PERTURBED FRIEDMANN COSMOLOGY WITHOUT ASSUMPTIONS ON THE STRESS: CONSISTENCY AND APPLICATION TO THE DARK ENERGY PROBLEM

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAI`I AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN PHYSICS AUGUST 2018

By
Kevin A. S. Croker

Dissertation Committee:
John Learned, Chairperson
Thomas Browder
Jeffery Kuhn
Brent Tully
Joel Weiner

Keywords: general relativity, cosmology, dark energy, gravastar
Copyright © 2018 by
Kevin A. S. Croker
To the memory of,

J. M. J. Madey,

inventor of the free-electron laser (FEL)
ACKNOWLEDGMENTS

I’ve had the fortune of being surrounded by experimental, observational, and instrumental specialists. For a theoretically inclined, naive kid, what initially seemed like an impediment has blossomed into an unimaginable boon. Writing an acknowledgments section is challenging, but not because you don’t know who to thank or what to thank them for. So many people, over a span of years, have lent their hands, hearts, and minds to me. If I do not mention you specifically below, if you are reading this, thank you.

Someone once told me that “travelers go places.” Tom Browder travels more than anyone I have ever met. His cultural savvy and charm are utterly delightful and, I hope, contagious. Without his guidance, encouragement, and support I would have never had the opportunities to study in Japan and meet so many wonderful people. I am glad to thank him for patiently and repeatedly reviewing early manuscript after manuscript. Most of these had completely incorrect physical content. His constant pressure to produce testable observational results ultimately led to the central results of this thesis.

This work would have never begun to converge without continual technical input and spiritual support from Joel Weiner. His nearly inexhaustible patience and dedication to instruction guided and reassured me at every step of the way. Thank you for taking the time to typeset so many clarifications for me! Joel was the first professional to earnestly and enthusiastically try to work with me to incorporate fundamental ideas from biology into quantitative cosmological models. This attitude, toward a stupid kid, from a tenured professor who didn’t even know him well, did so much to mend the previous soul-crushing dismissals of lesser men.

I got to share a semester with Nick Kaiser three doors down the hall, which changed my trajectory in life. I like to do physics like a martial artist likes to spar. Many physicists refuse to even enter the ring if you challenge them, but not Kaiser. We argued for days, upon weeks, upon months. Each time I would rebut, he would come at me from another angle. After months of this extremely strenuous process, without even realizing it, I had put together a thesis. I would have been expelled from the University if it weren’t for John Learned. His interest in my desire to approach old problems from different perspectives led him to accept me as his student. His faith in my ability to eventually produce something of value is ultimately the only reason I have been able to write this thesis. Finally, I want to thank Jeff Kuhn for teaching me GR, tending my spiritual wounds, and assuring me that I always had an out if I really needed one.

Neill Warrington, you taught me how to swim. You taught me how to kindle my inner fire again. Ryo Matsuda, you were the first person to refer to me as omae. Then you verified all my equations and became my first coauthor. I’m going to do my best to pull you back from that Machine Learning orgy. Brandon Wilson, you taught me the value of meditation. You pushed me, at a moment of great despair, to endure. That small push laid the foundation for all of my financial support the past five years. Mia Yoshimoto, you taught me kindness and how to cry in a good way. We also found that cat, which inhaled an entire asparagus stalk in three chomps. You shared with me a relationship that, every time I think back to it, I can’t help but grin like an idiot. Kurtis Nishimura, at a time when people I thought I could count on disappeared, you stepped in and worked with me side by side to make a conference happen. You made it possible for me to
keep my word to many others. It will be a privilege to help you keep yours. To everyone that has ever given a couch, a floor, or a bed to me, so that I could live as you do for a few days, thank you. This includes: J. Klacsmann, E. Francisco, D. Schainker, K. Lautenbach, H. Wong, Y. Yamaguchi, C. Jackson, C. Kala, D. Hiramatsu, A. Romero-Wolf, N. Fernandez, E. Mottola, J. Bramante, B. Schiff, and E. Mendez. I hope that, one day, we might get to work together on something great.

I’ve noticed in my life that I do my best work whenever there is someone I am excited to come home to. Over the past year, whenever a “great” idea completely failed, whenever Nick still just wasn’t convinced, Haruka Watanabe always had the hug. Whenever something worked, whenever there was cause to celebrate, no matter how small, she has been there to share it with. She has kept my life balanced and, through her grace, given me the strength and perspective to finally bring this chapter of my life to completion. Finally, Mom and Dad, I love you.
ABSTRACT

The origin of the accelerated late-time expansion of the universe has stimulated significant theoretical interest over the past two decades. Contemporary data remain consistent with the simplest explanation: a fixed Dark Energy density of vacuum $\rho_\Lambda \approx 0.7$ times the critical density required to make the universe spatially flat, today. Since the matter density decays with time, but $\rho_\Lambda$ remains constant, their approximate equality now requires tuning $\rho_\Lambda$ to at least 24 decimal places. This Coincidence Problem suggests that our understanding of the vacuum remains incomplete.

Evidence for $\rho_\Lambda$ comes from astrophysical observations interpreted within the context of perturbed Friedmann cosmology. We present a thorough re-derivation of Friedmann’s equations directly from the Einstein-Hilbert action. We identify an apparently overlooked commutation in the calculation of the stress-energy term. The resulting zero-order source necessarily includes an average contribution over all pressures, everywhere. This resolves the long-standing ambiguity of how to construct the isotropic and homogeneous source to Friedmann’s equations, given a matter distribution far from thermal equilibrium. The perturbed Friedmann cosmological framework remains unaltered at first order. This result opens a new avenue of attack on the Dark Energy problem: localized regions of Dark Energy can mimic a cosmological $\rho_\Lambda$.

It has been argued, from at least 1966, that complete stellar gravitational collapse should not produce a true Black Hole (BH). Instead, it should produce some sort of GEneric Object of Dark Energy (GEODE). We investigate a simple cosmological scenario where collapsed stellar matter decays into a cosmological GEODE contribution with $\langle P_s \rangle_V = (-1 + \chi) \langle \rho_s \rangle_V$. We find that our model mimics the required $\rho_\Lambda$ and remains consistent with existing astrophysical constraints for $0 \leq \chi < 0.06$. This resolves the Coincidence Problem. The model tightly couples the time-evolution of Dark Energy to the integrated BH mass-function. The resulting unambiguous observational signatures may be accessible to upcoming experiments.
# TABLE OF CONTENTS

Acknowledgments ........................................................................................................ iv

Abstract ......................................................................................................................... vi

List of Tables .................................................................................................................. x

List of Figures ................................................................................................................ xi

1 Introduction ................................................................................................................ 1

1.1 Layout ...................................................................................................................... 2

1.2 Conventions ............................................................................................................. 2

1.3 The standard cosmology ......................................................................................... 3

1.3.1 Differential geometry in kinematic context ......................................................... 3

1.3.2 Einstein’s Equations ............................................................................................ 5

1.3.3 Friedmann’s equations ......................................................................................... 6

1.3.4 Cosmography .................................................................................................... 7

1.4 The Dark Energy Problem ....................................................................................... 11

2 Friedmann cosmology from consistent application of the action principle ............ 13

2.1 Preliminaries ........................................................................................................... 13

2.1.1 The action principle for gravity ....................................................................... 14

2.2 The marble and the bowl: constrained classical variations ................................... 15

2.2.1 Application to a scalar field ............................................................................. 17

2.3 Zero-order Friedmann equations without stress-energy assumptions ................. 19

2.3.1 Significance of the volume $V$ ......................................................................... 22

2.3.2 Generalizations ............................................................................................... 22
2.3.3 Discussion ................................................................. 23

2.4 Model-restricted covariant conservation of stress-energy at zero order ......................... 24

3 Covariant cosmological perturbation theory without assumptions on the stress ............. 26

3.1 Preliminaries ............................................................... 26

3.1.1 Notation ................................................................. 27

3.1.2 Standard definition of the cosmological stress tensor ....................................... 28

3.2 A cosmological stress-tensor without assumptions ................................................. 30

3.2.1 The vierbein frame, with review ............................................. 30

3.2.2 The principal frame ...................................................... 32

3.2.3 Regeneration of the Bardeen stress-tensor ............................................. 34

3.2.4 Relation to the distribution function definition of the stress ............................ 36

3.2.5 Alternative definition of the null stress .......................................... 37

3.2.6 Summary ................................................................. 37

3.3 Scalar perturbations, in conformal Newtonian gauge ............................................. 38

3.3.1 Connection 1-forms ...................................................... 39

3.3.2 Riemann curvature and Ricci scalar .......................................... 41

3.3.3 Conformal term .......................................................... 45

3.3.4 Variation of the gravitational action ........................................ 46

3.3.5 Variation within the matter action .......................................... 47

3.3.6 Field Equations ........................................................ 48

3.3.7 Conservation conditions .................................................. 50

3.4 Conclusion ................................................................. 50

4 Late-time accelerated expansion from localized Dark Energy ................................. 51
LIST OF TABLES

4.1 Constraints on Population III star formation ........................................... 62

B.1 Astrophysical parameters for BH population estimation ................................. 77

B.2 Linear fits to $M_{\text{BH}}(M_*)$ ................................................................. 78

B.3 Metallicity dependence of remnant population parameters ............................ 79

B.4 Fraction of comoving stellar density collapsed into BH, based on metallicity and supernovae engine ............................................................................. 80

B.5 Adopted astrophysical parameters .............................................................. 81
4.1 Estimated instantaneous GEODE formation redshift $z_f$, consistent with present day DE density $\Omega_\Lambda$. The horizontal axis gives the deviation $\chi$ from a perfect de Sitter ($w_s = -1$) contribution to the cosmological source, due to the presence of GEODE crusts. Note that the onset of star formation $z_{early} \equiv 20$ encloses the viable region (green/light grey) for $\chi < 0.16$. The red (medium grey) region indicates $\Omega_\Lambda$ within a factor of 1.5 of the present day value, while the blue (dark grey) region indicates $\Omega_\Lambda$ within a factor of 0.5 of the present day value.

4.2 Estimated dark energy density induced by GEODE formation from onset of stellar formation at $z_{ee} \equiv 20$ until some cutoff $z_f$. This assumes a constant, but conservative, primordial rate of stellar density production $d\rho/dt \equiv 5.3 \times 10^{-3}$. This value is consistent with the more conservative $\gamma$-ray opacity constraints reported by Inoue et al. [2014] at $z \sim 6$ (see Table 4.1). Further assumptions include a fixed fraction of stellar density collapsing into BH and a constant amplification due to accretion. Note consistency with Figure 4.1 and convergence in the de Sitter limit $\chi \to 0$. Green (light grey) indicates production of the present-day $\Omega_\Lambda$, red (medium grey) indicates excessive production up to $3\Omega_\Lambda/2$, and blue (dark grey) indicates insufficient production no less than $\Omega_\Lambda/2$.

4.3 Time-evolution of GEODE $w_{eff}(a)$ estimated from the stellar population. Both rapid (orange/light grey) and delayed (blue/dark grey) stellar collapse scenarios are shown. Best-fit lines (dashed) assuming the Planck linear ansatz near $a = 1$, shown for $\chi = 3 \times 10^{-2}$ only. Colors (greyshades) indicate distinct $\chi$ (c.f. Figure 4.4). Note breakdown of the dark fluid (linear) approximation at $a = 0.9$.

4.4 GEODE constraint region (orange line/horizontal axis) estimated from the stellar population. Region is superimposed upon Planck constraints of the Dark Energy equation of state evolution. Overlaid “chevrons” represent positive BH density constraints for $a > 0.9$, which is the lower-bound $a$ for validity of the Taylor expansion. Colors (greyshades) indicate distinct $\chi$ (c.f. Figure 4.3). Chevron boundaries have thickness $\Delta \chi(1)/3\Omega_\Lambda \sim 10^{-5}$. The intersection of these vertical bands with the horizontal band gives viable $w_0 - w_a$ for any particular $\chi$ (c.f. Figure 4.5). Note that Planck+WBR is the Planck+WL+BAO/RSD combined constraint, where WL is weak lensing and BAO/RSD is baryon acoustic oscillations/redshift space distortions. BSH is a combined constraint from BAO, Type Ia Supernovae, and local measurements of the Hubble constant.
4.5 GEODE constraint region (orange/light grey) estimated from the stellar population, vertical axis magnified by $10^5$ compared to Figure 4.4. The region is superimposed upon Planck constraints of the Dark Energy equation of state evolution. Stellar collapse model indicated via label. Acronyms are defined in the caption of Figure 4.4. Planck $2\sigma$ contours have been removed for clarity. For illustration, permissible region for $\chi \equiv 3 \times 10^{-2}$ shown, with thickness $10^3$ smaller than indicated. The permissible region is the overlap of the orange band and the vertical orange line.

5.1 The fixed mass $M \equiv 1$ Nolan interior solution, joined to McVittie’s asymptotically arbitrary Robertson-Walker (RW) spacetime. The physical radius parameter $w_0/a$ is given on the horizontal axis. Physical distance $r/a$ from the spherical origin is given on the vertical axis in units of $w_0$, where $a$ is the RW scale factor. The object is embedded in a radiation-dominated FRW universe at $z = 5000$, with cosmological parameters taken from Planck. The top panels give energy density (top left) and pressure (top right) in units of the critical density scaled by $10^2$ for visualization. The radial boundary of the object is clearly visible in the energy density plot as a diagonal line. (Bottom) The equation of state $w \equiv P/\rho$ of these solutions. Physically significant transitions in $w$ are indicated by overlaid contours. Note that the exterior McVittie spacetime approaches $w = 1/3$ as $r \to \infty$, consistent with radiation-domination. Both solutions exhibit regions of $w_0$ space where the pressure diverges and changes sign at some fixed radius. The resulting “interiors” of both solutions are GEODEs.

5.2 The fixed mass $M \equiv 1$ Nolan interior solution, joined to McVittie’s asymptotically arbitrary Robertson-Walker (RW) spacetime. The physical radius parameter $w_0/a$ is given on the horizontal axis. Physical distance $r/a$ from the spherical origin is given on the vertical axis in units of $w_0$, where $a$ is the RW scale factor. The object is embedded in a dust-dominated FRW universe at $z = 20$, with cosmological parameters taken from Planck. The top panels give energy density (top left) and pressure (top right) in units of the critical density scaled by $10^2$ for visualization. The radial boundary of the object is clearly visible in the energy density plot as a diagonal line. (Bottom) The equation of state $w \equiv P/\rho$ of these solutions. Physically significant transitions in $w$ are indicated by overlaid contours. Note that the exterior McVittie spacetime approaches $w = 0$ as $r \to \infty$, consistent with matter-domination. Both solutions exhibit regions of $w_0$ space where the pressure diverges and changes sign at some fixed radius. The resulting “interiors” of both solutions are GEODEs.
5.3 The fixed mass $M \equiv 1$ Nolan interior solution, joined to McVittie’s asymptotically arbitrary Robertson-Walker (RW) spacetime. The physical radius parameter $w_0/a$ is given on the horizontal axis. Physical distance $r/a$ from the spherical origin is given on the vertical axis in units of $w_0$, where $a$ is the RW scale factor. The object is embedded in a pure dark energy-dominated FRW universe at $z = 0$, with cosmological parameters taken from Planck. The top panels give energy density (top left) and pressure (top right) in units of the critical density scaled by $10^2$ for visualization. The radial boundary of the object is clearly visible in the energy density plot as a diagonal line. (Bottom) The equation of state $w \equiv P/\rho$ of these solutions. Physically significant transitions in $w$ are indicated by overlaid contours. Note that the exterior McVittie spacetime approaches $w = -1$ as $r \to \infty$, consistent with pure dark energy-domination. In contrast to radiation and matter-domination, the solutions exhibit regions of negative pressure, without transition through a $P \to \infty$ region. Portions of the “interiors” of both solutions are GEODEs. (In grayscale, below the diagonal, the diagram transitions from positive to negative in the right to left direction only. From top to bottom, for $w_0 > r$, pressure switches sign from negative to positive.)
CHAPTER 1
INTRODUCTION

“This quest may be attempted by the weak
with as much hope as the strong. Yet such is oft
the course of deeds that move the wheels of the
world: small hands do them because they
must, while the eyes of the great are
elsewhere.”

Elrond, Eärendil’s son

Cosmology is the study of the history of the universe. Like the study of the history of anything, it provides the context and identity necessary to orient future-directed endeavors. A coherent understanding of the history of our own observable universe necessarily touches upon every energy scale, including those where our current models of gravity and matter falter. It is thus the hope of many that the insights necessary for reconciliation of gravity and quantum mechanics might be found in Cosmology.

At present, an international effort across decades has efficiently compressed ignorance of the Universe’s history into two unknowns. One unknown is Dark Matter, a material which only interacts through gravity. The observational evidence for its existence is diverse and substantial, and we know it comprises about 25% of all energy within the observable Universe. The other unknown is Dark Energy. It is a seemingly diffuse, but all-permeating, substance. Its properties are strange; we’ve never observed anything similar in a terrestrial experiment. At a hefty 70%, it dominates the energy of the observable Universe today. Despite vigorous research and many ingenious proposals, we still have no idea what it is.

At the dawn of the 21st century, when asked how he felt that, “some say . . . the 20th century was the century of physics, we are now entering the century of biology,” Stephen Hawking made a very curious response:

“I think the next century will be the century of complexity. We have already discovered the basic laws that govern matter and understand all the normal situations. We don’t know how the laws fit together, and what happens under extreme conditions. But I expect we will find a complete unified theory sometime this century. That will be the end of basic theory. But there is no limit to the complexity that we can build using those basic laws.” [Chui 2000]

As exhibited by [Kuhn 1970], this statement runs counter to the observed growth of our discipline, which is conceptually reductive by definition. It suffices to say that history provides numerous instances of this sort of thinking. In our cosmological context, a particularly sharp historical example is the luminiferous aether. The aether, like Dark Energy, was also an all-permeating substance with extraordinary properties. The experimental history of the aether hypothesis is enumerated at length by [Panofsky and Phillips 2005], which we may summarize as follows:

1
1. Researchers incorrectly attempted to transfer knowledge from an established realm (Newtonian fluids) into an emerging realm (Electrodynamics).

2. Experimental evidence against this transfer steadily mounted.

3. This did not initially lead to rejection of the transfer itself. Instead, each failure was interpreted as the need to add increasingly intricate and ornate features to try to salvage the transfer.

Researchers did not doubt the existence of the luminiferous aether. Experimental results were interpreted as demonstrations of its “complexity,” which resulted in a proliferation of unsatisfactory models.

The aether gradually became obsolete after Einstein and Marić, as recently clarified by Gagnon [2016 and references therein], succinctly explained the accumulated observational evidence by discarding the notion of absolute time. The scientific lesson from this episode is not that “there is no luminiferous aether.” It is, instead, the following

**Principle 1** appeals to complexity, or a proliferation of unsatisfactory models of a phenomenon, signal a fundamentally flawed assumption near the foundations of the contemporary theoretical framework.

In this light, there are many parallels between theoretical frameworks and production computer software. Both are constructed with some goal in mind at the outset. Both encounter unexpected realities during development and use, which force significant revisions. Both are very much living systems, developed collaboratively by humans. When a collaborative research endeavor becomes mired, and practitioners default to increasingly intricate approaches, the root cause is often a “bug.” Thus, small, incremental, additions to the current state-of-the-art may obscure or worsen the difficulty. What we really need to do is “debug” the framework and write a “patch.” In the following pages, we will apply these ideas to the Dark Energy problem of cosmology.

### 1.1 Layout

This work represents a synthesis of results from small-scale and large-scale gravitation theory. The required background for both of these settings is General Relativity (GR). We review the relevant aspects of GR and the standard cosmological framework in §1.3. The text is then divided into three chapters. Each chapter begins with further introductory and pedagogical material, designed to clarify the following technical expositions. We believe this is a more “user-friendly” presentation of the work, as it may be studied linearly.

### 1.2 Conventions

Unless otherwise indicated, throughout the text, we employ Misner et al. [1973] geometrodynamic units

\[
G \equiv 1 \equiv c
\]  

(1.1)
where $G$ is Newton’s constant and $c$ is the speed of light in vacuum. In the specification of actions, we also set $\hbar \equiv 1$.

Where there is overlap, we adopt the symbols and notation of Dodelson [2003]. Greek indices will include time, while Latin indices will be spatial. We will use Einstein summation notation, where a repeated index occurring in raised and lowered positions encodes a sum

$$T^\mu_\mu \equiv T^0_0 + \sum_{j\in\{x,y,z\}} T^j_j. \quad (1.2)$$

### 1.3 The standard cosmology

We will be debugging an existing framework, finding the bug, and patching the framework. We must thus introduce the existing framework first, in the customary way that obscures the bug. We will then describe the Dark Energy problem.

#### 1.3.1 Differential geometry in kinematic context

Every physically plausible gravitation theory at present is a geometric theory: kinematic concepts are modeled as objects defined on a differential manifold $M$. The essential idea is that $M$ is a pointset, onto which various additional structures are defined so that one can generalize the ideas of calculus. The construction is outlined as follows

1. Curves $\gamma : M \leftarrow [a, b] \subset \mathbb{R}$ can be composed with
2. Coordinates $x^\mu : \mathbb{R}^n \leftarrow M$ to produce
3. Ordinary real-valued functions $x^\nu \circ \gamma : \mathbb{R}^n \leftarrow X \leftarrow [a, b]$. Derivatives of these functions define
4. Linear objects called vectors attached to each point $p \in M$: $v_p \in T_pM$ and
5. Linear objects called covectors $w^*_p \in T^*_pM$, functions which take vectors to real numbers $w^*_p(v_p) : \mathbb{R} \leftarrow T_pM$

Metric theories of gravity posit at least one metric, a geometric object which observers can use to measure distances and times

$$g = g_{\mu\nu} dx^\mu \otimes dx^\nu. \quad (1.3)$$

Here $dx^\mu \in T^*M$ provide a set of 4 covector fields, a basis set of covectors at each point, which give a local notion of “where and when.” This is called a coframe field. Using the duality property, there are likewise 4 paired vector fields $e_\mu$, called the frame field. Changes in the frame field encode forces in the following
sense. Consider an observer’s spacetime trajectory $u$ along a curve $\gamma$

$$u \equiv \frac{dx}{d\tau} \equiv \frac{dx^\alpha(\tau)}{d\tau} e_\alpha(x^\beta(\tau))$$

(1.4)

where $x^\alpha(\tau)$ are some coordinate values in this representation along the curve $\gamma$. The observer’s acceleration is defined as the derivative along this curve, as parameterized by $\tau$. The product and chain rules give

$$\frac{d^2 x}{d\tau^2} = \frac{d^2 x^\alpha}{d\tau^2} e_\alpha + \frac{dx^\alpha}{d\tau} \frac{dx^\beta}{d\tau} \partial_\beta e_\alpha.$$ 

(1.5)

The term with $\partial_\beta e_\alpha$ can be seen to act like a non-inertial force. This derivative of the frame will, in general, be some new linear combination of $e_\alpha$

$$\partial_\beta e_\alpha \equiv \Gamma^\rho_{\alpha\beta} e_\rho.$$ 

(1.6)

The choice of a coframe basis is completely arbitrary, and may be adapted as convenient. For example, consider again our observer along $u^\alpha(\tau)$. If the observer only responds to gravity, then there is a privileged coframe basis $dy^\mu$ for this particular observer. We may identify the timelike covector of the coframe basis as dual to the observer’s worldline. Then it is always possible to choose a set of 3 orthonormal spacelike covectors so that

$$g\bigg|_{u^\alpha(\tau)} = \eta_{\mu\alpha} dy^\mu \otimes dy^\alpha$$ 

(1.7)

$$\Gamma\bigg|_{u^\alpha(\tau)} = 0$$

(1.8)

Thus, such an observer is in gravitational freefall: they perceive a special relativistic environment of flat space without accelerations. This property of the differential geometric framework was Einstein’s chief motivation for the geometric approach.

Physical measurements are made with a metric. Structurally, a metric can be regarded as a scaled angle cosine between two infinitesimal displacements $v, u \in TM$

$$g(u,v) : \mathbb{R} \leftarrow T^*M \times T^*M.$$ 

(1.9)

For example, if an observer moves along a trajectory $u^\alpha(\tau)$ and encounters a particle with 4-momentum $p^\mu$, the energy of the particle as perceived by this observer is

$$E = -g_{\mu\alpha} p^\mu u^\alpha.$$ 

(1.10)

This follows immediately from the definition of the 4-momentum and the coframe local to $u^\alpha$. Regarded as a function of a single infinitesimal displacement $v \in TM$, a metric gives a privileged map between $T^*M$ and
Thus, given any vector \( v \), we can produce a unique covector \( v^\ast \). It follows from the angle cosine interpretation that

\[
v^\ast [v] = g(\ , v)[v] = g(v, v),
\]

is the squared magnitude of \( v \). Note that this uniqueness is invariant to the choice of coframe basis \( dx^\mu \), because the map \( g \) exists independently of the basis chosen to represent it. In contrast, the canonical duality between \( TM \) and \( T^*M \) without a metric is entirely dependent on the choice of coframe basis. For a specific example, consider the local coframe above. By the canonical duality,

\[
\exists! e_\mu \in TM \text{ st } dy^\nu [e_\mu] = \delta^\nu_\mu
\]

but \( dx^\nu [e_\mu] \) need not be Kronecker in general.

As in the case of the energy, physically meaningful quantities must be free of reference to the arbitrary choice of coframe \( dx^\mu \) (but, of course, not the observer along \( u^\alpha \)!). Consider for example

\[
S \equiv \int_{\tau_0}^{\tau_0+\Delta\tau} \sqrt{g(u^\alpha(\tau), u^\beta(\tau))} \, d\tau.
\]

This is just the arc-length along the trajectory \( u^\alpha \). If this trajectory is a geodesic, we might as well compute this quantity in a coframe adapted to the trajectory. Then \( S = \Delta\tau \) is revealed to be the time elapsed on the free-fall observers’ own clock, since the integrand becomes unity. Because we observe our universe through massless phenomena, such as light and gravitational radiation, of particular interest is the case where \( u^\alpha \) is light-like, but non-zero

\[
g(u, u) \equiv 0.
\]

A photon’s 4-momentum, for example, exhibits this behaviour. This equation constrains the time elapsed and distance traveled by a photon, as perceived by non-comoving observers, such as humans.

