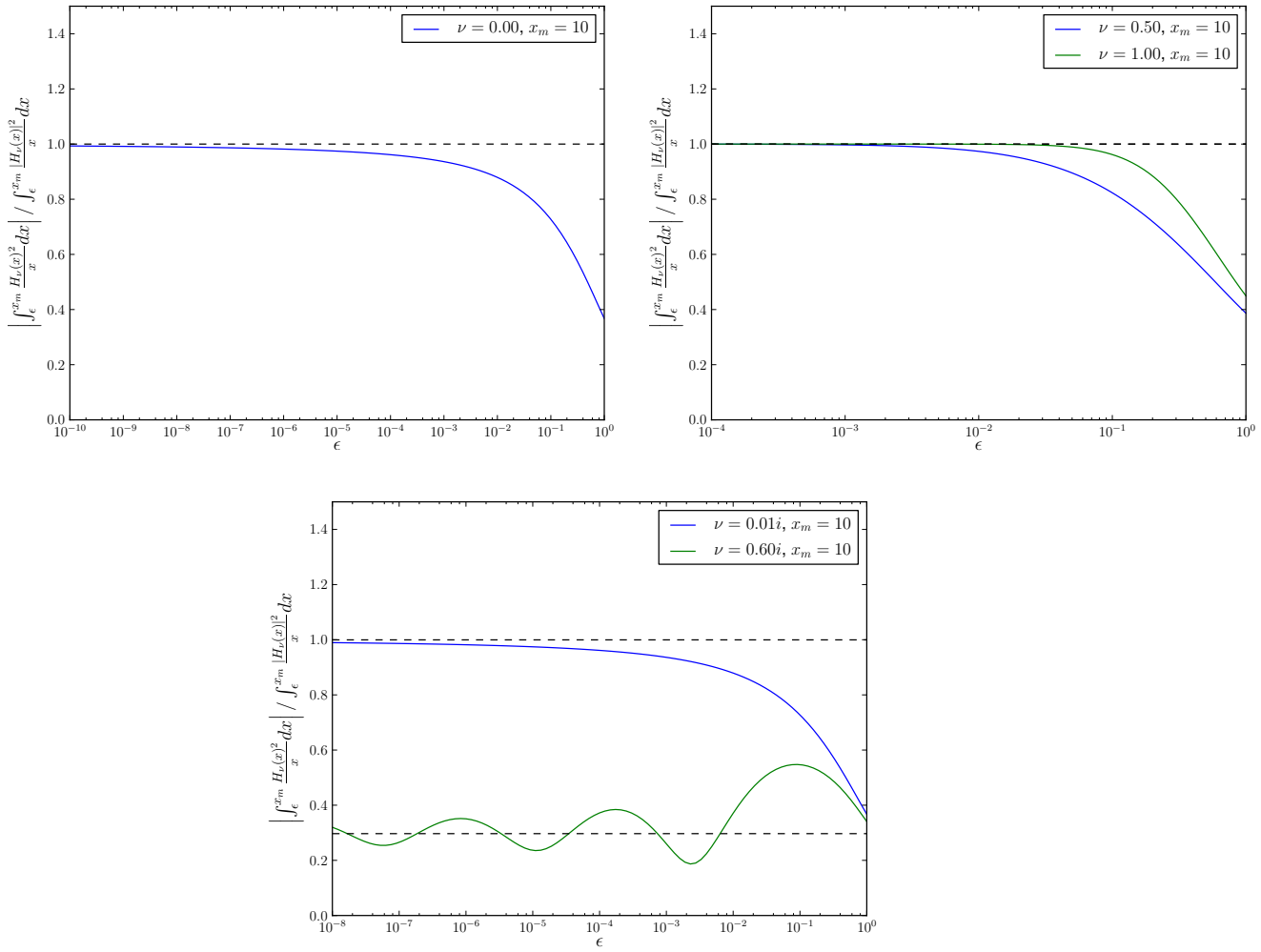


Figure C.2: Behaviour of $|F(\epsilon, x_m)|$ as a function of ϵ. The dependence on x_m is negligible for small ϵ.

Here $z_T = z(T)$ and all other quantities have been defined in Section C.9.2. With these definitions:

$$\lim_{z_T \to \infty} \frac{\langle \overline{y}_L^E, y_L^E \rangle_t}{\langle y_L^E, y_L^E \rangle_t} = \lim_{z_T \to \infty} \frac{\mathrm{Re}\left\{\mathcal{N}_L^2 \int_2^{z_T} I_1(z) dz\right\}}{|\mathcal{N}_L|^2 \int_2^{z_T} I_2(z) dz}. \tag{C.158}$$

The lower limit of these integrals is completely well-behaved, but they diverge in the limit where $z_T \to \infty$. Therefore, it suffices to study the integrands in this limit only. Using the

asymptotic behaviour of the hypergeometric function given in (C.113), it can be checked that **when ν is real**: (see Section C.9.2 for definition of γ and ξ)

$$I_1(z) \xrightarrow{z\to\infty} \gamma^2 e^{-2i\pi a} z^{\nu-1}, \qquad I_2(z) \xrightarrow{z\to\infty} |\gamma|^2 z^{\nu-1}. \tag{C.159}$$

Given that both quantities have the same scaling with z in this limit, their ratio must converge to a constant when $z_T \to \infty$:

$$\lim_{z_T\to\infty} \frac{\langle \overline{y}_L^E, y_L^E\rangle_t}{\langle y_L^E, y_L^E\rangle_t} = \frac{\mathrm{Re}\left\{|\mathcal{N}_L|^2 e^{2i\Theta} e^{-2i\pi a}\right\}}{|\mathcal{N}_L|^2} = \cos[\pi(\nu-a)], \tag{C.160}$$

where $\Theta = \frac{\pi}{2}[a + \mathrm{Re}(\nu)]$. Here we have used the fact that $\gamma = \overline{\gamma}$ when ν is real. **When ν is imaginary**, it follows from (C.113) that

$$I_1(z) \xrightarrow{z\to\infty} e^{-2i\pi a} z^{-1} \left[2\gamma\xi e^{\pi|\nu|} + \gamma^2 e^{i|\nu|\ln(z)} + \xi^2 e^{\pi|\nu|} e^{-i|\nu|\ln z}\right] \tag{C.161}$$

$$I_2(z) \xrightarrow{z\to\infty} z^{-1} \left[|\gamma|^2 + |\xi|^2 e^{2\pi|\nu|} + \gamma\overline{\xi} e^{i|\nu|\ln(z)} e^{\pi|\nu|} + \overline{\gamma}\xi e^{-i|\nu|\ln(z)} e^{\pi|\nu|}\right]. \tag{C.162}$$

Again, since both quantities have the same scaling with z in this limit, the ratio of their integrals converges to a constant as $z_T \to \infty$:

$$\lim_{z_T\to\infty} \frac{\langle \overline{y}_L^E, y_L^E\rangle_t}{\langle y_L^E, y_L^E\rangle_t} = \frac{\mathrm{Re}\left\{2\gamma\xi|\mathcal{N}_L|^2 e^{2i\Theta} e^{-2i\pi a}\right\}}{|\mathcal{N}_L|^2(|\gamma|^2 + |\xi|^2 e^{2\pi\nu})} \tag{C.163}$$

$$= \cos(\pi a)\mathrm{sech}(\pi|\nu|), \tag{C.164}$$

having used the fact that $\gamma = \overline{\xi}$ when ν is imaginary. Notice that

$$\cos\left[\pi(\nu-a)\right] = \cos(\pi L + \pi d/2 - \pi\nu) = (-1)^L \cos(\pi d/2 - \pi\nu) \tag{C.165}$$

$$= (-1)^L \cos(\pi D/2 - \pi\nu - \pi/2) = (-1)^L \sin(\pi D/2 - \pi\nu). \tag{C.166}$$

Similarly, $\cos(\pi a) = (-1)^L \sin(\pi D/2)$. Summarising our results:

$$r_L = \begin{cases} \sin\dfrac{D}{2}\pi\,\mathrm{sech}\,\pi|\nu| & \text{if } m \geq m_*, \\ \sin\left[\left(\dfrac{D}{2} - |\nu|\right)\pi\right] & \text{if } 0 < m \leq m_*. \end{cases} \tag{C.167}$$

C.11 Sprinkling into a diamond in dS^2

To produce a sprinkling $\mathcal{C}_M$ into a causal diamond M in dS^2, we need to pick a coordinate chart. The cosmological coordinates x_P^μ defined in (C.93) are well suited because they

have a conformally flat metric, which makes it particularly simple to compute the causal relation between points, given their coordinate values. Even though this chart only covers half of de Sitter space, there is no loss of generality because the symmetries of de Sitter space imply that any causal diamond can be isometrically mapped to a causal diamond entirely contained in the Poincaré patch.

Let M be a causal diamond between two points $p, q \in dS^2_P$ such that $p \prec q$. Denote the (timelike) geodesic distance between p and q by τ. Since any two causal diamonds with the same value of τ are isometric, we choose $x^\mu_p = (\eta_\tau, 0)$ and $x^\mu_q = (\ell^2/\eta_\tau, 0)$ with

$$\eta_\tau = -\ell e^{\tau/2\ell} < -\ell, \tag{C.168}$$

without loss of generality. To obtain a sprinkling $\mathcal{C}_M$ into M we first generate a uniform Poisson distribution of N points in the square $[0,1]^2$ using a Mersenne Twister algorithm [100]. We use Cartesian coordinates y_1, y_2 on $[0,1]^2$ and find an embedding $\varphi : [0,1]^2 \to R$, which for any subset $A \subset [0,1]^2$ satisfies

$$V_M \int_A dy_1 dy_2 = \int_{\varphi(A)} d^2x \sqrt{-g}. \tag{C.169}$$

The factor V_M on the left hand side guarantees that the embedding scales the volume correctly. Its value for the causal diamond of length τ is

$$V_M = 4\ell^2 \ln\cosh \frac{\tau}{2\ell}. \tag{C.170}$$

By inspection it can be shown that the embedding $\varphi : (y_1, y_2) \to (\eta, x)$ defined by

$$\begin{aligned} \eta &= \frac{-\ell e^{\tau/2\ell}}{1 + y_1(e^{\tau/\ell} - 1)}, \\ r &= (1 - 2y_2) \sinh \frac{\tau}{2\ell}, \end{aligned} \tag{C.171}$$

satisfies the above condition (C.169). By keeping only such points for which $|x| < \min(\eta_\tau - \eta, \eta - \ell^2/\eta_\tau)$ and recording the causal relations among them, we obtain a sprinkling $\mathcal{C}_M$ into M. Note that, as explained above, we also calculate the geodesic distance between any two points using the metric on the manifold, even though this data is not explicitly part of $\mathcal{C}_M$.

Appendix D

Supplementary material for Chapter 5

D.1 Gravitational Aether Fluid Equations

The full set of equations that define the Gravitational Aether theory are

$$(8\pi\tilde{G})^{-1}G_{\mu\nu} = T_{\mu\nu} - \frac{1}{4}T^{\alpha}{}_{\alpha}g_{\mu\nu} + \tilde{T}_{\mu\nu}, \tag{D.1}$$

$$\tilde{T}_{\mu\nu} = \tilde{p}(\tilde{u}_{\mu}\tilde{u}_{\nu} + g_{\mu\nu}), \qquad \tilde{u}^{\mu}\tilde{u}_{\mu} = -1, \tag{D.2}$$

$$\nabla^{\mu}T_{\mu\nu} = 0, \tag{D.3}$$

where $\tilde{T}_{\mu\nu}$, $\tilde{p}$, and $\tilde{u}_{\mu}$ are the energy-momentum tensor, pressure, and normalized four-velocity of the aether fluid, respectively. The energy-momentum tensor $T_{\mu\nu}$ of "conventional" matter (i.e. dust, radiation, humans, etc.) is assumed to be conserved. By applying ∇^{μ} to both sides of (D.1) and using $\nabla^{\mu}G_{\mu\nu} = 0$ and $\nabla^{\mu}T_{\mu\nu} = 0$, we find

$$\nabla^{\mu}\tilde{T}_{\mu\nu} = \frac{1}{4}\nabla_{\nu}T. \tag{D.4}$$

This equation can be cast as a relativistic version of the continuity and Euler equations for an incompressible fluid. Before doing so, let us go through some useful relativistic algebra. Consider any perfect fluid with pressure p, density ρ, and normalized four-velocity u^{μ}:

$$\mathrm{T}_{\mu\nu} = (\rho + \mathrm{p})\mathrm{u}_{\mu}\mathrm{u}_{\nu} + \mathrm{p}g_{\mu\nu}, \qquad \mathrm{u}^{\mu}\mathrm{u}_{\mu} = -1, \tag{D.5}$$

It can then be checked that [1]

$$\mathrm{u}^\nu \nabla^\mu \mathrm{T}_{\mu\nu} = -\mathrm{u}^\mu \nabla_\mu \rho - (\rho + \mathrm{p}) \nabla_\mu \mathrm{u}^\mu, \tag{D.6}$$

$$P^\sigma{}_\nu \nabla^\mu \mathrm{T}_{\mu\sigma} = (\nabla_\nu + \mathrm{u}_\nu \mathrm{u}^\sigma \nabla_\sigma)\mathrm{p} + (\rho + \mathrm{p})\mathrm{u}^\mu \nabla_\mu \mathrm{u}_\nu, \tag{D.7}$$

where $P^\sigma{}_\nu$ is the tensor which de-projects the component of any vector along u^ν: (i.e. $P^\sigma{}_\nu \mathrm{u}^\nu = 0$)

$$P^\sigma{}_\nu = \delta^\sigma{}_\nu + \mathrm{u}^\sigma \mathrm{u}_\nu. \tag{D.8}$$

Aether is a perfect fluid with zero density. Applying $\tilde{u}^\nu$ to both sides of (D.4) and using (D.6) with $\rho = 0$, $\mathrm{p} = \tilde{p}$, $\mathrm{u}^\mu = \tilde{u}^\mu$, we find

$$\tilde{p} \nabla_\mu \tilde{u}^\mu = -\frac{1}{4} \tilde{u}^\mu \nabla_\mu T. \tag{D.9}$$

This is the relativistic continuity equation for the aether fluid. Similarly, by applying $\tilde{P}^\sigma{}_\nu = \delta^\sigma{}_\nu + \tilde{u}^\sigma \tilde{u}_\nu$ to both sides of (D.4) and using (D.7), we get the relativistic Euler equation:

$$\tilde{p} \tilde{u}^\mu \nabla_\mu \tilde{u}_\nu = -\tilde{\nabla}^\perp_\nu (\tilde{p} - T/4), \tag{D.10}$$

where

$$\tilde{\nabla}^\perp_\nu = \nabla_\nu + \tilde{u}_\nu \tilde{u}^\sigma \nabla_\sigma. \tag{D.11}$$

Lemma 2. *Let the matter energy-momentum tensor $T_{\mu\nu} = \sum_i T^{(i)}_{\mu\nu}$ consist of a collection of co-moving perfect fluids $T^{(i)}_{\mu\nu} = (\rho_i + p_i) u_\mu u_\nu + p_i g_{\mu\nu}$, each of which is separately conserved, i.e. $\nabla^\mu T^{(i)}_{\mu\nu} = 0 \ \forall \ i$. Assume that all $T^{(i)}_{\mu\nu}$ have a constant equation of state, i.e. $p_i = w_i \rho_i$ where w_i is constant. Then,*

$$\tilde{u}^\mu = u^\mu, \qquad \tilde{p} = \frac{1}{4} \sum_i (1 + w_i)(3w_i - 1)\rho_i, \tag{D.12}$$

solves (D.4) *and the GA field equations* (D.1) *become*

$$(8\pi \tilde{G})^{-1} G_{\mu\nu} = \frac{3}{4} \sum_i (1 + w_i) T^{(i)}_{\mu\nu}. \tag{D.13}$$

[1] In arriving at these expressions, it is useful to note that $\mathrm{u}_\mu \nabla_\nu \mathrm{u}^\mu = 0$. This can be derived by applying ∇_ν to both sides of $\mathrm{u}^\mu \mathrm{u}_\mu = -1$.

