

Dark Matter and Dark Energy

Authors:

Eleanor Gleave

Henrik Sherling

Furqaan Yusaf

September 29, 2019

Contents

1	Introduction	7
1.1	The FRW Metric and The Expansion of the Universe	8
1.2	Statistical Mechanics of Cosmology	14
1.3	The Friedmann Equation	17
1.4	The Point of All This Laborious Stuff	20
2	Astrophysical Evidence for Dark Matter	23
2.1	Methods of Indirect Detection	24
2.2	Galactic Clusters and the Virial Theorem	26
2.3	Modified Newtonian Dynamics	32
2.4	Gravitational Lensing	32
2.5	Bullet Cluster	32
2.6	Rotation Curves	36
2.6.1	The Isothermal Sphere	37
2.7	Freeze-Out	38
2.8	N-body Simulations	46
2.9	Nucleosynthesis	47
2.9.1	Neutrons and Protons	48
2.9.2	Deuterium, Helium, and Heavier Nuclei	49
2.9.3	Cosmological Parameters	50
3	Candidates for Dark Matter	51
3.1	WIMPS and Thermal Relics	51
3.1.1	Cross-Sections	52
3.1.2	A model of dark matter	55
3.1.3	Self-Annihilating Dark Matter	56
3.2	Alternatives to Thermal Relics	57
3.2.1	Axions	57
3.2.2	Primordial Black Holes	61
3.2.3	Sterile neutrinos	61

4 The CMB Constraints	63
4.1 The Angular Power Spectrum	64
4.1.1 Spherical Harmonics	65
4.2 The Theory	66
4.2.1 The Surface of Last Scattering	66
4.3 Primary Causes of Anisotropies in the CMB	67
4.3.1 The Photon-Baryon Plasma	68
4.3.2 Perturbations in the Photon-Baryon Plasma	69
4.3.3 Baryonic Acoustic Oscillations	74
4.3.4 Recombination	75
4.4 Expansion, and Secondary Causes of Anisotropies in the CMB	78
4.4.1 Ionization	79
4.4.2 Sachs-Wolfe Effect	79
4.5 What Does the CMB Tell Us About the Universe?	82
4.5.1 The Angular Power Spectrum Graph	82
4.5.2 First Peak - The Curvature of the Universe	84
4.5.3 Damping	88
4.5.4 Cosmological Parameters	89
4.6 The Galaxy 2 Point Correlation Function	90
5 Detection	93
5.1 Detection of WIMPs	93
5.1.1 The Theory	93
5.1.2 Cryogenic Dark Matter Search (CDMS)	96
5.1.3 Noble Gas and Crystal Scintillators	97
5.1.4 Bubble Detectors	98
5.2 Detection of Axions	99
5.3 Detection of Primordial Black Holes	102
5.4 Detection of Sterile Neutrinos	103
6 Dark Energy	105
6.1 Late Universe Acceleration	105
6.2 The Cosmological Constant	107
6.2.1 The Constant Problem	110
6.2.2 The Coincidence Problem	112
6.3 Other Alternatives	118
6.3.1 Light Scalar Fields and Quintessence	118
6.3.2 Modified Gravity	119

6.4 Conclusion: Expanded Anthropic Reasoning; or, Why Physics Needs Philosophy	120
A Appendices	125
A.1 Spherical Bessel Functions	125
A.2 Spherical Harmonics	126
A.3 Angular Power Spectrum Derivation	129
A.4 The CMB Dipole	130

Recommended readings.

David Tong's lecture notes on Cosmology ([2018a](#)) and Statistical Physics ([2018b](#)) are good for those who want a more thorough introduction to some of the physics and mathematics of this course. They are very descriptive, but moderately hard.

Eugene Lim's lecture notes on General Relativity ([2019](#)) provide a very good introduction to the tensor algebra and relativistic principles needed in this course. They are similar in style to Tong's lecture notes. See also Lim's ([2018](#)) notes for additional stuff on Statistical Mechanics.

Mariangela Lisanti has excellent, but relatively difficult, lecture notes on Dark Matter Physics ([2018](#)). They cover rotation curves (section 2.6), n-body simulations (section 2.8), and freeze-out (section 2.7) among other things.

Joshua A. Frieman has the most accessible and comprehensive lecture notes on the specific topic of Dark Energy ([2008](#)). They summarize the mathematics, evidence, experiments, and models in a brief but easy-to-follow manner.

Beyond the works listed here, there are a lot of relevant references in the main body of the lecture notes. Make sure to check these out. Please feel free to submit corrections and questions on Keats.

Chapter 1

Introduction

It is possible to do pretty well in this course by reading these notes, following the derivations, and doing the exercises. But that will probably be both hard and fruitless: you will struggle with the notes and you will remember very little after the exam.