### 1.3.2 Einstein’s Equations

We have established that a metric theory of kinematics automatically encodes gravitational free-fall. Further, it reduces to kinematics without gravity when the metric is trivial. Einstein postulated that all mechanical phenomena ever are encoded on a 4 dimensional manifold with a unique symmetric metric \( g_{\mu\nu} \). The geometry of this manifold, encoded in certain combinations of derivatives of the metric called Riemann’s tensor
\( R^\alpha_{\mu\nu\rho} \) is determined through Einstein’s equations

\[
R^\alpha_{\mu\nu\rho} - \frac{1}{2} R^\alpha_{\rho\sigma\nu} g^\rho\sigma g_{\mu\nu} \equiv G_{\mu\nu} = 8\pi G T_{\mu\nu}. \tag{1.16}
\]

The new object appearing on the right-hand side is the stress-energy tensor \( T^{\mu\nu} \) (stress). The stress can be understood [e.g. Weinberg [1972, §2.8]] as the 4-momentum 4-current density, in direct analogy to the electromagnetic charge 4-current density \( J^\mu \). In this way, 4-momentum flux deforms spacetime.

### 1.3.3 Friedmann’s equations

We would like to construct an analytic model suitable for describing the universe on the largest scales: a cosmological model. One starts with the following

**Observation 1** *The contents of the universe appear isotropic and homogeneous on scales beyond 300Mpc.*

This suggests that a “coarse grained” model, where the stress has no position dependence, might be fruitful. This would require a non-trivial metric ansatz without position dependence. Such a metric is uniquely determined to be a Robertson-Walker (RW) metric

\[
ds^2 = -dt^2 + a^2(t) \gamma_{ij} dx^i dx^j \tag{1.17}
\]

where \( \gamma_{ij} \) is the metric of a 3-space of constant curvature \( \kappa \). In practice, it is found that \( \gamma_{ij} \equiv \delta_{ij} \) is an excellent approximation to reality. Substitution of the RW ansatz into Einstein’s equations produces Friedmann’s equations

\[
\left( \frac{\dot{a}}{a} \right)^2 + \frac{\kappa}{a^2} = -\frac{8\pi G}{3} T^0_0 \tag{1.18}
\]

\[
\frac{\ddot{a}}{a} = -\frac{4\pi G}{3} \left( -T^0_0 + T^j_j \right). \tag{1.19}
\]

Consider an observer at rest with respect to the coframe employed. By definition of the stress and the measured energy of a single particle given in Eqn. (1.10), they perceive an energy density

\[
\rho \equiv T^{\mu\nu} \delta_{\mu}^0 \delta_{\nu}^0. \tag{1.20}
\]

Since \( g_{00} = -1 \) in this coframe, it follows that

\[
T^0_0 = -\dot{\rho}(t). \tag{1.21}
\]
Since the RW model is manifestly isotropic and homogeneous in the coframe employed, there can be no preferred 4-momentum fluxes in this coframe. This forces the remaining stress components to be

\[ T^i_j \equiv \delta^i_j \mathcal{P}(t) \] (1.22)
\[ T^i_0 = 0. \] (1.23)

### 1.3.4 Cosmography

The function \( a(t) \) plays a pivotal role in the interpretation of astrophysical data. These data come to us through particles, photons, and gravitational radiation. Our first step in interpreting our observations is to understand how such phenomena propagate in a RW universe.

**Redshift**

Let \( p^\mu(\tau_0) \) by a 4-momentum at emission. We may determine the 4-momentum at some later time directly from Fermat’s principle

\[ \delta \int_{\tau_0}^{\tau} \sqrt{-g(p, p)} \, d\tau' \equiv 0. \] (1.24)

This is, of course, just the action principle with the infinitesimal arc length as the Lagrange function. We can eliminate the square-root by noting that

\[ \delta (L^2) = 2L\delta L. \] (1.25)

So, \( \delta (L^2) = 0 \implies \delta L = 0 \) as long as \( L \) contains no roots in the interval \([\tau_0, \tau]\). Note that, for light-like particles, the problem as posed becomes undefined. A light-like trajectory is an edge case, since variations in \( x(\tau) \) and \( t(\tau) \) must maintain the causal property of the path because we are doing a classical variation. To side-step this subtlety, we investigate the massive case, because \( L \equiv -g(p, p) = m^2 \) which guarantees the no-root property. Further, an always time-like trajectory allows independent infinitesimal variations in \( x(\tau) \) and \( t(\tau) \). We align the spatial coordinates so that the \( x \) direction lies along the particle’s spatial trajectory. Then

\[ m^2 = -g(p, p) = -\left(p^t\right)^2 + a(t)^2 (p^x)^2 \equiv L^2 \] (1.26)

An infinitesimal variation in trajectory corresponds to

\[ t(\tau) \to t(\tau) + \delta t \quad x(\tau) \to x(\tau) + \delta x, \] (1.27)
and induces a corresponding change to the arc length

\[ S(t + \delta t, x + \delta x) \equiv S(t, x) + \delta S = S(t, x) + \int \delta L \, d\tau'. \]  

(1.28)

Substitution of the varied trajectories into \( L \) gives

\[ \delta L^2 = -\left[ \frac{d}{d\tau} (t + \delta t) \right] + a(t + \delta t)^2 \left[ \frac{d}{d\tau} (x + \delta x) \right]^2. \]  

(1.29)

Let overset dot denote differentiation with respect to the parameter \( \tau \). Expansion to first order in small-quantities gives

\[ \delta L^2 = -\dot{t}^2 - 2\dot{t}\dot{\delta t} + \left( a^2(t) + 2a \frac{da}{dt} \delta t \right) \left[ \dot{x}^2 + 2i\dot{x}\delta x \right] + O(\delta^2) \]  

(1.30)

Since the variations are independent, we may focus on the energy only. Returning to the entire action and using \( \delta L^2 \) in place of \( \delta L \) we find

\[ S(t, x) + \delta S = \int_{\tau_0}^{\tau} \left( -\dot{t}^2 + a^2 \dot{x}^2 \right) \, d\tau' + \int_{\tau_0}^{\tau} 2a \frac{da}{dt} \dot{x}^2 \delta t - 2i\dot{\delta t} \, d\tau' \]  

(1.31)

\[ \delta S = \int_{\tau_0}^{\tau} 2a \frac{da}{dt} \dot{x}^2 \delta t - 2i\dot{\delta t} \, d\tau' \]  

(1.32)

Using the product rule, the resulting boundary term dies, leaving

\[ \delta S = \int_{\tau_0}^{\tau} 2 \left[ a \frac{da}{dt} \dot{x}^2 + i \right] \delta t \, d\tau', \]  

(1.33)

from which we conclude that

\[ -a \frac{da}{dt} \dot{x}^2 = \dot{i}. \]  

(1.34)

To solve this equation, use \( g(p, p) = -m^2 \) to replace \( \dot{x} \) with \( i \)

\[ \dot{i} = \frac{1}{a} \frac{da}{dt} \left[ \dot{i}^2 - m^2 \right], \]  

(1.35)

and then use a chain-rule to rewrite one of the time derivatives on the left to get a separable form

\[ \frac{-i \, d\dot{i}}{\dot{i}^2 - m^2} = \frac{da}{a}. \]  

(1.36)
The remarkable solution of this equation,

\[ i = E = \sqrt{\left(\frac{a_{\text{emit}}}{a}\right)^2 \left(E_0^2 - m^2\right) + m^2}, \]

has been repeatedly confirmed for almost a century by the following observation:


\textit{Observation 2} photons from cosmologically distant sources have spectra very similar to near sources, but translated toward lower energy.

This follows at once from the \( m \to 0 \) limit, where

\[ E = E_0 \frac{a_{\text{emit}}}{a}. \]

Switching to wavelength with the Planck relation gives

\[ \lambda = \lambda_0 \frac{a}{a_{\text{emit}}}, \]

and subtracting and then dividing by \( \lambda_0 \) gives the expected fractional shift in spectra

\[ z(a, a_{\text{emit}}) \equiv \frac{a}{a_{\text{emit}}} - 1 \]

or redshift. Since today is typically defined to be \( a \equiv 1 \), astronomers usually imply the quantity

\[ z(a_{\text{emit}}) = \frac{1}{a_{\text{emit}}} - 1. \]

The practical significance of this phenomena becomes clear if the time-evolution of \( a(t) \) is monotonic. In this case, redshift gives a well-defined notion of “when”

\[ a_{\text{emit}} = \frac{1}{1 + \frac{1}{z}}. \]

Luminosity and redshift

Now that we have established a notion of “when,” we require a notion of “where.” There are many ways to do this, but we will focus on the observed brightness of distant objects. Luminosity is energy per unit time (power) in the context of emissive phenomena. The flux of emitted objects at an observer is defined as the luminosity divided by the area of a 2-sphere of radius \( r \)

\[ F_e \equiv \frac{L_e}{4\pi r^2}, \]

centered on the source and grazing the observer. Here \( L_e \) is the intrinsic luminosity of the object. It is measured by observers that are
**Definition 1** *(spacetime proximal)* located in space relative to a phenomenon, and taking data over timescales, such that gravitational effects can be safely ignored.

Assuming a stable, monochromatic source for simplicity, it can be decomposed as

\[ L_e \equiv \frac{NE_e}{\Delta t_e}. \tag{1.44} \]

Here \( N \) is the number of objects, each with energy \( E_e \), emitted per unit self-time \( \Delta \tau \). Since our fiducial observers are spacetime proximal, we have that \( \Delta \tau \) is equal to coordinate time \( \Delta t_e \) plus corrections of order \( \beta^2 \).

The above result can be generalized to a flat RW universe model. We care about the flux at the receiver today

\[ F = \frac{L_0}{4\pi(a_0r)^2} \tag{1.45} \]

where \( L_0 \) is the measured luminosity at reception. The relevant radius is the physical length (i.e. integrated spacelike arc) from the source to the receiver, which is just the coordinate length times the scale factor

\[ \int_0^r ds = \int_{t=0}^t a(t_0)dr' = a_0 \int_0^r dr' = a_0r. \tag{1.46} \]

Since the scale factor at the present day is usually defined to be one

\[ a_0 \equiv 1, \tag{1.47} \]

the physical distance is the coordinate distance. The measured luminosity \( L_0 \), however, will be diminished from the intrinsic luminosity \( L_e \): the spatial separation between source and receiver increases between emission events *and* the local 4-momentum of the objects is altered consistent with §1.3.4. Let us focus on photons. In our decomposition of the luminosity, the redshift alters the energy observed today by

\[ E = a_eE_e. \tag{1.48} \]

To capture the effect of expansion between emissions, we will determine how the expansion affects the measured time delay \( \Delta t_r \) between absorptions. We assume that the source and the receiver are at rest with respect to the comoving grid, so that \( r \) is fixed for all time. Given that \( g(p, p) = 0 \), we have for any massless emissive phenomenon \( dt = a \, dr \). For emission at time \( t_e \) and reception at time \( t_r \), we have

\[ r = \int_{t_e}^{t_r} \frac{dt}{a(t)}. \tag{1.49} \]
This expression also holds at the next emission $\Delta t_e$ later

$$ r = \int_{t_e}^{t_e + \Delta t_e} \frac{dt}{a(t)}. \quad (1.50) $$

Subtracting these two expressions and rearranging gives

$$ \int_{t_e}^{t_e + \Delta t_e} \frac{dt}{a(t)} = \int_{t_e}^{t_e + \Delta t_r} \frac{dt}{a(t)}. \quad (1.51) $$

If $\Delta t$ is small compared to the timescales on which $a(t)$ changes at both emission and reception events, then the integrands are essentially constant and we find that

$$ \frac{\Delta t_e}{a_e} = \frac{\Delta t_r}{a_r}, \quad (1.52) $$

and the measured flux becomes

$$ F = \frac{L_e}{4 \pi r^2} \frac{a_e^2}{a^2} = \frac{L_e}{4 \pi r^2 (1 + z)^2}. \quad (1.53) $$

Note that this expression can be inverted to find $r$, giving a notion of “where.” In practice, this expression may be expressed entirely in terms of observables

$$ F(z) = \frac{L_e}{4 \pi (1 + z)^2} \left( \int_0^z H^{-1} dz' \right)^{-2} \quad (1.55) $$

$$ H = \frac{1}{a} \frac{da}{dt}. \quad (1.56) $$

In this way, if we have measured redshifts $\{z_i\}$ and fluxes $\{F_i\}$ from many objects of the same known intrinsic luminosity $L_e$, we may reconstruct the expansion history

$$ H(z) = \sqrt{\frac{4 \pi F(z)}{L_e}} \left( 1 + z \right)^2 \left( 1 + \frac{(1 + z)}{2} \frac{d}{dz} \ln F(z) \right)^{-1}. \quad (1.57) $$

### 1.4 The Dark Energy Problem

We may now present the Dark Energy Problem. It is known from astronomy that Type Ia Supernovae (SNIa) have well-defined intrinsic luminosities and spectra. Consistent spectra across events permits one to perform redshift measurements. Well-defined intrinsic luminosities, combined with measured fluxes, then allows one to reconstruct the expansion history. The Dark Energy problem of cosmology comes from the following

**Observation 3** The magnitude (i.e. flux)-redshift relation of SNIa is inconsistent with a mostly matter filled
universe. It is instead consistent with a universe mostly filled by an isotropic substance $P \propto -\rho$ which does not dilute as the universe expands.

The inconsistency is an accelerated growth rate, and this substance is called Dark Energy. To understand when this can come about, we can combine the two Friedman equations

$$\frac{1}{a^2} \frac{d^2 a}{dt^2} = -\frac{4\pi G}{3} (\rho + 3P).$$

(1.58)

Accelerated growth can only occur if $P < -\rho/3$. The simplest modification to GR which meets these criteria is to introduce a term $\Lambda g_{\mu\nu}$. This term acts as a fixed energy density of vacuum. Though consistent with very many observations to date (see however Riess et al. [2016]), it is a problem for the following reasons

1. **The coincidence problem** The DE density is $\sim 2$ times the matter density today. Since the energy density has decreased by at least $10^{24}$ since the Big Bang, this requires a fine-tuning of $\Lambda$ to at least $\sim 20$ decimal places.

2. **The vacuum catastrophe** Quantum field theory predicts incorrect, and possibly even inconsistent, vacuum energy densities.

   - *(Majority viewpoint)* Quantum field theory, as an effective field theory with Planck scale cutoff, predicts an egregious DE density $\sim 10^{120}$ times larger than the measured $\Lambda$ [Weinberg, 1989]
   - *(Minority viewpoint)* Quantum field theory, as an effective field theory, contributes exactly 0 energy density to the vacuum. [Brodsky and Shrock, 2011]

The Dark Energy problem has led to a proliferation of speculative models over the past two decades. All of them introduce additional complexity in the form of new fields, new parameters, or couplings to other poorly understood phenomena like Dark Matter. None of them fit the data more efficiently than the Cosmological Constant. More troubling than an overabundance of models, however, is that the Cosmological Constant has been argued sufficient via the “strong anthropic principle.” The “weak anthropic principle” was used successfully by Hoyle to predict a Carbon nuclear resonance: there would not be enough carbon for our implementation of life to exist without one. The “strong anthropic principle,” however, can be summarized as, “If the universe were slightly different, we would not be here to observe it.” This “principle” is not scientific because it is *conceptually obstructive* instead of *conceptually reductive*. It is clear that the Dark Energy problem has come to meet all the criteria required to trigger application of Principle [1] it is time to debug the framework presented in §1.3
CHAPTER 2
FRIEDMANN COSMOLOGY FROM CONSISTENT APPLICATION OF
THE ACTION PRINCIPLE

“... don’t EVER make the mistake that you can
design something better than what you get
from ruthless massively parallel
trial-and-error with a feedback cycle. That’s
giving your intelligence much too much
credit.”

Linus Torvalds, LKML, 30.11.2001

In the previous chapter, we reviewed the existing cosmographical (how to measure) and cosmological
(how to predict) framework and defined the Dark Energy problem. Consistent with our discussion, we antic-
ipate that the DE problem indicates a bug somewhere within this framework. Based on this anticipation, we
will now re-derive Friedmann’s equations with greater care. Instead of working from Einstein’s equations,
we will go one step more primitive. We return directly to the action principle, as we did when deriving the
photon redshift. The action principle, as an integral statement over all events, is uniquely well-positioned
to make definitive large-scale statements. This is useful, since cosmologies are the largest scale models
possible.

2.1 Preliminaries

The foundational modern cosmological treatments by [Peebles 1980], [Bardeen 1980], and [Kodama and
Sasaki 1984] have become textbook standard material [e.g. Dodelson 2003, Hu 2004]. All these treat-
ments assume a pre-existing RW geometry and define tensorial objects on this background. While the scale
factor remains determined dynamically, the assumption of prior geometry runs counter to the spirit of GR.
In particular, this approach is vulnerable to the criticisms of [Ellis 1984, §3.1], who laments that produc-
ing a position-independent metric model from an obviously position-dependent metric (only known on a
backward lightcone) is not uniquely defined. Our pragmatic approach is to make the following

Assumption 1 There is no prior geometry. There exists a well-defined metric \( g \) at all spacetime events on
the manifold \( M \). This metric respects Einstein’s equations.

In other words, we interpret the metric representation

\[
g_{\mu \nu}(\eta, x) \equiv \lim_{N \to \infty} a(\eta)^2 \left[ \eta_{\mu \nu} + \sum_{n=1}^{N} \epsilon^n h^{(n)}_{\mu \nu}(\eta, x) \right] \quad \epsilon \ll 1 \tag{2.1}
\]
as nothing more than a series representation for the actual \( g_{\mu \nu}(\eta, x) \) on some open submanifold \( \mathcal{U} \subset \mathcal{M} \), which exists by Assumption [1]. There are, of course, very many other series representations. Our particular choice of this form of the series is motivated by simplicity, observational agreement and convention. A poor choice would not converge in the sense that its predictions would disagree with observation at low orders.

Our purpose as cosmological model builders then is not to understand how much our model metric deviates from the actual metric. We do not need to know the actual metric, because we only fit our presumed model Eqn. (2.1) to observation. We codify our imperfect knowledge of the metric into a truncation of Eqn. (2.1), which we assume converges to the actual metric tensor. We will often suppress notation that we have restricted attention to the open submanifold \( \mathcal{U} \subset \mathcal{M} \).

### 2.1.1 The action principle for gravity

There is no principle more general, more succinct, and more experimentally verified than the action principle. Let \( \Phi(\eta, x) \) represent all other fields besides the metric. In the context of gravity, the action principle demands that demands that

\[
\delta \left\{ S_G \left[ g_{\mu \nu} \right] + S_M \left[ g_{\mu \nu}, \Phi(\eta, x) \right] \right\}_{\Phi} = 0 \quad (2.2)
\]

where \( S_G \) is a gravitational action, \( S_M \) is a matter action, and \( \Phi \) is held fixed for the variation. By the definition given in [Weinberg 1972, §12.2], the symmetric, rank-(2,0), stress tensor \( T^{\mu \nu} \) enters Einstein’s equations through

\[
\delta S_M \bigg|_{\Phi} = \frac{1}{2} \int_M T^{\mu \nu} \delta g_{\mu \nu} \sqrt{-g} \, d^4x. \quad (2.3)
\]

Note that we have flipped indices, relative to Weinberg’s definition, and so incur a minus sign through the metric variation

\[
\delta \left( \delta^\rho_{\rho} \right) = 0 = \delta \left( g^{\rho \mu} g_{\rho \nu} \right). \quad (2.4)
\]

If we acquire the matter action from a Lagrange density, then

\[
\delta S_M = \int_M \frac{\delta}{\delta g_{\mu \nu}} \left( \mathcal{L}_M \sqrt{-g} \right) \, d^4x \quad (2.5)
\]

\[
= \int_M \left[ \frac{1}{\sqrt{-g}} \delta \left( \mathcal{L}_M \sqrt{-g} \right) \right] \delta g_{\mu \nu} \sqrt{-g} \, d^4x \quad (2.6)
\]

and we see that, under the integral, one may identify

\[
T^{\mu \nu} = \frac{2}{\sqrt{-g}} \frac{\delta}{\delta g_{\mu \nu}} \left\{ \mathcal{L}_M \left[ \Phi(\eta, x), g_{\mu \nu} \right] \sqrt{-g} \right\}. \quad (2.7)
\]
The stress tensor components are, therefore, a mixture of the metric and the non-gravitational contributions \( \Phi(\eta, x) \). Note carefully that prescription of any particular gravitational model has no explicit effect on \( \Phi(\eta, x) \). Of course, the \( \Phi(\eta, x) \) are implicitly constrained by the eventual equations of motion. Any premature constraint, before equations of motion consistent with the action principle are determined, risks introducing subtle inconsistencies into the field equations.

2.2 The marble and the bowl: constrained classical variations

Since we will be using the action principle classically, it is helpful to pedagogically review this use in some detail. Our intent is not to bore or insult the reader, but to clearly and methodically lay the appropriate foundation for our subsequent analysis. Our discussion in this section is heavily based upon the classic text “Variational Principles of Classical Mechanics” by Lanczos [2012, reprinted].

The hallmark of the action principle is the temporary introduction of non-physical degrees of freedom, the variations, in order to glean information about the actual motion of the system. For all \( N \) degrees of freedom \( q_i(\eta) \) in some physical model, one temporarily introduces an additional degree of freedom called the variation \( \delta q_i(\eta) \). Requiring that the action be independent, to first order, of each variation gives \( N \) equations of motion for the \( q_i(\eta) \). Mathematically, this amounts to finding an implicit solution to an extremization problem over a space of functions.

To illustrate these concepts physically, consider a marble of mass \( m \) in a hemisphere bowl of radius \( R \) resting on table. Its Lagrangian is

\[
L = \frac{1}{2}m\ddot{x}^2 - U \\
= \frac{1}{2}m\ddot{x}^2 + mgR \cos \theta(\eta)
\]

where we have neglected spin of the marble. By symmetry, we already know that if there is no initial azimuthal velocity, the marble will never acquire one because \( U \) has no explicit azimuthal dependence. The solution of this problem, for small displacements, is simple harmonic motion about the bottom of the bowl.

Now, suppose that we lacked the advantage of an intuitive physical picture for this situation. We somehow identify that we have two degrees of freedom and write the Lagrangian of the marble as

\[
L = \frac{1}{2}m \left( \dot{x}^2 + \dot{y}^2 \right) - mgy.
\]

The Euler-Lagrange equations, in their generality, obscure the essential features of the classical variation procedure. Instead, we will proceed explicitly and introduce the variational degrees of freedom directly through the substitutions

\[
q_i(\eta) \rightarrow q_i(\eta) + \delta q_i(\eta) \quad \forall i \leq N
\]
into the action. We find that

\[ S + \delta S + \cdots \equiv \int \frac{1}{2} m \left[ (\dot{x} + \delta \dot{x})^2 + (\dot{y} + \delta \dot{y})^2 \right] - mg(y + \delta y) \, dt. \]  \hfill (2.12)

Expanding to first order gives

\[ \delta S = \int (m \ddot{x} \delta \dot{x} + m \ddot{y} \delta \dot{y} - m y \delta y) \, dt. \]  \hfill (2.13)

Integrating by parts to shift the time derivatives off of the variations gives

\[ \delta S = \int \left[ (-m \dddot{x}) \delta x - (m \dddot{y} + mg) \delta y \right] \, dt. \]  \hfill (2.14)

Note that we have discarded the boundary terms, since we are free to choose variations which vanish at the start and end times. If we demand that \( S \) be stable to the variations at first order, then we conclude that

\[ m \dddot{x} = 0 \quad m \dddot{y} = -mg \]  \hfill (2.15)

which is inertial motion in \( x \) and uniform field gravitational freefall in \( y \).

What went wrong? Lacking an intuitive physical picture for the situation, we forgot about the bowl. The bowl enforces a kinematic constraint

\[ x(\eta)^2 + y(\eta)^2 - R^2 = 0, \]  \hfill (2.16)

which reduces the number of physical degrees of freedom. This condition between the degrees of freedom is enforced by electromagnetic microphysics that we do not need to track since we are interested in the macrophysical behaviour of the system. If we substitute the variations into this constraint equation, we find that

\[ \left[ x(\eta) + \delta x(\eta) \right]^2 + \left[ y(\eta) + \delta y(\eta) \right]^2 - R^2 = 0. \]  \hfill (2.17)

If we use the original constraint between \( x \) and \( y \), we immediately conclude that

\[ 2\delta x^2 + 2\delta y^2 = O(\delta^2). \]  \hfill (2.18)

In other words, the kinematic constraint not only constrains the physical degrees of freedom, it also constrains the temporarily introduced, non-physical, variational degrees of freedom. In the case of the marble, we have that the variations must satisfy

\[ \delta x(\eta) = -\delta y(\eta). \]  \hfill (2.19)
If one repeats the variational procedure, enforcing the kinematic constraint in both the physical degrees of freedom and in the variations, one arrives at the correct equations of motion.

### 2.2.1 Application to a scalar field

Now consider a complex-valued scalar field \( a(\eta, x) \). For the sake of clarity, and to avoid mathematical questions of convergence, we will define our field on the spatial 3-torus \( a(\eta, x) : (S^1)^3 \times \mathbb{R} \to \mathbb{C} \). Let each torus radius be \( L \), then we may Fourier expand the field as

\[
a(\eta, x) = \lim_{N \to \infty} \sum_{\ell, m, n} |N| c_{\ell mn}(\eta) \exp \left( \frac{2\pi i \ell x}{L} \right) \exp \left( \frac{2\pi i m y}{L} \right) \exp \left( \frac{2\pi i n z}{L} \right).
\]  

(2.20)

Just to be clear: \( x, y, \) and \( z \) are not degrees of freedom, but coordinates. If we truncate the sum at some finite \( N \), then \( a(\eta, x) \) is entirely characterized by the \((2N + 1)^3\) physical degrees of freedom \( c_{\ell mn}(\eta) \). Compare this to the marble’s position, which had two degrees of freedom \( x(\eta) \) and \( y(\eta) \).