Proof. Conservation of each $T^{(i)}_{\mu\nu}$ is equivalent to (see (D.6) and (D.7))

$$u^\mu \nabla_\mu \rho_i + (\rho_i + p_i)\nabla_\mu u^\mu = 0, \tag{D.14}$$

$$\nabla^\perp_\nu p_i + (\rho_i + p_i) u^\mu \nabla_\mu u_\nu = 0, \tag{D.15}$$

where $\nabla^\perp_\nu = \nabla_\nu + u_\nu u^\sigma \nabla_\sigma$. With this at hand let us now turn to (D.1), which we showed is equivalent to (D.9) and (D.10). Let us show that (D.9) is satisfied when the aether pressure and four-velocity are given by (D.12). Note that $T^\alpha{}_\alpha = \sum_i (3p_i - \rho_i) = \sum_i (3w_i - 1)\rho_i$. Then

$$\begin{aligned}
\tilde{p}\nabla_\mu \tilde{u}^\mu + \frac{1}{4}\tilde{u}^\mu \nabla_\mu T &= \frac{1}{4}\sum_i (1+w_i)(3w_i-1)\rho_i \nabla_\mu u^\mu + \frac{1}{4}\sum_i (3w_i - 1) u^\mu \nabla_\mu \rho_i \\
&= \frac{1}{4}\sum_i (3w_i - 1)\left\{(1+w_i)\rho_i \nabla_\mu u^\mu + u^\mu \nabla_\mu \rho_i\right\} \\
&= \frac{1}{4}\sum_i (3w_i - 1)\left\{(\rho_i + p_i)\nabla_\mu u^\mu + u^\mu \nabla_\mu \rho_i\right\} \\
&= 0,
\end{aligned} \tag{D.16}$$

where the last line follows from (D.14). To see that (D.10) is also satisfied:

$$\begin{aligned}
\tilde{p}\tilde{u}_\mu \nabla^\mu \tilde{u}_\nu + \tilde{\nabla}^\perp_\nu (\tilde{p} - T/4) &= \frac{1}{4}\sum_i (1+w_i)(3w_i-1)\rho_i u^\mu \nabla_\mu u_\nu \\
&\quad + \frac{1}{4}\sum_i (1+w_i)(3w_i-1)\nabla^\perp_\nu \rho_i - \frac{1}{4}\sum_i (3w_i - 1)\nabla^\perp_\nu \rho_i \\
&= \frac{1}{4}\sum_i (1+w_i)(3w_i-1)\rho_i u^\mu \nabla_\mu u_\nu + (3w_i - 1) w_i \nabla^\perp_\nu \rho_i \\
&= \frac{1}{4}\sum_i (3w_i - 1)\left\{(\rho_i + p_i) u^\mu \nabla_\mu u_\nu + \nabla^\perp_\nu p_i\right\} \\
&= 0,
\end{aligned} \tag{D.17}$$

where the last line follows from (D.15). Finally, (D.13) is true because

$$\begin{aligned}(8\pi\tilde{G})^{-1}G_{\mu\nu} &= T_{\mu\nu} - \frac{1}{4}T^{\alpha}{}_{\alpha}g_{\mu\nu} + \tilde{T}_{\mu\nu} \\ &= \sum_i (1+w_i)\rho_i u_\mu u_\nu + p_i g_{\mu\nu} - \frac{1}{4}\sum_i (3w_i - 1)\rho_i g_{\mu\nu} \\ &\quad + \frac{1}{4}\sum_i (1+w_i)(3w_i-1)\rho_i(u_\mu u_\nu + g_{\mu\nu}) \\ &= \frac{3}{4}\sum_i (1+w_i)\left\{(1+w_i)\rho_i + w_i\rho_i g_{\mu\nu}\right\} \\ &= \frac{3}{4}\sum_i (1+w_i)T^{(i)}_{\mu\nu}.\end{aligned} \tag{D.18}$$

□

D.2 Conformal Newtonian Gauge

In the Conformal Newtonian Gauge the metric reads (see e.g. [92])

$$ds^2 = a^2(\tau)\{-(1+2\psi)d\tau^2 + (1-2\phi)\delta_{ij}dx^i dx^j\}. \tag{D.19}$$

We define the perturbation in any geometric quantity S by $\delta S = S - \overline{S}$, where $\overline{S}$ is the unperturbed value of S. The unperturbed components of the metric are

$$\overline{g}_{00} = -a^2, \qquad \overline{g}_{i0} = 0, \qquad \overline{g}_{ij} = a^2\delta_{ij}. \tag{D.20}$$

The unperturbed non-zero components of the Christoffel symbols are

$$\overline{\Gamma}^0_{00} = \frac{\dot{a}}{a}, \qquad \overline{\Gamma}^0_{ij} = \overline{\Gamma}^i_{0j} = \frac{\dot{a}}{a}\delta_{ij}. \tag{D.21}$$

where $\dot{}$ denotes differentiation with respect to conformal time τ. The unperturbed components of the Ricci tensor take the form

$$\overline{R}_{00} = 3\left[(\dot{a}/a)^2 - \ddot{a}/a\right], \qquad \overline{R}_{0i} = 0, \qquad \overline{R}_{ij} = \left[(\dot{a}/a)^2 + \ddot{a}/a\right]\delta_{ij}, \qquad , \tag{D.22}$$

using which it can be shown that the unperturbed Ricci scalar is

$$\overline{R} = 6\frac{\ddot{a}}{a^3}. \tag{D.23}$$

Finally, the components of the unperturbed Einstein tensor $G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}$ are

$$\overline{G}_{00} = 3(\dot{a}/a)^2, \qquad \overline{G}_{0i} = 0, \qquad \overline{G}_{ij} = \left[(\dot{a}/a)^2 - 2\ddot{a}/a\right]\delta_{ij}. \tag{D.24}$$

We now list the important perturbed geometric quantities. The metric perturbations $h_{\mu\nu} = g_{\mu\nu} - \overline{g}_{\mu\nu}$ are [2]

$$h_{00} = -2a^2\psi, \qquad h_{0i} = 0, \qquad h_{ij} = -2a^2\phi\delta_{ij} \tag{D.25}$$

$$h^{00} = 2a^{-2}\psi, \qquad h^{0i} = 0, \qquad h^{ij} = 2a^{-2}\phi\delta_{ij}. \tag{D.26}$$

The Christoffel symbol perturbations $\delta\Gamma^{\alpha}_{\mu\nu}$ are

$$\delta\Gamma^0_{00} = \dot{\psi}, \qquad \delta\Gamma^0_{0i} = \partial_i\psi, \qquad \delta\Gamma^0_{ij} = -\left[2\frac{\dot{a}}{a}(\psi+\phi) + \dot{\phi}\right]\delta_{ij} \tag{D.27}$$

$$\delta\Gamma^i_{00} = \partial_i\psi, \qquad \delta\Gamma^i_{0j} = -\dot{\phi}\delta_{ij}, \qquad \delta\Gamma^i_{jk} = \partial_i\phi\delta_{jk} - \partial_j\phi\delta_{ik} - \partial_k\phi\delta_{ji}. \tag{D.28}$$

The Ricci tensor and Ricci scalar perturbations are

$$\delta R_{00} = 3\left[\ddot{\phi} + (\dot{a}/a)(\dot{\psi}+\dot{\phi})\right] + \nabla^2\psi \tag{D.29}$$

$$\delta R_{0i} = 2\partial_i\dot{\phi} + 2\frac{\dot{a}}{a}\partial_i\psi \tag{D.30}$$

$$\delta R_{ij} = \partial_i\partial_j(\phi-\psi) + \left\{(\nabla^2 - \partial_\tau^2)\phi - 2\left[\ddot{a}/a + (\dot{a}/a)^2\right](\psi+\phi) - \frac{\dot{a}}{a}(\dot{\psi}+5\dot{\phi})\right\}\delta_{ij} \tag{D.31}$$

$$a^2\delta R = -6\ddot{\phi} + 4\nabla^2\phi - 2\nabla^2\psi - 12(\ddot{a}/a)\psi - 6(\dot{a}/a)(\dot{\psi}+3\dot{\phi}). \tag{D.32}$$

Finally, the Einstein tensor perturbations $\delta G_{\mu\nu}$ are

$$\delta G_{00} = -6(\dot{a}/a)\dot{\phi} + 2\nabla^2\phi \tag{D.33}$$

$$\delta G_{0i} = 2\partial_i\dot{\phi} + 2(\dot{a}/a)\partial_i\psi \tag{D.34}$$

$$\delta G_{ij} = \partial_i\partial_j(\phi-\psi) + \left\{2\ddot{\phi} - \nabla^2\phi + \left[4\ddot{a}/a - 2(\dot{a}/a)^2\right](\phi+\psi) + (\dot{a}/a)(4\dot{\phi}+2\dot{\psi}) + \nabla^2\psi\right\}\delta_{ij} \tag{D.35}$$

[2] The metric perturbations with upper indices $h^{\mu\nu}$ are defined so that $g^{\mu\nu}g_{\nu\sigma} = (\overline{g}^{\mu\nu}+h^{\mu\nu})(\overline{g}_{\nu\sigma}+h_{\nu\sigma}) = \delta^{\mu}{}_{\sigma}$, which implies $h^{\mu\rho} = -\overline{g}^{\mu\nu}h_{\nu\sigma}\overline{g}^{\sigma\rho}$.

Also useful to know are $\delta G^{\mu}{}_{\nu}$:

$$\delta G^{0}{}_{0} = 2a^{-2}\left\{-\nabla^2\phi + 3(\dot{a}/a)\left[\dot{\phi} + (\dot{a}/a)\psi\right]\right\} \tag{D.36}$$

$$\delta G^{0}{}_{i} = -2a^{-2}\partial_i\left[\dot{\phi} + (\dot{a}/a)\psi\right] = -\delta G^{i}{}_{0} \tag{D.37}$$

$$a^2\delta G^{i}{}_{j} = \partial_i\partial_j(\phi - \psi) + \left\{2\ddot{\phi} - \nabla^2\phi + \left[4\ddot{a}/a - 2(\dot{a}/a)^2\right]\psi + (\dot{a}/a)(4\dot{\phi} + 2\dot{\psi}) + \nabla^2\psi\right\}\delta_{ij}. \tag{D.38}$$

D.3 Aether Perturbations Through Equality

Here we derive Equations (5.48)–(5.50). In what follows, we will only need to use the unperturbed metric and Christoffel symbols, all of which are listed in Section D.2 of this Appendix.

Equation (D.9) can be written as

$$\nabla_\mu(\tilde{p}\tilde{u}^\mu) - \tilde{u}^\mu\nabla_\mu(\tilde{p} - T/4) = 0. \tag{D.39}$$

It follows from (5.47) and (5.38) that $\tilde{p} - T/4 = \delta\tilde{p}$. Then, perturbing (D.39) to first order, we find

$$\begin{aligned}
0 &= g^{\mu\nu}\nabla_\mu\left[\left(\frac{-\rho_b - \rho_{dm}}{4}\right)\left(u^{dm}_\nu + \delta\tilde{u}_\nu\right)\right] - \overline{u}^\mu\partial_\mu\delta\tilde{p} \\
&= -\frac{1}{4}g^{\mu\nu}\nabla_\mu(\rho_b u^{dm}_\nu) - \frac{1}{4}\overline{g}^{\mu\nu}\overline{\nabla}_\mu(\overline{\rho}_m\delta\tilde{u}_\nu) + \overline{g}^{\mu\nu}\overline{\nabla}_\mu(\delta\tilde{p}\overline{u}_\nu) - \overline{u}^\mu\partial_\mu\delta\tilde{p} \\
&= -\frac{1}{4}g^{\mu\nu}\nabla_\mu(\rho_b(u^{dm}_\nu - u^{b}_\nu)) - \frac{1}{4}\overline{g}^{\mu\nu}\overline{\nabla}_\mu(\overline{\rho}_m\delta\tilde{u}_\nu) + \overline{g}^{\mu\nu}\overline{\nabla}_\mu(\delta\tilde{p}\overline{u}_\nu) - \overline{u}^\mu\partial_\mu\delta\tilde{p} \\
&= -\frac{1}{4}\overline{g}^{\mu\nu}\overline{\nabla}_\mu(\overline{\rho}_b(\delta u^{dm}_\nu - \delta u^{b}_\nu)) - \frac{1}{4}\overline{g}^{\mu\nu}\overline{\nabla}_\mu(\overline{\rho}_m\delta\tilde{u}_\nu) + \overline{g}^{\mu\nu}\overline{\nabla}_\mu(\delta\tilde{p}\overline{u}_\nu) - \overline{u}^\mu\partial_\mu\delta\tilde{p},
\end{aligned} \tag{D.40}$$

where $\overline{\nabla}_\mu$ denotes the covariant derivative with respect to the unperturbed metric $\overline{g}_{\mu\nu}$[3], and in arriving at the second and third equalities we have used (5.42) and (5.44), respectively. If we let

$$\delta w_\nu = \delta u^{dm}_\nu - \delta u^{b}_\nu, \tag{D.41}$$

then to first order in perturbation theory (D.39) becomes

$$\delta\tilde{p}\overline{\nabla}^\mu\overline{u}_\mu = \frac{1}{4}\overline{g}^{\mu\nu}\overline{\nabla}_\mu(\overline{\rho}_b\delta w_\nu + \overline{\rho}_m\delta\tilde{u}_\nu). \tag{D.42}$$

[3] Concretely: $\overline{\nabla}_\mu V_\nu = \partial_\mu V_\nu - \overline{\Gamma}^{\alpha}_{\mu\nu}V_\alpha$.

Let us now turn to (D.10). We start by computing $\tilde{u}^\mu \nabla_\mu \tilde{u}_\nu$ to first order in perturbation theory:

$$\begin{aligned}\tilde{u}^\mu \nabla_\mu \tilde{u}_\nu &= g^{\alpha\beta}\tilde{u}_\alpha \nabla_\beta \tilde{u}_\nu \\ &= g^{\alpha\beta}(u^{dm}_\alpha + \delta\tilde{u}_\alpha)\nabla_\beta(u^{dm}_\nu + \delta\tilde{u}_\nu) \\ &= g^{\alpha\beta}u^{dm}_\alpha \nabla_\beta u^{dm}_\nu + \overline{g}^{\alpha\beta}\overline{u}_\alpha \overline{\nabla}_\beta \delta\tilde{u}_\nu + \overline{g}^{\alpha\beta}\delta\tilde{u}_\alpha \overline{\nabla}_\beta \overline{u}_\nu \\ &= \overline{g}^{\alpha\beta}\overline{u}_\alpha \overline{\nabla}_\beta \delta\tilde{u}_\nu + \overline{g}^{\alpha\beta}\delta\tilde{u}_\alpha \overline{\nabla}_\beta \overline{u}_\nu,\end{aligned} \tag{D.43}$$

where in arriving at the third equality we have used (5.43). Substituting this in (D.10) we find

$$\overline{\nabla}^\perp_\nu \delta\tilde{p} = \frac{\overline{\rho}_m}{4}\left(\overline{g}^{\alpha\beta}\overline{u}_\alpha \overline{\nabla}_\beta \delta\tilde{u}_\nu + \overline{g}^{\alpha\beta}\delta\tilde{u}_\alpha \overline{\nabla}_\beta \overline{u}_\nu\right), \tag{D.44}$$

We will now show that $\delta\tilde{u}_0 = \delta w_0 = 0$. To show $\delta\tilde{u}_0 = 0$, note that

$$\begin{aligned}-1 &= g^{\mu\nu}\tilde{u}_\mu \tilde{u}_\nu \\ &= g^{\mu\nu}(u^{dm}_\mu + \delta\tilde{u}_\mu)(u^{dm}_\nu + \delta\tilde{u}_\nu) \\ &= -1 + 2\overline{g}^{\mu\nu}\overline{u}_\mu \delta\tilde{u}_\nu,\end{aligned} \tag{D.45}$$

which implies $\overline{g}^{\mu\nu}\overline{u}_\mu \delta\tilde{u}_\nu = a^{-1}\delta\tilde{u}_0 = 0$, or $\delta\tilde{u}_0 = 0$. Also, to show $\delta w_0 = \delta u^{dm}_0 - \delta u^b_0 = 0$, note that

$$\begin{aligned}-1 &= g^{\mu\nu}u^{dm}_\mu u^{dm}_\nu \\ &= (\overline{g}^{\mu\nu} + h^{\mu\nu})(\overline{u}_\mu + \delta u^{dm}_\mu)(\overline{u}_\nu + \delta u^{dm}_\nu) \\ &= -1 + 2\overline{g}^{\mu\nu}\overline{u}_\mu \delta u^{dm}_\nu + h^{\mu\nu}\overline{u}_\mu \overline{u}_\nu,\end{aligned} \tag{D.46}$$

which implies

$$2a^{-1}\delta u^{dm}_0 + a^2 h^{00} = 0. \tag{D.47}$$

Similarly, using $g^{\mu\nu}u^b_\mu u^b_\nu = -1$ one finds

$$2a^{-1}\delta u^b_0 + a^2 h^{00} = 0. \tag{D.48}$$

By subtracting the two equations (D.47) and (D.48) we arrive at the desired result:

$$\delta w_0 = \delta u^{dm}_0 - \delta u^b_0 = 0. \tag{D.49}$$

It can now be checked that (D.42) is equivalent to

$$3\dot{a}\delta\tilde{p} = \frac{1}{4}\delta^{ij}(\overline{\rho}_b \partial_i \delta w_j + \overline{\rho}_m \partial_i \delta\tilde{u}_j). \tag{D.50}$$

Moreover, the i^{th} component of (D.44) takes the form

$$\partial_i \delta\tilde{p} = \frac{\overline{\rho}_m}{4} a^{-1} \delta\dot{\tilde{u}}_i. \tag{D.51}$$

Equations (5.48) and (5.49) then follow from (D.50) and (D.51), respectively, by letting

$$\delta\tilde{u}^i = a^{-2}\delta\tilde{u}_i, \qquad \delta w^i = a^{-2}\delta w_i. \tag{D.52}$$

Let us now turn to deriving (5.50). Applying $\delta^{ij}\partial_j$ to (5.49) and the Laplacian $\nabla^2 = \delta^{ij}\partial_i\partial_j$ to (5.48), $\delta\tilde{p}$ can be eliminated to arrive at

$$3\frac{\dot{a}}{a^3}\partial_\tau(a^2\Omega) - \nabla^2\Omega = \frac{\bar{\rho}_b}{\bar{\rho}_m}\nabla^2(\partial_i \delta w^i). \tag{D.53}$$

It then remains to show that $\partial_i \delta w^i = a^{-1}(\dot{\delta}_b - \dot{\delta}_{dm})$. To do so, note that

$$\begin{aligned} g^{\mu\nu}\nabla_\nu(u^{dm}_\mu - u^b_\mu) &= \overline{g}^{\mu\nu}\overline{\nabla}_\nu \delta w_\mu \\ &= a^{-2}\delta^{ij}\partial_i \delta w_j \\ &= \partial_i \delta w^i. \end{aligned} \tag{D.54}$$

It follows from (5.42) that

$$\begin{aligned} g^{\mu\nu}\nabla_\mu u^{dm}_\nu &= -\frac{g^{\mu\nu}u^{dm}_\mu \partial_\nu \rho_{dm}}{\rho_{dm}} \\ &= -\frac{(\overline{g}^{\mu\nu} + h^{\mu\nu})(\overline{u}_\mu + \delta u^{dm}_\mu)(\partial_\nu \overline{\rho}_{dm} + \partial_\nu \delta\rho_{dm})}{\overline{\rho}_{dm}(1 + \delta\rho_{dm}/\overline{\rho}_{dm})} \\ &= \frac{\dot{\overline{\rho}}_{dm}}{\overline{\rho}_{dm}}(ah^{00} - a^{-1} + a^{-2}\delta u^{dm}_0) - a^{-1}\frac{d}{d\tau}\left(\frac{\delta\rho_{dm}}{\overline{\rho}_{dm}}\right) \\ &= -3\frac{\dot{a}}{a}(ah^{00} - a^{-1} + a^{-2}\delta u^{dm}_0) - a^{-1}\dot{\delta}_{dm}, \end{aligned} \tag{D.55}$$

where in arriving at the last line we have used the homogeneous conservation equation $\dot{\overline{\rho}}_{dm} + 3\frac{\dot{a}}{a}\overline{\rho}_{dm} = 0$ and defined

$$\delta_{dm} = \frac{\delta\rho_{dm}}{\overline{\rho}_{dm}}. \tag{D.56}$$

The same derivation for baryons leads to

$$g^{\mu\nu}\nabla_\mu u^b_\nu = -3\frac{\dot{a}}{a}(ah^{00} - a^{-1} + a^{-2}\delta u^b_0) - a^{-1}\dot{\delta}_b, \tag{D.57}$$

where

$$\delta_b = \frac{\delta\rho_b}{\overline{\rho}_b}. \tag{D.58}$$

We then have that

$$\begin{aligned}\partial_i \delta w^i &= g^{\mu\nu}\nabla_\nu(u^{dm}_\mu - u^b_\mu) \\ &= -3\frac{\dot{a}}{a^3}(\delta u^{dm}_0 - \delta u^b_0) - a^{-1}(\dot{\delta}_{dm} - \dot{\delta}_b) \\ &= a^{-1}(\dot{\delta}_b - \dot{\delta}_{dm}),\end{aligned} \tag{D.59}$$

where in arriving at the last equality we have used (D.49).