To properly understand the nature and significance of dark matter you need to be very comfortable with concepts from statistical mechanics, particle physics, and relativity. If you don't know what a cross-section is, these notes will not help you; nor will they help you understand the metric or tensor algebra, both of which are needed to properly understand the concepts involved in this course.

This introduction will set up the mathematical and physical background for the rest of the course. It will unashamedly abbreviate the story and make some dodgy mathematical moves to cut relatively quickly to the chase. There will be almost no GR or Standard Model stuff in here, since that quickly gets too tedious to deal with faithfully while also having enough time to actually get to Dark Matter and Dark Energy, which is what we care about here.

This chapter will introduce some of the core notation and terminology for the main meat of the course. It will of aforementioned necessity be too brief to be read as a standalone introduction, and is instead intended to refresh memories and point to relevant independent research. The point of this section is essentially to show the threads that tie the other sections together; that is, to explain our assumptions (the cosmological principle), to show the FRW metric and the Friedmann equation, and to explore why our current data, in conjunction with these, lead us to believe in dark matter and dark energy.

Chapter 2 will cover the *indirect* evidence we have for believing that there is dark matter in the universe.

Chapter 3 will cover the main candidates we currently have for dark matter and

the (general lack of) evidence we have for them.

Chapter 4 will cover some important constraints on dark matter models given by independent cosmological facts inferred from the Cosmic Microwave Background radiation (CMB).

Chapter 5 will cover the detection methods we have for the candidates previously mentioned.

Chapter 6 will cover the little we know about dark energy and our possible explanations of it. This will be a brief chapter. The dark energy problem is really hard to state. It's not even clear to all physicists that there is a problem. We will discuss why.

And finally the appendices A will cover some of the mathematics needed to fully grasp the rest of the notes. Beware, however, that they are nowhere near a complete collection of the math you will need to understand what is really going on here.

1.1 The FRW Metric and The Expansion of the Universe

Dark Matter and Energy are subject matters of cosmology, so we begin with **The Cosmological Principle**: the universe is spatially homogeneous and isotropic on the largest scales. That is, it looks roughly identical at any given point (homogeneity) and in every direction (isotropy). This principle is key to constructing models that can be tested against observable data.

A minimal amount of General Relativity inevitably appears, so we might as well begin with an overview of the preliminaries. The spacetime interval is, in general, defined as

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu \quad (1.1)$$

which for approximately flat spacetimes yields $g_{\mu\nu} \approx \eta_{\mu\nu}$, such that

$$dl^2 \equiv ds^2 + c^2 dt^2 = \eta_{ij}dx^i dx^j \equiv dx^2 + dy^2 + dz^2 \quad (1.2)$$

where we have, for now, decided to focus only on the spatial $dl^2 = g_{ij}dx^i dx^j$ part of the metric for now. In spherical polar coordinates this is just (homework)

$$\begin{aligned} dl^2 &= \eta_{ij}dx^i dx^j = dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta d\varphi^2 \\ &\equiv dr^2 + r^2 d\Omega^2 \end{aligned} \quad (1.3)$$

But we want to work with a more general metric that can capture all homogeneous

and isotropic geometries, not just the flat ones. We label the homogenous and isotropic 3 dimensional metric by γ_{ij} .

There are three such spaces, and the above is simply one of them, namely a flat (Euclidean) geometry \mathbf{R}^3 , where parallel lines remain at an equal distance at all points. Here we have $\gamma_{ij} = \nu_{ij}$, as in special relativity (i.e. Minkowski) and other flat geometries.

But there are two other options: spheres \mathbf{S}^3 with positive curvature (where parallel lines converge) and hyperbolic geometries, or saddles, \mathbf{H}^3 with negative curvature (where parallel lines diverge). Fortunately, we can use a single metric to capture all of them:

$$dl^2 = \frac{dr^2}{1 + \frac{kr^2}{R^2}} + r^2 d\Omega^2 \text{ with } k = \begin{cases} +1 & \text{if spherical} \\ 0 & \text{if Euclidean} \\ -1 & \text{if hyperbolic} \end{cases} \quad (1.4)$$

for (r, θ, φ) coordinates of a sphere \mathbf{S}^3 in an \mathbf{R}^4 embedding, with $x_i x^i + k(w)^2 = R^2$. But if we want to do without the embedding (to avoid pretending that there is some bigger space in which our universe floats), we need a four-dimensional generalization of spherical polar coordinates:

$$\begin{aligned} x &= R \sin \chi \sin \theta \cos \varphi \\ y &= R \sin \chi \sin \theta \sin \varphi \\ z &= R \sin \chi \cos \theta \\ w &= R \cos \chi \end{aligned}$$

such that in (χ, θ, φ) we have

$$dl^2 = R^2 (d\chi^2 + S_k^2(\chi) d\Omega^2),$$

$$\text{with } S_k(\chi) = \begin{cases} \sin \chi & \text{for } k = +1 \\ \chi & \text{for } k = 0 \\ \sinh \chi - 1 & \text{for } k = -1 \end{cases} \quad (1.5)$$

where χ is a dimensionless coordinate and $S_k(\chi) = r/R$. It's a good exercise to try these substitutions at home to verify them. See [Tong \(2018a\)](#) for much more detail.