Suppose we wish to enforce a constraint on every truncated expansion of this field, namely that these partial sums are isotropic and homogeneous. In other words, the partial sums describing \( a \) can never have position dependence. This means that

\[
\sum_{(\ell, m, n) \neq 0} |N| c_{\ell mn}(\eta) \exp \left( \frac{2\pi i \ell x}{L} \right) \exp \left( \frac{2\pi i m y}{L} \right) \exp \left( \frac{2\pi i n z}{L} \right) = 0
\]

(2.21)

just like the marble’s

\[
x(\eta)^2 + y(\eta)^2 - R^2 = 0.
\]

(2.22)

In the case of our scalar field, however, orthogonality of the exponential basis forces all but one of the \( c_{\ell mn}(\eta) \) to vanish identically

\[
c_{\ell mn}(\eta) = 0 \quad \forall (\ell, m, n) \neq 0.
\]

(2.23)

Our constraint of isotropy and homogeneity has reduced a \((2N + 1)^3\) dimensional problem to a single degree of freedom required to describe the state of the scalar field \( a \). This is analogous to how the bowl reduced a two dimensional problem to a single degree of freedom required to describe the marble’s position. In more technical language, the Fourier machinery makes explicitly clear that an isotropic and homogeneous restriction applied to a scalar field is equivalent to an infinite dimensional, holonomic, constraint. The consequence is clear,

\[
a(\eta, x) = a(\eta).
\]

(2.24)
Now consider any variation $\delta a(\eta, x)$ to the scalar field $a$. Introduction and substitution of an additional $(2N + 1)^3$ variational degrees of freedom into the constraint Eqn. (2.21) gives

$$
\sum_{(\ell, m, n) \neq 0}^{[N]} [c_{\ell mn}(\eta) + \delta c_{\ell mn}(\eta)] \exp \left( \frac{2\pi i \ell x}{L} \right) \exp \left( \frac{2\pi imy}{L} \right) \exp \left( \frac{2\pi inz}{L} \right) = 0 \tag{2.25}
$$

As before, substitution of the constraint relation itself gives

$$
\sum_{(\ell, m, n) \neq 0}^{[N]} \delta c_{\ell mn}(\eta) \exp \left( \frac{2\pi i \ell x}{L} \right) \exp \left( \frac{2\pi imy}{L} \right) \exp \left( \frac{2\pi inz}{L} \right) = 0. \tag{2.26}
$$

Again, by orthogonality of the exponential basis, we conclude that

$$
\delta c_{\ell mn}(\eta) = 0 \quad \forall (\ell, m, n) \neq 0. \tag{2.27}
$$

Since $\delta a(\eta, x)$ is defined by its expansion

$$
\delta a(\eta, x) \equiv \sum_{(\ell, m, n)}^{[N]} \delta c_{\ell mn}(\eta) \exp \left( \frac{2\pi i \ell x}{L} \right) \exp \left( \frac{2\pi imy}{L} \right) \exp \left( \frac{2\pi inz}{L} \right) \tag{2.28}
$$

$$
= \delta c_{000}(\eta) \tag{2.29}
$$

we conclude that the non-physical variational degree of freedom must also be isotropic and homogeneous

$$
\delta a(\eta, x) = \delta a(\eta). \tag{2.30}
$$

If one fails to constrain the variation $\delta a$ to the field $a$ in this way, one will be solving the wrong problem: the marble will freely fall, instead of rolling around in the bowl.

Our illustration with a complex-valued scalar field was only to simplify the Fourier expansion. The arguments above immediately apply to a real-valued scalar field, since $\mathbb{R} \subset \mathbb{C}$. In the perturbed Friedmann-Robertson-Walker model, the real-valued $a(\eta)$ enters the model through the definition given in Eqn. (2.1) and thus can never acquire position dependence. By definition of the model, all position dependence is made to lie within the perturbation degrees of freedom $h_{\mu\nu}^{(n)}(\eta, x)$. These are not yet the physical degrees of freedom, since these are components in a tensorial object. For the purpose of constructing equations of motion, however, this very important subtlety can be addressed at the model builder’s leisure. Detailed discussion can be found in §3.3.

Contrary to intuition from Newtonian cosmology, there is no local scale factor. As should now be clear, the additional perturbation degrees of freedom $h_{\mu\nu}^{(n)}(\eta, x)$, which enter beyond zero order in the perturbative treatment, are logically and mathematically distinct from any variations $\delta a(\eta)$ to the scale factor $a(\eta)$. When one performs a variational analysis at linear order, one will of course consider variations $\delta h_{\mu\nu}^{(1)}(\eta, x)$ to the
perturbation degrees of freedom $h_{\mu\nu}^{(1)}(\eta, \mathbf{x})$, all with position dependence.

\section*{2.3 Zero-order Friedmann equations without stress-energy assumptions}

The gravitational action in GR is given by

$$S_G \equiv \frac{1}{16\pi G} \int \mathcal{U} R \sqrt{-g} \, d^4x. \quad (2.31)$$

To determine the most appropriate definition of $\mathcal{U}$, we note ample observational evidence that the universe began in a nearly singular hot and dense point. It is not clear how far back in time one can “safely” apply GR, though, and we have also assumed spatially flat slices. In order to guarantee that the action is well-defined, we will temporarily restrict ourselves to an arbitrary, simply connected, bounded, open spatial volume $\mathcal{V} \subset \mathbb{R}^3$

$$\mathcal{U} \equiv \mathcal{V} \times (\eta_i, \eta_f). \quad (2.32)$$

Here $\eta_i$ is some arbitrary initial time and $\eta_f$ can either be unbounded, in the case of matter domination, or an asymptotic value in the case of Dark Energy domination.

We now work only to $O(1)$, so that we derive the Friedmann equation. Since

$$g_{\mu\nu} = a^2(\eta) \left[ \eta_{\mu\nu} + O(\epsilon) \right] \quad (2.33)$$

we have that

$$\sqrt{-g} = a^4 + O(\epsilon) \quad (2.34)$$
$$\delta g^{\mu\nu} = -2\eta^{\mu\nu} a^{-3} \delta a + O(\epsilon). \quad (2.35)$$

Note that we can exploit that the flat RW metric is a conformal rescaling of flat space to write

$$R = -6\eta^{\mu\nu} a^{-3} \partial_\mu \partial_\nu a + O(\epsilon) \quad (2.36)$$

where the derivatives are simple partials. Substitution into the definition of the gravitational action $S_G$ gives

$$S_G = -\frac{3}{8\pi G} \int_{\mathcal{U}} \eta^{\mu\nu} a(\partial_\mu \partial_\nu a) \, d^4x + O(\epsilon). \quad (2.37)$$

Performing the variation gives

$$\delta S_G = -\frac{3}{8\pi G} \int_{\mathcal{U}} \left[ \delta a(\partial_\mu \partial_\nu a) + a(\partial_\mu \partial_\nu \delta a) \right] d^4x \quad (2.38)$$
plus terms $O(\epsilon)$. Integrating by parts twice gives

$$\delta S_G = -\frac{3}{4\pi G} \int_{\mathcal{U}} \delta a \partial^\mu \partial_\mu a \, d^4 x + B + O(\epsilon) \tag{2.39}$$

where we have separated off the boundary term

$$B = -\frac{3}{8\pi G} \int_{\mathcal{U}} \partial_\mu \left[ \eta^{\mu \nu} \left[ a \delta a \partial_\nu a - \delta a \partial_\nu a \right] \right] d^4 x. \tag{2.40}$$

Given any $\mathcal{U}$ as defined in Eqn. (2.32), the variations $\delta a$ can always be chosen to have compact support within some open subset of $\mathcal{U}$. So, for purposes of determining the equations of motion, we may set $B = 0$ without loss of generality.

Now we come to the most relevant technical point for the physics. As methodically reviewed in §2.2, only variations consistent with the constraints imposed on the system are permitted. In our setting, constraint is not imposed by physical forces, like the ultimately electromagnetic forces which keep a marble rolling along the surface of a bowl. Instead, the constraints are the RW model symmetries: isotropy and homogeneity. In other words, the variations $\delta a(\eta)$ of the scale factor $a(\eta)$ must continue to respect this model symmetry. Thus, we must commute everything through the spatial integrals

$$\delta S_G = -\frac{3}{4\pi G} \int \delta a \partial^\mu \partial_\mu a \, d\eta \int_{\mathcal{V}} d^3 x + O(\epsilon). \tag{2.41}$$

Here we have used Fubini’s theorem to write the compactly supported integral as an iterated integral. Note that the spatial integration is just the volume $\mathcal{V}$

$$\delta S_G = -\frac{3}{4\pi G} \int \delta a \partial^\mu \partial_\mu a \, \mathcal{V} d\eta + O(\epsilon). \tag{2.42}$$

To compute the variation of the matter action, first note that

$$T^{\mu \nu} \delta g_{\mu \nu} = 2a \eta_{\mu \nu} T^{\mu \nu} \delta a + O(\epsilon) \tag{2.43}$$

$$= 2a^{-1} \delta a \left[ g_{\mu \nu} - O(\epsilon) \right] T^{\mu \nu} + O(\epsilon) \tag{2.44}$$

$$= 2a^{-1} T_{\mu}^\mu \delta a + O(\epsilon) \tag{2.45}$$

where, using square brackets to denote functional dependence,

$$T_{\mu}^\mu = T_{\mu}^\mu \left[ a^2 \eta_{\mu \nu}, \Phi(\eta, x) \right] + O(\epsilon) \tag{2.46}$$

is the trace of the model stress-tensor: that built from the RW model metric and all non-gravitational fields
\( \Phi(\eta, x) \). Substitution into the variation of the matter action gives

\[
\delta S_M = \int \int_V a^3 T^{\mu}_{\mu}(\eta, x) \delta a \, d^3 x \, d\eta + O(\epsilon). \tag{2.47}
\]

Again, we must commute the variation \( \delta a \) through the spatial integrals, consistent with the constraints imposed by homogeneity and isotropy. If we fail to perform this commutation, “the marble will not remain in the bowl” and our stress contribution will be inconsistent with our gravitational contribution. This commutation gives

\[
\delta S_M = \int a^3 \delta a \int_V T^{\mu}_{\mu}(\eta, x) \, d^3 x \, d\eta + O(\epsilon). \tag{2.48}
\]

The trace of the stress-tensor cannot be commuted because it is explicitly position dependent: it depends on all non-gravitational fields \( \Phi(\eta, x) \). As emphasized in §2.1 these fields are not explicitly constrained by our (or any) choice of metric model.

We now combine the variation of the gravitational action Eqn. (2.42) with that of the matter action Eqn. (2.48). The action principle as stated in Eqn (2.2) gives

\[
\int \delta a \left[ -\frac{3}{4\pi G} \partial^\mu \partial_\mu a \mathcal{V} + a^3 \int_V T^{\mu}_{\mu}(\eta, x) \, d^3 x \right] \, d\eta = 0, \tag{2.49}
\]

which implies the following equation of motion

\[
\frac{3}{4\pi G} \partial^\mu \partial_\mu a \, \mathcal{V} = a^3 \int_V T^{\mu}_{\mu}(\eta, x) \, d^3 x. \tag{2.50}
\]

Dividing by the constants on the left naturally reveals

\[
\partial^\mu \partial_\mu a = \frac{4\pi G}{3} \frac{1}{a^3} \int_V T^{\mu}_{\mu}(\eta, x) \, d^3 x \tag{2.51}
\]

\[= \frac{4\pi G}{3} a^3 \left\langle T^{\mu}_{\mu}(\eta, x) \right\rangle_V \, d^3 x. \tag{2.52}\]

Note the spatial-slice average, over the volume \( \mathcal{V} \), of the trace of the stress tensor. This average comes directly from the integral statement of the action principle. Performing the derivatives gives Friedmann’s equation

\[
\frac{d^2 a}{d\eta^2} = \frac{4\pi G}{3} a^3 \left( \rho(\eta, x) - \sum_{i=1}^3 P_i(\eta, x) \right)_V. \tag{2.53}
\]
expressed in conformal time. One may introduce the definitions

\[ \langle \rho(\eta, x) \rangle_\mathcal{V} \equiv \rho(\eta) \]  

\[ \left( \sum_{i=1}^{3} P_i(\eta, x) \right)_{\mathcal{V}} \equiv 3P(\eta) \]  

to arrive at the familiar statement of Friedmann’s equations in conformal time.

### 2.3.1 Significance of the volume \( \mathcal{V} \)

Note that the spatial average contains the arbitrary, but finite, volume \( \mathcal{V} \). We may understand \( \mathcal{V} \) as follows. Fix \( \mathcal{V} \) and pick some point \( P \in \mathcal{U} \). Compute the spatial average of the stress trace over a sphere of radius \( \mathcal{V}^{1/3} \) centered at \( P \). The resulting source will produce some dynamics for \( a(\eta) \). With the same \( \mathcal{V} \), repeat this procedure at a distinct point \( Q \in \mathcal{U} \). If the spatial average centered at \( Q \) does not agree with that performed at \( P \), the dynamics for \( a(\eta) \) will not agree. We conclude that the equation of motion for \( a(\eta) \) is not well-defined below some critical value of \( \mathcal{V} \). In other words, well-defined field equations for \( a(\eta) \) require an averaging volume large enough so that the averaged trace is position independent. Observations tell us that, at the present epoch, such a volume certainly exists: it is around 300Mpc. Note that the parameter \( \mathcal{V} \) is purely cosmological. It is conceptually distinct from any “lower-bound length scale” used to construct a fluid approximation of particle contributions to \( T^{\mu\nu}(\eta, x) \) [c.f. Landau and Lifshitz 1959, §1]. Note that the stress-tensor, written as integrals over distribution functions, is well-defined down to the Planck scale and need not be anywhere near thermal equilibrium [e.g. Stewart 1971].

### 2.3.2 Generalizations

The most general RW metric admits a gauge degree of freedom \( b(\tau) \)

\[ ds^2 = -b(\tau)^2 d\tau^2 + a(\tau)^2 dx^i dx^j \gamma_{ij} \]  

and \( \gamma_{ij} \) can be the metric for any 3-space of constant curvature. In the previous computation, we have “gauge fixed” by first defining

\[ b(\tau) \, d\tau \equiv dt \]  

and then switching to conformal time. If one keeps the \( b(\tau) \) degree of freedom, and varies with respect to it too, the typical Friedmann energy equation is also obtained [c.f. Suzuki et al. 1996, Eqn. (9)]. That calculation cannot be done as rapidly, however, because we cannot directly write the metric representation as a conformal rescaling of flat space. In this case, one does not obtain the trace immediately. The resulting equations reveal that one must compute the average of the diagonal stress-tensor components, as perceived in the RW coordinate frame. This is consistent with the conservation equation, which is suitably derived in
Finally, the restriction to flat spatial slices is not necessary. Including a $\gamma_{ij}$ with non-vanishing constant spatial curvature leads to the same conclusions with respect to computation of the source terms.

### 2.3.3 Discussion

Friedmann’s equation, as unambiguously derived in Eqn. (2.53), is rather remarkable for three reasons:

1. We made no assumptions whatsoever about the spatial distribution or magnitude of the non-gravitational fields $\Phi(\eta, \mathbf{x})$, except that they have a continuum representation.

2. There is no ambiguity averaging the trace, since it is a general coordinate scalar.

3. Pressures everywhere source the Friedmann equation. This will necessarily include the interiors of all compact objects.

Small scales can, in principle, affect large scales in GR. This result may be somewhat counter-intuitive, but it follows directly from consistently enforcing the RW model constraints during variation of both the gravitational and matter actions. It is also clear in hindsight from symmetry arguments alone. A RW spacetime has no notion of inside or outside of anything, because every point is indistinguishable. This means that the source term to a RW spacetime cannot contain any implicit notions of interior or exterior. This is why a Friedmann source based upon Birkhoff intuition, such as the “swiss-cheese model” of Einstein and Straus [1945, 1946], is inconsistent perturbatively. One cannot use an intuited first-order metric to determine the source for the zero-order metric.

Incidentally, §2.3.1 lucidly refutes the premise of Rácz et al. [2017]. In that work, the authors perform exactly the localized averaging procedure given in §2.3.1 to get very many distinct $a(\eta)$ behaviours. They then average these into a global scale factor. As we have shown, such a procedure is manifestly not consistent with GR. Our result agrees with the conclusions of Kaiser [2017] and Buchert [2017].

To further illustrate how the “actual” source shapes the gravitational model, consider a $T^{\mu\nu}$ which is only finite in spatial extent with finite masses. For concreteness, take a sphere of dust in an otherwise empty universe. Now let us naively assume the RW metric as the starting place for our analysis of this system. Starting from Einstein’s equations, the system would be inconsistent. The LHS has no position dependence, but the RHS is not translation invariant. Working directly from the action, however, the spatial average which emerges to source the zero-order equations will produce zero on the RHS. This is because the average must be taken over all space, to guarantee a position independent average. This drives the result to zero for our particular $T^{\mu\nu}$. Note that this remains true for delta-like point particle sources. This zero value will force a pure spatial curvature solution for Friedmann’s equations at zero order. In the case of spatially flat RW slices, only one solution exists with spatial curvature $k = 0$, it is Minkowski (i.e. constant scale factor, which can be absorbed into the coordinates trivially). This is the correct, and consistent, zero-order solution for the $T^{\mu\nu}$ considered.

We emphasize that the choice of the metric model is heavily influenced by observation. We use RW because the universe appears isotropic and homogeneous on large scales. Suppose we tried to model a stress
distribution of infinite extent, from a Minkowski metric with perturbations. In this case, our procedure will simply regenerate the well-known weak field equations. The equations, however, will fail in the familiar way: the source terms for the various metric potentials will approach $O(1)$ on large scales. This is not an inconsistency in our procedure, just a limitation on applicability imposed by a poor choice of metric model.

### 2.4 Model-restricted covariant conservation of stress-energy at zero order

In the previous section, we considered variations in the single physical degree of freedom $a(\eta)$. These variations $\delta a(\eta)$ were additional (non-physical) degrees of freedom, temporarily introduced in order to ascertain the actual equations of motion.

In this section, we consider deviations in the metric tensor components which appear when one transforms between frames. The metric is a tensor object, and so its components are covariant, not invariant. It is clear that a general coordinate scalar, such as the matter action, cannot change under coordinate transformations. It is shown in Weinberg [1972, §12.3] that this restriction leads to

$$\Delta S_M = 0 = \int_\mathcal{U} \epsilon^4 \nabla_\nu T^\nu_\lambda (\eta, x) \sqrt{-g} \, d^4x \quad (2.58)$$

where $\epsilon^4$ is an arbitrary (infinitesimal) vector field. Note that we deviate slightly from Weinberg and use notation $\Delta$ to clearly distinguish between the variational “differential” used in the previous section. We also have restricted the action to $\mathcal{U}$ so that we may interpret and apply our results to those of the previous section. Since $\epsilon^4$ is arbitrary, the above equation immediately leads to the familiar statement of covariant conservation of stress-energy.

In §2.3, we found that the relevant quantity for the Friedmann model was the volume averaged trace

$$\left\langle T^\mu_\mu \right\rangle_V = \left\langle T^0_0 (\eta, x) \right\rangle_V + \left\langle T^k_k (\eta, x) \right\rangle_V \quad (2.59)$$

where we have exploited linearity of the integral and the components are in the preferred RW frame. The appropriate conservation constraint for these volume-averaged quantities will again follow directly from the matter action. Noting that Eqn. (2.58) is true for arbitrary $\epsilon^4$, it is certainly true for position-independent $\epsilon^4(\eta)$. Then, commuting though the temporal integral as before

$$0 = \int \epsilon^4(\eta) d^4 \int_V \left[ \partial_\nu T^\nu_\lambda (\eta, x) + \Gamma^\nu_\nu T^\rho_\lambda (\eta, x) - \Gamma^\nu_\rho T^\nu_\lambda (\eta, x) \right] d^3x \, d\eta + O(\epsilon). \quad (2.60)$$

This is permissible because the preferred RW frame is a function of $\eta$ alone. We now expand the derivative term and apply the divergence theorem on the spatial slice

$$0 = \int \epsilon^4(\eta) d^4 \left\{ \int_V \left[ \partial_0 T^0_\lambda (\eta, x) + \Gamma^\nu_\nu T^\rho_\lambda (\eta, x) - \Gamma^\nu_\rho T^\nu_\lambda (\eta, x) \right] d^3x + \int_{\partial\mathcal{V}} T^k_\lambda (\eta, x) \, d^2x_k \right\} d\eta. \quad (2.61)$$
The boundary $\partial V$ is a spatial 2-surface, in contrast to the boundary $\partial M$ which was $\{\eta_i, \eta_f\} \times \partial V$. Thus, we cannot eliminate this term through restriction of $\epsilon^4(\eta)$. Yet, recall the physical significance of the volume $\mathcal{V}$: large enough so that the spatial-slice averages are position-independent. We may thus discard this boundary term on physical grounds, as any anisotropy over the 2-surface would privilege certain points. Considering only $\lambda = 0$ gives

$$0 = \int d^4\epsilon^0 \int_\mathcal{V} \left[ \partial_0 T^0_0 (\eta, x) + \Gamma^\nu_\rho T^\rho_0 (\eta, x) - \Gamma^\rho_0 T^\nu_\rho (\eta, x) \right] d^3 x d\eta. \quad (2.62)$$

A standard calculation gives

$$\Gamma^0_0 = H_\eta\delta^0_\nu \quad \Gamma^k_0 = H_\eta\delta^k_\nu \quad \Gamma^\nu_\nu = 4H_\eta\delta^0_\rho \quad (2.63)$$

where $H_\eta = a^{-1} da/d\eta$. All of these connection components are dependent on $\eta$ alone, so that

$$\frac{\partial}{\partial \eta} \langle T^0_0 (\eta, x) \rangle_\mathcal{V} + 3H_\eta \langle T^0_0 (\eta, x) \rangle_\mathcal{V} - H_\eta \langle T^k_0 (\eta, x) \rangle_\mathcal{V} = 0, \quad (2.64)$$

which is the appropriate conservation of stress-energy statement under the RW model constraint.
CHAPTER 3
COVARIANT COSMOLOGICAL PERTURBATION THEORY
WITHOUT ASSUMPTIONS ON THE STRESS

“And it ought to be remembered that there is nothing more difficult to take in hand, more perilous to conduct, or more uncertain in its success, than to take the lead in the introduction of a new order of things. Because the innovator has for enemies all those who have done well under the old conditions, and lukewarm defenders in those who may do well under the new. This coolness arises partly from fear of the opponents, who have the laws on their side, and partly from the incredulity of men, who do not readily believe in new things until they have had a long experience of them.”

N. Machiavelli, “The Prince”, §VI, ¶5

3.1 Preliminaries

In the last Chapter, we showed how the metric model, through the action principle, “digested” the stress-tensor into an appropriate source for the gravitational field equations at zero-order. Of course, we regenerated Friedmann’s equations. Our new result was that the isotropic and homogeneous source must be spatial slice averages over the diagonal stress-tensor components on $\mathcal{U}$, as perceived in natural RW coordinates. Since we made no assumptions about the stress-tensor a priori, we concluded that these averages necessarily probe the interiors of compact objects. One might object, however, that strong gravitational interiors would render the averages ill-defined. An example objection goes as

“One does not have the interior metric in cosmological perturbation theory, so one cannot well-prescribe the interior stress. The stress tensor for cosmology must be defined as a homogeneous and isotropic fluid plus perturbations which enter at the same order as the metric perturbations. Anything else is inconsistent.”

This chapter is devoted to explicitly addressing, and dissolving, such criticisms. The essential point will again be that one’s choice of gravitational model cannot have any influence on what actually exists in Nature. We will first show that the usual definition of the stress-tensor in covariant linear perturbation theory can be significantly generalized. We can then arrive at precisely Bardeen’s definition, under very clear assumptions.
We next make a case study of the scalar modes in Newtonian gauge. We will show that the action formalism
at first order automatically conspires to produce the correct first-order sources. In other words, selection of a
perturbed RW gravity model is sufficient to “shape” the form of stress-tensor appropriately to produce self-
consistent first-order gravitational field equations too. Technical material will be pedagogically presented
within section introductions, as required.

3.1.1 Notation

We first introduce notation, which we hope will clarify the following discussion. Let $g^{(N)}$ denote the power
series representation in $\epsilon$, which approximates the unique metric $g$ that satisfies the full Einstein equations
on some open submanifold $U \subset M$. Define the series representation in the typical RW coordinate basis.
Thus

$$
g \equiv \lim_{N \to \infty} g(\epsilon, N) \tag{3.1}$$

$$
g^{(N)}(\eta) \equiv a^2(\eta) \left[ \eta_{\mu\nu} + \sum_{n=1}^{N} \epsilon^n h^{(n)}_{\mu\nu}(\eta, x) \right] \partial^\mu \otimes \partial^\nu. \tag{3.2}$$

To prevent a proliferation of notation inside expressions, we will indicate frame (i.e. coordinate system)
basis vectors by a symbol inside curly braces to the left of the expression, offset by a colon. Indices will thus
be the “abstract index notation” of Wald, indicating sums and rank. For example, if $\{\hat{e}_\mu\}$ is an orthonormal
frame for $g^{(1)}$, then the following covariant equations are equivalent

$$
\{\hat{e}_\mu\} : g^{(1)}_{\mu\nu} = \text{diag}(-1, 1, 1, 1)_{\mu\nu} \tag{3.3}
$$

$$
\{\partial_\mu\} : g^{(1)}_{\mu\nu} = a^2(\eta) \left[ \eta_{\mu\nu} + \epsilon h^{(1)}_{\mu\nu} \right]. \tag{3.4}
$$

Expressions that are invariant in principle (i.e. scalars with respect to general coordinate transformations)
will not include indication of any frame. We will refer to the frame $\{\partial_\mu\}$, in which the metric model is
specified, as “the natural frame.” This is often elsewhere called “the coordinate frame;” but we feel such
language to be quite ambiguous.