D.4 PPN notations

The metric components are in terms of particular potential functions, thus defining the PPN parameters:

$$\begin{aligned}g_{00} &= -1 + 2U - 2\beta U^2 - 2\xi\phi_W + (2\gamma + 2 + \alpha_3 + \zeta_1 - 2\xi)\phi_1 \\ &\quad + 2(3\gamma - 2\beta + 1 + \zeta_2 + \xi)\phi_2 + 2(1 + \zeta_3)\phi_3 \\ &\quad + 2(3\gamma + 3\zeta_4 - 2\xi)\phi_4 - (\zeta_1 - 2\xi)A \qquad \text{(D.60)} \\ g_{ij} &= (1 + 2\gamma U)\delta_{ij} \qquad \text{(D.61)} \\ g_{0i} &= -\frac{1}{2}(4\gamma + 3 + \alpha_1 - \alpha_2 + \zeta_1 - 2\xi)V_i \\ &\quad - \frac{1}{2}(1 + \alpha_2 - \zeta_1 + 2\xi)W_i \qquad \text{(D.62)}\end{aligned}$$

The potentials are all of the form

$$F(x) = G_N \int \mathrm{d}^3 y \, \frac{\rho(y) f}{|x - y|} \tag{D.63}$$

and the correspondences $F : f$ are given by

$$\begin{aligned}&U : 1 \qquad \phi_1 : u_i u_j \qquad \phi_2 : U \qquad \phi_3 : \Pi \qquad \phi_4 : p/\rho \\ &\phi_W : \int \mathrm{d}^3 z \, \rho(z) \frac{(x-y)_j}{|x-y|^2}\left[\frac{(y-z)_j}{|x-z|} - \frac{(x-z)_j}{|y-z|}\right] \\ &A : \frac{(v_i(x-y)_i)^2}{|x-y|^2} \\ &V_i : u^i \qquad W_i : \frac{u_j(x_j - y_j)(x^i - y^i)}{|x-y|^2}.\end{aligned} \tag{D.64}$$

Appendix E

Supplementary material for Chapter 6

E.1 Technical Details

E.1.1 Likelihood Function

Given an underlying temperature field[1]

$$\delta T(\widehat{\boldsymbol{n}}) = \sum_{\ell=0}^{\ell_{max}} \sum_{m=-\ell}^{\ell} \delta T_{\ell m} Y_{\ell m}(\widehat{\boldsymbol{n}}), \tag{E.1}$$

the value of the discretized temperature map at pixel i and frequency band ν is given by

$$\delta T_{i\nu} = \sum_{\ell=0}^{\ell_{max}} \sum_{m=-\ell}^{\ell} \delta T_{\ell m} B_{\ell\nu} W_{\ell} Y_{\ell m}(\widehat{\boldsymbol{n}}_i), \tag{E.2}$$

where $B_{\ell\nu}$ is the isotropicized beam transfer function for the mode ℓ and frequency channel ν, and W_ℓ is the isotropicized pixel transfer function. In case of the primordial anisotropies, because $\delta T^{\rm PA}_{i\nu}$ are linear functionals of $\delta T^{\rm PA}(\widehat{\boldsymbol{n}})$, they are correlated Gaussian random variables with zero mean and two-point function

$$[\mathbf{C_S}]_{i\nu,i'\nu'} \equiv \langle \delta T^{\rm PA}_{i\nu} \delta T^{\rm PA}_{i'\nu'} \rangle = \sum_{\ell=0}^{\ell_{max}} \left(\frac{2\ell+1}{4\pi} \right) C_\ell B_{\ell\nu} B_{\ell\nu'} W_\ell^2 P_\ell(\widehat{\boldsymbol{n}}_i \cdot \widehat{\boldsymbol{n}}_{i'}), \tag{E.3}$$

[1] We follow the HEALPix conventions for spherical harmonic transforms [70].

where P_ℓ is the ℓ-th degree Legendre polynomial and we have used $\langle \delta T^{\rm PA}_{lm} \delta T^{\rm PA}_{l'm'} \rangle = C_l \delta_{ll'} \delta_{mm'}$.

Finally, the log-likelihood probability $-\chi^2/2$ of jointly measuring the CMB temperature values $\{\delta T_{i\nu}\}^{\nu\in L_\nu}_{i\in L_p}$ given the tSZ contribution $\{\delta T^{\rm SZ}_{i\nu}\}^{\nu\in L_\nu}_{i\in L_p}$ and the primary CMB fluctuations $\{\delta T^{\rm PA}_{i\nu}\}^{\nu\in L_\nu}_{i\in L_p}$ is

$$
\begin{aligned}
&-\frac{1}{2}\chi^2(\{\delta T_{i\nu}\}^{\nu\in L_\nu}_{i\in L_p}|\{\delta T^{\rm SZ}_{i\nu}\}^{\nu\in L_\nu}_{i\in L_p},\{\delta T^{\rm PA}_{i\nu}\}^{\nu\in L_\nu}_{i\in L_p}) = \\
&\quad -\frac{1}{2}\boldsymbol{\delta T}^{\rm PA\,T}\mathbf{C}_S{}^{-1}\boldsymbol{\delta T}^{\rm PA} - \frac{1}{2}(\boldsymbol{\delta T}-\boldsymbol{\delta T}^{PA}-\boldsymbol{\delta T}^{\rm SZ})^T\mathbf{C}_N^{-1}(\boldsymbol{\delta T}-\boldsymbol{\delta T}^{\boldsymbol{PA}}-\boldsymbol{\delta T}^{\rm SZ}),
\end{aligned}
\tag{E.4}
$$

where $\mathbf{C_N}$ is the noise covariance matrix and L_p (L_ν) is the set containing pixels (frequency channels) we wish to use in our analysis.[2] After integrating over all possible primary fluctuations $\boldsymbol{\delta T}^{\rm PA}$, which can be done analytically, the log-likelihood takes the form

$$
-\frac{1}{2}\chi^2(\{\delta T_{i\nu}\}^{\nu\in L_\nu}_{i\in L_p}|\{\delta T^{\rm SZ}_{i\nu}\}^{\nu\in L_\nu}_{i\in L_p}) = -\frac{1}{2}(\boldsymbol{\delta T}-\boldsymbol{\delta T}^{\rm SZ})^T\mathbf{C}^{-1}(\boldsymbol{\delta T}-\boldsymbol{\delta T}^{\rm SZ}),
\tag{E.6}
$$

with

$$
\mathbf{C} = \mathbf{C}_S + \mathbf{C}_N.
\tag{E.7}
$$

This proves Equation (6.10).

E.1.2 Masking

All quantities used (and not defined) here are introduced in Sections 6.2.3 and 6.2.4.

Let $L_{\bar{p}}$ (L_p respectively) denotes the set of all masked (unmasked respectively) pixels, so that $L = L_{\bar{p}} \cup L_p$ contains all pixels on the sky. In what will follow, $\mathbf{C_{S_f}}$ ($\mathbf{C_S}$) will denote the signal covariance matrix defined on L (L_p). The same notation will be used for the noise covariance matrix. Consider now the full covariance matrix $\mathbf{C_f} = \mathbf{C_{S_f}} + \mathbf{C_{N_f}}$. It is related to $\mathbf{C} = \mathbf{C_S} + \mathbf{C_N}$ via $\mathbf{C} = \mathbf{C_P}^T\mathbf{C_f}\mathbf{C_P}$, where $[\mathbf{C_P}]_{i\nu,j\nu'} = P_{ij}\delta_{\nu\nu'}$ is a projection matrix with $i \in L$ and $j \in L_p$, such that all components of $\mathbf{P}$ are zero except $P_{ii} = 1$ for $i \in L_p$. Construct from $\mathbf{C_f}$ another matrix $\tilde{\mathbf{C}}_\mathbf{f}$ with the same entries, except that

[2] The matrix notation used here is explicitly:

$$
\delta T^{\rm PA\,T}\mathbf{C_S}{}^{-1}\delta T^{\rm PA} = \sum_{\nu,\nu'\in L_\nu}\sum_{i,i'\in L_p}[\mathbf{C_S}{}^{-1}]_{i\nu,i'\nu'}\delta T^{\rm PA}_{i\nu}\delta T^{\rm PA}_{i'\nu'},
\tag{E.5}
$$

where we are considering $\mathbf{C_S}$ as a matrix whose rows and columns are labeled by $i\nu$.

$[\tilde{\mathbf{C}}_\mathbf{f}]_{i\nu,i\nu} = x$ for all $i \in L_{\bar{p}}$. Then, it can be shown from the definition of the inverse of a matrix that [3]

$$\mathbf{C}^{-1} = \lim_{x\to\infty} \mathbf{C_P}^T \tilde{\mathbf{C}}_\mathbf{f}^{-1} \mathbf{C_P}. \tag{E.9}$$

Because $[\mathbf{C}_{\mathbf{N_f}}]_{i\nu,i'\nu'} = [\tilde{\mathbf{N}}_\nu]_{ii'}\delta_{\nu\nu'}$, we may construct $\tilde{\mathbf{C}}_\mathbf{f}$ as follows: $[\tilde{\mathbf{C}}_\mathbf{f}]_{i\nu,i'\nu'} = [\tilde{\mathbf{N}}_\nu]_{ii'}\delta_{\nu\nu'} + \mathbf{C}_{\mathbf{S_f}}$, where $\tilde{\mathbf{N}}_\nu$ has the same elements as $\mathbf{N}_\nu$ except $[\tilde{\mathbf{N}}_\nu]_{ii} = x$ for all $i \in L_{\bar{p}}$. Since $[\mathbf{N}_\nu]_{ii'} = n^2_{i\nu}\delta_{ii'}$ is diagonal, it follows that $[\tilde{\mathbf{N}}_\nu^{-1}]_{ii'} = 1/[\tilde{\mathbf{N}}_\nu]_{ii'}$, which then implies

$$\lim_{x\to\infty} \tilde{\mathbf{N}}_\nu^{-1} = \mathbf{M}\mathbf{N}_\nu^{-1}\mathbf{M}, \tag{E.10}$$

where $\mathbf{M} = \mathbf{P}\mathbf{P}^T$ is the masking matrix defined in Section 6.2.4. Moreover, we show in Section E.1.3 of this Appendix that [4]

$$[\tilde{\mathbf{C}}_\mathbf{f}^{-1}]_{i\nu,i'\nu'} = \left[\tilde{\mathbf{N}}_\nu^{-1}\delta_{\nu,\nu'} - \tilde{\mathbf{N}}_\nu^{-1}\mathbf{B}_\nu \mathbf{S}^{1/2}\left\{1 + \mathbf{S}^{1/2}\left(\sum_{\mu\in L_\nu} \mathbf{B}_\mu \tilde{\mathbf{N}}_\mu^{-1}\mathbf{B}_\mu\right)\mathbf{S}^{1/2}\right\}^{-1}\mathbf{S}^{1/2}\mathbf{B}_{\nu'}\tilde{\mathbf{N}}_{\nu'}^{-1}\right]_{ii'}. \tag{E.11}$$

Using the above two equations we find

$$\lim_{x\to\infty} [\tilde{\mathbf{C}}_\mathbf{f}^{-1}]_{i\nu,i'\nu'} = [\mathbf{G}_{\nu,\nu'}]_{ii'}, \tag{E.12}$$

where $\mathbf{G}_{\nu,\nu'}$ is given by (6.23b).

Finally, let $[\boldsymbol{V_f}]_{i\nu}$ and $[\boldsymbol{W_f}]_{i\nu}$ be two vectors defined on every pixel on the sky (i.e. $i \in L$). Also, let V and W be the corresponding vectors defined only on the unmasked

[3] As a simple example, consider the case where there are only two pixels, one frequency channel, and one of the pixels is masked out: $L_p = \{1\}$, $L_{\bar{p}} = \{2\}$. In this case $\mathbf{C_P} = \mathbf{P} = \begin{bmatrix}1\\0\end{bmatrix}$, $\mathbf{C} = [\mathbf{C_f}]_{11}$, $\tilde{\mathbf{C}}_\mathbf{f} = \begin{bmatrix}[\mathbf{C_f}]_{11} & [\mathbf{C_f}]_{12}\\ [\mathbf{C_f}]_{21} & x\end{bmatrix}$ and $\tilde{\mathbf{C}}_\mathbf{f}^{-1} = \frac{1}{x[\mathbf{C_f}]_{11} - [\mathbf{C_f}]_{12}^2}\begin{bmatrix}x & -[\mathbf{C_f}]_{12}\\ -[\mathbf{C_f}]_{12} & [\mathbf{C_f}]_{11}\end{bmatrix}$. Equality (E.9) can now be easily verified:

$$\lim_{x\to\infty} \mathbf{C_P}^T \tilde{\mathbf{C}}_\mathbf{f}^{-1} \mathbf{C_P} = \lim_{x\to\infty} \frac{x}{x[\mathbf{C_f}]_{11} - [\mathbf{C_f}]_{12}^2} = \frac{1}{[\mathbf{C_f}]_{11}} = \mathbf{C}^{-1}. \tag{E.8}$$

[4] Note that the proof presented in Appendix E.1.3 proceeds in the exact same fashion when $\mathbf{N}_\nu$ is replaced by $\tilde{\mathbf{N}}_\nu$.

pixels: $\boldsymbol{V} = {\mathbf{C_P}}^T \boldsymbol{V_f}$, $\boldsymbol{W} = {\mathbf{C_P}}^T \boldsymbol{W_f}$. Then

$$\boldsymbol{V}^T \mathbf{C}^{-1} \boldsymbol{W} = \lim_{x \to \infty} {\boldsymbol{V_f}}^T \mathbf{C_P} {\mathbf{C_P}}^T \tilde{\mathbf{C}}_\mathbf{f}^{-1} \mathbf{C_P} {\mathbf{C_P}}^T \boldsymbol{W_f} \tag{E.13}$$

$$= \sum_{i,i' \in L} \sum_{\nu,\nu' \in L_\nu} [\mathbf{M} \mathbf{G}_{\nu,\nu'} \mathbf{M}]_{ii'} [\boldsymbol{V_f}]_{i\nu} [\boldsymbol{W_f}]_{i'\nu'} \tag{E.14}$$

$$= \sum_{i,i' \in L} \sum_{\nu,\nu' \in L_\nu} [\mathbf{G}_{\nu,\nu'}]_{ii'} [\boldsymbol{V_f}]_{i\nu} [\boldsymbol{W_f}]_{i'\nu'}, \tag{E.15}$$

where we have used the fact that $[\mathbf{C_P} {\mathbf{C_P}}^T]_{i\nu,i'\nu'} = M_{ii'} \delta_{\nu\nu'}$, as well as $\mathbf{M}^2 = \mathbf{M}$, which combined with (E.12) implies $\mathbf{M} \mathbf{G}_{\nu,\nu'} \mathbf{M} = \mathbf{G}_{\nu,\nu'}$. This justifies the equality between the corresponding equations in (6.15) and (6.22).

E.1.3 Covariance Matrix Re-loaded

All quantities used (and not defined) here are introduced in Sections 6.2.3 and 6.2.4. Also, all matrices are defined on the entire sky.