Technical sidenote: not all spaces constructed from the above metrics obey the cosmological principle, but all spaces that do obey it can be captured with these metrics.

And a note for those who want to read more: $K \equiv k/R^2$ in many textbooks and lecture notes, such that we define the curvatures as $K = 0$ for flat, $K > 0$ for

positive, and $K < 0$ for negative. This is pretty intuitive, and at times in these notes we will simply switch to this notation for convenience, as it (sometimes) lets us simplify the algebra.

Recall that we have left out the time-coordinate from the metric. But since our universe is $3 + 1$ dimensional, not just three-dimensional, we need to put it back in. That leaves us with a general cosmological metric of

$$ds^2 = -c^2 dt^2 + \gamma_{ij} dx^i dx^j \quad (1.6)$$

In these notes there will be a lot of phrases like “the current time” or “the time when”, etc. That sounds like a strictly speaking meaningless phrase in a general relativistic (or even a special relativistic) universe, because of course we know that time slices differently for different observers due to velocity (SR) and acceleration (GR). Simultaneity itself is relative on these theories, so there is no privileged way of slicing time into distinct planes. That’s why we need co-moving coordinates to do cosmology.

The general idea of co-moving coordinates is to privilege a perspective on the universe that is the most useful for tracking the large-scale changes that we can investigate with the assumptions of the cosmological principle. That is, we want a slicing of time that yields a homogeneous and isotropic spatial structure of the universe. We want to create a set of **hyperslicings** (or **foliations**) Σ_t labeled by a time coordinate t such that at each t the (spatial and 3 dimensional) hypersurface we are left with is itself homogeneous and isotropic. That is what γ_{ij} is doing above and in fig. 1.1:

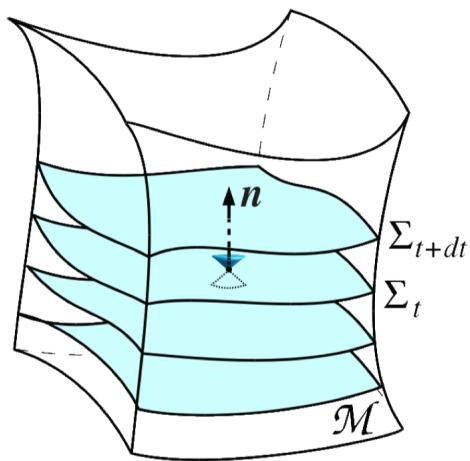


Figure 1.1: Foliation of a manifold \mathcal{M} into hyperslicings Σ_t , from (cite) Gourgonhol via Lim.

But we have left out a crucial option from the above equation, to allow for the last bit of freedom under the regime of the cosmological principle: the hypersurfaces themselves can, if they stay isotropic and homogeneous, conceivably evolve with time. This implies that we can add a dimensionless *scale factor* $a(t)$ to get the famous Friedmann-Robertson-Walker (FRW) metric:

$$\begin{aligned} ds^2 &= -c^2 dt^2 + a^2(t) \gamma_{ij} dx^i dx^j \\ &\equiv -c^2 dt^2 + a^2(t) \left(\frac{dr^2}{1 + \frac{kr^2}{R^2}} + r^2 d\Omega^2 \right) \end{aligned} \quad (1.7)$$

This metric, $a^2(t)\gamma_{ij}$, is invariant under the transformation

$$\begin{aligned} a &\rightarrow \lambda a \\ r &\rightarrow \lambda^{-1}r \\ R &\rightarrow \lambda^{-1}R \end{aligned}$$

which means that we are free to demand that the scale factor is 1 at the current time t_0 :

$$a_0 = a(t_0) \equiv 1 \quad (1.8)$$

Note that the present time will always be denoted by the subscript 0; and it will sometimes be confusing that we refer to the past with subscripts of greater value than 0. Keep this in mind when reading the equations below.

The physical distance between an arbitrarily chosen origin and a point (r, θ, φ) or (χ, θ, φ) is given by

$$l = a(t) \int_0^r \frac{dr}{\sqrt{1 + kr^2/R^2}} = a(t)R\chi \quad (1.9)$$

where we have defined

$$dl^2 \equiv a^2(t) \gamma_{ij} dx^i dx^j \quad ; \quad l = \int dl \quad (1.10)$$

Thus we have

$$\begin{cases} \dot{a}(t) > 0 & \text{for an expanding universe} \\ \dot{a}(t) = 0 & \text{for a static universe} \\ \dot{a}(t) < 0 & \text{for a contracting universe} \end{cases} \quad (1.11)$$

But note that the universe does not expand from, stay static at, or shrink to an origin. This holds for *any* origin, so all of space must be changing size simultaneously.