If necessary, we will denote the rank of a tensor via $(n, m)$. Here $n$ is the contravariant rank and $m$ is the
covariant rank. A rank $(1, 1)$ object may be regarded as a linear transformation, so its matrix representation
can be matrix multiplied. Such matrices will be represented with parenthesis $(\cdot)$. We will place the column
index in 2nd position, and it will be covariant, consistent with matrix multiplication acting from the left. For
example,

$$
\Lambda^\beta_\gamma u^\gamma \equiv \gamma \begin{pmatrix} 1 & \beta \\ \beta & 1 \end{pmatrix} \begin{pmatrix} E \\ p \end{pmatrix} = \gamma \begin{pmatrix} E + \beta p \\ \beta E + p \end{pmatrix} = u^\beta. \tag{3.5}
$$
Physicists frequently omit the basis from their computations

\[ \{ \hat{e}_\mu \} : A^\sigma \equiv A^\sigma \hat{e}_\sigma. \]  (3.6)

If we have a linear transformation between bases

\[ \hat{e}_\sigma = V^\mu_\sigma \partial_\mu, \]  (3.7)

substitution gives

\[ A^\sigma \hat{e}_\sigma = A^\sigma V^\mu_\sigma \partial_\mu. \]  (3.8)

Written in “physicist,” this relation flips around

\[ \{ \hat{e}_\sigma \rightarrow \partial_\mu \} : A^\mu = V^\mu_\sigma A^\sigma. \]  (3.9)

Other symmetric bilinear forms (i.e. rank (0, 2) objects) cannot be matrix multiplied, yet autism compels us to align them visually into tables. Such matrices will be represented with *splayed* square brackets \[ \cdot \]. This notation is non-standard, but it would have saved this author much trouble. Remember, Wu Tang is for the children.

### 3.1.2 Standard definition of the cosmological stress tensor

Before we present our definition of the cosmological stress tensor, we review the standard definition. This modern definition originates from [Bardeen 1980], but was somewhat elided in presentation. It is rephrased and more clearly presented in [Kodama and Sasaki 1984], promulgated by [Hu 2004], and used implicitly by [Dodelson 2003]. We omit the typical spatial harmonic decomposition for clarity. The standard definition begins by assuming that \( g^{(1)} \) describes the actual universe. One begins by diagonalizing the actual stress-tensor, approximated as a timelike stress-tensor \( T_{\mu\nu} \), with respect to \( g^{(1)} \)

\[ (T_{\mu\sigma} g^{\sigma\nu}_{(1)} - \lambda \delta^\nu_{\mu}) u^\mu = 0. \]  (3.10)

Selecting only the timelike unit vector of the stress eigenframe, and calling this \( u^\mu \), one may write

\[ \{ \partial_\mu \} : T^{\mu\nu} = \rho u^\mu u_\nu + P_\perp(u)^{\mu}_{\sigma} P_\perp(u)^{\nu}_{\chi} T^{\sigma\chi}. \]  (3.11)

where \( \rho \) is the timelike eigenvalue and \( P_\perp(u) \) is the orthogonal projector to \( u \)

\[ \{ \partial_\mu \} : P_\perp(u)^{\mu}_{\sigma} \equiv \delta^\mu_{\sigma} + u^\mu u_\sigma. \]  (3.12)
One then defines a density perturbation $\delta(\eta, \mathbf{x})$ from an isotropic and homogeneous background density $\rho_0$ as

$$\rho \equiv \rho_0(\eta)(1 + \delta).$$

(3.13)

Next, one defines a 3-velocity through $u$

$$\{\partial_\mu\} : u^\mu \equiv u^0 \left( \begin{array}{c} 1 \\ \mathbf{v} \end{array} \right)$$

(3.14)

and let

$$\beta \equiv \sqrt{v^k v^j \delta_{kj}}.$$  

(3.15)

Normality of $u$ demands that

$$g^{(1)}_{\mu\nu} u^\mu u^\nu = -1.$$  

(3.16)

One uses this to write the orthogonal projector through $O(\epsilon)$ and $O(\beta)$. Let $P_0(\eta)$ be an isotropic and homogeneous pressure. Substitution of the doubly-truncated orthogonal projector into Eqn. (3.11) gives

$$\{\partial_\mu\} : \mathcal{T}^0_{0} = -\rho_0(1 + \delta)$$

(3.17)

$$\{\partial_\mu\} : \mathcal{T}^0_{j} = (\rho_0 + P_0)\left(v_j - B_j\right)$$

(3.18)

$$\{\partial_\mu\} : \mathcal{T}^{j}_{0} = -(\rho_0 + P_0) v^j.$$  

(3.19)

We have suppressed coordinate dependence since it is given explicitly in the previous exposition. The projector places no constraints on the spatial portion, which is just defined as

$$\{\partial_\mu\} : \mathcal{T}^{i}_{j} \equiv P_0 \left[ \delta^{i}_j + \pi_L(\eta, \mathbf{x})\delta^{i}_j + \pi_T(\eta, \mathbf{x}) \delta^{i}_j \right],$$

(3.20)

with $\pi_T$ traceless and symmetric. Note that the density perturbation $\delta$, the anisotropic stress $\pi_T$, and the isotropic pressure perturbation $\pi_L$ are, by fiat, regarded as $O(\epsilon)$ or $O(\beta)$ quantities.

**Discussion**

We feel there are numerous less than desirable features of the standard definition. All of these objections have a common origin: typical applications and interpretation of the perturbed RW framework at early times are very different than at late times. Yet these must be the same cosmology.

1. **The metric perturbation parameter $\epsilon$ explicitly appears in the stress definition.** This is most apparent in the above definition, where $\delta, \pi_L, \pi_T$ are defined to be $O(\epsilon)$. Further, $\epsilon$ is often conflated with
the dimensionless velocity $\beta$. Repeatedly, expansions are truncated at linear order in both parameters. This guarantees that the matter sector remains linear and a Fourier treatment of the perturbations remains decoupled. Observationally, however, it is clear that a linear expansion in the velocity is no longer reasonable below $z \sim 20$. The gravitational sector, however, works whenever and wherever $\epsilon$ is small. This is independent of $\beta$.

2. **Lightlike $\beta \sim 1$ flux is explicitly excluded.** There is no lightlike contribution to the standard stress tensor in cosmology. This is sidestepped in the kinetic theory construction of the stress tensor, since isotropic radiation behaves as a perfect fluid. In the early universe, perturbations in massless particles enter via their fractional temperature perturbations away from isotropic radiation. This technique, however, will not work for $z \lesssim 20$. Our experience from linearized flatspace GR is that arbitrary velocities $\beta \leq 1$ should pose no problems. This is not a novelty, since many exotic sources have regions of strong lightlike flux. It would be desirable to have precise understanding of such contributions, far from the strong gravity which leads to their production, with realistic asymptotic boundary conditions.

3. **The stress definition is slaved to the RW model.** As model-builders, our choice of a gravitational model for convenience cannot have any bearing on what actually exists in Nature.

### 3.2 A cosmological stress-tensor without assumptions

In the previous section, we reviewed the standard construction of the cosmological stress tensor and discussed its assumptions. In this section, we will construct a stress tensor suitable for cosmology, with far fewer assumptions. At each step of our construction, we will address one of the objections raised in the previous section. Our strategy will be straightforward. Every metric admits orthonormal frames. We will first construct the most natural orthonormal frame for any given metric model. Any two orthonormal frames of a specific metric model are related by a Lorentz transformation: a boost and a rotation. The diagonalization of the stress, presented by Kodama and Sasaki [1984], can provide a well-defined orthonormal frame. We will diagonalize the most general timelike contribution to the stress with respect to the specific metric model. Since both frames are orthonormal with respect to the same metric model, there is a natural (model-independent) notion of the relative velocity between them. This introduces $\beta$ in a model-independent way, addressing criticism 1. Then we introduce a directed lightlike flux, in the model’s natural orthonormal frame. This addresses criticism 2. The metric model then enters explicitly only at the final stage, shaping the stress in a natural way. This addresses criticism 3. As our aim is clarity and precision, the exposition will again be pedagogical. We beg the readers’ indulgence.

#### 3.2.1 The vierbein frame, with review

We begin by decoupling the description of the stress from the particular metric model. To do this, we will define quantities relative to an approximate orthonormal frame of the true metric $g$. The lowered index
metric, presented as a matrix, explicitly gives the angle cosines of the frame. For example,

$$\{ \partial_\mu \} : g_{\mu \nu} = \begin{bmatrix} \partial_t \cdot \partial_t & \ldots & \partial_z \cdot \partial_t \\ \vdots & \ddots & \vdots \\ \partial_t \cdot \partial_z & \ldots & \partial_z \cdot \partial_z \end{bmatrix}$$

(3.21)

where centered dot denotes the inner product wrt \( g \). For clarity, this notation was defined in 3.1.1. Each column of this object gives the overlap of a particular frame vector with all other frame vectors. For example, the timelike vector of the frame has representation

$$\{ \partial_\mu \} : \partial_t = \begin{bmatrix} \partial_t \cdot \partial_t \\ \partial_t \cdot \partial_x \\ \partial_t \cdot \partial_y \\ \partial_t \cdot \partial_z \end{bmatrix}.$$  

(3.22)

One may explicitly perform the Gram-Schmidt orthonormalization process on any model’s representation to arrive at an orthonormal frame \( \{ \hat{e}_{\mu} \} \) of \( g \)

$$\{ \hat{e}_{\mu} \} : g_{\mu \nu} = \begin{bmatrix} \hat{e}_t \cdot \hat{e}_t & \ldots & \hat{e}_z \cdot \hat{e}_t \\ \vdots & \ddots & \vdots \\ \hat{e}_t \cdot \hat{e}_z & \ldots & \hat{e}_z \cdot \hat{e}_z \end{bmatrix} \equiv \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{3.23}$$

One may fix this frame uniquely, up to cyclic permutation of the spatial vectors, by considering only positively-oriented (i.e. right-handed) and future-directed frames. We will call this \( \{ \hat{e}_{\mu} \} \) frame the “vierbein frame.” The orthonormalization procedure expresses the new frame as linear combinations of the model’s “natural” frame

$$\{ \partial_\mu \} : \hat{e}_\sigma = V^\mu_\sigma \partial_\mu.$$  

(3.24)

The numbers \( V^\mu_\sigma \) express a linear transformation, which changes basis. This change of basis is from an orthonormal frame, a concept common to any metric, to a model-dependent frame, in which the gravitational field equations are expressed and solved. In this way, we will use \( V^\mu_\sigma \) in the final step to “shape” the stress.
The vierbein frame for gauge-free perturbed RW

For illustration of the above discussion, and for later use, we now explicitly write the vierbein frame for $g^{(1)}$. We omit the scale factor as it represents a trivial rescaling. In covariant linear perturbation theory,

\[ \{ \partial_\mu \} : h^{(1)\mu\nu} = \begin{pmatrix} -2A & -B_x & -B_y & -B_z \\ -B_x & 2H_L + H_T^{Txx} & H_T^{Txy} & H_T^{Txz} \\ -B_y & H_T^{Txy} & 2H_L + H_T^{Tyy} & H_T^{Tyz} \\ -B_z & H_T^{Txz} & H_T^{Tyz} & 2H_L + H_T^{Tzz} \end{pmatrix}. \] (3.25)

Applying the Gram-Schmidt procedure to $g^{(1)}$ and discarding terms $O(\epsilon^2)$ produces the linear transformation from the vierbein frame into the model’s natural frame

\[ \{ \hat{e}_\sigma \rightarrow \partial_\mu \} : V^\nu_\sigma = \begin{pmatrix} 1 - A & 0 & 0 & 0 \\ B_x & 1 - H_L - H_T^{Txx}/2 & -2H_T^{Txy} & -2H_T^{Txz} \\ B_y & H_T^{Txy} & 1 - H_L - H_T^{Tyy}/2 & -2H_T^{Tyz} \\ B_z & H_T^{Txz} & H_T^{Tyz} & 1 - H_L - H_T^{Tzz}/2 \end{pmatrix} + O(\epsilon^2). \] (3.26)

3.2.2 The principal frame

The stress tensor is a symmetric bilinear form. The eigenvalues $\{ \lambda \}$ of a symmetric bilinear form are only well-defined with respect to some other symmetric bilinear form. In the framework of GR, this other symmetric bilinear form is the unique metric $\left| T_{\mu\nu} - \lambda g_{\mu\nu} \right| = 0$. (3.27)

The physical eigenvalues in GR, therefore, are those of the actual stress tensor with respect to the unapproximated metric. Though we do not know the unapproximated metric, we may still write the eigensystem approximately

\[ \{ \partial_\nu \} : T_{\tau\nu} \left[ g^{(1)\tau\mu} + O(\epsilon^2) \right] u^\nu - \lambda u^\mu = 0. \] (3.28)

We may write this in the vierbein frame defined above

\[ \{ \hat{e}_\nu \} : T_{\tau\nu} \left[ \text{diag}(-1, 1, 1, 1) T^{\tau\mu} + O(\epsilon^2) \right] u^\nu - \lambda u^\mu = 0 \] (3.29)

We now assume that $T$ is of Serge (Hawking-Ellis) Classification Type I: 1 timelike and 3 spacelike eigenvectors. This is sufficient to encode all subluminal material which satisfies the dominant energy condition, isotropic radiation, and more. Solution of this eigensystem gives a unique, positively-oriented, future-directed, orthogonal frame with respect to $g^{(1)}$ (up to cyclic permutation of the spatial basis), only if all eigenvalues $\{ \lambda \}$ are distinct. Repeated eigenvalues introduce spurious 3+1 rotational degrees of freedom.
In general, these spurious degrees of freedom can be encoded into a Lorentz transformation plus a spatial rotation $S^\mu_\nu$

$$ \left( S^\mu_\sigma u^\sigma \right) \left( S^\chi_\tau u^\tau \right) \eta_{\chi \nu} \propto \delta^\mu_\nu. \quad (3.30) $$

We absorb any such $S^\mu_\sigma$ required to orthogonalize $\{ u_\mu \}$ into $\{ u_\mu \}$ and simply write

$$ \left\{ u_\mu \right\} : T^I_{\tau \nu} = \text{diag}(\rho, P_1, P_2, P_3)_{\tau \nu} + O(\epsilon^2). \quad (3.31) $$

This frame will be referred to as “the principal frame.” We may form an orthonormal frame $\{ \hat{u}_\mu \}$ from the principle frame by dividing by $\sqrt{\lambda}$. Note that the principal frame need not agree with the vierbein frame. Since both $\{ \hat{u}_\mu \}$ and $\{ \hat{e}_\nu \}$ are orthonormal with respect to the same metric $g^{(1)}$, however, they are related by a Lorentz transformation

$$ \Lambda^\mu_\nu(\mathbf{v}) \hat{u}_\mu \equiv \hat{e}_\nu. \quad (3.32) $$

This defines a 3-velocity $\mathbf{v}$, which we will write in component form as $v_k$. We use the 3 rotational degrees of freedom to choose the same arbitrary cyclic permutation of the spatial basis, which appeared in the definition of $\{ \hat{e}_\mu \}$. The Type I stress-tensor then, as perceived in the vierbein frame is

$$ \left\{ \hat{e}_\nu \right\} : T^{I\mu}_\nu = \Lambda^\tau_\mu(\mathbf{v}) \Lambda^\sigma_\nu(\mathbf{v}) \text{diag}(\rho, P_1, P_2, P_3)_{\tau \sigma} + O(\epsilon^2). \quad (3.33) $$

In order to incorporate the possibility of lightlike flux, we now introduce a tensor of Serge Classification Type II

$$ \left\{ \hat{e}_\mu \right\} : T^{II\mu}_\nu \equiv \rho_\ell \ell_\mu \ell_\nu + O(\epsilon^2). \quad (3.34) $$

Here $\rho_\ell$ is the radiation density perceived by an observer who sees timelike flux moving with velocity $\mathbf{v}$. The future-directed null covector $\ell_\mu$ is defined relative to the vierbein frame spatial axes

$$ \left\{ \hat{e}_\mu \right\} : \ell_\mu \equiv \left( -1, \ell_1, \ell_2, \ell_3 \right) \quad (3.35) $$

$$ \delta^{ij} \ell_i \ell_j \equiv 1. \quad (3.36) $$

Note that the spatial components $\ell_j$ encode the direction of radiation flux as perceived by an observer who sees timelike flux moving with velocity $\mathbf{v}$. As discussed by Maruno and Visser in [Lobo (2017) §9.2.1], this does not represent the most general Type II stress tensor. We have excluded two further eigenvalue combinations. We exclude an eigenvalue combination which represents an additive cosmological constant to the null flux, since dark energy can already be encoded as Type I stress. We further exclude an eigenvalue combination which represents an additive EM field density aligned with the null flux, again because such
a contribution can already be encoded as Type I stress. Finally, we exclude Type III and Type IV Serge Classification stress tensors as unphysical.

We may now write our most general stress tensor, as perceived in the vierbein frame

\[ T_{\mu\nu} = \Lambda_{\mu}^\tau(v)\Lambda^\sigma_{\nu}(v)\text{diag}(\rho, \mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3)_{\sigma\tau} + \rho_{\tau\sigma}\ell_{\tau} + O(\epsilon^2). \]  

(3.37)

Counting degrees of freedom: we have 4 Type I eigenvalues, 3 velocity components relative to the vierbein frame, 1 radiation energy density, and 2 angles encoding lightlike flux relative to the vierbein frame. This gives a total of 10 degrees of freedom, as expected for a generic symmetric rank-(0,2) object in 3+1 dimensions.

### 3.2.3 Regeneration of the Bardeen stress-tensor

We now demonstrate that the stress tensor of Eqn. (3.37) reduces to the standard one, under appropriate assumptions. Let

\[ \beta \equiv \|v\| \]  

(3.38)

\[ \gamma \equiv \frac{1}{\sqrt{1 - \beta^2}} \]  

(3.39)

where \( v \) is the Lorentz transformation’s 3-velocity, as defined in Eqn. (3.32). We first present the explicit representation of the Type I stress-tensor in the vierbein frame

\[ \{\hat{e}_v\} : T^{I}_{00} = \gamma^2 \left( \rho + \sum_k v_k^2 \mathcal{P}_k \right) \]  

(3.40)

\[ \{\hat{e}_v\} : T^{I}_{0j} = \gamma v_j \left[ \gamma \rho + \mathcal{P}_j + \frac{\gamma^2}{\gamma + 1} \sum_k v_k^2 \mathcal{P}_k \right] \]  

(3.41)

\[ \{\hat{e}_v\} : T^{I}_{kj} = \delta_{kj} \mathcal{P}_j + \gamma^2 v_k v_j \left[ \rho + \frac{\mathcal{P}_k + \mathcal{P}_j}{\gamma + 1} + \frac{\gamma^2}{(\gamma + 1)^2} \sum_n q_n^2 \mathcal{P}_n \right]. \]  

(3.42)

To linear order in the velocity, these become

\[ \{\hat{e}_v\} : T^{00}_1 = \rho \]  

(3.43)

\[ \{\hat{e}_v\} : T^{0j}_1 = -v_j (\rho + \mathcal{P}_j) \]  

(3.44)

\[ \{\hat{e}_v\} : T^{kj}_1 = \delta^{kj} \mathcal{P}_j. \]  

(3.45)
where we have raised both indices. We now use $V^\mu_\sigma$ to view the stress contribution in the model’s natural frame. Regarding $\epsilon < \beta \ll 1$, we find

\[
\{\partial_\nu\} : T^{0\,0}_I = -\rho \tag{3.46}
\]

\[
\{\partial_\nu\} : T^{0\,j}_I = (v_j - B_j)(\rho + \mathcal{P}_j) \tag{3.47}
\]

\[
\{\partial_\nu\} : T^{j\,0}_I = -\nu^j(\rho + \mathcal{P}^j) \tag{3.48}
\]

\[
\{\partial_\nu\} : T^{k\,j}_I = \delta^k_j \mathcal{P}_j + \mathcal{H}_k(\mathcal{P}_k - \mathcal{P}_j) \times \begin{cases} 1 & k < j \\ 2 & k > j \end{cases}. \tag{3.49}
\]

To cast this into the standard formalism, first define

\[
\rho \equiv \rho_0(\eta)[1 + \delta] \tag{3.50}
\]

\[
\mathcal{P}_j \equiv \mathcal{P}_0(\eta)[1 + \xi_j]. \tag{3.51}
\]

In the standard formalism, deviations from $\mathcal{P}_0$ are regarded as “small.” Substituting into Eqns. (3.46)–(3.49), and keeping only leading order corrections (linear in either $\beta$ or $\epsilon$) we have

\[
\{\partial_\nu\} : T^{0\,0}_I = -\rho_0(1 + \delta) \tag{3.52}
\]

\[
\{\partial_\nu\} : T^{0\,j}_I = (v_j - B_j)(\rho_0 + \mathcal{P}_0) \tag{3.53}
\]

\[
\{\partial_\nu\} : T^{j\,0}_I = -\nu^j(\rho_0 + \mathcal{P}_0) \tag{3.54}
\]

\[
\{\partial_\nu\} : T^{k\,j}_I = \delta^k_j \mathcal{P}_0(1 + \xi^k). \tag{3.55}
\]

Now, we may irreducibly decompose $\delta^k_j(1 + \xi^k)$ under spatial rotations. Note that we can always do this in the vierbein frame, but it is not necessarily useful in more general settings. To this order of perturbation, performing the decomposition in the natural frame is equivalent. We define

\[
\pi_L \equiv \frac{1}{3} \sum_n \xi^n \tag{3.56}
\]

\[
\pi_T^k \, j = \frac{\delta^k_j}{3} \left( 2\xi^k - \sum_{n \neq k} \xi^n \right). \tag{3.57}
\]

Then, we may write

\[
\delta^k_j(1 + \xi^k) = \delta^k_j(1 + \pi_L) + \pi_T^k \, j. \tag{3.58}
\]

Substitution of this expression into Eqn. (3.55) gives Eqns. (3.52)–(3.55) identical to [Bardeen] [1980] Eqns. (2.17)–(2.20), except that the spatial coordinate system automatically diagonalizes the anisotropic stress. In the standard formalism, deviations from isotropic radiation are assumed to be small. This eliminates the
need for a Type II contribution.

### 3.2.4 Relation to the distribution function definition of the stress

One can also define the stress-tensor through use of the Boltzmann formalism. Here one considers a distribution function over phase space perturbed from one of the quantum gases (or just Boltzmann). This is eminently reasonable after the Big Bang, but it is not useful at late times. Even though the distribution function formalism given by [Stewart 1971] remains completely valid, one does not know about which distribution function to perturb.

Still, it is useful to show how our definition is the same one used by [Dodelson 2003 §4], though he does not make this explicitly clear. Dodelson uses the Boltzmann/distribution function formalism to describe the stress-energy

\[ T^{ab} \propto \sum_i \int \frac{p^a p^b}{\sqrt{m^2 + p^2}} f_i(p, x, t) \, p^2 \sin(\theta_p) \, dp d\theta_p d\phi_p \]

where \( f_i(p, x, t) \) is the distribution function for species \( i \), pre-weighted for spin-multiplicity, and

\[
\begin{align*}
    p^0 &\equiv E & p^x &\equiv p \cos(\phi_p) \sin(\theta_p) \\
    p^y &\equiv p \sin(\phi_p) \sin(\theta_p) & p^z &\equiv p \cos(\theta_p).
\end{align*}
\]

The use of an orthonormal frame is encoded in the following Dodelson definitions

\[
\begin{align*}
    \delta_{ij} p^i p^j &\equiv g_{ij} p^i p^j & \text{Dodelson: (4.13)} \\
    -m^2 &\equiv P^\mu P^\nu g_{\mu\nu} & \text{Dodelson: (4.64)} \\
    E &\equiv \sqrt{p^2 + m^2} & \text{Dodelson: (4.65)}
\end{align*}
\]

where \( P^\mu \) are regarded as 4-momenta in the natural frame. Defining \( p^0 \equiv E \), it follows at once that

\[
p^a p^b \eta_{ab} \equiv P^\mu P^\nu g_{\mu\nu}.
\]

To see that it is our vierbein frame, note that Dodelson defines the velocity of Dark Matter to be

\[
v^j \propto \int d^3p f_{dm} \frac{p^j}{E} \quad \text{Dodelson: (4.71)}.
\]

Metric influence enters only upon return to the natural frame.
3.2.5 Alternative definition of the null stress

This section is included for completeness, but the material presented here is not used elsewhere in this thesis. In some sense, it is more natural to define the lightlike flux relative to an observer comoving with the timelike flux. In principle, any observer, in the instrumentalist’s sense of the word, would necessarily measure the lightlike flux relative to herself at rest. In this case, the $T^\mu_\nu$ contribution also becomes boosted.

With this definition of the lightlike flux, we find

$$\{\hat{e}_\nu\} : T^\mu_0 = \rho r \gamma^2 (1 + v \cdot \hat{\ell})^2$$  \hspace{1cm} (3.67)

$$\{\hat{e}_\nu\} : T^\mu_j = \rho r \gamma (1 + v \cdot \hat{\ell}) \left[\gamma v_j + \hat{\ell}_j + v_j \frac{\gamma^2}{\gamma + 1} v \cdot \hat{\ell}\right]$$  \hspace{1cm} (3.68)

$$\{\hat{e}_\nu\} : T^\mu_k = \rho r \left[\ell_j \ell_k + (v \hat{\ell})_{(k,j)} \right] \left[\gamma + \frac{\gamma^2}{\gamma + 1} v \cdot \hat{\ell}\right]^2\right]$$  \hspace{1cm} (3.69)

To linear order in the velocity, these become

$$\{\hat{e}_\nu\} : T^\mu_0 = \rho r (1 + 2v \cdot \hat{\ell})$$  \hspace{1cm} (3.70)

$$\{\hat{e}_\nu\} : T^\mu_j = \rho r \left[v_j + \hat{\ell}_j (1 + v \cdot \hat{\ell})\right]$$  \hspace{1cm} (3.71)

$$\{\hat{e}_\nu\} : T^\mu_k = \rho r \left[\ell_j \ell_k + (v \hat{\ell})_{(k,j)}\right].$$  \hspace{1cm} (3.72)

3.2.6 Summary

We now list what we believe are advantages of our approach, relative to the standard one

- The metric perturbation parameter $\epsilon$ does not contaminate the description of the stress.