Using the definition of matrices $\mathbf{B}_\nu$ and $\mathbf{S}$ as given in Section 6.2.4, it may be checked that

$$[\mathbf{C_S}]_{i\nu,i'\nu'} \simeq [\mathbf{B}_\nu \mathbf{S} \mathbf{B}_{\nu'}]_{ij}\,, \tag{E.16}$$

where we have used

$$A_{pix} \sum_{i \in L} Y_{lm}(\widehat{\boldsymbol{n}}_i) \bar{Y}_{l'm'}(\widehat{\boldsymbol{n}}_i) \simeq \int Y_{lm}(\widehat{\boldsymbol{n}}) \bar{Y}_{l'm'}(\widehat{\boldsymbol{n}}) d^2 n = \delta_{ll'} \delta_{mm'}. \tag{E.17}$$

It then follows that

$$[\mathbf{C}]_{i\nu,i'\nu'} = [\mathbf{B}_\nu \mathbf{S} \mathbf{B}_{\nu'}]_{ii'} + [\mathbf{N}_\nu]_{ii'}\, \delta_{\nu\nu'}. \tag{E.18}$$

Let us now prove the following

$$[\mathbf{C}^{-1}]_{i\nu,i'\nu} = [\mathbf{G}_{\nu,\nu'}]_{ii'}, \tag{E.19}$$

where

$$\mathbf{G}_{\nu,\nu'} = \mathbf{N}_\nu^{-1} \delta_{\nu,\nu'} - \mathbf{C_P} \mathbf{N}_\nu^{-1} \mathbf{B}_\nu \mathbf{S}^{1/2} \mathbf{D}^{-1} \mathbf{S}^{1/2} \mathbf{B}_{\nu'} \mathbf{N}_{\nu'}^{-1}, \tag{E.20}$$

$$\mathbf{D} = 1 + \mathbf{S}^{1/2} \left(\sum_{\nu \in L_\nu} \mathbf{B}_\nu \mathbf{N}_\nu^{-1} \mathbf{B}_\nu \right) \mathbf{S}^{1/2}. \tag{E.21}$$

That this is true may be easily checked:

$$\begin{aligned}
\sum_{j\in L}\sum_{\mu\in L_\nu}[\mathbf{C}]_{i\nu,j\mu}[\mathbf{G}_{\mu,\rho}]_{jk} &= \sum_{j\in L}\sum_{\mu\in L_\nu}[\mathbf{B}_\nu\mathbf{S}\mathbf{B}_\nu]_{ij}[\mathbf{N}_\mu^{-1}]_{jk}\delta_{\mu,\rho} + [\mathbf{N}_\nu]_{ij}[\mathbf{N}_\mu^{-1}]_{jk}\delta_{\mu,\rho}\delta_{\nu\mu} \\
&\quad - \sum_{j\in L}\sum_{\mu\in L_\nu}[\mathbf{N}_\nu]_{ij}[\mathbf{N}_\mu^{-1}\mathbf{B}_\nu\mathbf{S}^{1/2}\mathbf{D}^{-1}\mathbf{S}^{1/2}\mathbf{B}_\nu\mathbf{N}_\rho^{-1}]_{jk}\delta_{\nu\mu} \\
&\quad - \sum_{j\in L}\sum_{\mu\in L_\nu}[\mathbf{B}_\nu\mathbf{S}\mathbf{B}_\nu]_{ij}[\mathbf{N}_\mu^{-1}\mathbf{B}_\nu\mathbf{S}^{1/2}\mathbf{D}^{-1}\mathbf{S}^{1/2}\mathbf{B}_\nu\mathbf{N}_\rho^{-1}]_{jk} \\
&= [\mathbf{B}_\nu\mathbf{S}\mathbf{B}_\nu\mathbf{N}_\rho^{-1}]_{ik} + [\mathbf{N}_\nu\mathbf{N}_\rho^{-1}]_{ik}\delta_{\nu\rho} - [\mathbf{B}_\nu\mathbf{S}^{1/2}\mathbf{D}^{-1}\mathbf{S}^{1/2}\mathbf{C}_\mathbf{f}\mathbf{N}_\rho^{-1}]_{ik} \\
&\quad - \left[\mathbf{B}_\nu\mathbf{S}^{1/2}\left\{1+\mathbf{S}^{1/2}\left(\sum_{\mu\in L_\nu}\mathbf{B}_\nu\mathbf{N}_\rho^{-1}\mathbf{B}_\nu\right)\mathbf{S}^{1/2}\right\}\mathbf{D}^{-1}\mathbf{S}^{1/2}\mathbf{B}_\nu\mathbf{N}_\rho^{-1}\right]_{ik} \\
&\quad + [\mathbf{B}_\nu\mathbf{S}^{1/2}\mathbf{D}^{-1}\mathbf{S}^{1/2}\mathbf{B}_\nu\mathbf{N}_\rho^{-1}]_{ik} \\
&= [\mathbf{B}_\nu\mathbf{S}\mathbf{B}_\nu\mathbf{N}_\rho^{-1}]_{ik} + \delta_{ik}\delta_{\nu\rho} - [\mathbf{B}_\nu\mathbf{S}^{1/2}\mathbf{D}\mathbf{D}^{-1}\mathbf{S}^{1/2}\mathbf{B}_\nu\mathbf{N}_\rho^{-1}]_{ik} \\
&= \delta_{ik}\delta_{\nu\rho}, \qquad \text{(E.22)}
\end{aligned}$$

where we have used $\mathbf{S^{1/2}S^{1/2}} = \mathbf{S}$, which may be checked using (E.17).

E.1.4 Fitting the Monopole and the Dipole

Let us briefly discuss how any possible residual monopole and dipole CMB components can be accounted for in our framework. We denote monopole and dipole contributions by $\delta T^{\ell=0}$ and $\delta T^{\ell=1}$, respectively. They take the form

$$\delta T^{\ell=0}(\widehat{n}) = a_{00}Y_{00}(\widehat{n}), \qquad \text{(E.23a)}$$

$$\delta T^{\ell=1}(\widehat{n}) = a_{10}Y_{10}(\widehat{n}) + 2\mathrm{Re}(a_{11})\mathrm{Re}(Y_{11}(\widehat{n})) - 2\mathrm{Im}(a_{11})\mathrm{Im}(Y_{11}(\widehat{n})). \qquad \text{(E.23b)}$$

These components should be added to the tSZ signal:

$$\delta T^{\mathrm{SZ}}(\widehat{\boldsymbol{n}},\nu) \to \delta T^{\mathrm{SZ}}(\widehat{\boldsymbol{n}},\nu) + \delta T^{\ell=0}(\widehat{n}) + \delta T^{\ell=1}(\widehat{n}). \qquad \text{(E.24)}$$

This may be conveniently done by making the following definitions:

$$\mathbb{P}_{N_b+1} = a_{00}, \qquad \mathbb{P}_{N_b+2} = a_{10}, \qquad \mathbb{P}_{N_b+3} = \mathrm{Re}(a_{11}), \qquad \mathbb{P}_{N_b+4} = \mathrm{Im}(a_{11}), \qquad \text{(E.25a)}$$

and

$$t^{(\nu)}_{N_b+1}(\widehat{n}) = Y_{00}(\widehat{n}), \qquad t^{(\nu)}_{N_b+2}(\widehat{n}) = Y_{10}(\widehat{n}), \qquad \text{(E.25b)}$$

$$t^{(\nu)}_{N_b+3}(\widehat{n}) = 2\mathrm{Re}(Y_{11}(\widehat{n})), \qquad t^{(\nu)}_{N_b+4}(\widehat{n}) = -2\mathrm{Im}(Y_{11}(\widehat{n})), \qquad \text{(E.25c)}$$

where N_b is the total number of radial bins. The statistical machinery developed in Section 6.2.3 now goes through exactly the same way, except that $N_b \to N_b+4$. Once the matrix $\boldsymbol{\alpha}$ is found (see (6.15)), which is now $(N_b+4)\times(N_b+4)$ dimensional, the pressure covariance matrix becomes the restriction of its inverse to the bins of physical interest: $[\mathbf{C}_{\mathbb{P}}]_{kk'} = [\boldsymbol{\alpha}^{-1}]_{kk'}$ where $k, k' \in \{1, \ldots, N_b\}$.

E.2 Robustness Tests

As it was shown in Section 6.2.4, the most important part of our analysis is solving a linear equation of the form $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$, where $\mathbf{A}$ is a very large ($\sim 10^6 \times 10^6$) matrix. In order to be certain that our numerical methods are correct, we perform two tests.

E.2.1 Simulations

We create sky maps with known tSZ amplitudes (i.e. the quantities of interest $\mathbb{P}_k$ and other parameters such as the monopole and dipole anisotropies) and see if the outcome of the pipeline matches with what is inputted. More specifically, we generate N random realizations of the CMB primary anisotropies, add the tSZ signal with known amplitudes $\widehat{\mathbb{P}}_k$, and finally add random detector noise. The outcome of every fitting procedure is the set of values $\mathbb{P}_k^{(i)}$, with $i \in \{1, \ldots, N\}$ denoting the i-th simulation, and the covariance matrix $[\mathbf{C}_{\mathbb{P}}]_{k,k'}$ (see Section 6.2.3). The covariance matrix $[\mathbf{C}_{\mathbb{P}}]_{k,k'}$ does not change from one simulation to the other since it only depends on the tSZ templates, and the detector noise and primary CMB covariance matrices. If $\mathbb{P}_k^{(i)}$ really are realizations of a gaussian random variable with mean $\widehat{\mathbb{P}}_k$ and variance $[\mathbf{C}_{\mathbb{P}}]_{k,k}$, then their mean $\overline{\mathbb{P}}_k = \frac{1}{N}\sum_{i=1}^{N} \mathbb{P}_k^{(i)}$ should converge to $\widehat{\mathbb{P}}_k$ as N becomes large. More specifically, the expected error in determining the true value of the mean is $\sqrt{< (\overline{\mathbb{P}}_k - \widehat{\mathbb{P}}_k)^2 >} = \sqrt{[\mathbf{C}_{\mathbb{P}}]_{k,k}/N}$. Similarly the best estimator of the variance $\sigma_k^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbb{P}_k^{(i)} - \overline{\mathbb{P}}_k)^2$ should converge to $[\mathbf{C}_{\mathbb{P}}]_{k,k'}$, with an expected error of $\sqrt{< (\sigma_k^2 - [\mathbf{C}_{\mathbb{P}}]_{k,k})^2 >} = \sqrt{\frac{2}{N-1}}[\mathbf{C}_{\mathbb{P}}]_{k,k}$. Figure E.1 shows the results of our simulations for a few templates (i.e. values of k). As it can be seen, all estimators converge to the values computed by our pipeline.

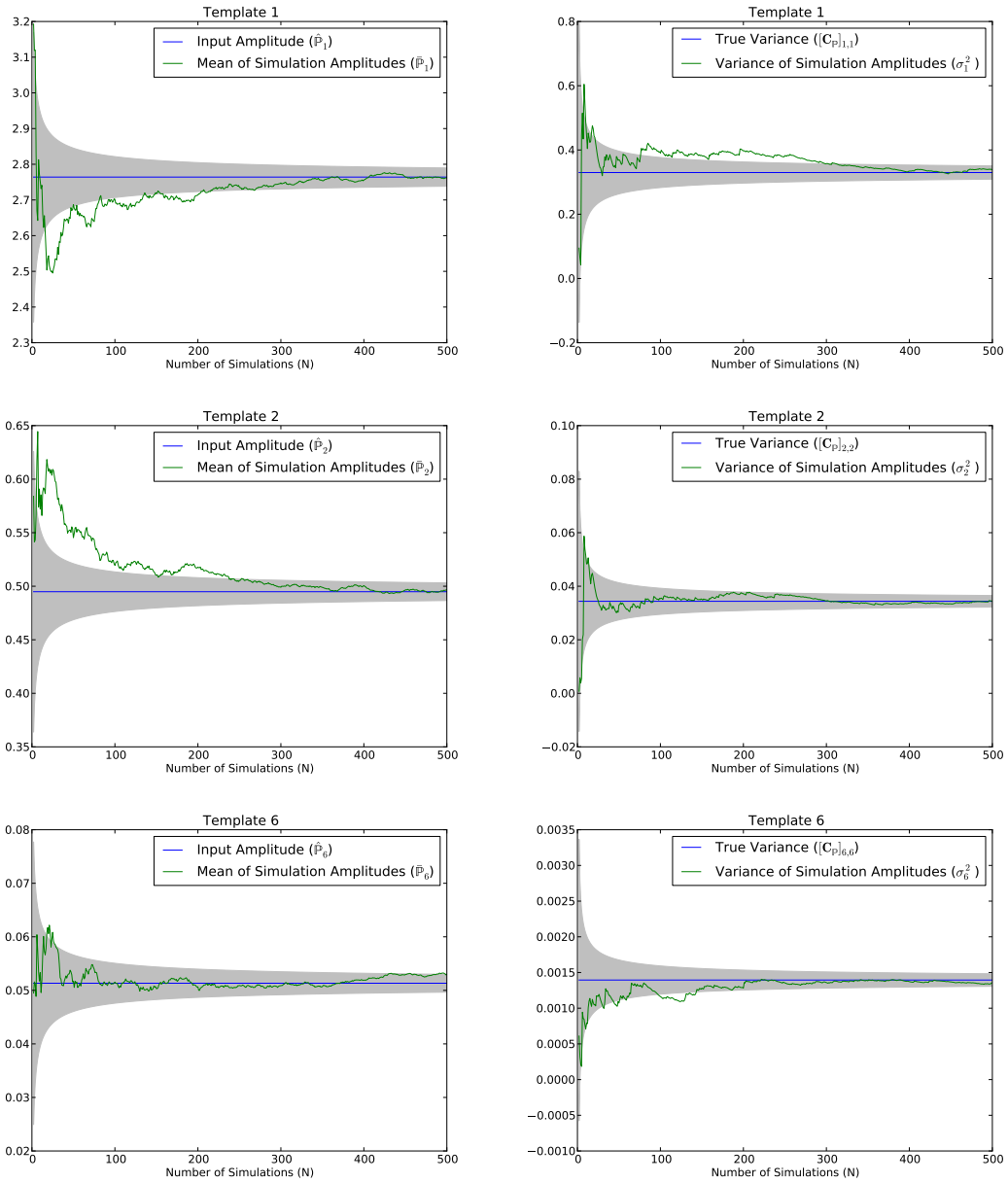


Figure E.1: Testing the pipeline with random primary CMB+noise simulations. The shaded area represents the expected error in the quantity of interest (see text for more details.)

E.2.2 Different Resolutions

We performed our analysis on all MCXC clusters using WMAP7 sky maps at two different HEALPix resolutions of 9 ($N_{side} = 512$) and 10 ($N_{side} = 1024$). The modified self-similar scaling of pressure with mass is used for these measurements (i.e. $\delta = 0.12$ in Equation (6.9)). Figure E.2 shows the results. The null chi-squared for the $N_{side} = 512$ ($N_{side} = 1024$) measurement is $\chi^2_0 = 246.8$ ($\chi^2_0 = 272.1$), corresponding to a $\mathbf{14.68}\sigma$ ($\mathbf{15.50}\sigma$) detection.

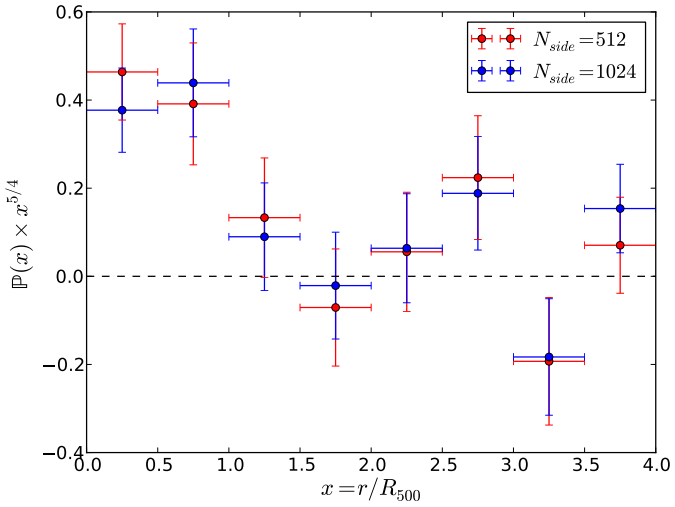


Figure E.2: Results of our analysis of WMAP7 sky maps at two different resolutions.

E.3 Effect of Uncertainty in Mass of Clusters

Our entire analysis depends crucially on the self-similarity length/mass scales of clusters. Therefore, it is important to investigate how our measurements are affected by the uncertainty present in determining masses of clusters. [121] use 62 clusters from the Early Release SZ (ESZ) sample [120] which also belong to the MCXC catalogue [122]. To get an idea for the degree of uncertainty present in mass estimates, we compare ESZ and MCXC masses of these clusters. This is shown in Figure E.3a. The ESZ mass estimates are systematically higher on average by about 12 percent. To investigate how such systematics

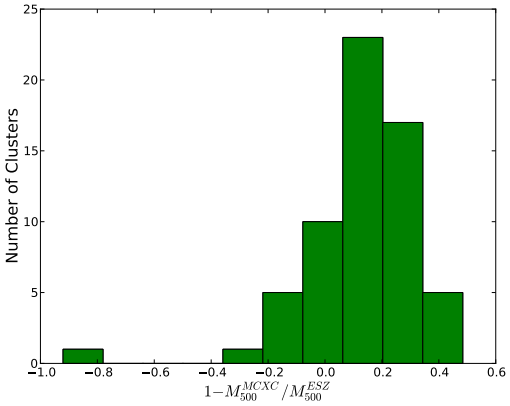
affect our pressure measurements, we randomly changed masses of all MCXC clusters according to the distribution in Figure E.3a. (We used the standard self-similar scaling to create our templates, i.e. we set $\delta = 0$ in Equation (6.9).) The resulting pressure profile is shown in Figure (E.3b). In the first three bins, where there is signal, the pressure values decrease systematically. This difference, however, is at most at the 1σ level. As shown in Figure E.3c, this is no longer the case for gas mass fraction, which decreases by about **20** percent on average.

E.4 All vs. Resolved MCXC clusters

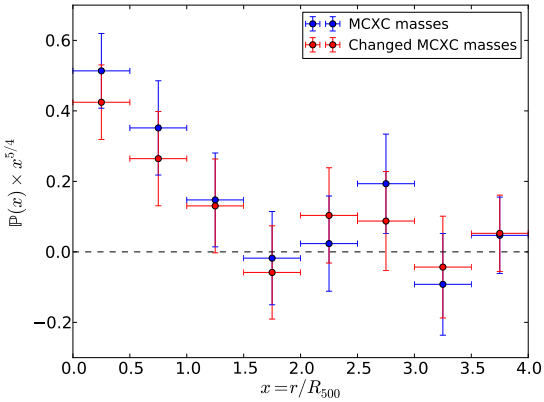
Let us provide a quantitative explanation for how statistical information is lost when unresolved clusters are not accounted for. The discussion that will follow is based on measurements presented in Figure 6.3b, i.e. the modified scaling of pressure with mass ($\delta = 0.12$ in Equation (6.9)). The same analysis for the standard self-similar scaling (i.e. measurements presented in Figure 6.3b) gives the same results. Let λ_n and $\boldsymbol{T_n}$ denote the eigenvalues and eigenvectors of $\mathbf{C}_{\mathbb{P}}$, respectively. We choose the labels n such that $\lambda_1 < \lambda_2 < \cdots < \lambda_{N_b}$, where $N_b = 8$ is the total number of radial bins. Since $\mathbf{C}_{\mathbb{P}}$ is a positive symmetric matrix, its eigenvalues are positive and its eigenvectors are real. The null chi-squared can be re-expressed as

$$\chi_0^2 = \widehat{\mathbb{P}}^T \mathbf{C}_{\mathbb{P}}^{-1} \widehat{\mathbb{P}} = \sum_{n=1}^{N_b} (\widehat{\mathbb{P}}^T \boldsymbol{T_n})^2 / \lambda_n. \tag{E.26}$$

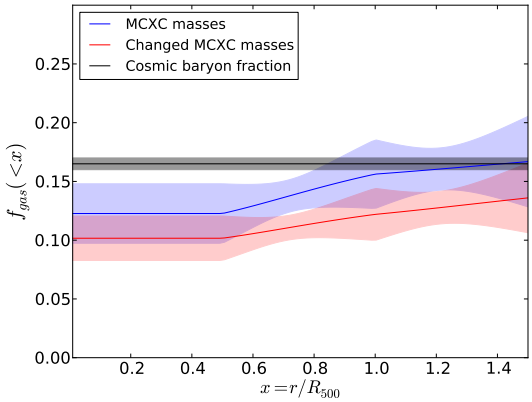
Figure E.4 shows the eigenvalues λ_n and the contribution $(\widehat{\mathbb{P}}^T \boldsymbol{T_n})^2/\lambda_n$ of the different eigenmodes to χ_0^2. In the case of the resolved clusters, the modes with the three largest eigenvalues are responsible for most of the contribution to χ_0^2. For the whole MCXC sample, however, all eigenmodes contribute more or less equally. Eigenvectors corresponding to larger eigenvalues carry most of their weight from the inner bins. To see this, we have plotted the components of all eigenvectors in Figure E.5. We denote the k^{th} component of the eigenvector $\boldsymbol{T_n}$ by $T_{n,k}$. As in the text, $k = 1 \ldots N_b$ labels the radial bins around clusters with $k = 1$ and $k = N_b$ corresponding to the inner and outer-most bins, respectively. It is clear from Figure E.5 that for eigenvectors with larger eigenvalues, the components corresponding to the inner bins dominate, and vice versa. Therefore, in the case of resolved MCXC clusters, the fact that most of the contribution to χ_0^2 comes from modes with the three largest eigenvalues indicates that mostly the inner bins are contributing to the signal. Whereas for the whole MCXC sample, there is also contribution from outer bins. This



(a) ESZ vs. MCXC mass estimates of 62 clusters common to both catalogues [122].