- We regenerate the familiar formalism by applying very specific assumptions to an extremely generic stress-tensor. This means we can use this same generic stress-tensor at other times and places in the universe where these assumptions no longer apply. In particular, cosmological and weak-field effects in the vicinity of $O(\epsilon)$ density local and luminal flows can now be considered. Most importantly for the Dark Energy problem, cosmological and weak-field effects from $O(1)$ density local and luminal flows, at sufficient distance, can also now be considered. This aligns with our experience of with the flat-space linearized gravity formalism, which fully generalizes SR to weak gravitational fields.

This final point is highly relevant. In §2.3 we showed that the dynamics of the scale factor is determined from spatial slice averages of all energy densities and pressures, even those interior to compact objects. We have now established that localized $O(1)$ contributions are readily described by the same stress tensor which reduces to the standard definition under assumptions appropriate at early times.
3.3 Scalar perturbations, in conformal Newtonian gauge

In this section, we will verify that our procedure regenerates the standard first-order field scalar equations. Again, we work directly from the action. In order to arrive at field equations linear in the perturbation variables, we must compute the action to quadratic order in these variables [e.g., Mukhanov [1988]]. This is because the variation in a degree of freedom is itself a distinct degree of freedom, which factors out of the final expression when determining equations of motion. The new result in this section is that the standard perturbation orders for the overdensity $\delta$ and the isotropic pressure perturbation $\pi_l$ emerge automatically from the action formalism. This section primarily exists to demonstrate internal consistency to those who are unfamiliar working directly from the action. Much of this section is a rather lengthy computation of the Ricci scalar done with differential forms. We show this approach explicitly for completeness and pedagogy: there are seemingly scant physics examples of such computations. All computations have been verified with the CAS Maxima.

Consistent with the previous discussion, we fix Bardeen’s longitudinal/Newtonian gauge so that the metric is diagonal, and consider only the scalar perturbations

$$g_{\mu\nu} \equiv a^2 \left[ \eta_{\mu\nu} + 2 \left( -A d\eta^2 + H_L dx^2 \right) \right] \equiv a^2 \bar{g}_{\mu\nu}. \quad (3.73)$$

We will regard $A(x, \eta)$ and $H_L(x, \eta)$ as $O(\epsilon)$ quantities. We will omit explicit indication of the perturbation order in this section. As discussed, we will require the Lagrangian to quadratic order in the field variables. Again, we exploit that we have a conformal rescaling of the barred metric to write

$$R = a^{-2} \bar{R} - 6 \bar{g}^{\mu\nu} a^{-3} \bar{\nabla}_\mu \bar{\nabla}_\nu a. \quad (3.75)$$

We will determine $\bar{R}$ and the covariant derivative terms separately. First, we express the metric in terms of the vierbein coframe

$$\bar{g}(, ) = \eta_{\mu\nu} \theta^\mu \theta^\nu \quad (3.76)$$

where the vierbein coframe basis 1-forms are related to the natural coframe basis 1-forms by

$$\theta^0 \equiv \sqrt{1 + 2A} \, d\eta$$

$$\theta^i \equiv \sqrt{1 + 2H_L} \, dx^i. \quad (3.77)$$

$$\theta^\mu \equiv \sqrt{1 + 2A} \, d\eta$$

$$\theta^\mu \equiv \sqrt{1 + 2H_L} \, dx^i. \quad (3.78)$$
3.3.1 Connection 1-forms

To determine the connection 1-forms $\Omega^\mu_\nu$, we first take the exterior derivative of the coframe basis

$$d\theta^0 = \frac{\partial_\mu A}{\sqrt{1 + 2A}} \, dx^\mu \wedge d\eta$$

$$d\theta^i = \frac{\partial_\mu H_L}{\sqrt{1 + 2H_L}} \, dx^\mu \wedge dx^i. \tag{3.79}$$

Note that these quantities are already at least first-order. Rewriting in terms of the coframe basis gives

$$d\theta^0 = (1 - 2A - H_L) \partial_\mu A \, \theta^i \wedge \theta^0 \tag{3.80}$$

$$d\theta^i = (1 - 2H_L - A) \partial_\mu H_L \, \theta^0 \wedge \theta^i$$

$$+ (1 - 3H_L) \partial_k H_L \, \theta^k \wedge \theta^i, \tag{3.81}$$

valid through second order. We now exploit that GR is torsion free

$$D \theta^\mu = d\theta^\mu + \Omega^\mu_\nu \wedge \theta^\nu = 0 \tag{3.82}$$

where “$D$” denotes the exterior covariant derivative. We immediately have that

$$d\theta^\mu = -\Omega^\mu_\nu \wedge \theta^\nu. \tag{3.83}$$

Before proceeding, we note some features of the connection 1-forms. It follows that, in the vierbein frame,

$$\Omega^0_\nu = -\Omega^\nu_0 \tag{3.84}$$

$$\Omega^i_k = -\Omega^k_i \tag{3.85}$$

$$\Omega^0_0 = 0. \tag{3.86}$$

which is of remarkable utility. Raising an index with the Minkowski representation of the metric gives

$$\Omega^0_k = \Omega^k_0$$

$$\Omega^i_k = -\Omega^k_i$$

$$\Omega^0_0 = 0. \tag{3.87}$$

Now we can match coefficients to determine the connection 1-forms. Combining Eqns. (3.82) and (3.84) we have that

$$\left[(1 - 2H_L - A) \partial_\mu H_L \, \theta^i = \Omega^i_0 \right] \wedge \theta^0 \tag{3.88}$$

$$\left[(1 - 3H_L) \partial_k H_L \, \theta^i = \Omega^i_k \right] \wedge \theta^k. \tag{3.89}$$
While combining Eqns. (3.81) and (3.84) we have that
\[
(1 - 2A - H_L)\partial_k A \theta^0 = \Omega^0_k \wedge \theta^k. \tag{3.91}
\]

Let
\[
\Omega^0_k \equiv c^0_{k\mu} \theta^\mu \tag{3.92}
\]
where the \(c\)'s are 0-forms. Then it follows from Eqn. (3.91) that
\[
c^0_{kj} = c^0_{jk} \quad k \neq j \tag{3.93}
\]
\[
c^0_{0k} = (1 - 2A - H_L)\partial_k A \tag{3.94}
\]
but \(c^0_{kk}\) is left unconstrained. From Eqn. (3.86) it follows that
\[
c^k_{0j} = c^j_{0k} \quad k \neq j. \tag{3.95}
\]

Let
\[
\Omega^i_k \equiv c^i_{k\mu} \theta^\mu. \tag{3.96}
\]
Then it follows from Eqns. (3.89) and (3.86) that
\[
c^0_{kk} = (1 - 2H_L - A)\partial_0 H_L \tag{3.97}
\]
and from Eqn. (3.90) that
\[
c^i_{k\mu} = c^i_{\mu k} \quad i \neq \mu \neq k \tag{3.98}
\]
\[
c^i_{k\bar{i}} = (1 - 3H_L)\partial_k H_L. \tag{3.99}
\]

Here, bar above an index holds it fixed (i.e. no summations). From Eqn. (3.87), we have that
\[
c^i_{kk} = -c^k_{i\bar{k}}. \tag{3.100}
\]

To resolve the remaining coefficients, note that repeated use of Eqns. (3.87), (3.96), and (3.98) gives
\[
c^i_{kj} = -c^k_{ij} = -c^k_{ji} = \cdots = -c^i_{kj} = 0 \tag{3.101}
\]
and similarly

\[ c^0_{kj} = c^j_{0k} = c^j_{0j} = c^j_{k0} = -c^k_{j0} = \cdots = 0. \] (3.102)

The connection 1-forms may now be read off from Eqns. (3.92) and (3.96)

\[ \Omega^0_k = (1 - 2A - H_L) \partial_k A \theta^0 + (1 - 2H_L - A) \partial_0 H_L \theta^k \] (3.103)
\[ \Omega^i_k = (1 - 3H_L) \left( \partial_k H_L \theta^i - \partial_i H_L \theta^k \right). \] (3.104)

Note that \( \Omega^i_k \) satisfies Eqn. (3.87) under exchange of indices because both are spacelike. Note that one cannot exchange indices in \( \Omega^0_k \) to determine \( \Omega^k_0 \) because one index is timelike. Substitution into \( \Omega^k_0 \equiv c^k_{0\mu} \theta^\mu \) gives the correct (i.e. equal) result, as it must by construction.

### 3.3.2 Riemann curvature and Ricci scalar

The components of the Riemann curvature tensor are equal to the components of the exterior covariant derivative of the connection 1-forms

\[ D\Omega^\mu = \frac{1}{2} \bar{R}^\mu_{\nu\sigma\rho} \theta^\sigma \wedge \theta^\rho \]
\[ = \bar{R}^\mu_{\nu\sigma\rho} \theta^\sigma \wedge \theta^\rho + \cdots \] (3.105)

This follows from the definition of the exterior covariant derivative and identifying the \( c \)'s above as the components of the Christoffel connection. Since

\[ D\Omega^\mu = d\Omega^\mu + \Omega^\mu_\rho \wedge \Omega^\rho \] (3.106)

we require the exterior derivatives of the connection 1-forms. These are the lengthiest contribution, which we now compute. We begin with

\[ d\Omega^0_k = -\partial_k A \ d[2A + H_L] \wedge \theta^0 \\
+ (1 - 2A - H_L) \ d[\partial_k A \wedge \theta^0] \\
+ \partial_k A \ d\theta^0 \\
- \partial_0 H_L \ d[2H_L + A] \wedge \theta^k \\
+ (1 - 2H_L - A) \ d[\partial_0 H_L] \wedge \theta^k \\
+ \partial_0 H_L \ d\theta^k \] (3.107)
Noting that \( dx^0 = (1 - A) \theta^0 \) and \( dx^j = (1 - H_L) \theta^j \) to sufficient order, we process each line

\[
d\Omega^0_k = -\partial_k A \partial_j (2A + H_L) \theta^j \wedge \theta^0 \\
+ (1 - 2A - 2H_L) \partial_k A \theta^j \wedge \theta^0 \\
+ \partial_k A \partial_j A \theta^j \wedge \theta^0 \\
- \partial_0 H_L \left[ \partial_0 (2H_L + A) \theta^j \wedge \theta^k + \partial_j (2H_L + A) \theta^j \wedge \theta^k \right] \\
+ \left[ (1 - 2H_L - 2A) \partial_{00} H_L \theta^0 \wedge \theta^k \\
+ (1 - 3H_L - A) \partial_{0j} H_L \theta^j \wedge \theta^k \right] \\
+ \partial_0 H_L \left[ \partial_0 H_L \theta^0 \wedge \theta^k + \partial_j H_L \theta^j \wedge \theta^k \right]
\] (3.108)

We compute the curvature 2-form

\[
D\Omega^0_k = d\Omega^0_k + \Omega^0_j \wedge \Omega^j_k \\
= d\Omega^0_k - \partial_0 H_L \partial_j H_L \theta^j \wedge \theta^k \\
+ \delta^j_k \left[ \partial_j A \partial_k H_L \theta^0 \wedge \theta^j - \partial_j A \partial_j H_L \theta^0 \wedge \theta^k \right]
\] (3.109)

where we have introduced the negated identity

\[
\delta^j_k \equiv (1 - \delta^j_k).
\] (3.110)

to exclude terms from a sum. Combining Eqns. (3.109) and (3.108), grouping under basis 2-forms, gives

\[
D\Omega^0_k = \\
\left[ \partial_k A \partial_j (A + H_L) + (2H_L + 2A - 1) \partial_k A + \delta^j_k \partial_j A \partial_k H_L \right] \theta^0 \wedge \theta^j \\
\left[ (1 - 3H_L - A) \partial_{0j} H_L - \partial_{0j} H_L \partial_j (2H_L + A) \right] \theta^j \wedge \theta^k \\
\left[ (1 - 2H_L - 2A) \partial_{00} H_L - \partial_0 H_L \partial_0 (H_L + A) - \delta^j_k \partial_j A \partial_j H_L \right] \theta^0 \wedge \theta^k
\] (3.111)

from which the Riemann components can be directly read off.

For the other curvature 2-form, we again compute the exterior derivative of the appropriate connection 1-forms

\[
d\Omega^j_k = -d[3H_L] \wedge \left[ \partial_k H_L \theta^j - \partial_j H_L \theta^k \right] \\
+ (1 - 3H_L) \left[ d[\partial_k H_L] \wedge \theta^j - d[\partial_j H_L] \wedge \theta^k \right] \\
+ \partial_k H_L d\theta^j - \partial_j H_L d\theta^k
\] (3.112)
Again, we process this line by line

\[
\begin{align*}
\text{d} \Omega^i_k &= -3 \partial_\mu H_L \theta^\mu \wedge \left[ \partial_k H_L \theta^j - \partial_i H_L \theta^k \right] \\
&+ \left( 1 - 3 H_L - A \right) \partial_{\theta^0} H_L \theta^0 \wedge \theta^i \\
&+ \left( 1 - 4 H_L \right) \partial_{\theta^1} H_L \theta^1 \wedge \theta^j \\
&\frac{1}{2} \left[ \partial_k H_L \left( \partial_{\theta^0} H_L \theta^0 \wedge \theta^i + \partial_{\theta^1} H_L \theta^1 \wedge \theta^j \right) \right] \\
&- \left( 1 - 3 H_L - A \right) \partial_{\theta^0} H_L \theta^0 \wedge \theta^k \\
&- \left( 1 - 4 H_L \right) \partial_{\theta^1} H_L \theta^1 \wedge \theta^k \\
&+ \left[ \partial_k H_L \left( \partial_{\theta^0} H_L \theta^0 \wedge \theta^i + \partial_{\theta^1} H_L \theta^1 \wedge \theta^j \right) \right] \\
&\frac{1}{2} \left[ \partial_i H_L \left( \partial_{\theta^0} H_L \theta^0 \wedge \theta^k + \partial_{\theta^1} H_L \theta^1 \wedge \theta^k \right) \right].
\end{align*}
\] (3.113)

The first line is equal to \(-3\) times the final line, so

\[
\begin{align*}
\text{d} \Omega^i_k &= (1 - 3 H_L - A) \partial_{\theta^0} H_L \theta^0 \wedge \theta^i \\
&+ (1 - 4 H_L \partial_{\theta^1} H_L \theta^1 \wedge \theta^j \\
&- (1 - 3 H_L - A) \partial_{\theta^0} H_L \theta^0 \wedge \theta^k \\
&- (1 - 4 H_L \partial_{\theta^1} H_L \theta^1 \wedge \theta^k \\
&- 2 \partial_k H_L \left( \partial_{\theta^0} H_L \theta^0 \wedge \theta^i + \partial_{\theta^1} H_L \theta^1 \wedge \theta^j \right) \\
&+ 2 \partial_i H_L \left( \partial_{\theta^0} H_L \theta^0 \wedge \theta^k + \partial_{\theta^1} H_L \theta^1 \wedge \theta^k \right)
\end{align*}
\] (3.114)

We compute the curvature 2-form

\[
D \Omega^i_k = \text{d} \Omega^i_k + \Omega^i_\mu \wedge \text{d} \Omega^\mu_k
\]

\[
= \text{d} \Omega^i_k + \left[ \partial_i A \theta^0 + \partial_0 H_L \theta^i \right] \wedge \left[ \partial_k A \theta^0 + \partial_0 H_L \theta^k \right] \\
+ \left( \partial^0 H_L \right)^2 \theta^j \wedge \theta^k + \partial_0 H_L \partial_k A \theta^0 \wedge \theta^0
\] (3.115)
and combine Eqn. (3.114) and Eqn. (3.115), grouping under basis 2-forms to get

\[
DQ^i_k = \left[ (1 - 3H_L - A)\partial_{k0}H_L - 2\partial_k H_L \partial_0 H_L - \partial_0 H_L \partial_k A \right] \theta^0 \wedge \theta^i
\]

\[
+ 2\partial_k H_L \partial_j H_L - (1 - 4H_L)\partial_{kj} H_L + \phi^0_k \partial_k H_L \partial_j H_L \right] \theta^j \wedge \theta^i
\]

\[
[2\partial_k H_L \partial_0 H_L + \partial_i A \partial_0 H_L - (1 - 3H_L - A)\partial_{i0} H_L] \theta^0 \wedge \theta^k
\]

\[
[2\partial_i H_L \partial_0 H_L + \phi^0_i \partial_i H_L \partial_0 H_L - (1 - 4H_L)\partial_{ij} H_L] \theta^j \wedge \theta^k
\]

\[
[(\partial_0 H_L)^2 - \phi^0_k \phi^i_k \partial_j H_L^2] \theta^i \wedge \theta^k
\]

(3.116)

from which the Riemann components can be directly read off. Note that terms containing \( j \) sum against the 2-form basis, so care must be taken to include all desired terms.

Since we only require the Ricci scalar for the Lagrangian

\[ \bar{R} = \bar{R}^{\mu\nu}_a \rho^{\mu\nu} \]

\[ = 2 \sum_k \bar{R}^{00}_{k0k} + \sum_k \bar{R}^i_{kik} \]

(3.117)

(3.118)

where the second line follows by the symmetries of Riemann. By inspection of Eqns. (3.111) and (3.116) we find

\[
\bar{R}^{00}_{k0k} = \partial_k A \partial_{k}(A + H_L) - \partial_0 H_L \partial_0 (A + H_L)
\]

\[ + (2H_L + 2A - 1)(\partial_{kk} A - \partial_{00} H_L) - \phi^0_k \partial_k A \partial_j H_L
\]

\[
\bar{R}^i_{kik} = 2(\partial_k H_L)^2 + 2(\partial_i H_L)^2 + (\partial_0 H_L)^2
\]

\[ - (1 - 4H_L)(\partial_{kk} H_L + \partial_{ij} H_L) - \phi^0_k \phi^i_k (\partial_j H_L)^2.
\]

(3.119)

(3.120)

Assembling these and cleaning up, we arrive at the Ricci scalar

\[
\bar{R} = 2|\nabla A|^2 - 6\partial_0 A \partial_0 H_L + 2(2H_L + 2A - 1)(\nabla^2 A - 3\partial_{00} H_L)
\]

\[ - 2\nabla H_L \cdot \nabla A + 6|\nabla H_L|^2 + 4(4H_L - 1)\nabla^2 H_L.
\]

(3.121)

Note that this expression reduces to the correct Ricci scalar at first order, as can be verified against Dodelson (5.17) with \( a \equiv 1 \). It is convenient, for the variation, to expand the Ricci scalar, grouping by like terms

\[
\bar{R} = \left[ (-2\nabla^2 A) + 2|\nabla A|^2 + 4A\nabla^2 A \right] +
\]

\[
\left[ (6\partial_{00} H_L - 4\nabla^2 H_L) - 12\partial_{00} H_L H_L + 16H_L \nabla^2 H_L + 6|\nabla H_L|^2 \right] +
\]

\[
4H_L \nabla^2 A - 6\partial_0 A \partial_0 H_L - 12A\partial_{00} H_L - 2\nabla H_L \cdot \nabla A.
\]

(3.122)

Note that the first order terms are offset in parenthesis.
3.3.3 Conformal term

In this section, we compute the conformal term, which contains covariant derivatives. First, note that

\[ \bar{\nabla}_\beta f = e_\beta f \]  

(3.123)

where \( e_\beta \) is a vierbein frame basis vector

\[ e_\beta [\theta] = \delta^\alpha_\beta. \]  

(3.124)

Together, these imply that

\[ e_0 = \frac{1}{\sqrt{1 + 2A}} \frac{\partial}{\partial \eta} \]  

(3.125)

and so

\[ \bar{\nabla}_\beta a(\eta) = \frac{1}{\sqrt{1 + 2A}} a' \delta^0_\beta. \]  

(3.126)

because a covariant derivative of a scalar is just a regular derivative. Keep in mind that we are in the vierbein frame basis, so the metric representation is Minkowski. Continuing with the second covariant derivative

\[ \bar{\nabla}_\beta \bar{\nabla}_\beta a = e_\alpha \left[ \frac{a' \delta^0_\beta \eta^\beta}{\sqrt{1 + 2A}} - \frac{\eta^\beta a' \delta^0_\rho \rho^\beta}{\sqrt{1 + 2A}} \right] \]  

(3.127)

Expanding out to second order gives

\[ \bar{\nabla}_\beta \bar{\nabla}_\beta a = a' (1 - 4A) - a'' (1 - 2A + 4A^2) \]  

\[ - 3a' \partial_0 H_L (1 - 2A - 2H_L). \]  

(3.128)

Again, it is convenient for the variation to expand this expression, grouping by like terms

\[ 6a \bar{\nabla}_\beta \bar{\nabla}_\beta a = -6aa'' + 6a [a' A' + 2a'' A - 3a' \partial_0 H_L] \]  

\[ + 6a \left[ 6a' \partial_0 H_L (A + H_L) - 4a' A' A - 4a'' A^2 \right]. \]  

(3.129)

We have also multiplied by \( 6a \) to produce the final form required for the Lagrangian.

Integration by parts in the action

It is possible to perform this computation much faster. One can integrate by parts within the action, before substitution of the metric ansatz. Such a procedure was performed in \( \S 2.3 \). We explicitly computed the
covariant derivatives in this section for illustration of such calculations in the vierbein frame.

### 3.3.4 Variation of the gravitational action

We can now write the gravitational action in the vierbein frame

\[
S_G = \frac{1}{16\pi G} \int_M \left( a^{-2} \bar{R} - 6a^{-3} \bar{\nabla}^\alpha \bar{\nabla}_\alpha a \right) a^4 \wedge \theta^\mu. \tag{3.130}
\]

Since the vierbein coframe depends on \( A \) and \( H_L \), we note that

\[
\bigwedge_{\mu} \theta^\mu = \sqrt{1 + 2A(1 + 2H_L)^{3/2}} \bigwedge_{\mu} dx^\mu
\]

\[
= \left[ 1 + A + 3H_L + \frac{3}{2} H_L^2 - \frac{1}{2} A^2 + 3AH_L \right] \bigwedge_{\mu} dx^\mu
\]

\[
= \sqrt{-g} \bigwedge_{\mu} dx^\mu. \tag{3.131}
\]

Since we are working only to quadratic order, it is most efficient to multiply everything together first, and then perform the variation. Also, since our derivatives are with respect to the natural frame, to massage these derivatives off of the variations and into boundary terms, we will need to consider the integration in \( dx^\mu \). We will proceed with each term in the Lagrangian separately.

Proceeding termwise through the action integrand, the Ricci scalar is already first order. The term which requires variation is

\[
a^2 \bar{R} \bigwedge \theta^\mu = \mathcal{L}_R \bigwedge_{\mu} dx^\mu =
\]

\[
a^2 \left( \bar{R} + (3H_L + A) \left( 6\partial_{00} H_L - 2\nabla^2 A - 4\nabla^2 H_L \right) \right) \bigwedge_{\mu} dx^\mu. \tag{3.132}
\]

Expanding and grouping gives

\[
\mathcal{L}_R = a^2 \left( 6\partial_{00} H_L (H_L + 1) + 4(H_L - 1)\nabla^2 H_L + 6|\nabla H_L|^2 \right)
\]

\[
+ a^2 \left( 2|\nabla A|^2 + 2(A - 1)\nabla^2 A \right)
\]

\[
- a^2 \left( 2H_L \nabla^2 A + 4A\nabla^2 H_L + 6\partial_0 A \partial_0 H_L \right.
\]

\[
+ 6A \partial_{00} H_L + 2\nabla H_L \cdot \nabla A \bigg). \tag{3.133}
\]
Performing the variations and removing boundary terms gives
\[
\frac{\delta \mathcal{L}_R}{\delta A} = a^2 \left[ 12 H_0 \partial_0 H_L - 4 \nabla^2 H_L \right] \tag{3.134}
\]
\[
\frac{\delta \mathcal{L}_R}{\delta H_L} = a^2 \left[ 6 H_0 (4 \partial_0 H_L - 2 \partial_0 A) - 4 \nabla^2 H_L - 4 \nabla^2 A 
+ 12 \partial_{00} H_L + 12 \left( \frac{a''}{a} + H_0^2 \right) \right] (1 + H_L - A) \right) \tag{3.135}
\]

The final portion of the gravitational action is the conformal term. Multiplying Eqn. (3.129) by the volume form Eqn. (3.131) gives
\[
6a \bar{\nabla}_\mu \bar{\nabla}_\nu a \int d^\mu \theta \equiv \mathcal{L}_\varphi \int d^\mu x \tag{3.136}
\]
where
\[
\mathcal{L}_\varphi = -6aa'' + 6a (a' \partial_0 A + a'' A - 3a'' H_L - 3a' \partial_0 H_L) 
+ \left[ 18aa' A (\partial_0 H_L - \partial_0 A) - 9aa'' (A^2 + H_L^2) \right] + 18aa' H_L (\partial_0 A - \partial_0 H_L) + 18aa'' H_L A \right) \tag{3.137}
\]

Performing the variation and removing boundary terms gives
\[
\frac{\delta \mathcal{L}_\varphi}{\delta A} = 6a^2 H_0^2 (3A - 3H_L - 1) \tag{3.138}
\]
\[
\frac{\delta \mathcal{L}_\varphi}{\delta H_L} = 6a^2 H_0^2 (3H_L - 3A + 3). \tag{3.139}
\]

### 3.3.5 Variation within the matter action

Since the matter contribution is, by definition, already varied, we only require expressions to first order in the perturbation variables. The required expression is
\[
T^\mu_\nu \delta g^\mu_\nu \sqrt{-g}. \tag{3.140}
\]

Since we have defined the stress in the vierbein frame, we require
\[
T^\sigma^\tau V^\nu_\sigma V^\mu_\tau \delta g^\mu_\nu a^4 \sqrt{-g}, \tag{3.141}
\]
where \(V^\mu_\sigma\) is given by Eqn. (3.26) to the appropriate order, with \(H^T_{ij} = B \equiv 0\) for the Newtonian gauge. Note that we have restored the omitted \(a^{-1}\) on \(V^\mu_\sigma\), and used the resulting \(a^{-2}\) to convert \(\delta g^\mu_\nu\) to \(\delta \bar{g}^\mu_\nu\). This works because \(a\) is held fixed at first-order. By direct variation of the metric, we have that
\[
\delta \bar{g}^\mu_\nu = -2 \delta A \delta^0_\mu \delta^0_\nu + 2 \delta H_L \delta^i_\mu \delta^j_\nu. \tag{3.142}
\]
Combination with Eqn. (3.131) for $\sqrt{-g}$ gives

$$-2a^4T^{00}(1 - A + 3H_L) \delta A + 2a^4(1 + A + H_L) \delta H_L \sum_i T^{ii} \tag{3.143}$$

so that applying the $1/2$ prefactor gives

$$\frac{\delta S_M}{\delta A} = -a^4(1 - A + 3H_L) \rho(\eta, x) \tag{3.144}$$

$$\frac{\delta S_M}{\delta H_L} = a^4(1 + A + H_L) \sum_i P_i(\eta, x). \tag{3.145}$$

### 3.3.6 Field Equations

To determine the field equations, we write

$$\frac{\delta \mathcal{L}_R}{\delta x} - \frac{\delta \mathcal{L}_R}{\delta x} = -16\pi G \frac{\delta S_M}{\delta x} \tag{3.146}$$

for $A$ and $H_L$. Note that the minus sign enters on the right hand side because we have moved the term to the right.

**Field equation from $A$**

At first-order in $\delta A$, we produce the equation

$$a^2 \left[ 12H_\eta \partial_\eta H_L - 4\nabla^2 H_L \right] - 6a^2H^2_\eta (3A - 3H_L - 1) = 16\pi Ga^4[1 - A + 3H_L]\rho(\eta, x) \tag{3.147}$$

To clean this up, first divide through by $4a^2$

$$3H_\eta \partial_\eta H_L - \nabla^2 H_L - \frac{3}{2}H^2_\eta(3A - 3H_L - 1) = 4\pi Ga^2(1 - A + 3H_L)\rho(\eta, x) \tag{3.148}$$

and separate off $-3H^2_\eta A$ on the left hand side

$$3H_\eta \partial_\eta H_L - 3H^2_\eta A - \nabla^2 H_L + \frac{3}{2}H^2_\eta (1 - A + 3H_L) = 4\pi Ga^2(1 - A + 3H_L)\rho(\eta, x). \tag{3.149}$$

Now we group like terms in $(1 - A + 3H_L)$ to the right

$$3H_\eta \partial_\eta H_L - 3H^2_\eta A - \nabla^2 H_L = (1 - A + 3H_L) \left[ 4\pi Ga^2 \rho(\eta, x) - \frac{3}{2}H^2_\eta \right] \tag{3.150}$$

and substitute the zero-order equation

$$H^2_\eta = \frac{8\pi G}{3} a^2 \langle \rho \rangle V \tag{3.151}$$

48
on the right to arrive at

\[ 3H_H \partial_0 H_L - 3H_H^2 A - \nabla^2 H_L = 4\pi Ga^2 (1 - A + 3H_L) \left[ \rho(\eta, x) - \langle \rho \rangle_V \right]. \quad (3.152) \]

Finally, factor out a $3H_H$ on the left and use the definition of the density fluctuation $\delta \rho(\eta, x)$

\[ \delta \rho(\eta, x) \equiv \frac{\rho(\eta, x) - \langle \rho \rangle_V}{\langle \rho \rangle_V}, \quad (3.153) \]

to write the equation in the form of Dodelson (5.27)

\[ -\nabla^2 H_L + 3H_H \left( \partial_0 H_L - AH_H \right) = 4\pi Ga^2 \delta \rho(\eta, x) \langle \rho \rangle_V + 4\pi Ga^2 \langle \rho \rangle_V \delta \rho(\eta, x)(3H_L - A). \quad (3.154) \]

Apart from a single new term, the equation is identical. To clarify the role of this new term, recall that $A$ and $H_L$ are $O(\epsilon)$, so we have schematically

\[ O(\epsilon) = 4\pi Ga^2 \delta \rho(\eta, x) \langle \rho \rangle_V [1 + O(\epsilon)], \quad (3.155) \]

which implies that

\[ \delta \rho \langle \rho \rangle_V \sim O(\epsilon). \quad (3.156) \]

In other words, we have automatically derived the consistency condition between the overdensity and the potentials. We conclude that the new term is implicitly higher order, instead of explicitly. The result is that we have regenerated the standard equation in conformal Newtonian gauge without making any assumptions a priori about the stress. This should be unsurprising because the usual definition of the stress is

\[ T^{00} \equiv - \langle \rho \rangle_V \left[ 1 + \epsilon \delta \rho(\eta, x) \right]. \quad (3.157) \]

When the zero-order terms automatically introduce a $\langle \rho \rangle_V$, this cancels the one in the stress definition leaving only the $\delta \rho$ term.

**Field equation from $H_L$**

At first order in $\delta H_L$, we produce the equation

\[ H_H \partial_0 (2H_L - A) - \frac{1}{3} \nabla^2 (H_L + A) + \partial_{00} H_L + \left( \frac{a''}{a} - \frac{H_H^2}{2} \right) (1 + H_L - A) \]

\[ = -\frac{4\pi G}{3} a^2 (1 + A + H_L) \sum_i P_i(\eta, x) \quad (3.158) \]
where we have combined terms and divided by $12a^2$. This equation is not commonly used, as it is more convenient to work with the equation got from variation of a traceless metric degree of freedom $H_T$, which we have set to zero before finding equations of motion. In any case, Eqn. (3.152) minus $1/3$ of Eqn. (3.158) regenerates [Hu 2004 §2.7, Eqn. (17d)]

$$\left[2 \frac{a''}{a} - 2H^2_\eta + H_\eta \frac{d}{d\eta} + \frac{\nabla^2}{3}\right] A - \left[\frac{d}{d\eta} + H_\eta\right] H_L' = 4\pi Ga^2 \left(\pi_L \langle P\rangle_N + \frac{1}{3}\delta_\rho \langle \rho\rangle_N\right)$$ (3.159)

under the appropriate perturbed isotropic pressure assumption

$$P_i = P_0(1 + \pi_L).$$ (3.160)

There is also an additional term

$$4\pi Ga^2 \left\{\delta_\rho \langle \rho\rangle_N (A - 3H_L) - (A + H_L) \sum_i \delta P_i \langle P_i \rangle_N\right\}$$ (3.161)

where $\delta P_i$ is an overpressure defined like Eqn. (3.153). Again, by inspection, this term is implicitly $O(\epsilon^2)$ and should be dropped. Note again how the overdensity, and now also the overpressure, automatically appear.

### 3.3.7 Conservation conditions

Since scalar perturbations are characterized by two degrees of freedom, Eqns. (3.152) and (3.158) reproduce the standard formalism, when combined with the conservation conditions [Hu 2004 §2.7]. We have shown in §3.2.3 that our stress-tensor is equivalent to that of Bardeen, in the $\beta, \epsilon$ linear regime. Since the connection coefficients are unchanged because the metric model is unchanged, this suffices to produce the same two conservation equations given by [Hu 2004 §2.7, Eqn. (18),(19)].

### 3.4 Conclusion

We have carefully generalized the cosmological stress-tensor and shown how it reduces to the standard definition under typical assumptions. Using the action formalism, we have regenerated the familiar covariant linear perturbation equations for scalar modes in the conformal Newtonian gauge. We have confirmed that the action principle automatically produces appropriate sources for the field equations relating $A$ and $H_L$: combinations of the overdensity $\delta$ and the isotropic pressure perturbation $\pi_L$. Consistency of the field equations then reveals these sources to be implicitly $O(\epsilon)$. This result strongly supports the conclusions of Chapter §2 with respect to the appropriate zero-order Friedmann source. In order to reproduce the standard results, we have restricted the stress to linear order in $\beta$. If one is not interested in a Fourier treatment of the field equations, then this restriction is almost certainly not necessary.
CHAPTER 4
LATE-TIME ACCELERATED EXPANSION FROM LOCALIZED DARK ENERGY

“Don’t worry about people stealing an idea. If it’s original, you will have to ram it down their throats.”

Howard H. Aiken, Inventor of the Harvard Mark I protocomputer

In Chapter §2, we made a careful application of the action principle to Friedmann cosmology, making no a priori assumptions about the stress-tensor. We found that, contrary to widespread intuition, pressures interior to compact objects are necessarily averaged to produce the Friedmann source. In Chapter §3, we carefully decoupled the gravitational model from the definition of the stress tensor. We further verified that $O(1)$ localized densities and pressures are perfectly viable within a cosmological stress tensor. Taken together, these results open an entirely new avenue of attack on the Dark Energy problem: “exotic” compact objects. Perhaps surprisingly, GR solutions that describe strong, localized, Dark Energy are ubiquitous. [Gliner 1966] is usually recognized as the first researcher to suggest that “$\mu$-vacuum,” localized regions of cosmological constant, could be physically realized during gravitational collapse. Amazingly, [Mazur and Mottola 2015] have traced such behaviours all the way back to Schwarzschild’s second solution from 1916. It is somewhat curious that these GEneric Objects of Dark Energy (GEODEs) are labeled “exotic”:

- they appear as BHs to local exterior observers
- they are free of physical (i.e. curvature) singularities
- they are often free of event horizons
- they are comprised of observationally confirmed material

In this chapter, we will consider whether a cosmologically averaged contribution of GEODEs could give rise to accelerated late-time expansion. We will conclude that they certainly can, without running afoul of existing astrophysical constraint. We then develop some observational signatures of such scenarios, and make some predictions for current and upcoming experiments.

4.1 Preliminaries

The simplest GEODE is the following exact, spherically symmetric, GR solution

$$ds^2 = -\zeta(r)dt^2 + \zeta(r)^{-1}dr^2 + r^2 \left[ d\theta^2 + \sin^2 \theta \, d\phi^2 \right]$$

(4.1)
with

\[
\zeta(r) \equiv \begin{cases} 
1 - (r/2M)^2 & r \leq 2M \\
1 - 2M/r & r > 2M.
\end{cases}
\] (4.2)

This solution is a Schwarzschild BH exterior for \( r > 2M \), and a vacuum with cosmological constant interior to \( r < 2M \). This solution is static, consistent with the intuition that material under tension stays bound. Note that the interior energy density correctly integrates to \( M \). This means that a vacuum observer at infinity will perceive a point mass \( M \). The interior region contains no physical singularity, but the infinite pressure gradient at \( r = 2M \) in Eqn. (4.1) motivates more sophisticated GR solutions.

Dymnikova [1992, Eqn. (14)] produced a “G-lump” model without an infinite pressure gradient. About a decade later, Mazur and Mottola [2004] argued a “gravastar” from QFT motivations, and Chapline [2003] argued a very similar object from condensed matter motivations. A common feature of such objects is a “crust” region of finite thickness slightly beyond \( r = 2M \), which surrounds the de Sitter core. These objects need not have a horizon: the interior may be in causal contact with the outside universe. In fact, this ability to resolve the BH information paradox sparked renewed interest in exotic objects during the “firewall” debates initiated by Almheiri et al. [2013].

The gravastar GEODE, in particular, has received significant theoretical attention. Theoretical checks on gravastar stability to perturbations [e.g. Visser and Wiltshire, 2004; Lobo, 2006; DeBenedictis et al., 2006] and stability to rotation [e.g. Chirenti and Rezzolla, 2008; Uchikata and Yoshida, 2015; Maggio et al., 2017] have found large ranges of viable parameters describing the crust. Phenomenological checks from distinct merger ringdown signatures, such as an altered quasinormal mode spectrum [e.g. Chirenti and Rezzolla, 2007] have been computed. In addition, possible optical signatures from the crust itself [e.g. Broderick and Narayan, 2007], accretion disks [e.g. Harko et al., 2009], and even direct lensing [Sakai et al., 2014] have been investigated.

The first direct detection of gravitational radiation from massive compact object mergers has greatly stimulated critical analysis of exotic object scenarios. Bhagwat et al. [2016, 2017] show that the existing aLIGO interferometer and the proposed Cosmic Explorer and Einstein telescopes can resolve the higher harmonics, which could probe the existence and properties of GEODEs. On the other hand, Chirenti and Rezzolla [2016] have already claimed, using the ringdown, that GW150914 likely did not result in a gravastar GEODE final state. Yunes et al. [2016] criticize the Chirenti and Rezzolla [2016] claim as premature, yet argue that ringdown decay faster than the light crossing time of the remnant poses the more severe challenge for any GEODE model.

In all phenomenology to date, however, GEODE observational signatures depend critically on the unknown properties of the presumed crust. As we have shown in Chapter 4, cosmological dynamics must be sensitive to the interior structure of such objects, if they exist in Nature. We will exploit this result to develop complementary observational signatures dominated instead by the common properties of their cores. For our investigation, consistent with the theoretical arguments against classical BH boundaries and interiors,
we will replace all BHs with GEODEs.

4.1.1 Notation and conventions

In this chapter, we will frequently work with observational quantities. For consistency with existing literature, and to keep the notation clean, we will adjust slightly the notation of this chapter, relative to the other chapters.

Use of comoving densities

In the astrophysical literature, the use of comoving densities is standard. A comoving density is defined, in the context of Friedmann cosmology, as

\[ \rho_{\text{comoving}} \equiv \rho_{\text{physical}}(\eta)^3 \]

In this chapter, if we need to explicitly denote a physical density, we will overset it with a tilde.

Use of Hubble time and critical density

It is customary to express densities as fractions \( \Omega \) of the critical density at \( a = 1 \)

\[ \rho_{\text{cr}} \equiv \frac{3H_0^2}{8\pi G} \]

where \( H_0 \) is the Hubble constant at \( a = 1 \). Note that a tilde here changes nothing, because \( a = 1 \). Unless otherwise noted, we take the time unit as the reciprocal present-day Hubble constant \( H_0^{-1} \) and the density unit as the critical density today \( \rho_{\text{cr}} \). As before, we define the speed of light \( c \equiv 1 \).

Nomenclature

For efficiency, we will often describe quantities with scale factor or redshift dependence as time-dependent. We will also freely use either scale factor \( a \) or redshift \( z \), depending on which produces the more compact quantitative expression. Though we replace all BHs with GEODEs, for consistency with existing astrophysical literature, we will often refer to these objects as BHs.

Organization

The rest of this chapter is organized as follows. In §4.2, we construct a simple model of a GEODE population’s cosmological contribution. This produces quantitative predictions relating the DE density to the observed BH population. In §4.3, we demonstrate that stellar collapse between redshift \( 8 < z < 20 \) can readily produce enough GEODEs to account for the present-day DE density. In §4.4, we analyze the GEODE-induced DE at late times (\( z < 5 \)) in the dark fluid framework described in Ade et al. [2016]. Here, we
produce predictions for the DE equation of state parameters $w_0$ and $w_a$ in terms of present-day BH population observables. We also verify that present-day BH population parameters are consistent with Planck and DES constraint of $w_0$ and $w_a$.

A few appendices present material complementary to the phenomenological results of this chapter. Appendix A estimates the effect of pressures within large, virialized systems, like clusters. Appendix B develops a simple model for the BH population in terms of the stellar population.

### 4.2 GEODE contribution to the Friedmann equations

We wish to model the cosmological effect of a GEODE population in a way that does not depend on the details of the specific GEODE model. To guide our efforts, we feel it best to quote, at length, Gliner [1966]

“The meaning of a negative pressure is that the internal volume forces in the matter are not forces of repulsion (as they are for the media accessible to observation, which consist of particles), but forces of attraction. This implies the assumption that the usual mechanisms which oppose the merging of particles of matter are annulled . . . If, as has been assumed, the vacuum-like state is actually the lowest possible state of matter, a $\mu$-vacuum [Dark Energy] is the only possible final result of the process of contraction under the action of forces of negative pressure. The hypothetical process we have considered is interesting in connection with the problem of the final state of matter which has undergone gravitational collapse. According to the ideas developed here this state is a $\mu$-vacuum. The hypothesis that transitions between ordinary and $\mu$-vacuum states of matter are possible raises questions about the quantum conservation laws, primarily the law of conservation of the baryon number.”

Such ideas have now become mainstream research topics [e.g. Bassett et al., 2006]. During Big Bang ignition or reheating, the inflationary Dark Energy state decays into baryons during an $O(1)$ gravitational expansion. Gliner essentially proposes the time-reversal of reheating: baryons “coalesce” into a Dark Energy state during $O(1)$ gravitational collapse.

This suggests a model-independent way to characterize the formation and growth of a GEODE population for $z \lesssim 20$. We consider a two-component model of the zero-order comoving energy density and pressure

\[
\langle \rho \rangle_V \equiv \rho_{\text{dust}}(\eta) + \rho_s(\eta) \\
\sum_i \langle P_i \rangle_V \equiv 0 + 3w_s(\eta)\rho_s(\eta).
\]  

The dust, described by $\rho_{\text{dust}}$, will encode both dark matter and baryonic contributions. As usual, and as justified in Appendix A, we may approximate dust as pressureless. The GEODE contribution will be encoded by an equation of state $w_s$ and a density $\rho_s$, initially defined to be zero. To capture the effect of GEODE
formation, we simply begin depleting baryonic matter by $\Delta_b(\eta)$

$$\rho_{\text{dust}}(\eta) \equiv \rho_{\text{dust}}^0 - \Delta_b(\eta). \quad (4.7)$$

Here $\rho_{\text{dust}}^0$ is a constant, since the comoving dust density remains fixed. The cumulative density of baryons converted to GEODEs is $\Delta_b(\eta)$. We may use conservation of stress-energy to relate $\Delta_b(\eta)$ to $\rho_s(\eta)$. The zero-order covariant conservation of stress-energy statement Eqn. (2.64) becomes

$$-\partial_0 \left\{ a^{-3} \left[ \rho_{\text{dust}}^0 - \Delta_b(\eta) + \rho_s(\eta) \right] \right\} - 3H_\eta a^{-3} \left[ \rho_{\text{dust}}^0 - \Delta_b(\eta) + \rho_s(\eta) \right] = 3w_s\rho_s(\eta)H_\eta a^{-3} \quad (4.8)$$

where we have added appropriate factors of $a^{-3}$ to move from comoving densities to physical densities required for the conservation statement. Performing the derivative and multiplying through by $a^3$ gives

$$\partial_0 \rho_s + 3w_sH_\eta \rho_s = \partial_0 \Delta_b. \quad (4.9)$$

This should be recognized as an out-of-equilibrium decay scenario [c.f. Kolb and Turner [1994] §5.3], with the peculiarity that the “decay” is non-exponential.

Before we can solve Eqn. (4.9), we must characterize $w_s$. Both the G-lump and gravastar have a transition region, which ultimately encloses a DE interior. The possible nature of such “crusts” has been studied extensively by Visser and Wiltshire [2004], Martin-Moruno et al. [2012] and others. There is consensus that substantial freedom exists in the construction of GEODE models. This freedom, while constrained, permits a range of crust thicknesses and equations of state. We will consider the following structure

- $r \leq 2M$: de Sitter interior
- $2M < r \leq 2M + \ell$: crust region
- $r > 2M + \ell$: vacuum exterior

where $M$ is a GEODE mass and $\ell$ is a crust thickness. Note that we have simplified the model by placing the inner radius at the Schwarzschild radius. The cosmological contribution from a GEODE population thus consists of two components

$$\rho_s \equiv \rho_{\text{int}} + \rho_{\text{crust}} \quad (4.10)$$
$$P_s \equiv -\rho_{\text{int}} + P_{\text{crust}} \quad (4.11)$$

where interior and crust contributions come from averages over the population. We may define an equation of state for the GEODE source

$$w_s \equiv \frac{-\rho_{\text{int}} + P_{\text{crust}}}{\rho_{\text{int}} + \rho_{\text{crust}}}, \quad (4.12)$$
Mazur and Mottola [2004] assert that most of the energy density is expected to reside within each core and that \( \mathcal{P} \sim \rho \) within each crust. It is thus reasonable to expand the aggregate quantities and discard higher-order terms

\[
w_s = -1 + \frac{\mathcal{P}_{\text{crust}}}{\rho_{\text{int}}} \left( 1 - \frac{\rho_{\text{crust}}}{\rho_{\text{int}}} + \ldots \right) \tag{4.13}
\]

\[
w_s \approx -1 + \frac{\mathcal{P}_{\text{crust}}}{\rho_{\text{int}}}. \tag{4.14}
\]

This expression successfully isolates ignorance of the crust to a small dimensionless parameter, which we now define as

\[
\chi(a) \equiv \frac{\mathcal{P}_{\text{crust}}}{\rho_{\text{int}}}. \tag{4.15}
\]

To avoid confusion, we emphasize now that \( w_s \) is not the Dark Energy equation of state as constrained by Planck. Switching from conformal time \( \eta \) to scale factor \( a \) in Eqn. (4.9) gives

\[
\frac{d\rho_s}{da} + 3w_s \frac{\rho_s}{a} = \frac{d\Delta_b}{da} \tag{4.16}
\]

This equation can be separated with an integrating factor

\[
\mu(a, a_c) \equiv \exp \left( 3 \int_{a_c}^a \frac{w_s(a')}{a'} da' \right) \tag{4.17}
\]

\[
= \frac{1}{a^3} \exp \left( 3 \int_{a_c}^a \frac{\chi(a')}{a'} da' \right) \tag{4.18}
\]

where we have defined the proportionality to be unity and \( a_c \) is a cutoff below which there are no GEODEs. If we omit the cutoff parameter for \( \mu \), then we implicitly assume \( a_c \). The resulting energy density is

\[
\rho_s = \frac{1}{\mu(a)} \int_0^a \frac{d\Delta_b}{da'} \mu(a') \, da', \tag{4.19}
\]

where the lower-bound does not depend on \( a_c \) because \( \Delta_b/da \) vanishes by definition before \( a_c \). For simplicity, we will study the phenomenology of GEODEs with the single parameter \( \chi \) held fixed.

To develop some intuition for Eqn. (4.19), consider a sequence of instantaneous conversions

\[
\frac{d\Delta_b}{da} = \sum_n Q_n \delta(a - a_n) \tag{4.20}
\]

where \( Q_n \) is the comoving density of baryons instantaneously converted at \( a_n \). Substitution into Eqn. (4.19)
gives
\[ \rho_s = \sum_{n}^{a_n \leq a} Q_n \frac{\mu(a_n)}{\mu(a)}. \] (4.21)

Explicitly, at the instant of the \( m \)-th conversion \( a_m \), we have the following comoving density of DE
\[ \rho_s(a_m) = Q_m + \sum_{n}^{a_n \leq a_m} Q_n \frac{\mu(a_n)}{\mu(a_m)}. \] (4.22)

Thus, at the instant of conversion, \( \rho_s \) agrees with a spatial average over the newly formed GEODEs’ interiors. In other words, a quantity of baryonic mass has been converted to an equal quantity of DE. Beyond the instant of conversion, however, the contribution due to GEODEs dilutes more slowly than the physical volume expansion. For example, in the de Sitter limit of \( \chi \rightarrow 0 \), we find that \( \mu \rightarrow a^{-3} \) and
\[ \tilde{\rho}_s(a) = \sum_{n}^{a_n \leq a} \frac{Q_n}{a_n^3} \quad (\chi \equiv 0). \] (4.23)

In this case, the converted physical baryon density becomes “frozen in” as DE at the time of conversion.

### 4.2.1 Relation of the model to observables

The fundamental observable for any population study of GEODEs is the epoch dependent BH mass function (BHMF)
\[ N_{BH}(a) \equiv \int_{0}^{\infty} \frac{dN_{BH}}{dM_{BH}} dM_{BH}, \] (4.24)

where \( N_{BH}(a) \) is the number count of BHs at \( a \). Fishbach and Holz [2017] have shown that even a few mergers permit useful constraint of \( \frac{dN_{BH}}{dM_{BH}} \). Kovetz et al. [2017] anticipate the BHMF can be measured to greater than 10% accuracy after a few years of aLIGO at design sensitivity. As detailed by Dwyer et al. [2015], the proposed Cosmic Explorer interferometer could place excellent constraints on the BHMF: they expect \( \sim 10^5 \) events per year, and out to redshift \( z \sim 10 \). We must relate this observable quantity to \( \rho_s \). Since our model is defined at zero-order, however, we are unable to localize the energy responsible for \( \rho_s \) \( a \) priori. We consider two scenarios, a local interpretation and a non-local interpretation.

#### Local interpretation

Consider again the \( \chi \rightarrow 0 \) limit. To locally reproduce the behaviour of Eqn. (4.23), each GEODEs’ mass \( Q_n \) must increase such that
\[ Q_n \propto a^3, \] (4.25)
to keep the physical density constant as the objects themselves dilute in number density. The assertion here is that GEODEs experience a cosmological blueshift. There is ample precedent for this sort of behaviour in RW cosmology: photons cosmologically redshift. The Strong Equivalence Principle [e.g. Will 1993 §3.3] is not violated, because a quasilocal Lorentz frame cannot be defined on cosmological timescales. In fact, there is already precedent in GR for cosmologically coupled evolution within compact object models. For example, Nolan [1993] constructs an $O(1)$ interior source which joins correctly to the arbitrary FRW-embedded point mass solution of McVittie [1933]. Its interior densities and pressures evolve cosmologically\footnote{In the case of Nolan’s solution, they evolve so as to keep the mass of the object fixed. This follows from constraints imposed by McVittie in the construction of his original solution. His specific intent was to keep the mass fixed.} Such behaviour is also conceivable within scalar-tensor theories.