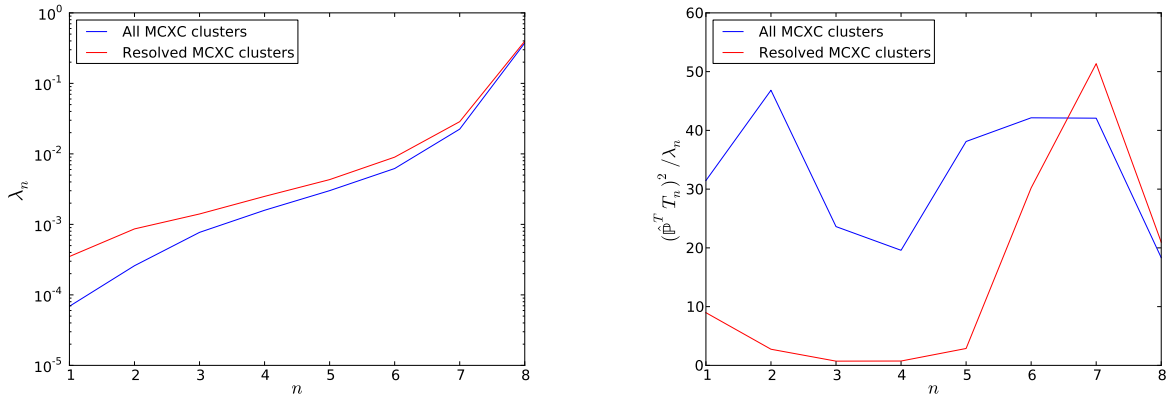


(b) Effect of cluster mass uncertainties on the universal pressure profile.



(c) Effect of cluster mass uncertainties on gas mass profile.

Figure E.3: Effect of systematic uncertainties associated with the mass of clusters. (a) Difference between ESZ and MCXC mass estimates of 62 clusters common to both catalogues [122]. (b) Effect of changing MCXC masses on the universal pressure profile. The entire MCXC sample is used and the standard self-similar model is assumed ($\delta = 0$ in Equation (6.9)). The blue data points show the result of our measurements using MCXC mass estimates. The red points show measurements for which the MCXC masses are randomly changed according to the distribution shown in (a). (c) Gas mass fraction, computed using Equation (6.44), corresponding to pressure measurements in (b). The shaded areas represent the standard deviation in the measurement of f_{gas} as given by Equation (6.45).



(a) Eigenvalues of the covariance matrix $\mathbf{C}_{\mathbb{P}}$.

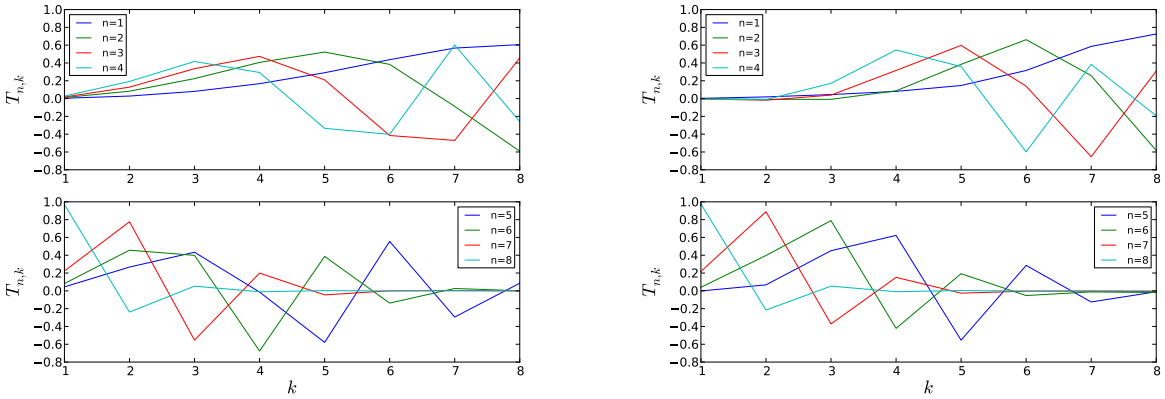
(b) Contribution of eigenmodes of $\mathbf{C}_{\mathbb{P}}$ to χ^2_0.

Figure E.4: Spectrum of the covariance matrix and the contribution of different eigenmodes to χ^2_0.

analysis reassures our intuition that even unresolved clusters contribute to the tSZ signal in the outskirts of the ICM.

E.5 Pressure Measurements: the Exact Numbers

Here we report exact numbers corresponding to the measurement of $\mathbb{P}$, i.e. the best fit universal pressure value $\widehat{\mathbb{P}}$ and its associated covariance matrix $\mathbf{C}_{\mathbb{P}}$. As before, δ quantifies deviation from the standard self similar model (see Equation (6.9)). We give in Table E.1 (E.2) $\widehat{\mathbb{P}}$ and $\mathbf{C}_{\mathbb{P}}$ in the case of all clusters (resolved clusters), with $\delta = 0$. Similarly, Table E.3 (E.4) shows our measurements in the case of all clusters (resolved clusters), with $\delta = 0.12$. In Tables E.5−E.8, we show the result of our analysis on mass bins $1-4$ introduced in Table 6.1a, respectively (with $\delta = 0$). Similarly, Tables E.9−E.11 provide the best fit universal pressure values and the corresponding covariance matrix for mass bins $1-3$ introduced in Table 6.1b, respectively (with $\delta = 0$).



(a) All MCXC clusters.

(b) Resolved MCXC clusters.

Figure E.5: Components $T_{n,k}$ of the eigenvectors $\boldsymbol{T_n}$ of the covariance matrix. Here n labels different eigenvectors, chosen so that the eigenvalues satisfy $\lambda_1 < \lambda_2 < \cdots < \lambda_{N_b}$. The label k runs over all radial bins.

Table E.1: Pressure measurement of all MCXC clusters with the standard self-similar model ($\delta = 0$ in Equation (6.9)).

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7	Bin 8
Pressure	2904.845	503.878	111.528	-8.831	8.515	54.610	-21.088	8.990

(a) Best fit pressure values $\widehat{\mathbb{P}}$ ($\times 10^{-3}$)

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7	Bin 8
Bin 1	359.047	-82.359	16.095	-2.989	0.872	-0.540	0.220	-0.016
Bin 2	-82.359	36.703	-12.792	2.632	-0.470	0.052	-0.023	0.018
Bin 3	16.095	-12.792	10.142	-4.167	0.720	-0.010	-0.067	0.031
Bin 4	-2.989	2.632	-4.167	4.320	-1.928	0.298	0.010	-0.017
Bin 5	0.872	-0.470	0.720	-1.928	2.405	-1.162	0.200	-0.008
Bin 6	-0.540	0.052	-0.010	0.298	-1.162	1.577	-0.786	0.132
Bin 7	0.220	-0.023	-0.067	0.010	0.200	-0.786	1.094	-0.476
Bin 8	-0.016	0.018	0.031	-0.017	-0.008	0.132	-0.476	0.432

(b) Covariance matrix $\mathbf{C}_{\mathbb{P}}$ ($\times 10^{-3}$)

Table E.2: Pressure measurement of resolved MCXC clusters with the standard self-similar model ($\delta = 0$ in Equation (6.9)).

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7	Bin 8
Pressure	3156.051	652.465	163.366	2.700	48.387	78.075	-14.598	14.782

(a) Best fit pressure values $\widehat{\mathbb{P}}$ ($\times 10^{-3}$)

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7	Bin 8
Bin 1	385.004	-76.888	18.073	-2.839	1.073	-0.664	-0.018	-0.199
Bin 2	-76.888	42.837	-11.045	3.328	-0.439	-0.052	-0.177	-0.240
Bin 3	18.073	-11.045	11.960	-3.454	0.846	-0.087	-0.154	-0.136
Bin 4	-2.839	3.328	-3.454	5.082	-1.641	0.291	-0.054	-0.141
Bin 5	1.073	-0.439	0.846	-1.641	2.785	-1.022	0.214	-0.084
Bin 6	-0.664	-0.052	-0.087	0.291	-1.022	1.812	-0.677	0.102
Bin 7	-0.018	-0.177	-0.154	-0.054	0.214	-0.677	1.281	-0.466
Bin 8	-0.199	-0.240	-0.136	-0.141	-0.084	0.102	-0.466	0.745

(b) Covariance matrix $\mathbf{C}_{\mathbb{P}}$ ($\times 10^{-3}$)

Table E.3: Pressure measurement of all MCXC clusters with the modified self-similar model ($\delta = 0.12$ in Equation (6.9)).

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7	Bin 8
Pressure	2819.487	515.355	97.656	-15.799	4.248	50.901	-22.110	8.053

(a) Best fit pressure values $\widehat{\mathbb{P}}$ ($\times 10^{-3}$)

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7	Bin 8
Bin 1	352.226	-82.197	16.578	-3.117	0.820	-0.496	0.211	-0.018
Bin 2	-82.197	36.333	-12.687	2.618	-0.467	0.035	-0.018	0.023
Bin 3	16.578	-12.687	9.842	-4.026	0.683	0.001	-0.075	0.036
Bin 4	-3.117	2.618	-4.026	4.114	-1.830	0.269	0.019	-0.019
Bin 5	0.821	-0.467	0.683	-1.830	2.273	-1.093	0.179	-0.004
Bin 6	-0.496	0.035	0.001	0.268	-1.093	1.487	-0.740	0.121
Bin 7	0.211	-0.018	-0.075	0.019	0.179	-0.740	1.034	-0.451
Bin 8	-0.018	0.023	0.036	-0.019	-0.004	0.121	-0.451	0.405

(b) Covariance matrix $\mathbf{C}_{\mathbb{P}}$ ($\times 10^{-3}$)

Table E.4: Pressure measurement of resolved MCXC clusters with the modified self-similar model ($\delta = 0.12$ in Equation (6.9)).

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7	Bin 8
Pressure	3095.361	672.793	159.082	0.880	46.378	75.526	-16.262	8.843

(a) Best fit pressure values $\widehat{\mathbb{P}}$ ($\times 10^{-3}$)

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7	Bin 8
Bin 1	378.015	-77.464	18.369	-2.951	1.042	-0.619	-0.015	-0.185
Bin 2	-77.463	42.440	-11.029	3.331	-0.412	-0.058	-0.166	-0.237
Bin 3	18.369	-11.029	11.612	-3.340	0.811	-0.060	-0.153	-0.126
Bin 4	-2.951	3.331	-3.340	4.854	-1.554	0.263	-0.036	-0.140
Bin 5	1.042	-0.412	0.811	-1.554	2.639	-0.959	0.193	-0.072
Bin 6	-0.619	-0.058	-0.060	0.263	-0.959	1.713	-0.638	0.094
Bin 7	-0.015	-0.166	-0.153	-0.036	0.193	-0.638	1.214	-0.442
Bin 8	-0.185	-0.237	-0.126	-0.140	-0.072	0.094	-0.442	0.711

(b) Covariance matrix $\mathbf{C}_{\mathbb{P}}$ ($\times 10^{-3}$)

Table E.5: Pressure measurement of Mass-bin 1 of all MCXC clusters, as defined in Table 6.1a ($\delta = 0$ in Equation (6.9)).

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7	Bin 8
Pressure	2864.897	143.372	165.443	82.563	57.910	-22.757	53.503	37.140

(a) Best fit pressure values $\widehat{\mathbb{P}}$ ($\times 10^{-3}$)

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7	Bin 8
Bin 1	1351.387	-266.076	30.865	0.085	1.519	-0.727	-0.089	0.199
Bin 2	-266.075	117.062	-37.249	5.581	0.115	0.093	-0.088	-0.025
Bin 3	30.866	-37.249	34.247	-13.720	2.080	0.136	0.038	-0.086
Bin 4	0.084	5.582	-13.720	16.293	-7.307	1.239	0.070	-0.055
Bin 5	1.520	0.115	2.080	-7.307	9.706	-4.636	0.856	-0.025
Bin 6	-0.728	0.093	0.136	1.239	-4.636	6.452	-3.190	0.559
Bin 7	-0.089	-0.088	0.038	0.070	0.856	-3.190	4.424	-1.869
Bin 8	0.199	-0.025	-0.086	-0.055	-0.025	0.559	-1.869	1.832

(b) Covariance matrix $\mathbf{C}_{\mathbb{P}}$ ($\times 10^{-3}$)

Table E.6: Pressure measurement of Mass-bin 2 of all MCXC clusters, as defined in Table 6.1a ($\delta = 0$ in Equation (6.9)).

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7	Bin 8
Pressure	2546.299	226.786	253.102	18.308	-79.667	204.195	-123.737	-7.187

(a) Best fit pressure values $\widehat{\mathbb{P}}$ ($\times 10^{-3}$)

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7	Bin 8
Bin 1	2504.452	-719.122	230.700	-84.783	37.468	-19.648	10.328	-3.445
Bin 2	-719.123	311.542	-137.925	54.001	-22.478	10.146	-4.269	1.271
Bin 3	230.702	-137.925	99.347	-53.731	22.694	-9.263	3.633	-0.946
Bin 4	-84.785	54.002	-53.731	46.988	-27.135	11.806	-4.740	1.355
Bin 5	37.469	-22.479	22.695	-27.135	26.089	-15.997	6.810	-1.871
Bin 6	-19.648	10.146	-9.264	11.806	-15.997	16.275	-9.554	2.819
Bin 7	10.328	-4.269	3.633	-4.740	6.809	-9.554	9.325	-3.905
Bin 8	-3.445	1.271	-0.946	1.355	-1.871	2.819	-3.905	2.673

(b) Covariance matrix $\mathbf{C}_{\mathbb{P}}$ ($\times 10^{-3}$)

Table E.7: Pressure measurement of Mass-bin 3 of all MCXC clusters, as defined in Table 6.1a ($\delta = 0$ in Equation (6.9)).

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7	Bin 8
Pressure	2614.687	599.174	-119.300	-23.585	9.475	161.234	-148.958	139.492

(a) Best fit pressure values $\widehat{\mathbb{P}}$ ($\times 10^{-3}$)

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7	Bin 8
Bin 1	816.335	-142.117	21.841	-11.060	7.061	-4.547	1.150	0.328
Bin 2	-142.117	90.898	-31.204	6.198	-3.173	1.336	-0.726	0.394
Bin 3	21.841	-31.204	33.339	-13.358	2.133	-0.887	0.478	-0.212
Bin 4	-11.060	6.198	-13.358	16.314	-6.910	1.096	-0.559	0.262
Bin 5	7.061	-3.173	2.133	-6.911	9.575	-4.265	0.700	-0.205
Bin 6	-4.547	1.336	-0.887	1.096	-4.265	6.279	-2.804	0.305
Bin 7	1.150	-0.726	0.478	-0.559	0.700	-2.804	4.368	-1.827
Bin 8	0.328	0.394	-0.212	0.262	-0.205	0.305	-1.827	1.997

(b) Covariance matrix $\mathbf{C}_{\mathbb{P}}$ ($\times 10^{-3}$)

Table E.8: Pressure measurement of Mass-bin 4 of all MCXC clusters, as defined in Table 6.1a ($\delta = 0$ in Equation (6.9)).

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7	Bin 8
Pressure	4522.423	811.544	140.014	-63.051	-16.867	10.731	60.649	-94.528

(a) Best fit pressure values $\widehat{\mathbb{P}}$ ($\times 10^{-3}$)

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7	Bin 8
Bin 1	4322.774	-1224.992	307.886	-40.992	-10.662	8.444	-1.387	-0.802
Bin 2	-1224.992	434.752	-126.640	19.097	2.756	-3.787	0.775	0.135
Bin 3	307.888	-126.641	59.055	-15.870	0.103	1.643	-0.966	0.207
Bin 4	-40.993	19.097	-15.870	13.835	-4.773	-0.305	0.650	-0.317
Bin 5	-10.663	2.756	0.103	-4.773	6.955	-2.669	-0.122	0.219
Bin 6	8.445	-3.788	1.643	-0.305	-2.669	4.461	-1.854	0.077
Bin 7	-1.387	0.775	-0.966	0.650	-0.122	-1.854	3.298	-1.417
Bin 8	-0.803	0.135	0.207	-0.316	0.219	0.077	-1.417	1.642

(b) Covariance matrix $\mathbf{C}_{\mathbb{P}}$ ($\times 10^{-3}$)

Table E.9: Pressure measurement of Mass-bin 1 of resolved MCXC clusters, as defined in Table 6.1b ($\delta = 0$ in Equation (6.9)).