This interpretation leads to a number of observational consequences. Immediately, given some fiducial volume $V$, we conclude that

$$
\rho_s(a) = \frac{1}{V} \int_0^\infty M_{BH} \frac{dN_{BH}}{dM_{BH}} dM_{BH}.
$$

In other words, the DE density is proportional to first moment of the BHMF, at fixed epoch. Consider a $2M_\odot$ BH produced at $z = 2$, a bit before the peak of star formation. Suppose we observe this BH in a binary merger at $z = 0.18$. Its mass at merger becomes

$$
2M_\odot \frac{(1 + 2)^3}{(1 + 0.18)^3} = 33.6 M_\odot.
$$

We see that a local GEODE blueshift could provide a novel explanation for lower mass gap. Now consider a $100M_\odot$ BH produced at $z = 20$ and observed today

$$
100M_\odot(1 + 20)^3 = 0.9 \times 10^6 M_\odot.
$$

Completely neglecting accretion, such an object need only merge with 3 other similar objects to reproduce the mass of the supermassive BH Sgr A* at the center of our own galaxy. We see that a local GEODE blueshift could readily produce the observed masses of supermassive BHs. Such a process may significantly affect the growth and evolution of galaxies. It can be modeled and falsified. Most directly, Eqn. (4.25) directly impacts the shape of the BHMF. It will become top-heavy, relative to the underlying Salpeter distribution initially inherited from the stellar IMF. A thorough treatment of these consequences is beyond the scope of the present work.
Non-local interpretation

Another interpretation is that energy beyond the initially converted baryonic mass resides in the gravitational field. In other words, local GEODE masses remain (cosmologically) fixed. Under this interpretation,

\[ \Delta_b(a) = \frac{1}{V} \int_0^\infty M_{BH} \frac{dN_{BH}}{dM_{BH}} \, dM_{BH} \]  

(4.29)

and so \( \Delta_b \) is essentially the first moment of the BHMF. Given this interpretation, Eqn. (4.14) must be interpreted as a motivation for \( \chi \), and nothing more.

4.3 Resolution of the Coincidence Problem

The prediction given in Eqn. (4.19) tightly correlates the DE density to the matter density. This suggests a natural resolution to the coincidence problem: GEODEs are responsible for all of the cosmological dark energy. Since late-time accelerated expansion is a zero-order effect, the phenomenology is independent of one’s interpretation of the decay Eqn. (4.9). That such a small amount of baryons could conspire to produce a zero-order effect is plausible because the \( \mu(a, a_c) \approx a^{-3} \) in Eqn. (4.19) strongly amplifies the effect of conversion at early times. The earliest possible time that stellar collapse GEODEs could begin to contribute is shortly after the birth of Population III stars. We will regard that time, the Dark Ages, as beginning at \( z_{early} \equiv 20 \).

At present, there are large uncertainties in Population III star production and Eqn. (4.19) depends on the rate of depletion. To bound the consequences of conversion to GEODEs, we will approach the question in two complementary ways. We first investigate a “burst” production, where all baryonic matter that will ever become a GEODE is converted at one instant in time. Then we will investigate a very slow, but constant, production which is quenched near reionization.

4.3.1 Technique I: Instantaneous formation epoch from present-day BH density

In this section, we will determine an instantaneous formation time \( a_f \) for all GEODEs such that the correct \( \Omega_\Lambda \) is obtained. Note all current BH population models do not consider the possibility of cosmological effects. So, we can use literature constraint of \( \Omega_{BH} \) as a faithful proxy to \( \Delta_b/\rho_{cr} \). When gravitational wave observatories have constrained this density directly, this estimate can be checked anew depending on the interpretations discussed in §4.2.1. Until then, we will use the stellar population based BH model developed in Appendix B.

We now assume that all GEODEs are produced at a single moment in time \( a_f \)

\[ \frac{d\Delta_b}{da} = \Omega_{BH}\delta(a - a_f). \]  

(4.30)
Figure 4.1 Estimated instantaneous GEODE formation redshift \( z_f \), consistent with present day DE density \( \Omega_\Lambda \). The horizontal axis gives the deviation \( \chi \) from a perfect de Sitter (\( w_s = -1 \)) contribution to the cosmological source, due to the presence of GEODE crusts. Note that the onset of star formation \( z_{\text{early}} \equiv 20 \) encloses the viable region (green/light grey) for \( \chi < 0.16 \). The red (medium grey) region indicates \( \Omega_\Lambda \) within a factor of 1.5 of the present day value, while the blue (dark grey) region indicates \( \Omega_\Lambda \) within a factor of 0.5 of the present day value.

We take the total cosmological density in GEODEs to be the present-day values

\[
\Omega_{\text{BH}} = \begin{cases} 
3.02 \times 10^{-4} & \text{(rapid)} \\
3.25 \times 10^{-4} & \text{(delayed)}
\end{cases}
\]  

(4.31)

Here, rapid and delayed refer to the stellar collapse model as detailed by Fryer et al. [2012]. These values are likely slight underestimates due to accretion, which is neglected in Appendix B. Note that these values are consistent with the fraction of stars that collapse to BH, as estimated by Brown and Bethe [1994]. Substituting the approximation Eqn. (4.30) into Eqn. (4.19) we find

\[
z_f = \left( \frac{\Omega_\Lambda}{\Omega_{\text{BH}}} \right)^{1/(3(1-\chi))} - 1,
\]  

(4.32)
Figure 4.2 Estimated dark energy density induced by GEODE formation from onset of stellar formation at $z_{\text{re}} \equiv 20$ until some cutoff $z_f$. This assumes a constant, but conservative, primordial rate of stellar density production $d\rho/dt = 5.3 \times 10^{-3}$. This value is consistent with the more conservative $\gamma$-ray opacity constraints reported by Inoue et al. [2014] at $z \sim 6$ (see Table 4.1). Further assumptions include a fixed fraction of stellar density collapsing into BH and a constant amplification due to accretion. Note consistency with Figure 4.1 and convergence in the de Sitter limit $\chi \rightarrow 0$. Green (light grey) indicates production of the present-day $\Omega_\Lambda$, red (medium grey) indicates excessive production up to $3\Omega_\Lambda/2$, and blue (dark grey) indicates insufficient production no less than $\Omega_\Lambda/2$.

where we have again fixed $\chi$, and used that $\mu(1, a_{\text{c}}) = 1$ for any fixed $\chi$. The instantaneous formation epoch is displayed in Figure 4.1.

4.3.2 Technique II: Constant formation from onset, rate consistent with $\gamma$-ray opacity

In this section, we will assume a slow, but constant, rate of Dark Ages GEODE production. We will use astrophysical constraints, distinct from those of §4.3.1 determined from $\gamma$-ray opacity reported by Inoue et al. [2014]. We will assume that a constant fraction of stellar density collapses to GEODEs between stellar onset $z_{\text{early}}$ up until some cutoff time $z_f$. We will determine the value of $z_f$ such that the depleted baryon density $\Delta_b$ yields the observed $\Omega_\Lambda$ today. We will further assume that the stellar density formation rate during $z_f < z < z_{\text{early}}$ is constant.
Table 4.1. Constraints on Population III star formation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value (natural units)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \frac{d\rho_g}{dt}, z \approx 20 )</td>
<td>(&lt; 5.4 \times 10^{-2} )</td>
<td>[Inoue et al. 2014, Fig. 2]</td>
</tr>
<tr>
<td>( \frac{d\rho_g}{dt}, z \approx 6 )</td>
<td>(&lt; 5.4 \times 10^{-3} )</td>
<td>[Inoue et al. 2014, Fig. 2]</td>
</tr>
<tr>
<td>( A )</td>
<td>(&lt; 22.7 )</td>
<td>[Abel et al. 2002]</td>
</tr>
<tr>
<td>( \Xi )</td>
<td>(&lt; 10^{-1} )</td>
<td>Table B.4</td>
</tr>
</tbody>
</table>

Note. — data are used to constrain an approximate GEODE formation epoch, which suffices to produce all present-day dark energy density \( \Omega_\Lambda \) by some cutoff redshift \( z_f \). Variables defined in the text. The results of this analysis can be found in Figure 4.2.

Under these assumptions, we find that

\[
\Omega_\Lambda = \int_{a_{\text{early}}}^{a_f} a^{-3(1-\chi)} A \Xi \frac{d\rho_g}{da} \, da. \quad (4.33)
\]

Here we have introduced a constant collapse fraction \( \Xi \) and a factor \( A \) to account for accretion processes. We will take \( A \approx 10 \). This is reasonable, given that [Li et al. 2007] have shown that accretion can be very efficient at primordial times. With the constant star formation rate assumption, we may use the chain rule to write

\[
\Omega_\Lambda = A \Xi \frac{d\rho_g}{dt} \int_{a_{\text{early}}}^{a_f} \frac{da}{H a^{-\Lambda-3\chi}}. \quad (4.34)
\]

We present reasonable data for the above parameters in Table 4.1 where the formation rates listed are upper limits. Given the matter-dominated Hubble factor

\[
H = \frac{\sqrt{\Omega_m}}{a^{3/2}}, \quad (4.35)
\]

we may integrate Eqn. (4.34) and solve for the cutoff time. The result is

\[
a_f = \left[ \frac{3 \Omega_\Lambda \sqrt{\Omega_m}}{A \Xi} \left( \frac{d\rho_g}{dt} \right)^{-1} \left( \chi - \frac{1}{2} \right) + a_{\text{early}}^{3(\chi-1/2)} \right]^{1/3(\chi-1/2)}. \quad (4.36)
\]

The behavior of Eqn. (4.36), converted to redshift, is shown in Figure 4.2.

Note that we have taken the 10× more conservative bound established at \( z \approx 6 \). If we take \( \Xi \sim 0.01 \), there is no viable space. This is consistent with the stellar model of Appendix B where the primordial collapse fraction is \( \Xi \sim 0.1 \). It is not surprising that the collapse remnant model’s primordial \( \Xi \) is larger than the value of [Brown and Bethe 1994], because their analysis depends on iron cores, which are not present.
in Population III stars. Note that the fraction of matter density converted to GEODE material under our assumptions,

$$\Omega_{BH} = A \Xi \frac{d\rho_g}{dt} \int_{a_{early}}^{a_f} \frac{da}{Ha} < 6.5 \times 10^{-4}, \quad (4.37)$$

is consistent with the present-day $\Omega_{BH}$ estimates, which do not consider possible cosmological effects on the constituent object masses.

### 4.3.3 Discussion

In this section, we have considered whether GEODEs alone could account for the present-day observed DE density. This is plausible because the DE induced by GEODE formation is amplified by $\sim 1/a^3$. We used two complementary techniques, both with respect to formation rate and observational constraint. Through both techniques, we have estimated an approximate formation epoch for GEODEs, which suffices to produce the present-day observed $\Omega_\Lambda$. Encouragingly, both techniques agree that sufficient production is viable over a large range of $\chi$. This resolves the coincidence problem [e.g. Amendola and Tsujikawa 2010, §6.4] of a narrow permissible value for $\Omega_\Lambda$. Furthermore, both techniques converge toward $z_f \sim 12$ and produce excluded regions for GEODE $\chi$

$$\chi \lesssim 1.6 \times 10^{-1} \quad \text{(instantaneous)} \quad (4.38)$$
$$\chi \lesssim 6.0 \times 10^{-2} \quad \text{(constant cutoff)} \quad (4.39)$$

Above these thresholds, resolution of the coincidence problem by GEODEs is disfavored. As we will soon see, this exclusion is consistent with those based on late-time behavior and existing Planck constraint.

This result is very useful for many reasons. Most importantly, the viable region lies squarely in the middle of the epoch of Population III star formation. This suggests that a naturally emerging population of GEODEs can produce the correct DE density. In addition, the baryonic density $\Delta_b/\rho_{cr}$ converted to induce this effect is $\sim 0.1\%$ of $\Omega_m$. This is well within uncertainties on the Planck best-fit value for $\Omega_m$. Since $z_f \ll z_{CMB}$, production of a sufficient population can occur without breaking agreement with precision CMB astronomy. In other words, CMB anisotropies are not altered at any level. Finally, GEODE formation at $z < 20$ can establish a present-day valued DE density after star formation has begun. Thus, initial conditions, and therefore results, of precision N-body simulations are not altered. For the subsequent discussion, we interpret these estimates in the following way: if GEODEs are to resolve the coincidence problem, we expect that most of the dark energy density will be established by stellar and accretion physics taking place during $5 \leq z \leq 20$. 

63
4.4 Late-time observational consequences

In this section, we will focus on the late-time (z < 5) observational consequences of Eqn. (4.19). Toward the end of this epoch, DE has already been constrained by many experiments, e.g. Planck and Dark Energy Survey (DES). In the case of Planck, DE is constrained through three parameters: an energy density Ω_Λ and two coefficients of a linear Taylor expansion about a = 1. These coefficients, w_0 and w_a, describe a DE equation of state w_{\text{eff}}

\[ w_{\text{eff}}(a) \equiv w_0 + w_a(1 - a), \]  

(4.40)

which evolves in time assuming [see Ade et al., 2016, §3.1] the dynamics a minimally coupled scalar field. This evolution assumption leads to the following conservation condition

\[ \frac{d\rho_s}{dt} \equiv -3w_{\text{eff}}H\rho_s. \]  

(4.41)

In the GEODE scenario, however, baryons are coupled to DE through stellar collapse and subsequent accretion. If we wish to investigate compatibility of the GEODE scenario under existing observational constraints, we must reinterpret the time-dependence of \( \rho_s \) as the dynamics of a decoupled fluid. This leads to an effective equation of state parameter, as indicated by the notation used in Eqn. (4.41). Subtracting Eqn. (4.9) from Eqn. (4.41), switching to scale factor, substituting Eqn. (4.19), and solving for \( w_{\text{eff}} \) gives

\[ w_{\text{eff}} = w_s - \frac{a\Delta'_{b}}{3\rho_s}, \]  

(4.42)

where prime denotes derivative with respect to a. Note immediately that \( a\Delta'_{b}/3\rho_s \leq 0 \) always. In other words, phantom energy is a possible “symptom” of using the decoupled fluid assumption to describe a GEODE formation scenario. In this setting, there is no phantom energy, just an incorrect assumption about the underlying physics.

Local interpretation \( w_{\text{eff}} \)

We digress for a moment to discuss a feature of the local interpretation. The local interpretation bypasses the need to work with the depleted baryon density \( \Delta_0 \), because \( \rho_s \) becomes proportional to the first moment of the BHMF. Substitution of Eqn. (4.26) into Eqn. (4.41) gives

\[ w_{\text{eff}} = -\frac{a}{3} \frac{d}{da} \ln \left[ \int_0^\infty M_{\text{BH}} \frac{dN_{\text{BH}}}{dM_{\text{BH}}} dM_{\text{BH}} \right] \]  

(local interpretation).

(4.43)

The observationally relevant quantity, the BHMF, can be directly input.
Interpretation-independent relations

In the non-local interpretation, $\Delta_b$ becomes proportional to the first moment of the BHMF. In this section, we develop interpretation-independent relations between $w_{\text{eff}}$ and $\Delta_b$. These are not only useful for direct experimental constraint of the non-local interpretation. As mentioned earlier, existing and simple BH population models traditionally assume no cosmological effects. Under this assumption, and neglecting accretion, the distribution of mass within BHs directly describes $\Delta_b$.

Using the definition of $\mu$ in Eqn. (4.17) and linearity of the integral, we may re-express this relative to the present-day value of $\Omega_\Lambda$

$$w_{\text{eff}} = w_s - \frac{d\Delta_b}{da}\mu(a, 1) \left[ \frac{3}{a} \left( \frac{\Omega_\Lambda}{a} - \int_a^1 \frac{d\Delta_b}{da'}\mu(a', 1) \, da' \right) \right]^{-1}. \quad (4.44)$$

This form is advantageous as it samples only values very near to the present epoch, where constraint and systematics are well-characterized. The first-order Taylor expansion in $(1-a)$ of Eqn. (4.44) does not contain $\mu$ at all

\begin{align*}
^{(0)}w_{\text{eff}} &= -1 + \left( \chi - \frac{\Delta'_b}{3\Omega_\Lambda} \right) \bigg|_{a=1} \quad (4.45) \\
^{(1)}w_{\text{eff}} &= \frac{\Omega_\Lambda(\Delta''_b - 2\Delta'_b) - \Delta'^2_b}{3\Omega_\Lambda^2} \left. + \left( \frac{\chi\Delta'_b}{\Omega_\Lambda} - \chi' \right) \right|_{a=1}. \quad (4.46)
\end{align*}

The coefficient Eqns. (4.45) and (4.46) will give a region which can be immediately compared against any Planck-style $w_0 - w_a$ constraint diagram. The results of §4.3 suggest that most of $\Omega_\Lambda$ should be already produced by $z \sim 5$. In this case, we may approximate Eqn. (4.44) as

$$w_{\text{eff}} \approx w_s - \Delta'_b\mu(a, 1) \frac{a}{3\Omega_\Lambda}. \quad (4.47)$$

We have shown that knowledge of the depleted baryon density and the present-day dark energy density completely determines $w_{\text{eff}}$. We already have a wealth of cosmological observations, however, with upcoming experiments [e.g. Levi et al., 2013, DESI] set to constrain $w_{\text{eff}}$. We may thus make predictions by inverting the above procedure. Solving for $d\Delta_b/da$ by differentiating Eqn. (4.44) and re-integrating, we find

$$\frac{d\Delta_b}{da} = 3\Omega_\Lambda \frac{w_s(a) - w_{\text{eff}}(a)}{a} \exp \left[ 3 \int_a^1 \frac{w_{\text{eff}}(a') \, da'}{a'} \right]. \quad (4.48)$$

Note that we have removed $\mu$ entirely through use of Eqn. (4.17). Given the particularly simple $w_{\text{eff}}$ assumed by Planck, we may express Eqn. (4.48) in closed form

$$\frac{d\Delta_b}{da} = 3\Omega_\Lambda \left[ \chi - (1 + w_0 + w_a) + w_a a \right] \frac{\exp \left[ 3w_a(a - 1) \right]}{a^{3(w_0 + w_a) + 1}}. \quad (4.49)$$
Figure 4.3 Time-evolution of GEODE $w_{\text{eff}}(a)$ estimated from the stellar population. Both rapid (orange/light grey) and delayed (blue/dark grey) stellar collapse scenarios are shown. Best-fit lines (dashed) assuming the Planck linear ansatz near $a = 1$, shown for $\chi = 3 \times 10^{-2}$ only. Colors (greyshades) indicate distinct $\chi$ (c.f. Figure 4.4). Note breakdown of the dark fluid (linear) approximation at $a = 0.9$.

4.4.1 Estimation of $w_{\text{eff}}$ from the comoving stellar density

In the GEODE scenario, the physical origin of DE is completely different than the internal dynamics of a scalar field. The linear model Eqn. (4.40), based on the dark fluid assumption, may not be the most suitable for characterization of a late-time GEODE DE contribution. In this section, we use a BH population model, developed from the stellar population in Appendix B to investigate $w_{\text{eff}}$ in the GEODE scenario. This BH population model estimates $\Delta b$, and is therefore appropriate for determining consistency of both interpretations.

The BH population model of Appendix B does not account for accretion, so we must remain outside of any epoch where BH accretion is significant. Consistent with the simulations of Li et al. [2007], we consider only $z < 5$. To begin evaluation of Eqn. (4.44), we require $\mu(a, 1)$, which requires $\chi$. For fixed $\chi$, the required integrations become trivial. We find

$$\mu(a, 1) = \frac{1}{a^3(1-\chi)}$$  

(4.50)
and predict for Eqn. (4.47)
\[
\text{weff} \approx -1 + \chi - \frac{\Delta'_b}{3\Omega_A a^{2-3\chi}}.
\] (4.51)

This relation is displayed in Figure 4.3. The range corresponds to “rapid” and “delayed” models of stellar collapse, as determined by Fryer et al. [2012]. The collapse model determines the distribution of remnant masses, which evolves in time with metallicity. Given \(\Delta_b\), Eqn. (4.51) provides a single parameter fit, with sub-percent precision, viable for \(z < 5\). It is clear that the Planck linear ansatz is only useful for
\[
a > 0.9 \quad z < 0.11
\] (4.52)
with departures growing significantly worse as \(\chi \rightarrow 0\).

### 4.4.2 Planck and DES constraint of \(\chi\) at the present epoch

Physically, \(d\Delta_b/da \geq 0\) must be satisfied wherever ansatz Eqn. (4.40) is valid. From Eqn. (4.48), this constraint becomes an upper bound
\[
\text{weff}^{(0)} \leq \chi^{(1)} - 1.
\] (4.53)

Regardless of any time-evolution in \(\chi\), we may use Eqns. (4.45) and (4.46) to constrain \(\chi\) now. Eliminating \(\chi\) from these equations, we find a permissible line through \(w_0 - w_a\) space
\[
\text{weff}^{(1)} = \frac{\Delta'_b}{\Delta''_b} \left( \text{weff}^{(0)} + \frac{\Delta''_b + \Delta'_b}{3\Omega_A} - \chi^{(1)} \right).
\] (4.54)

We will use the BH population model, developed from the stellar population in Appendix B, to estimate the derivatives of \(\Delta_b\). From Eqn. (B.18), we numerically find that
\[
\Delta'_b^{(1)} = \begin{cases} 5.364 \times 10^{-5} & \text{(rapid)} \\ 6.135 \times 10^{-5} & \text{(delayed)} \end{cases}
\] (4.55)

\[
\Delta''_b^{(1)} = \begin{cases} -2.214 \times 10^{-4} & \text{(rapid)} \\ -2.447 \times 10^{-4} & \text{(delayed)} \end{cases},
\] (4.56)

which will produce a permissible band in \(w_0 - w_a\) space. For \(\chi' \equiv 0\), this band is displayed in Figures 4.4 and 4.5. Evidently, Planck data disfavor GEODEs with large \(\chi\)
\[
0 < \chi^{(1)} \leq 3 \times 10^{-2} \quad \chi'(1) \equiv 0.
\] (4.57)

Within this range, however, GEODEs are consistent with Planck best fit constraints.
By inspection of Eqn. (4.54), $\chi'$ simply translates the constraint region vertically. If we are to remain consistent with $d\Delta_b/da \geq 0$ for $0.9 < a < 1$, we must have

$$w_{\text{eff}}^{(1)} \leq -10w_{\text{eff}}^{(0)} + 10[\chi(a) - 1].$$  \hspace{1cm} (4.58)

This bound, and the positivity bounds given in Eqn. (4.53) are displayed in Figure 4.4 for a variety of $\chi$.

Recent results constraining the $w$CDM model from the Dark Energy Survey (DES) [Abbott et al., 2017] cannot be immediately applied to constrain $\chi$. This is because $w_{\text{eff}}$ induced by a GEODE population changes in time even if $\chi$ remains fixed for all time. Since we predict only small changes in $w_{\text{eff}}$, however, it is reasonable to expect that the $w$CDM model will approximate the GEODE scenario at late times. Indeed,
Figure 4.5 GEODE constraint region (orange/light grey) estimated from the stellar population, vertical axis magnified by $10^5$ compared to Figure 4.4. The region is superimposed upon Planck constraints of the Dark Energy equation of state evolution. Stellar collapse model indicated via label. Acronyms are defined in the caption of Figure 4.4. Planck 2σ contours have been removed for clarity. For illustration, permissible region for $\chi \equiv 3 \times 10^{-2}$ shown, with thickness $10^5$ smaller than indicated. The permissible region is the overlap of the orange band and the vertical orange line.

their reported value of $w$ [Abbott et al., 2017, Eqn. VII.5] becomes the following constraint on $\chi$

$$\chi < 4 \times 10^{-2} \quad \left( w_0 = -1.00^{+0.04}_{-0.05} \right),$$

which is consistent with the constraints reported in §4.3.3.
CHAPTER 5
SUMMARY AND FUTURE DIRECTIONS

"‘Replay’ is to play again. ‘Rewrite’ is to write again. What then is research, but to search once more?"
Anonymous

Efforts to understand the observed accelerated late-time expansion of the universe have led to the Dark Energy (DE) problem. Data strongly suggest the uniform presence of a material with fixed $P \sim -\rho_m(a = 1)$, the matter density at the present epoch. Such a coincidence problem is highly suspicious, and motivates the search for an underlying physical explanation.

We have approached the DE problem as an exercise in “debugging.” In other words, we did not seek an omission in the formalism of General Relativity, but instead sought a subtle error in its application. The approach proved fruitful. Careful application of the action principle identified an overlooked commutation. This commutation directly affected the source, which determines the overall expansion of the Universe. We verified that our action approach remains consistent with the existing foundations of theoretical cosmology. We concluded that the actual DE distribution could be non-uniform, yet still produce a uniform contribution to the expansion rate.

As it so happens, strong, localized, and gravitationally bound regions of DE are ubiquitous in General Relativity. We have termed all such solutions GEneric Objects of Dark Energy (GEODEs). We considered the impact on cosmological evolution if all complete stellar gravitational collapse scenarios result in GEODEs. This led to numerous quantitative predictions that allow constraint of any GEODE population via gravitational wave astronomy. During the dark ages ($8.9 \lesssim z \lesssim 20$) we showed, via two complementary approaches, that existing astrophysical data support formation of a population of GEODEs that can account for all of the present-day DE density. We further demonstrated consistency with state-of-the-art observations from the Planck and DES collaborations.