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7	Bin 8
Pressure	2970.143	-124.936	58.889	102.331	97.592	54.978	75.993	94.229

(a) Best fit pressure values $\widehat{\mathbb{P}}$ ($\times 10^{-3}$)

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7	Bin 8
Bin 1	1375.649	-236.216	47.033	-2.420	2.119	-1.271	-1.240	-0.716
Bin 2	-236.214	129.290	-31.665	8.192	-0.684	-0.393	-0.512	-0.764
Bin 3	47.036	-31.667	38.119	-11.362	2.634	-0.653	-0.387	-0.557
Bin 4	-2.420	8.192	-11.361	17.521	-6.171	1.401	-0.481	-0.385
Bin 5	2.120	-0.685	2.634	-6.171	10.149	-4.066	1.023	-0.461
Bin 6	-1.271	-0.392	-0.653	1.401	-4.066	6.674	-2.656	0.468
Bin 7	-1.240	-0.512	-0.387	-0.481	1.023	-2.656	4.630	-1.700
Bin 8	-0.716	-0.764	-0.557	-0.385	-0.461	0.468	-1.700	2.578

(b) Covariance matrix $\mathbf{C}_{\mathbb{P}}$ ($\times 10^{-3}$)

Table E.10: Pressure measurement of Mass-bin 2 of resolved MCXC clusters, as defined in Table 6.1b ($\delta = 0$ in Equation (6.9)).

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7	Bin 8
Pressure	2482.111	747.796	-29.912	-32.605	2.269	180.980	-115.173	74.383

(a) Best fit pressure values $\widehat{\mathbb{P}}$ ($\times 10^{-3}$)

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7	Bin 8
Bin 1	682.399	-103.800	28.794	-8.645	4.760	-4.283	0.648	-0.521
Bin 2	-103.799	88.006	-19.527	7.560	-3.040	0.400	-1.156	-0.276
Bin 3	28.794	-19.527	31.489	-8.677	2.714	-1.348	-0.072	-0.722
Bin 4	-8.645	7.560	-8.676	15.190	-4.710	1.212	-0.803	-0.205
Bin 5	4.760	-3.040	2.714	-4.710	8.745	-3.022	0.850	-0.533
Bin 6	-4.283	0.400	-1.348	1.212	-3.022	5.652	-1.930	0.398
Bin 7	0.648	-1.156	-0.072	-0.803	0.850	-1.930	3.906	-1.247
Bin 8	-0.521	-0.276	-0.722	-0.205	-0.533	0.398	-1.247	2.349

(b) Covariance matrix $\mathbf{C}_{\mathbb{P}}$ ($\times 10^{-3}$)

Table E.11: Pressure measurement of Mass-bin 3 of resolved MCXC clusters, as defined in Table 6.1b ($\delta = 0$ in Equation (6.9)).

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7	Bin 8
Pressure	3737.228	1220.556	334.228	7.300	34.974	23.814	7.113	-101.074

(a) Best fit pressure values $\widehat{\mathbb{P}}$ ($\times 10^{-3}$)

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7	Bin 8
Bin 1	3842.297	-1117.928	295.139	-49.881	-6.830	7.859	-1.586	-0.158
Bin 2	-1117.928	417.687	-116.036	26.214	2.461	-3.534	0.417	-1.168
Bin 3	295.139	-116.036	58.291	-14.939	1.537	1.550	-1.007	-0.033
Bin 4	-49.882	26.214	-14.939	14.707	-4.168	0.113	0.563	-0.664
Bin 5	-6.829	2.461	1.537	-4.168	6.989	-2.360	0.175	0.093
Bin 6	7.858	-3.534	1.550	0.113	-2.360	4.449	-1.665	0.203
Bin 7	-1.586	0.417	-1.007	0.563	0.175	-1.665	3.262	-1.232
Bin 8	-0.158	-1.168	-0.033	-0.664	0.093	0.203	-1.232	2.008

(b) Covariance matrix $\mathbf{C}_{\mathbb{P}}$ ($\times 10^{-3}$)

References

[1] G. Aad, T. Abajyan, B. Abbott, J. Abdallah, S. Abdel Khalek, A. A. Abdelalim, O. Abdinov, R. Aben, B. Abi, M. Abolins, and et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716:1–29, September 2012, 1207.7214.

[2] T. Adam, N. Agafonova, A. Aleksandrov, O. Altinok, P. Alvarez Sanchez, A. Anokhina, S. Aoki, A. Ariga, T. Ariga, D. Autiero, A. Badertscher, A. Ben Dhahbi, A. Bertolin, C. Bozza, T. Brugière, R. Brugnera, F. Brunet, G. Brunetti, S. Buontempo, B. Carlus, F. Cavanna, A. Cazes, L. Chaussard, M. Chernyavsky, V. Chiarella, A. Chukanov, G. Colosimo, M. Crespi, N. D'Ambrosio, G. De Lellis, M. De Serio, Y. Déclais, P. del Amo Sanchez, F. Di Capua, A. Di Crescenzo, D. Di Ferdinando, N. Di Marco, S. Dmitrievsky, M. Dracos, D. Duchesneau, S. Dusini, T. Dzhatdoev, J. Ebert, I. Efthymiopoulos, O. Egorov, A. Ereditato, L. S. Esposito, J. Favier, T. Ferber, R. A. Fini, T. Fukuda, A. Garfagnini, G. Giacomelli, M. Giorgini, M. Giovannozzi, C. Girerd, J. Goldberg, C. Göllnitz, D. Golubkov, L. Goncharova, Y. Gornushkin, G. Grella, F. Grianti, E. Gschwendtner, C. Guerin, A. M. Guler, C. Gustavino, C. Hagner, K. Hamada, T. Hara, R. Enikeev, M. Hierholzer, A. Hollnagel, M. Ieva, H. Ishida, K. Ishiguro, K. Jakovcic, C. Jollet, M. Jones, F. Juget, M. Kamiscioglu, J. Kawada, S. H. Kim, M. Kimura, E. Kiritsis, N. Kitagawa, B. Klicek, J. Knuesel, K. Kodama, M. Komatsu, U. Kose, I. Kreslo, C. Lazzaro, J. Lenkeit, A. Ljubicic, A. Longhin, A. Malgin, G. Mandrioli, J. Marteau, T. Matsuo, V. Matveev, N. Mauri, A. Mazzoni, E. Medinaceli, F. Meisel, A. Meregaglia, P. Migliozzi, S. Mikado, D. Missiaen, P. Monacelli, K. Morishima, U. Moser, M. T. Muciaccia, N. Naganawa, T. Naka, M. Nakamura, T. Nakano, Y. Nakatsuka, D. Naumov, V. Nikitina, F. Nitti, S. Ogawa, N. Okateva, A. Olchevsky, O. Palamara, A. Paoloni, B. D. Park, I. G. Park, A. Pastore, L. Patrizii, E. Pennacchio, H. Pessard, C. Pistillo, N. Polukhina, M. Pozzato, K. Pretzl, F. Pupilli, R. Rescigno, F. Riguzzi, T. Roganova, H. Rokujo, G. Rosa, I. Rostovtseva, A. Rubbia, A. Russo, V. Ryasny,

O. Ryazhskaya, O. Sato, Y. Sato, Z. Sahnoun, A. Schembri, J. Schuler, L. Scotto Lavina, J. Serrano, I. Shakiryanova, A. Sheshukov, H. Shibuya, G. Shoziyoev, S. Simone, M. Sioli, C. Sirignano, G. Sirri, J. S. Song, M. Spinetti, L. Stanco, N. Starkov, S. Stellacci, M. Stipcevic, T. Strauss, S. Takahashi, M. Tenti, F. Terranova, I. Tezuka, V. Tioukov, P. Tolun, N. T. Trani, S. Tufanli, P. Vilain, M. Vladimirov, L. Votano, J.-L. Vuilleumier, G. Wilquet, B. Wonsak, J. Wurtz, V. Yakushev, C. S. Yoon, J. Yoshida, Y. Zaitsev, S. Zemskova, and A. Zghiche. Measurement of the neutrino velocity with the OPERA detector in the CNGS beam. *Journal of High Energy Physics*, 10:93, October 2012, 1109.4897.

[3] N. Afshordi. Gravitational Aether and the thermodynamic solution to the cosmological constant problem. *ArXiv e-prints*, July 2008, 0807.2639.

[4] N. Afshordi. Dark Energy, Black Hole Entropy, and the First Precision Measurement in Quantum Gravity. *ArXiv e-prints*, March 2010, 1003.4811.

[5] 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*, 8:137, August 2012, 1205.1296.

[6] N. Afshordi, M. Buck, F. Dowker, D. Rideout, R. D. Sorkin, and Y. K. Yazdi. A ground state for the causal diamond in 2 dimensions. *Journal of High Energy Physics*, 10:88, October 2012, 1207.7101.

[7] N. Afshordi, D. J. H. Chung, M. Doran, and G. Geshnizjani. Cuscuton cosmology: Dark energy meets modified gravity. *Physical Review D*, 75(12):123509, June 2007, astro-ph/0702002.

[8] N. Afshordi, D. J. H. Chung, and G. Geshnizjani. Causal field theory with an infinite speed of sound. *Physical Review D*, 75(8):083513, April 2007, hep-th/0609150.

[9] N. Afshordi, Y.-T. Lin, D. Nagai, and A. J. R. Sanderson. Missing thermal energy of the intracluster medium. *Monthly Notices of the Royal Astronomical Society*, 378:293–300, June 2007, astro-ph/0612700.

[10] N. Afshordi, Y.-T. Lin, and A. J. R. Sanderson. Wilkinson Microwave Anisotropy Probe Constraints on the Intracluster Medium. ApJ, 629:1–14, August 2005, arXiv:astro-ph/0408560.

[11] N. Agarwal, R. Bean, J. Khoury, and M. Trodden. Cascading cosmology. *Physical Review D*, 81(8):084020, April 2010, 0912.3798.

[12] B. Allen. Vacuum states in de Sitter space. Phys. Rev. D, 32:3136–3149, December 1985.

[13] S. W. Allen, A. E. Evrard, and A. B. Mantz. Cosmological Parameters from Observations of Galaxy Clusters. ARA&A, 49:409–470, September 2011, 1103.4829.

[14] R. Amanullah, C. Lidman, D. Rubin, G. Aldering, P. Astier, K. Barbary, M. S. Burns, A. Conley, K. S. Dawson, S. E. Deustua, M. Doi, S. Fabbro, L. Faccioli, H. K. Fakhouri, G. Folatelli, A. S. Fruchter, H. Furusawa, G. Garavini, G. Goldhaber, A. Goobar, D. E. Groom, I. Hook, D. A. Howell, N. Kashikawa, A. G. Kim, R. A. Knop, M. Kowalski, E. Linder, J. Meyers, T. Morokuma, S. Nobili, J. Nordin, P. E. Nugent, L. Östman, R. Pain, N. Panagia, S. Perlmutter, J. Raux, P. Ruiz-Lapuente, A. L. Spadafora, M. Strovink, N. Suzuki, L. Wang, W. M. Wood-Vasey, N. Yasuda, and T. Supernova Cosmology Project. Spectra and Hubble Space Telescope Light Curves of Six Type Ia Supernovae at $0.511 < z < 1.12$ and the Union2 Compilation. *The Astronomical Journal*, 716:712–738, June 2010, 1004.1711.

[15] M. Arnaud, E. Pointecouteau, and G. W. Pratt. Calibration of the galaxy cluster M{500}-Y{X} relation with XMM-Newton. A&A, 474:L37–L40, November 2007, 0709.1561.

[16] M. Arnaud, G. W. Pratt, R. Piffaretti, H. Böhringer, J. H. Croston, and E. Pointecouteau. The universal galaxy cluster pressure profile from a representative sample of nearby systems (REXCESS) and the Y_{SZ} - M_{500} relation. A&A, 517:A92, July 2010, 0910.1234.

[17] S. Aslanbeigi and M. Buck. A preferred ground state for the scalar field in de Sitter space. *Journal of High Energy Physics*, 8:39, August 2013, 1306.3231.

[18] S. Aslanbeigi, G. Lavaux, A. Hajian, and N. Afshordi. An optimal and model-independent measurement of the intracluster pressure profile - I. Methodology and first applications. MNRAS, 435:1788–1808, October 2013, 1304.6457.

[19] S. Aslanbeigi, G. Robbers, B. Z. Foster, K. Kohri, and N. Afshordi. Phenomenology of gravitational aether as a solution to the old cosmological constant problem. Phys. Rev. D, 84(10):103522, November 2011, 1106.3955.

[20] S. Aslanbeigi, M. Saravani, and R. D. Sorkin. Generalized causal set d'Alembertians. *Journal of High Energy Physics*, 6:24, June 2014, 1403.1622.

[21] F. Atrio-Barandela, A. Kashlinsky, D. Kocevski, and H. Ebeling. Measurement of the Electron-Pressure Profile of Galaxy Clusters in 3 Year Wilkinson Microwave Anisotropy Probe (WMAP) Data. ApJ, 675:L57–L60, March 2008, 0802.3716.

[22] E. Aver, K. A. Olive, and E. D. Skillman. A new approach to systematic uncertainties and self-consistency in helium abundance determinations. *Journal of Cosmology and Astroparticle Physics*, 5:3, May 2010, 1001.5218.

[23] N. Battaglia, J. R. Bond, C. Pfrommer, and J. L. Sievers. On the Cluster Physics of Sunyaev-Zel'dovich and X-Ray Surveys. II. Deconstructing the Thermal SZ Power Spectrum. ApJ, 758:75, October 2012, 1109.3711.

[24] D. M. T. Benincasa and F. Dowker. Scalar Curvature of a Causal Set. *Physical Review Letters*, 104(18):181301, May 2010, 1001.2725.

[25] C. L. Bennett, D. Larson, J. L. Weiland, N. Jarosik, G. Hinshaw, N. Odegard, K. M. Smith, R. S. Hill, B. Gold, M. Halpern, E. Komatsu, M. R. Nolta, L. Page, D. N. Spergel, E. Wollack, J. Dunkley, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, and E. L. Wright. Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results. *ArXiv e-prints*, December 2012, 1212.5225.

[26] S. Bhattacharya, S. Habib, K. Heitmann, and A. Vikhlinin. Dark Matter Halo Profiles of Massive Clusters: Theory vs. Observations. *ArXiv e-prints*, December 2011, 1112.5479.

[27] E. Bianchi and C. Rovelli. Why all these prejudices against a constant? *ArXiv e-prints*, February 2010, 1002.3966.

[28] BICEP2 Collaboration, P. A. R Ade, R. W. Aikin, D. Barkats, S. J. Benton, C. A. Bischoff, J. J. Bock, J. A. Brevik, I. Buder, E. Bullock, C. D. Dowell, L. Duband, J. P. Filippini, S. Fliescher, S. R. Golwala, M. Halpern, M. Hasselfield, S. R. Hildebrandt, G. C. Hilton, V. V. Hristov, K. D. Irwin, K. S. Karkare, J. P. Kaufman, B. G. Keating, S. A. Kernasovskiy, J. M. Kovac, C. L. Kuo, E. M. Leitch, M. Lueker, P. Mason, C. B. Netterfield, H. T. Nguyen, R. O'Brient, R. W. Ogburn, IV, A. Orlando, C. Pryke, C. D. Reintsema, S. Richter, R. Schwarz, C. D. Sheehy, Z. K. Staniszewski, R. V. Sudiwala, G. P. Teply, J. E. Tolan, A. D. Turner, A. G. Vieregg, C. L. Wong, and K. W. Yoon. BICEP2 I: Detection Of B-mode Polarization at Degree Angular Scales. *ArXiv e-prints*, March 2014, 1403.3985.

[29] M. Birkinshaw. The Sunyaev-Zel'dovich effect. Phys. Rep., 310:97–195, March 1999, arXiv:astro-ph/9808050.

[30] N. D. Birrell and P. C. W. Davies. *Quantum Fields in Curved Space*. April 1984.

[31] Joseph J. Bisognano and Eyvind H. Wichmann. On the duality condition for a hermitian scalar field. *Journal of Mathematical Physics*, 16(4):985–1007, 1975.

[32] H. Böhringer, P. Schuecker, G. W. Pratt, M. Arnaud, T. J. Ponman, J. H. Croston, S. Borgani, R. G. Bower, U. G. Briel, C. A. Collins, M. Donahue, W. R. Forman, A. Finoguenov, M. J. Geller, L. Guzzo, J. P. Henry, R. Kneissl, J. J. Mohr, K. Matsushita, C. R. Mullis, T. Ohashi, K. Pedersen, D. Pierini, H. Quintana, S. Raychaudhury, T. H. Reiprich, A. K. Romer, P. Rosati, K. Sabirli, R. F. Temple, P. T. P. Viana, A. Vikhlinin, G. M. Voit, and Y.-Y. Zhang. The representative XMM-Newton cluster structure survey (REXCESS) of an X-ray luminosity selected galaxy cluster sample. A&A, 469:363–377, July 2007, arXiv:astro-ph/0703553.

[33] L. Bombelli, J. Henson, and R. D. Sorkin. Discreteness Without Symmetry Breaking:. a Theorem. *Modern Physics Letters A*, 24:2579–2587, 2009, gr-qc/0605006.

[34] L. Bombelli, J. Lee, D. Meyer, and R. D. Sorkin. Space-time as a causal set. *Physical Review Letters*, 59:521–524, August 1987.

[35] P. Bonifacio, P. Molaro, T. Sivarani, R. Cayrel, M. Spite, F. Spite, B. Plez, J. Andersen, B. Barbuy, T. C. Beers, E. Depagne, V. Hill, P. François, B. Nordström, and F. Primas. First stars VII - Lithium in extremely metal poor dwarfs. *Astronomy & Astrophysics*, 462:851–864, February 2007, astro-ph/0610245.

[36] R. Bousso, A. Maloney, and A. Strominger. Conformal vacua and entropy in de Sitter space. Phys. Rev. D, 65(10):104039, May 2002, hep-th/0112218.