Our approach, however, considered only the simplest possible model of how a population of GEODEs could non-trivially affect the expansion history. We modeled the GEODE population’s contribution cosmologically as

$$P_s(\eta) = (-1 + \chi)\rho_s(\eta) \quad \chi \lesssim 10^{-2}$$

(5.1)

to the stress-tensor. In this way, we could evade details about the local structure of GEODEs and make model-independent predictions. After conversion from baryons into GEODEs during stellar collapse, the energy density decays more slowly than the volume expansion. This counterintuitive consequence of Einstein’s Equations resolved the coincidence problem while remaining consistent with state-of-the-art constraint. Unfortunately, a question of interpretation remains. Do individual GEODEs gain energy and
blueshift, analogously to the well-known photon redshift? Fortunately, such behaviour leads to numerous observational consequences and can easily be falsified with future studies and observations.

5.1 Dynamic GEODEs and future directions

Do any of the existing GEODE models exhibit the dynamic behaviour required for a local interpretation? The necessity of a dynamic GEODE is supported by the original and prescient discussion of Gliner [1966]. Energized vacuum is still vacuum, and so admits no notion of privileged frame. This means that isotropic media with $P = -\rho$ cannot rotate. The observed BH mergers to date all have dimensionless spin $> 0.5$, as determined by fits to Kerr spacetime waveform templates. McClintock et al. [2006] have claimed that BH dimensionless spins, measured in X-ray binary BHs, can approach 1. At present, there is no known GEODE analogue to the Kerr spacetime. Concerning specific GEODE models, the stress-tensor of Dymnikova’s “G-lump” can be numerically integrated over its volume. The result is that the object would contribute a net positive pressure. Mottola has made similar claims about the compressed Schwarzschild uniform density interior solution. Even though both objects are asymptotically flat, the fundamental obstacle seems to be the requirement that the objects remain radially static.

As it so happens, there is at least one known dynamic GEODE solution: the original mass point of McVittie [1933], embedded into an arbitrary RW universe. This can be seen at once in Figs. 5.1, 5.2, 5.3. These models have two free parameters, a mass and a radius $w_0$. We fix the mass to be one, and consider the object’s behavior in the coordinate $r$, given in units of $w_0$. These figures display the the energy density, pressure, and equation of state of the McVittie exterior solution joined to the constant density interior solution of Nolan [1993]. To the author’s limited knowledge, it does not seem that the GEODE character of McVittie’s solution, or of Nolan’s, has been appreciated in the literature. Note the presence of an infinite pressure “wall”, very similar to that of Schwarzschild’s interior solution, when compressed below $9/8$ of its Schwarzschild radius. While these GEODEs are dynamic outside of $w_0$, as mentioned before, McVittie enforces that his object’s mass remain static. The construction of a Kerr-like GEODE may provide clues to resolution of both these theoretical difficulties.

---

1 pointed out by Bjorken 2018, priv. dinner
2 priv. correspondence, 2018
Figure 5.1 The fixed mass $M \equiv 1$ Nolan interior solution, joined to McVittie’s asymptotically arbitrary Robertson-Walker (RW) spacetime. The physical radius parameter $w_0/a$ is given on the horizontal axis, where $a$ is the RW scale factor. Physical distance $r/a$ from the spherical origin is given on the vertical axis in units of $w_0$. The object is embedded in a radiation-dominated FRW universe at $z = 5000$, with cosmological parameters taken from Planck. The top panels give energy density (top left) and pressure (top right) in units of the critical density scaled by $10^2$ for visualization. The radial boundary of the object is clearly visible in the energy density plot as a diagonal line. (Bottom) The equation of state $w \equiv P/\rho$ of these solutions. Physically significant transitions in $w$ are indicated by overlaid contours. Note that the exterior McVittie spacetime approaches $w = 1/3$ as $r \to \infty$, consistent with radiation-domination. Both solutions exhibit regions of $w_0$ space where the pressure diverges and changes sign at some fixed radius. The resulting “interiors” of both solutions are GEODEs. (In grayscale, the diagram transitions from positive to negative in both top to bottom and right to left directions.)
Figure 5.2 The fixed mass $M \equiv 1$ Nolan interior solution, joined to McVittie’s asymptotically arbitrary Robertson-Walker (RW) spacetime. The physical radius parameter $w_0/a$ is given on the horizontal axis, where $a$ is the RW scale factor. Physical distance $r/a$ from the spherical origin is given on the vertical axis in units of $w_0$. The object is embedded in a dust-dominated FRW universe at $z = 20$, with cosmological parameters taken from Planck. The top panels give energy density (top left) and pressure (top right) in units of the critical density scaled by $10^2$ for visualization. The radial boundary of the object is clearly visible in the energy density plot as a diagonal line. (Bottom) The equation of state $w \equiv P/\rho$ of these solutions. Physically significant transitions in $w$ are indicated by overlaid contours. Note that the exterior McVittie spacetime approaches $w = 0$ as $r \to \infty$, consistent with matter-domination. Both solutions exhibit regions of $w_0$ space where the pressure diverges and changes sign at some fixed radius. The resulting “interiors” of both solutions are GEODEs. (In grayscale, the diagram transitions from positive to negative in both top to bottom and right to left directions.)
Figure 5.3 The fixed mass $M \equiv 1$ Nolan interior solution, joined to McVittie’s asymptotically arbitrary Robertson-Walker (RW) spacetime. The physical radius parameter $w_0/a$ is given on the horizontal axis, where $a$ is the RW scale factor. Physical distance $r/a$ from the spherical origin is given on the vertical axis in units of $w_0$. The object is embedded in a pure dark energy-dominated FRW universe at $z = 0$, with cosmological parameters taken from Planck. The top panels give energy density (top left) and pressure (top right) in units of the critical density scaled by $10^2$ for visualization. The radial boundary of the object is clearly visible in the energy density plot as a diagonal line. (Bottom) The equation of state $w \equiv P/\rho$ of these solutions. Physically significant transitions in $w$ are indicated by overlaid contours. Note that the exterior McVittie spacetime approaches $w = -1$ as $r \to \infty$, consistent with pure dark energy-domination. In contrast to radiation and matter-domination, the solutions exhibit regions of negative pressure, without transition through a $P \to \infty$ region. Portions of the “interiors” of both solutions are GEODEs. (In grayscale, below the diagonal, the diagram transitions from positive to negative in the right to left direction only. From top to bottom, for $w_0 > r$, pressure switches sign from negative to positive.)
APPENDIX A
SYSTEMS WITH LARGE VELOCITY DISPERSION

We wish to compute an order of magnitude pressure contribution from large, virialized systems, like rich clusters. Pressure supported clusters are regarded as ideal gases

\[ P_c = nkT \]  \hspace{1cm} (A.1)

where \( P_c \) is the pressure, \( n \) is the number density, \( k \) is Boltzmann’s constant, and \( T \) is the temperature. Let \( m \) be a typical mass, then

\[ P_c = \frac{\rho_c kT}{m} . \]  \hspace{1cm} (A.2)

Assuming a monoatomic gas, the average energy is

\[ E = \frac{3}{2} kT . \]  \hspace{1cm} (A.3)

One can construct a Newtonian kinetic energy from the velocity dispersion \( \sigma \)

\[ E = \frac{1}{2} m \sigma^2 . \]  \hspace{1cm} (A.4)

Combining Eqns. (A.3) and (A.4) gives the temperature

\[ kT = \frac{m \sigma^2}{3} . \]  \hspace{1cm} (A.5)

To determine a reasonable upper bound, rich clusters can have \( \sigma \sim 10^3 \text{ km/s} \) [Struble and Rood 1999]. Since we work in natural units, \( \sigma^2 \rightarrow (\sigma/c)^2 \sim 10^{-5} \). Thus, the kinetic pressure of a rich cluster goes as

\[ P_c \lesssim \frac{10^{-5}}{3} \rho_c . \]  \hspace{1cm} (A.6)

From this, we would expect \( \sim 10^{-5} \) corrections to \( H \), relative to the matter’s rest mass density contribution.
APPENDIX B
ESTIMATION OF BARYON DENSITY DEPLETED BY STELLAR COLLAPSE

In this Appendix, we develop a relation between the stellar population and the baryons initially consumed in production of a GEODE. A quantitative model will allow us to check consistency over a diverse set of existing astrophysical data. Throughout, we will treat BHs as classical, not GEODEs. Therefore, the density of the BH population will reflect the density of baryons consumed during stellar collapse. We will formally outline the procedure, and then construct a usable model.

Denote the comoving coordinate number density of galaxies with mass in the range $dM_g$ by

$$\frac{dn_g}{dM_g} dM_g \quad \text{(B.1)}$$

Denote the number count of BHs with mass in the range $dM_{BH}$ in a galaxy of mass $M_g$ as

$$\frac{dN_{BH}(M_g, M_{BH})}{dM_{BH}} dM_{BH} \quad \text{(B.2)}$$

We re-express this relation in terms of stars. Following Fryer and Kalogera [2001], we use the fact that BHs are sampled from the same distribution as stars to write

$$\frac{dN_{BH}}{dM_{BH}} = \frac{dN_*}{dM_*} \frac{dM_*}{dM_{BH}} \quad \text{(B.3)}$$

Assembling Eqns. (B.1), (B.2), and (B.3), the number count of BHs with mass in the range $dM_{BH}$ across galaxies in the mass range $dM_g$ is

$$\frac{dn_g}{dM_g} \frac{dN_*}{dM_*} (M_g, M_*) \frac{dM_*}{dM_{BH}} dM_{BH} dM_g \quad \text{(B.4)}$$

To form a population average comoving coordinate density over some observable $X(M_{BH})$, we multiply by the observable and integrate over all BH masses and galaxy masses

$$\langle X \rangle \equiv \int_{m_g}^{\infty} \frac{dn_g}{dM_g} \int_{m_{BH}}^{\infty} X(M_{BH}) \frac{dN_*}{dM_*} (M_g, M_*) \frac{dM_*}{dM_{BH}} dM_{BH} dM_g \quad \text{(B.5)}$$

Note that accretion effects do not change the number of objects, but that mergers do. Mergers also affect the total mass through radiative losses, which can be around 5% of the final remnant [e.g. Abbott et al., 2016]. In the interest of simplicity, we do not attempt to take account of these effects.
Table B.1. Astrophysical parameters for BH population estimation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Units</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>-2.35</td>
<td>-</td>
<td>Salpeter [1955]</td>
</tr>
<tr>
<td>$R$</td>
<td>0.27</td>
<td>-</td>
<td>Madau and Dickinson [2014]</td>
</tr>
<tr>
<td>$m_c$</td>
<td>0.1</td>
<td>$M_\odot$</td>
<td>Chabrier [2003]</td>
</tr>
<tr>
<td>$m_{TOV}$</td>
<td>3</td>
<td>$M_\odot$</td>
<td>Kalogera and Baym [1996]</td>
</tr>
</tbody>
</table>

To produce a quantitatively useful model from Eqn. (B.5), we note that

$$\int_{m_c}^{\infty} \frac{dN_*(M_g, M_*)}{dM_*} M_* dM_* \equiv (1 - R) M_g$$

where $R$ is the fraction of stellar mass returned to the host galaxy upon star death. Note that $m_c$ is a cutoff mass for the stellar distribution. We next assume a power-law for this distribution

$$\frac{dN_*(M_g, M_*)}{dM_*} \equiv B M_*^\alpha$$

with the normalization $B$ determined by Eqn. (B.6). The result is

$$\frac{dN_*(M_g, M_*)}{dM_*} = -(2 + \alpha) \frac{M_g(1 - R)}{m_c^{\alpha + 2}} M_*^\alpha.$$  (B.8)

Since Chabrier [2003] finds a flattening of the stellar IMF around $0.1 M_\odot$, we will take

$$m_c \equiv 0.1 M_\odot$$  (B.9)

to maintain the validity of our power-law assumption. Reasonable values for $R$ and $\alpha$ are given in Table B.1.

Inserting Eqn. (B.8) into Eqn. (B.5) and integrating over the host galaxy masses, we find

$$\langle X \rangle = -\rho_g(a)(1 - R) \left( \frac{2 + \alpha}{m_c^{\alpha + 2}} \right) \int_{m_{BH}}^{\infty} X(M_{BH}) M_*^\alpha \frac{dM_*}{dM_{BH}} \ dM_{BH}$$

where $\rho_g(a)$ is the comoving stellar density [e.g. Madau and Dickinson [2014, Fig. 11]. This can now be simplified by integrating over progenitor masses $M_*$ instead of BH masses

$$\langle X \rangle = -\rho_g(a)(1 - R) \left( \frac{2 + \alpha}{m_c^{\alpha + 2}} \right) \int_{0}^{\infty} X [M_{BH}(M_*)] M_*^\alpha \ dM_*.$$  (B.11)

We set the lower limit of integration to zero because the remnant mass as a function of progenitor mass $M_{BH}(M_*)$ will automatically cut off. To proceed further, we require this function and information about the temporal evolution of the IMF exponent $\alpha(a)$.  

77
Table B.2. Linear fits to $M_{\text{BH}}(M_*)$

| Metallicity | Model   | Slope $k$ | y-Intercept $|b|$ ($M_\odot$) | $\chi^2_\nu$ |
|-------------|---------|-----------|-----------------------------|-------------|
| $Z = Z_\odot$ | rapid   | 0.19 ± 0.06 | 0.038 ± 1.8 | 9.2 |
|             | delayed | 0.23 ± 0.03 | 1.1 ± 0.96 | 2.3 |
| $Z = 0$     | rapid   | 1.2 ± 0.15 | 18 ± 3.7 | 34 |
|             | delayed | 1.2 ± 0.22 | 18 ± 5.3 | 70 |

Note. — parameters characterize a coarse fit $M_{\text{BH}} = kM_* - |b|$ to $M_{\text{BH}}(M_*)$ based on the stellar collapse simulations of Fryer and Kalogera [2001].

B.1 Core-collapse simulation constraint of $M_{\text{BH}}(M_*)$

The BH remnant distribution as a function of the progenitor star mass has been recently estimated through stellar collapse simulations by Fryer et al. [2012, Fig. 4]. While these authors provide complete metallicity dependent fits, we consider a linear ansatz fit to their data for simplicity

$$M_{\text{BH}}(M_*) = kM_* - |b|. \tag{B.12}$$

We give the fit parameters in Table B.2. For ansatz Eqn. (B.12), $b < 0$ and $k > 0$ are the only physically allowed values. From Madau and Dickinson [2014, Fig. 14], the metallicity $Z(z)$ is reasonably well-approximated by an exponential. The simplest assumption is that $k, |b|,$ and $m_*$ are also exponentials

$$\{k, |b|, m_*\}(z) \equiv \{B\} \exp (\{\zeta\}z). \tag{B.13}$$

Here $\{B\}$ are the normalizations and $\{\zeta\}$ are redshift “time constants.” We present these fit parameters in Table B.3 to $k, |b|,$ and $m_*$ for both rapid and delayed supernovae models. Finally, we include mass ranges forbidden due to the pair-production instability. Detailed numerical integration of the exact fits of Fryer et al. [2012] within Eqn. (B.11), removing contributions from forbidden regions, give values that are $\sim 0.5$ of those from the approximations employed here. Since we have neglected entirely the effects of accretion, we do not feel this to be a serious omission.

The most important case for our concerns is

$$X(M_{\text{BH}}) \equiv M_{\text{BH}}^q, \quad q \in \mathbb{Q}. \tag{B.14}$$
Table B.3. Metallicity dependence of remnant population parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Model</th>
<th>Normalization $B$</th>
<th>$\zeta$-constant $\xi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$</td>
<td>rapid</td>
<td>0.19</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>delayed</td>
<td>0.23</td>
<td>0.083</td>
</tr>
<tr>
<td>$</td>
<td>b</td>
<td>$</td>
<td>rapid</td>
</tr>
<tr>
<td></td>
<td>delayed</td>
<td>1.11</td>
<td>0.14</td>
</tr>
<tr>
<td>$m_*$</td>
<td>Eqn. (B.16)</td>
<td>11</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Note. — Fits are exponential $B \exp(\zeta z)$, assuming $Z(z = 20) \equiv 0$ (i.e. neglecting primordial metal abundances)

For example, $q = 1$ will estimate the cosmological $\rho_{BH}$. Substituting Eqn. (B.12) into Eqn. (B.11), we find

$$\langle M_B^q \rangle = -\rho_g(a)(1 - R)\frac{(2 + \alpha)}{m_c^{\alpha+2}} \int_{m_*}^{\infty} (k M_* - |b|)^\alpha M_*^{\alpha + 1} \, dM_*$$

where we have switched to integration over the progenitors. In the relevant simulations of Fryer et al. [2012], the lower bound for the progenitors is $11 M_\odot$. For our purposes, we will demand that the remnant object exceed the Tolman-Oppenheimer-Volkoff limit $m_{TOV}$ for neutron-degeneracy pressure supported systems. Inspection of Fryer et al. [2012] Fig. 4] gives the following reasonable cutoffs

$$m_* = \begin{cases} 11 M_\odot & Z = Z_\odot \\ 16 M_\odot & Z = 0 \end{cases}$$

where $Z$ is the stellar metallicity. Note that

$$m_0 \equiv |b|/k$$

where $Z$ is the stellar metallicity. Note that

$$m_0 \equiv |b|/k$$

This cutoff is aphysical because it corresponds to a massless remnant, but is theoretically useful to estimate systematics from Eqn. (B.16).

### B.1.1 Cosmological BH density

The $q = 1$ case corresponds to the cosmological comoving coordinate BH mass density. Usefully, $q = 1$ can be integrated by hand in closed form

$$\rho_{BH}(a) = (1 - R) \left[ k \left( \frac{m_*}{m_c} \right)^{\alpha+2} \frac{|b|}{m_c} \left( \frac{\alpha + 2}{\alpha + 1} \right) \left( \frac{m_*}{m_c} \right)^{\alpha + 1} \right] \rho_g(a)$$

$$\equiv \Xi(a) \rho_g(a)$$
Table B.4. Fraction of comoving stellar density collapsed into BH, based on metallicity and supernovae engine

<table>
<thead>
<tr>
<th>Metallicity</th>
<th>Model</th>
<th>$m_0$ ($M_\odot$)</th>
<th>$\Xi(m_*)$ (%)</th>
<th>$\Xi(m_0)$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Z = Z_0$</td>
<td>rapid</td>
<td>0.15</td>
<td>2.6</td>
<td>4 ($m_{TOV}$)</td>
</tr>
<tr>
<td></td>
<td>delayed</td>
<td>4.8</td>
<td>2.8</td>
<td>3.2</td>
</tr>
<tr>
<td>$Z = 0$</td>
<td>rapid</td>
<td>15.7</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>delayed</td>
<td>15.2</td>
<td>12</td>
<td>11</td>
</tr>
</tbody>
</table>

Note. — Astrophysical parameters used are given in Table B.1. For comparison, [Brown and Bethe 1994] predict a late-time ($Z \sim Z_0$) collapse fraction $\Xi \sim 1\%$.

where we have defined the collapse fraction $\Xi(a)$. Time variation in $\Xi$ can come from $\alpha$, $k$, $|b|$, and $m_*$. For the collapse models detailed in Table B.2, we give cutoff parameters and collapse fractions in Table B.4 based on the astrophysical parameters given in Table B.1. We make no effort to propagate errors due to the coarseness of the fit.

To quantify the sensitivity of these data to the cutoffs given in Eqn. (B.16), note that the integrand remains finite for $q = 1$. Thus, we may naturally remove the cutoff and set $m_* \equiv m_0$. We have done this except where $m_0 < m_{TOV}$, where we take $m_* \equiv m_{TOV}$. In all cases, the solar metallicity results are consistent with $\Xi \sim 1\%$ as computed by [Brown and Bethe 1994] via entirely different means.

We have assumed zero time lag between the stellar population and the BH population. Let us justify this simplification for $0.2 \ll z < 20$. This places us within matter domination, so we may approximate

$$H \approx \frac{\sqrt{\Omega_m}}{a^{3/2}}. \quad (B.20)$$

Considering the differentials, we see that

$$da = \frac{\sqrt{\Omega_m}}{a^{1/2}} \, dt \quad (B.21)$$

and so we may approximate the lag in scale $a_L$ relative to the lag in time $\tau_L$ as

$$a_L \approx \frac{\sqrt{\Omega_m}}{a^{1/2}} \tau_L. \quad (B.22)$$

According to dimensional arguments, for stars massive enough to collapse to BH, $\tau_L \lesssim 10^7$ years at metal-
Table B.5. Adopted astrophysical parameters

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Value</th>
<th>Units</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G$</td>
<td>$6.67 \times 10^{-11}$</td>
<td>m$^3$ kg$^{-1}$ s$^{-2}$</td>
<td>Patrignani et al. [2016]</td>
</tr>
<tr>
<td>$c$</td>
<td>$3.00 \times 10^8$</td>
<td>m s$^{-1}$</td>
<td>Patrignani et al. [2016]</td>
</tr>
<tr>
<td>$M_\odot$</td>
<td>$1.99 \times 10^{30}$</td>
<td>kg</td>
<td>Patrignani et al. [2016]</td>
</tr>
<tr>
<td>$H_0$</td>
<td>69.3</td>
<td>km s$^{-1}$ Mpc$^{-1}$</td>
<td>Planck Collaboration et al. [2016]</td>
</tr>
<tr>
<td>$M_\odot$ Mpc$^{-3}$ yr$^{-1}$</td>
<td>0.277</td>
<td>$\rho_c H_0$</td>
<td>-</td>
</tr>
</tbody>
</table>

Note. — Values used to construct unit conversions between $H_0 = \rho_\ell \equiv 1$ and common literature units.

licity $Z_\odot$. Given our choice in time unit, this becomes the upper bound

$$\tau_L \lesssim \frac{6 \times 10^{14}}{5 \times 10^{17}} \text{ s} \sim 10^{-3}. \tag{B.23}$$

Note that this is a conservative upper bound on primordial star lifetimes. To justify the neglect of the time lag, we must have that

$$\rho_g(a - a_L(a)) = \rho_g[a(1 - a_L/a)] \approx \rho_g(a) \tag{B.24}$$

and so we require that $a_L/a \ll 1$. A sufficient condition is then

$$\frac{a_L}{a} < \frac{\sqrt{\Omega_m}}{a^{3/2}} 10^{-3} \ll 1, \tag{B.25}$$

which implies

$$z \ll 148. \tag{B.26}$$

For representative numbers, at $z = 20$ the time lag is a 5% correction, but at $z = 5$ the time lag is a 0.8% correction. Most importantly, at the peak of star formation near $z \sim 2$, the time lag is a 0.2% correction. We may thus consider

$$\rho_{BH}(a) = \Xi(a) \rho_g(a) \quad (z < 5) \tag{B.27}$$

to good precision.

In practice, we will be interested in the late-time behavior of these quantities. Since $\rho_{BH}$ depends on the entire conversion history, neglecting accretion could have a substantial effect on $\rho_{BH}$. Since Li et al. [2007]
find that accretion effects diminish significantly below \( z < 5 \), it is more convenient to examine \( \frac{d\rho_{\text{BH}}}{da} \)

\[
\frac{d\rho_{\text{BH}}}{da} = -\Xi(a) \frac{d\rho_{g}}{dz} \frac{1}{a^2} + \frac{d\Xi}{da} \rho_{g}.
\] (B.28)

A global fit, performed by [Madau and Dickinson (2014)] to very many stellar surveys gives the comoving stellar density as

\[
\frac{d\rho_{g}}{dz} = 0.277 (R - 1) \psi \left( \frac{1 + z}{2.9} \right)^{5.6} \] (B.29)

\[
\psi \equiv 0.015 \left( 1 + \frac{(1 + z)}{2.9} \right)^{2.7}
\] (B.30)

\[
R \equiv 0.27.
\] (B.31)

The factor of 0.277 converts units as listed in Table B.5. We determine \( H \) given the Planck assumed dark energy equation of state Eqn. (4.40)

\[
H(z) \equiv \sqrt{\frac{\Omega_m (1 + z)^3 + \Omega_{\Lambda} (1 + z)^3 [1 + w_{\text{eff}(0)}]}{3 w_{\text{eff}} z}} \] (B.32)

We neglect corrections due to \( \Omega_m \) being very slightly decreased by baryon consumption, since this decrease is well within Planck best-fit uncertainties.

### B.1.2 Expectation values with generic \( q \)

It is useful to define notation to compress and de-dimensionalize

\[
\lambda(q) = q + \alpha + 1
\] (B.33)

\[
r \equiv \frac{m_0}{m_*}.
\] (B.34)

We may now write the de-dimensionalized Eqn. (B.15)

\[
\left\langle M_{\text{BH}}^q \right\rangle = \rho_{g}(a)(R - 1) \frac{2 + \alpha}{m_e^{\alpha + 2}} \frac{K^{\lambda(q)}}{K^{\alpha + 1}} \int_0^\infty (z - 1)^{q} z^\alpha \, dz.
\] (B.35)

Notice that \( z \geq 1 \) always, and so the integrand remains real for all \( q, \alpha \in \mathbb{R} \). We then perform an inversion \( z \rightarrow 1/x \) to make the domain of integration finite

\[
\left\langle M_{\text{BH}}^q \right\rangle = \rho_{g}(a)(R - 1) \frac{2 + \alpha}{m_e^{\alpha + 2}} \frac{K^{\lambda(q)}}{K^{\alpha + 1}} \int_0^1 (1 - x)^{q} x^{-\lambda(q) - 1} \, dx.
\] (B.36)
Define the following notation

\[ 2F_1 \{q\} \equiv 2F_1 (-q, -\lambda(q); -(q + \alpha); r). \]  

(B.37)

We may then write

\[ \int_0^r (1 - x)q x^{-\lambda(q)-1} \, dx = -\frac{2F_1 \{q\}}{\lambda(q)} r^{-\lambda(q)}. \]  

(B.38)

and finally we find

\[ \langle M_{\text{BH}} \rangle = \rho_g(a)(1 - R) \frac{(2 + \alpha) m^{\lambda(q)}_e k^q}{m^{\alpha+2}_e} \frac{2F_1 \{q\}}{\lambda(q)}. \]  

(B.39)

Values for \( 2F_1 \{q\} \) can be computed in any modern CAS.


88