[37] J. Bros, H. Epstein, and U. Moschella. The lifetime of a massive particle in a de Sitter universe. *Journal of Cosmology and Astroparticle Physics*, 2:3, February 2008, hep-th/0612184.

[38] J. Bros, H. Epstein, and U. Moschella. Particle Decays and Stability on the de Sitter Universe. *Annales Henri Poincaré*, 11:611–658, June 2010, 0812.3513.

[39] R. Brunetti and K. Fredenhagen. Microlocal Analysis andInteracting Quantum Field Theories:Renormalization on Physical Backgrounds. *Communications in Mathematical Physics*, 208:623–661, 2000, arXiv:math-ph/9903028.

[40] R. Brunetti, K. Fredenhagen, and S. Hollands. A remark on alpha vacua for quantum field theories on de Sitter space. *Journal of High Energy Physics*, 5:63, May 2005, hep-th/0503022.

[41] T. S. Bunch and P. C. W. Davies. Quantum field theory in de Sitter space - Renormalization by point-splitting. *Royal Society of London Proceedings Series A*, 360:117–134, March 1978.

[42] C. P. Burgess and D. Hoover. UV sensitivity in supersymmetric large extra dimensions: The Ricci-flat case. *Nuclear Physics B*, 772:175–204, June 2007, hep-th/0504004.

[43] J. E. Carlstrom, G. P. Holder, and E. D. Reese. Cosmology with the Sunyaev-Zel'dovich Effect. ARA&A, 40:643–680, 2002, arXiv:astro-ph/0208192.

[44] S. M. Carroll. The Cosmological Constant. *Living Reviews in Relativity*, 4:1, February 2001, astro-ph/0004075.

[45] S. M. Carroll, W. H. Press, and E. L. Turner. The cosmological constant. ARA&A, 30:499–542, 1992.

[46] I. Ciufolini and E. C. Pavlis. A confirmation of the general relativistic prediction of the Lense-Thirring effect. Nature, 431:958–960, October 2004.

[47] R. H. Cyburt, B. D. Fields, and K. A. Olive. The NACRE thermonuclear reaction compilation and big bang nucleosynthesis. *New Astronomy*, 6:215–238, June 2001, astro-ph/0102179.

[48] R. H. Cyburt, B. D. Fields, and K. A. Olive. An update on the big bang nucleosynthesis prediction for ^{7}Li: the problem worsens. *Journal of Cosmology and Astroparticle Physics*, 11:12, November 2008, 0808.2818.

[49] C. de Rham, G. Dvali, S. Hofmann, J. Khoury, O. Pujolàs, M. Redi, and A. J. Tolley. Cascading Gravity: Extending the Dvali-Gabadadze-Porrati Model to Higher Dimension. *Physical Review Letters*, 100(25):251603, June 2008, 0711.2072.

[50] C. de Rham, G. Gabadadze, L. Heisenberg, and D. Pirtskhalava. Cosmic acceleration and the helicity-0 graviton. *Physical Review D*, 83(10):103516, May 2011, 1010.1780.

[51] C. de Rham, S. Hofmann, J. Khoury, and A. J. Tolley. Cascading gravity and degravitation. *Journal of Cosmology and Astroparticle Physics*, 2:11, February 2008, 0712.2821.

[52] G. D. Domínguez and S. E. Trione. On the Laplace transforms of retarded, Lorentz-invariant functions. *Advances in Mathematics*, 31(1):51–62, January 1979.

[53] M. Doran. CMBEASY: an object oriented code for the cosmic microwave background. *Journal of Cosmology and Astroparticle Physics*, 10:11, October 2005, astro-ph/0302138.

[54] M. Doran and C. M. Müller. Analyse this! A cosmological constraint package for CMBEASY. *Journal of Cosmology and Astroparticle Physics*, 9:3, September 2004, astro-ph/0311311.

[55] F. Dowker. Causal Sets and Discrete Spacetime. In J.-M. Alimi and A. Füzfa, editors, *Albert Einstein Century International Conference*, volume 861 of *American Institute of Physics Conference Series*, pages 79–88, November 2006.

[56] F. Dowker and L. Glaser. Causal set d'Alembertians for various dimensions. *ArXiv e-prints*, May 2013, 1305.2588.

[57] F. Dowker, S. Johnston, and R. D. Sorkin. Hilbert spaces from path integrals. *Journal of Physics A Mathematical General*, 43(27):A275302, July 2010, 1002.0589.

[58] J. Dunkley, R. Hlozek, J. Sievers, V. Acquaviva, P. A. R. Ade, P. Aguirre, M. Amiri, J. W. Appel, L. F. Barrientos, E. S. Battistelli, J. R. Bond, B. Brown, B. Burger, J. Chervenak, S. Das, M. J. Devlin, S. R. Dicker, W. Bertrand Doriese, R. Dünner, T. Essinger-Hileman, R. P. Fisher, J. W. Fowler, A. Hajian, M. Halpern, M. Hasselfield, C. Hernández-Monteagudo, G. C. Hilton, M. Hilton, A. D. Hincks, K. M. Huffenberger, D. H. Hughes, J. P. Hughes, L. Infante, K. D. Irwin, J. B. Juin, M. Kaul, J. Klein, A. Kosowsky, J. M. Lau, M. Limon, Y.-T. Lin, R. H. Lupton, T. A. Marriage, D. Marsden, P. Mauskopf, F. Menanteau, K. Moodley, H. Moseley, C. B. Netterfield, M. D. Niemack, M. R. Nolta, L. A. Page, L. Parker, B. Partridge, B. Reid, N. Sehgal, B. Sherwin, D. N. Spergel, S. T. Staggs, D. S. Swetz, E. R. Switzer, R. Thornton, H. Trac, C. Tucker, R. Warne, E. Wollack, and Y. Zhao. The Atacama Cosmology Telescope: Cosmological Parameters from the 2008 Power Spectrum. *The Astronomical Journal*, 739:52, September 2011, 1009.0866.

[59] G. Dvali, S. Hofmann, and J. Khoury. Degravitation of the cosmological constant and graviton width. *Physical Review D*, 76(8):084006, October 2007, hep-th/0703027.

[60] H. K. Eriksen, I. J. O'Dwyer, J. B. Jewell, B. D. Wandelt, D. L. Larson, K. M. Górski, S. Levin, A. J. Banday, and P. B. Lilje. Power Spectrum Estimation from High-Resolution Maps by Gibbs Sampling. ApJS, 155:227–241, December 2004, arXiv:astro-ph/0407028.

[61] C. W. F. Everitt, D. B. Debra, B. W. Parkinson, J. P. Turneaure, J. W. Conklin, M. I. Heifetz, G. M. Keiser, A. S. Silbergleit, T. Holmes, J. Kolodziejczak, M. Al-Meshari, J. C. Mester, B. Muhlfelder, V. G. Solomonik, K. Stahl, P. W. Worden, Jr., W. Bencze, S. Buchman, B. Clarke, A. Al-Jadaan, H. Al-Jibreen, J. Li, J. A. Lipa, J. M. Lockhart, B. Al-Suwaidan, M. Taber, and S. Wang. Gravity Probe B: Final Results of a Space Experiment to Test General Relativity. *Physical Review Letters*, 106(22):221101, June 2011, 1105.3456.

[62] C. J. Fewster and R. Verch. On a recent construction of 'vacuum-like' quantum field states in curved spacetime. *Classical and Quantum Gravity*, 29(20):205017, October 2012, 1206.1562.

[63] R. Flauger, J. C. Hill, and D. N. Spergel. Toward an Understanding of Foreground Emission in the BICEP2 Region. *ArXiv e-prints*, May 2014, 1405.7351.

[64] B. Z. Foster and T. Jacobson. Post-Newtonian parameters and constraints on Einstein-aether theory. *Physical Review D*, 73(6):064015, March 2006, gr-qc/0509083.

[65] M. Fukuma, S. Sugishita, and Y. Sakatani. Propagators in de Sitter space. Phys. Rev. D, 88(2):024041, July 2013, 1301.7352.

[66] S. A. Fulling. Remarks on positive frequency and Hamiltonians in expanding universes. *General Relativity and Gravitation*, 10:807–824, July 1979.

[67] L. Glaser. Fixing factors for the causal set d'Alembertian in d dimensions. *ArXiv e-prints*, November 2013, 1311.1701.

[68] K. M. Gorski, A. J. Banday, C. L. Bennett, G. Hinshaw, A. Kogut, G. F. Smoot, and E. L. Wright. Power Spectrum of Primordial Inhomogeneity Determined from the Four-Year COBE DMR Sky Maps. ApJ, 464:L11, June 1996, arXiv:astro-ph/9601063.

[69] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann. HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. ApJ, 622:759–771, April 2005, astro-ph/0409513.

[70] K. M. Gorski, B. D. Wandelt, E. Hivon, F. K. Hansen, and A. J. Banday. *The HEALPix Primer (version 2.15a)*, 2010.

[71] J. B. Hartle. Spacetime Quantum Mechanics and the Quantum Mechanics of Spacetime. *ArXiv General Relativity and Quantum Cosmology e-prints*, April 1993, gr-qc/9304006.

[72] S. W. Hawking. Particle Creation by Black Holes. *Advanced Series in Astrophysics and Cosmology*, 8:85–106, June 1993.

[73] S. W. Hawking, A. R. King, and P. J. McCarthy. A new topology for curved spacetime which incorporates the causal, differential, and conformal structures. *Journal of Mathematical Physics*, 17:174–181, 1976.

[74] J. Henson. The causal set approach to quantum gravity. *ArXiv General Relativity and Quantum Cosmology e-prints*, January 2006, gr-qc/0601121.

[75] 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. *ArXiv e-prints*, December 2012, 1212.5226.

[76] S. Hollands and R. M. Wald. Local Wick Polynomials and Time Ordered Productsof Quantum Fields in Curved Spacetime. *Communications in Mathematical Physics*, 223:289–326, 2001, gr-qc/0103074.

[77] S. Hollands and R. M. Wald. Existence of Local Covariant Time Ordered Products of Quantum Fields in Curved Spacetime. *Communications in Mathematical Physics*, 231:309–345, 2002, gr-qc/0111108.

[78] Cullan Howlett, Antony Lewis, Alex Hall, and Anthony Challinor. CMB power spectrum parameter degeneracies in the era of precision cosmology. *JCAP*, 1204:027, 2012, 1201.3654.

[79] R. Ilie, G. B. Thompson, and D. D. Reid. A numerical study of the correspondence between paths in a causal set and geodesics in the continuum. *Classical and Quantum Gravity*, 23:3275–3285, May 2006, gr-qc/0512073.

[80] L. Iorio. Will the recently approved LARES mission be able to measure the Lense-Thirring effect at 1%? *General Relativity and Gravitation*, 41:1717–1724, August 2009, 0803.3278.

[81] A. H. Jaffe, A. Balbi, J. R. Bond, J. Borrill, P. G. Ferreira, D. Finkbeiner, S. Hanany, A. T. Lee, B. Rabii, P. L. Richards, G. F. Smoot, R. Stompor, C. D. Winant, and J. H. P. Wu. Determining Foreground Contamination in Cosmic Microwave Background Observations: Diffuse Galactic Emission in the MAXIMA-I Field. ApJ, 615:55–62, November 2004, arXiv:astro-ph/0301077.

[82] D. P. Jatkar, L. Leblond, and A. Rajaraman. Decay of massive fields in de Sitter space. Phys. Rev. D, 85(2):024047, January 2012, 1107.3513.

[83] S. Johnston. FAST TRACK COMMUNICATION: Particle propagators on discrete spacetime. *Classical and Quantum Gravity*, 25(20):202001, October 2008, 0806.3083.

[84] S. Johnston. Feynman Propagator for a Free Scalar Field on a Causal Set. *Physical Review Letters*, 103(18):180401, October 2009, 0909.0944.

[85] S. Johnston. *Quantum Fields on Causal Sets*. PhD thesis, PhD Thesis, 2010, October 2010.

[86] F. Kamiab and N. Afshordi. Neutron stars and the cosmological constant problem. *Physical Review D*, 84(6):063011, September 2011, 1104.5704.

[87] E. Komatsu, K. M. Smith, J. Dunkley, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. R. Nolta, L. Page, D. N. Spergel, M. Halpern, R. S. Hill, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. ApJS, 192:18, February 2011, 1001.4538.

[88] A. V. Kravtsov and S. Borgani. Formation of Galaxy Clusters. ARA&A, 50:353–409, September 2012, 1205.5556.

[89] P. Lagogiannis, A. Maloney, and Y. Wang. Odd-dimensional de Sitter Space is Transparent. *ArXiv e-prints*, June 2011, 1106.2846.

[90] G. Lavaux, N. Afshordi, and M. J. Hudson. First measurement of the bulk flow of nearby galaxies using the cosmic microwave background. *ArXiv e-prints*, July 2012, 1207.1721.

[91] Antony Lewis, Anthony Challinor, and Anthony Lasenby. Efficient computation of CMB anisotropies in closed FRW models. *Astrophys. J.*, 538:473–476, 2000, astro-ph/9911177.

[92] Chung-Pei Ma and Edmund Bertschinger. Cosmological perturbation theory in the synchronous and conformal newtonian gauges. *The Astrophysical Journal*, 455:7, 1995.

[93] Y.-Z. Ma, G. Hinshaw, and D. Scott. WMAP Observations of Planck ESZ Clusters. *ArXiv e-prints*, March 2013, 1303.4728.

[94] D. B. Malament. The class of continuous timelike curves determines the topology of spacetime. *Journal of Mathematical Physics*, 8:1399, 1977.

[95] D. Marolf and I. A. Morrison. Infrared stability of de Sitter space: Loop corrections to scalar propagators. Phys. Rev. D, 82(10):105032, November 2010, 1006.0035.

[96] D. Marolf, I. A. Morrison, and M. Srednicki. Perturbative S-matrix for massive scalar fields in global de Sitter space. *Classical and Quantum Gravity*, 30(15):155023, August 2013, 1209.6039.

[97] T. A. Marriage, V. Acquaviva, P. A. R. Ade, P. Aguirre, M. Amiri, J. W. Appel, L. F. Barrientos, E. S. Battistelli, J. R. Bond, B. Brown, B. Burger, J. Chervenak, S. Das, M. J. Devlin, S. R. Dicker, W. Bertrand Doriese, J. Dunkley, R. Dünner, T. Essinger-Hileman, R. P. Fisher, J. W. Fowler, A. Hajian, M. Halpern, M. Hasselfield, C. Hernández-Monteagudo, G. C. Hilton, M. Hilton, A. D. Hincks, R. Hlozek, K. M. Huffenberger, D. Handel Hughes, J. P. Hughes, L. Infante, K. D. Irwin, J. Baptiste Juin, M. Kaul, J. Klein, A. Kosowsky, J. M. Lau, M. Limon, Y.-T. Lin, R. H. Lupton, D. Marsden, K. Martocci, P. Mauskopf, F. Menanteau, K. Moodley, H. Moseley, C. B. Netterfield, M. D. Niemack, M. R. Nolta, L. A. Page, L. Parker, B. Partridge, H. Quintana, E. D. Reese, B. Reid, N. Sehgal, B. D. Sherwin, J. Sievers, D. N. Spergel, S. T. Staggs, D. S. Swetz, E. R. Switzer, R. Thornton, H. Trac, C. Tucker, R. Warne, G. Wilson, E. Wollack, and Y. Zhao. The Atacama Cosmology Telescope: Sunyaev-Zel'dovich-Selected Galaxy Clusters at 148 GHz in the 2008 Survey. ApJ, 737:61, August 2011, 1010.1065.

[98] D. J. E. Marsh, D. Grin, R. Hlozek, and P. G. Ferreira. Tensor Detection Severely Constrains Axion Dark Matter. *ArXiv e-prints*, March 2014, 1403.4216.

[99] J. C. Mather, D. J. Fixsen, R. A. Shafer, C. Mosier, and D. T. Wilkinson. Calibrator Design for the COBE Far-Infrared Absolute Spectrophotometer (FIRAS). ApJ, 512:511–520, February 1999, arXiv:astro-ph/9810373.

[100] M. Matsumoto and T. Nishimura. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 8(1):3–30, 1998.

[101] J. Meléndez and I. Ramírez. Reappraising the Spite Lithium Plateau: Extremely Thin and Marginally Consistent with WMAP Data. *The Astrophysical Journal, Letters*, 615:L33–L36, November 2004, astro-ph/0409383.

[102] J.-B. Melin, J. G. Bartlett, J. Delabrouille, M. Arnaud, R. Piffaretti, and G. W. Pratt. The galaxy cluster Y_{SZ} - L_X and Y_{SZ} - M relations from the WMAP 5-yr data. A&A, 525:A139, January 2011, 1001.0871.

[103] M. J. Mortonson and U. Seljak. A joint analysis of Planck and BICEP2 B modes including dust polarization uncertainty. *ArXiv e-prints*, May 2014, 1405.5857.

[104] E. Mottola. Particle creation in de Sitter space. Phys. Rev. D, 31:754–766, February 1985.

[105] V. F. Mukhanov, H. A. Feldman, and R. H. Brandenberger. Theory of cosmological perturbations. Phys. Rep., 215:203–333, June 1992.

[106] J. Myrheim. STATISTICAL GEOMETRY. August 1978. CERN-TH-2538.

[107] D. Nagai, A. V. Kravtsov, and A. Vikhlinin. Effects of Galaxy Formation on Thermodynamics of the Intracluster Medium. ApJ, 668:1–14, October 2007, arXiv:astro-ph/0703661.

[108] K. Nakamura and Particle Data Group. Review of Particle Physics. *Journal of Physics G Nuclear Physics*, 37(7):075021, July 2010.

[109] A. Narimani, N. Afshordi, and D. Scott. How does pressure gravitate? Cosmological constant problem confronts observational cosmology. *ArXiv e-prints*, June 2014, 1406.0479.

[110] J. F. Navarro, C. S. Frenk, and S. D. M. White. A Universal Density Profile from Hierarchical Clustering. ApJ, 490:493, December 1997, arXiv:astro-ph/9611107.

[111] W. J. Percival, B. A. Reid, D. J. Eisenstein, N. A. Bahcall, T. Budavari, J. A. Frieman, M. Fukugita, J. E. Gunn, Ž. Ivezić, G. R. Knapp, R. G. Kron, J. Loveday, R. H. Lupton, T. A. McKay, A. Meiksin, R. C. Nichol, A. C. Pope, D. J. Schlegel, D. P. Schneider, D. N. Spergel, C. Stoughton, M. A. Strauss, A. S. Szalay, M. Tegmark,

M. S. Vogeley, D. H. Weinberg, D. G. York, and I. Zehavi. Baryon acoustic oscillations in the Sloan Digital Sky Survey Data Release 7 galaxy sample. *Monthly Notices of the Royal Astronomical Society*, 401:2148–2168, February 2010, 0907.1660.

[112] S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, I. M. Hook, A. G. Kim, M. Y. Kim, J. C. Lee, N. J. Nunes, R. Pain, C. R. Pennypacker, R. Quimby, C. Lidman, R. S. Ellis, M. Irwin, R. G. McMahon, P. Ruiz-Lapuente, N. Walton, B. Schaefer, B. J. Boyle, A. V. Filippenko, T. Matheson, A. S. Fruchter, N. Panagia, H. J. M. Newberg, W. J. Couch, and T. S. C. Project. Measurements of Ω and Λ from 42 High-Redshift Supernovae. *The Astronomical Journal*, 517:565–586, June 1999, astro-ph/9812133.

[113] M. Pettini, B. J. Zych, M. T. Murphy, A. Lewis, and C. C. Steidel. Deuterium abundance in the most metal-poor damped Lyman alpha system: converging on $\Omega_{b,0}\mathrm{h}^2$. *Monthly Notices of the Royal Astronomical Society*, 391:1499–1510, December 2008, 0805.0594.

[114] R. Piffaretti, M. Arnaud, G. W. Pratt, E. Pointecouteau, and J.-B. Melin. The MCXC: a meta-catalogue of x-ray detected clusters of galaxies. A&A, 534:A109, October 2011, 1007.1916.

[115] T. Plagge, B. A. Benson, P. A. R. Ade, K. A. Aird, L. E. Bleem, J. E. Carlstrom, C. L. Chang, H.-M. Cho, T. M. Crawford, A. T. Crites, T. de Haan, M. A. Dobbs, E. M. George, N. R. Hall, N. W. Halverson, G. P. Holder, W. L. Holzapfel, J. D. Hrubes, M. Joy, R. Keisler, L. Knox, A. T. Lee, E. M. Leitch, M. Lueker, D. Marrone, J. J. McMahon, J. Mehl, S. S. Meyer, J. J. Mohr, T. E. Montroy, S. Padin, C. Pryke, C. L. Reichardt, J. E. Ruhl, K. K. Schaffer, L. Shaw, E. Shirokoff, H. G. Spieler, B. Stalder, Z. Staniszewski, A. A. Stark, K. Vanderlinde, J. D. Vieira, R. Williamson, and O. Zahn. Sunyaev-Zel'dovich Cluster Profiles Measured with the South Pole Telescope. ApJ, 716:1118–1135, June 2010, 0911.2444.

[116] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. I. R. Alves, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, H. Aussel, and et al. Planck 2013 results. I. Overview of products and scientific results. *ArXiv e-prints*, March 2013, 1303.5062.

[117] Planck collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday, and et al. Planck 2013 results. XV. CMB power spectra and likelihood. *ArXiv e-prints*, March 2013, 1303.5075.

[118] Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday, and et al. Planck 2013 results. XVI. Cosmological parameters. *ArXiv e-prints*, March 2013, 1303.5076.

[119] Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday, and et al. Planck 2013 Results. XXIV. Constraints on primordial non-Gaussianity. *ArXiv e-prints*, March 2013, 1303.5084.

[120] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. Balbi, A. J. Banday, R. B. Barreiro, and et al. Planck early results. VIII. The all-sky early Sunyaev-Zeldovich cluster sample. A&A, 536:A8, December 2011, 1101.2024.

[121] Planck Collaboration V. Planck intermediate results. V. Pressure profiles of galaxy clusters from the Sunyaev-Zeldovich effect. A&A, 550:A131, February 2013, 1207.4061.

[122] Planck Collaboration XI. Planck early results. XI. Calibration of the local galaxy cluster Sunyaev-Zeldovich scaling relations. A&A, 536:A11, December 2011, 1101.2026.

[123] Planck HFI Core Team VI. Planck early results. VI. The High Frequency Instrument data processing. A&A, 536:A6, December 2011, 1101.2048.

[124] S. Planelles, S. Borgani, K. Dolag, S. Ettori, D. Fabjan, G. Murante, and L. Tornatore. Baryon census in hydrodynamical simulations of galaxy clusters. MNRAS, March 2013, 1209.5058.

[125] M. Pospelov and J. Pradler. Big Bang Nucleosynthesis as a Probe of New Physics. *Annual Review of Nuclear and Particle Science*, 60:539–568, November 2010, 1011.1054.

[126] G. W. Pratt, H. Böhringer, J. H. Croston, M. Arnaud, S. Borgani, A. Finoguenov, and R. F. Temple. Temperature profiles of a representative sample of nearby X-ray galaxy clusters. A&A, 461:71–80, January 2007, arXiv:astro-ph/0609480.

[127] G. W. Pratt, J. H. Croston, M. Arnaud, and H. Böhringer. Galaxy cluster X-ray luminosity scaling relations from a representative local sample (REXCESS). A&A, 498:361–378, May 2009, 0809.3784.

[128] C. Prescod-Weinstein, N. Afshordi, and M. L. Balogh. Stellar black holes and the origin of cosmic acceleration. *Physical Review D*, 80(4):043513, August 2009, 0905.3551.

[129] M. Reed and B. Simon. *Methods of Modern Mathematical Physics: Functional Analysis*. Academic Press, 1981.

[130] A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner, B. Leibundgut, M. M. Phillips, D. Reiss, B. P. Schmidt, R. A. Schommer, R. C. Smith, J. Spyromilio, C. Stubbs, N. B. Suntzeff, and J. Tonry. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *The Astronomical Journal*, 116:1009–1038, September 1998, astro-ph/9805201.

[131] A. G. Riess, L. Macri, S. Casertano, H. Lampeitl, H. C. Ferguson, A. V. Filippenko, S. W. Jha, W. Li, and R. Chornock. A 3% Solution: Determination of the Hubble Constant with the Hubble Space Telescope and Wide Field Camera 3. *The Astronomical Journal*, 730:119, April 2011, 1103.2976.

[132] D. H. Rudd and D. Nagai. Nonequilibrium Electrons and the Sunyaev-Zel'Dovich Effect of Galaxy Clusters. ApJ, 701:L16–L19, August 2009, 0907.1287.

[133] M. L. Ruggiero and A. Tartaglia. Gravitomagnetic effects. *Nuovo Cimento B Serie*, 117:743, July 2002, gr-qc/0207065.

[134] M. Saravani and S. Aslanbeigi. On the Causal Set-Continuum Correspondence. *ArXiv e-prints*, March 2014, 1403.6429.

[135] A. Schild. Discrete space-time and integral lorentz transformations. *Phys. Rev.*, 73:414–415, Feb 1948.

[136] H. J. Schmidt. On the de Sitter space-time - the geometric foundation of inflationary cosmology. *Fortschritte der Physik*, 41:179–199, 1993.

[137] N. Sehgal, H. Trac, V. Acquaviva, P. A. R. Ade, P. Aguirre, M. Amiri, J. W. Appel, L. F. Barrientos, E. S. Battistelli, J. R. Bond, B. Brown, B. Burger, J. Chervenak, S. Das, M. J. Devlin, S. R. Dicker, W. Bertrand Doriese, J. Dunkley, R. Dünner, T. Essinger-Hileman, R. P. Fisher, J. W. Fowler, A. Hajian, M. Halpern, M. Hasselfield, C. Hernández-Monteagudo, G. C. Hilton, M. Hilton, A. D. Hincks, R. Hlozek, D. Holtz, K. M. Huffenberger, D. H. Hughes, J. P. Hughes, L. Infante, K. D. Irwin, A. Jones, J. Baptiste Juin, J. Klein, A. Kosowsky, J. M. Lau, M. Limon, Y.-T. Lin,

R. H. Lupton, T. A. Marriage, D. Marsden, K. Martocci, P. Mauskopf, F. Menanteau, K. Moodley, H. Moseley, C. B. Netterfield, M. D. Niemack, M. R. Nolta, L. A. Page, L. Parker, B. Partridge, B. Reid, B. D. Sherwin, J. Sievers, D. N. Spergel, S. T. Staggs, D. S. Swetz, E. R. Switzer, R. Thornton, C. Tucker, R. Warne, E. Wollack, and Y. Zhao. The Atacama Cosmology Telescope: Cosmology from Galaxy Clusters Detected via the Sunyaev-Zel'dovich Effect. ApJ, 732:44, May 2011, 1010.1025.

[138] U. Seljak, A. Slosar, and P. McDonald. Cosmological parameters from combining the Lyman-α forest with CMB, galaxy clustering and SN constraints. *Journal of Cosmology and Astroparticle Physics*, 10:14, October 2006, astro-ph/0604335.

[139] A. Serebrov, V. Varlamov, A. Kharitonov, A. Fomin, Y. Pokotilovski, P. Geltenbort, J. Butterworth, I. Krasnoschekova, M. Lasakov, R. Tal'daev, A. Vassiljev, and O. Zherebtsov. Measurement of the neutron lifetime using a gravitational trap and a low-temperature Fomblin coating. *Physics Letters B*, 605:72–78, January 2005, nucl-ex/0408009.

[140] Jonathan R Shewchuk. An introduction to the conjugate gradient method without the agonizing pain. Technical report, Pittsburgh, PA, USA, 1994.

[141] M. S. Smith, L. H. Kawano, and R. A. Malaney. Experimental, computational, and observational analysis of primordial nucleosynthesis. *The Astrophysical Journal, Supplement*, 85:219–247, April 1993.

[142] R. D. Sorkin. Spacetime and causal sets. In J. C. D'Olivo, E. Nahmad-Achar, M. Rosenbaum, M. P. Ryan, Jr., L. F. Urrutia, and F. Zertuche, editors, *Relativity and Gravitation*, page 150, 1991.

[143] R. D. Sorkin. Quantum Mechanics as Quantum Measure Theory. *Modern Physics Letters A*, 9:3119–3127, 1994, gr-qc/9401003.

[144] R. D. Sorkin. Causal Sets: Discrete Gravity (Notes for the Valdivia Summer School). *ArXiv General Relativity and Quantum Cosmology e-prints*, September 2003, gr-qc/0309009.

[145] R. D. Sorkin. Does Locality Fail at Intermediate Length-Scales. *ArXiv General Relativity and Quantum Cosmology e-prints*, March 2007, gr-qc/0703099.

[146] R. D. Sorkin. Toward a "fundamental theorem of quantal measure theory". *ArXiv e-prints*, April 2011, 1104.0997.

[147] M. Spradlin, A. Strominger, and A. Volovich. Les Houches Lectures on De Sitter Space. *ArXiv High Energy Physics - Theory e-prints*, September 2001, hep-th/0110007.

[148] Z. Staniszewski, P. A. R. Ade, K. A. Aird, B. A. Benson, L. E. Bleem, J. E. Carlstrom, C. L. Chang, H.-M. Cho, T. M. Crawford, A. T. Crites, T. de Haan, M. A. Dobbs, N. W. Halverson, G. P. Holder, W. L. Holzapfel, J. D. Hrubes, M. Joy, R. Keisler, T. M. Lanting, A. T. Lee, E. M. Leitch, A. Loehr, M. Lueker, J. J. McMahon, J. Mehl, S. S. Meyer, J. J. Mohr, T. E. Montroy, C.-C. Ngeow, S. Padin, T. Plagge, C. Pryke, C. L. Reichardt, J. E. Ruhl, K. K. Schaffer, L. Shaw, E. Shirokoff, H. G. Spieler, B. Stalder, A. A. Stark, K. Vanderlinde, J. D. Vieira, O. Zahn, and A. Zenteno. Galaxy Clusters Discovered with a Sunyaev-Zel'dovich Effect Survey. ApJ, 701:32–41, August 2009, 0810.1578.

[149] M. Sun, N. Sehgal, G. M. Voit, M. Donahue, C. Jones, W. Forman, A. Vikhlinin, and C. Sarazin. The Pressure Profiles of Hot Gas in Local Galaxy Groups. ApJ, 727:L49, February 2011, 1012.0312.

[150] M. Sun, G. M. Voit, M. Donahue, C. Jones, W. Forman, and A. Vikhlinin. Chandra Studies of the X-Ray Gas Properties of Galaxy Groups. ApJ, 693:1142–1172, March 2009, 0805.2320.

[151] R. A. Sunyaev and Y. B. Zeldovich. The Observations of Relic Radiation as a Test of the Nature of X-Ray Radiation from the Clusters of Galaxies. *Comments on Astrophysics and Space Physics*, 4:173, November 1972.

[152] G. 't Hooft. Quantum Gravity: A Fundamental Problem and Some Radical Ideas. In M. Lévy & S. Deser, editor, *NATO ASIB Proc. 44: Recent Developments in Gravitation*, page 323, 1979.

[153] E. A. Tagirov. Consequences of field quantization in de Sitter type cosmological models. *Annals of Physics*, 76:561–579, 1973.

[154] I. Thompson. NIST Handbook of Mathematical Functions, edited by Frank W.J. Olver, Daniel W. Lozier, Ronald F. Boisvert, Charles W. Clark. *Contemporary Physics*, 52:497–498, September 2011.

[155] W. G. Unruh. Notes on black-hole evaporation. *Phys. Rev. D*, 14(4):870–892, Aug 1976.

[156] K. Vanderlinde, T. M. Crawford, T. de Haan, J. P. Dudley, L. Shaw, P. A. R. Ade, K. A. Aird, B. A. Benson, L. E. Bleem, M. Brodwin, J. E. Carlstrom, C. L. Chang, A. T. Crites, S. Desai, M. A. Dobbs, R. J. Foley, E. M. George, M. D. Gladders, N. R. Hall, N. W. Halverson, F. W. High, G. P. Holder, W. L. Holzapfel, J. D. Hrubes, M. Joy, R. Keisler, L. Knox, A. T. Lee, E. M. Leitch, A. Loehr, M. Lueker, D. P. Marrone, J. J. McMahon, J. Mehl, S. S. Meyer, J. J. Mohr, T. E. Montroy, C.-C. Ngeow, S. Padin, T. Plagge, C. Pryke, C. L. Reichardt, A. Rest, J. Ruel, J. E. Ruhl, K. K. Schaffer, E. Shirokoff, J. Song, H. G. Spieler, B. Stalder, Z. Staniszewski, A. A. Stark, C. W. Stubbs, A. van Engelen, J. D. Vieira, R. Williamson, Y. Yang, O. Zahn, and A. Zenteno. Galaxy Clusters Selected with the Sunyaev-Zel'dovich Effect from 2008 South Pole Telescope Observations. ApJ, 722:1180–1196, October 2010, 1003.0003.

[157] A. Vikhlinin, A. Kravtsov, W. Forman, C. Jones, M. Markevitch, S. S. Murray, and L. Van Speybroeck. Chandra Sample of Nearby Relaxed Galaxy Clusters: Mass, Gas Fraction, and Mass-Temperature Relation. ApJ, 640:691–709, April 2006, arXiv:astro-ph/0507092.

[158] R. M. Wald. *Quantum field theory in curved spacetime and black hole thermodynamics.* 1994.

[159] R. M. Wald. The Formulation of Quantum Field Theory in Curved Spacetime. *ArXiv e-prints*, July 2009, 0907.0416.

[160] B. D. Wandelt, D. L. Larson, and A. Lakshminarayanan. Global, exact cosmic microwave background data analysis using Gibbs sampling. Phys. Rev. D, 70(8):083511, October 2004, astro-ph/0310080.

[161] S. Weinberg. The cosmological constant problem. *Reviews of Modern Physics*, 61:1–23, January 1989.

[162] S. Weinberg. *Cosmology.* Oxford University Press, 2008.

[163] C. M. Will. The Confrontation between General Relativity and Experiment. *Living Reviews in Relativity*, 9:3, March 2006, gr-qc/0510072.

[164] R. Williamson, B. A. Benson, F. W. High, K. Vanderlinde, P. A. R. Ade, K. A. Aird, K. Andersson, R. Armstrong, M. L. N. Ashby, M. Bautz, G. Bazin, E. Bertin, L. E. Bleem, M. Bonamente, M. Brodwin, J. E. Carlstrom, C. L. Chang, S. C. Chapman, A. Clocchiatti, T. M. Crawford, A. T. Crites, T. de Haan, S. Desai, M. A. Dobbs,

J. P. Dudley, G. G. Fazio, R. J. Foley, W. R. Forman, G. Garmire, E. M. George, M. D. Gladders, A. H. Gonzalez, N. W. Halverson, G. P. Holder, W. L. Holzapfel, S. Hoover, J. D. Hrubes, C. Jones, M. Joy, R. Keisler, L. Knox, A. T. Lee, E. M. Leitch, M. Lueker, D. Luong-Van, D. P. Marrone, J. J. McMahon, J. Mehl, S. S. Meyer, J. J. Mohr, T. E. Montroy, S. S. Murray, S. Padin, T. Plagge, C. Pryke, C. L. Reichardt, A. Rest, J. Ruel, J. E. Ruhl, B. R. Saliwanchik, A. Saro, K. K. Schaffer, L. Shaw, E. Shirokoff, J. Song, H. G. Spieler, B. Stalder, S. A. Stanford, Z. Staniszewski, A. A. Stark, K. Story, C. W. Stubbs, J. D. Vieira, A. Vikhlinin, and A. Zenteno. A Sunyaev-Zel'dovich-selected Sample of the Most Massive Galaxy Clusters in the 2500 deg^2 South Pole Telescope Survey. ApJ, 738:139, September 2011, 1101.1290.

[165] Y.-Y. Zhang, T. F. Laganá, D. Pierini, E. Puchwein, P. Schneider, and T. H. Reiprich. Star-formation efficiency and metal enrichment of the intracluster medium in local massive clusters of galaxies. A&A, 535:A78, November 2011, 1109.0390.