It is not that *some* distance becomes larger or smaller; all distances do, at once.

An important qualification: when we refer to space, here, we are strictly speaking not talking about *space in general* but rather *our hypersurface* picked out by our *foliation* Σ_t and the requirements of homogeneity and isotropy. There are many alternative spaces that we could be talking about, but that's for a cosmology course — we won't (for the most part) need to talk about those here.

But note that this metric, which we will just call “the FRW metric” from now on, is not Lorentz-invariant. The co-moving coordinates picked out by our hyperslicing pick out, as mentioned, a preferred rest frame for the universe; things will not look the same from other relatively moving or accelerating frames, so care must be taken when applying the lessons of SR and GR.

We can see this by considering a trajectory $\mathbf{x}(t)$ in co-moving coordinates with one in *physical* coordinates $\mathbf{x}_{\text{phys}}(t) = a(t)\mathbf{x}(t)$ from (1.9) and getting

$$\begin{aligned}\mathbf{v}_{\text{phys}}(t) &= \dot{a}(t)\mathbf{x}(t) + a(t)\dot{\mathbf{x}}(t) \\ &\equiv \frac{\dot{a}}{a}\mathbf{x}_{\text{phys}}(t) + a(t)\dot{\mathbf{x}}(t) \\ &\equiv H\mathbf{x}_{\text{phys}}(t) + \mathbf{v}_{\text{peculiar}}\end{aligned}\tag{1.12}$$

where $\mathbf{v}_{\text{peculiar}}$ is the *peculiar velocity* of a galaxy relative to the privileged co-moving cosmological coordinate frame (normally due to gravity), and

$$H(t) = \frac{\dot{a}(t)}{a(t)}\tag{1.13}$$

is the *Hubble parameter* (or, misleadingly, the *Hubble constant*), which today has the value

$$H_0 \approx 70 \text{ km s}^{-1} \text{ Mpc}^{-1}\tag{1.14}$$

That is, a galaxy at 1 Mpc^{-1} away from us is moving away from us at 1 km s^{-1} purely due to the expansion of space itself.

Of course, we are losing out on a lot of interesting information when we hyperslice time according to a preferred frame. Things will look very different from disparate frames of reference, and our descriptions of the fundamental processes of the universe might too. *But not on the largest scales.* Cosmology cares about the largest-scale evolution of the universe, and the FRW metric gives us just that; the assumption of isotropy and homogeneity simply rules out any finer-grained interest. We will build in more variety as we go along, but for now, and for the most part, we will neither need nor want it.

Recall from special relativity that light travels at $ds = 0$. Consider a photon

traveling along the radial direction, i.e. $d\Omega = 0$. Then we have

$$ds^2 = -c^2 dt^2 + a^2(t) \left(\frac{dr^2}{1 + \frac{kr^2}{R^2}} \right) = 0 \quad (1.15)$$

and thus

$$cdt = \pm \frac{a(t) dr^2}{\sqrt{1 + \frac{kr^2}{R^2}}} \quad (1.16)$$

integrating to get (recall that t_0 is now and t_1 is the past)

$$c \int_{t_1}^{t_0} \frac{dt}{a(t)} = \int_0^{r_1} \frac{dr^2}{\sqrt{1 + \frac{kr^2}{R^2}}} = c \int_{t_1 + \delta t_1}^{t_0 + \delta t_0} \frac{dt}{a(t)} \quad (1.17)$$

if we suppose that a signal is emitted from r_1 at $t_1 + \delta t_1$ and received at $r = 0$ at $t_0 + \delta t_0$. It follows that

$$\frac{\delta t_1}{a(t_1)} = \frac{\delta t_0}{a(t_0)} = \frac{\delta t_0}{a_0} \equiv \frac{\delta t_0}{1} = \delta t_0 \quad (1.18)$$

But we know that $t = \lambda/c$ in general, so

$$\lambda_0 = \frac{\lambda_1}{a(t_1)} \quad (1.19)$$

which is probably more intuitively viewed as $\lambda_1 = a(t_1)\lambda_0$, because here it is easy to see that

$$\begin{cases} a(t_1) \in (0, 1) & \text{entails a redshift of } \lambda_1 \\ a(t_1) \notin (0, 1) & \text{entails a blueshift of } \lambda_1 \end{cases} \quad (1.20)$$

and since we have take care to defined $a_0 = 1$, and observed that $H_0 > 0$, then we know that $a(t_1)$ (which, recall, is in the past) must be $a(t_1) \in (0, 1)$. Conversely, and in line with the history of experiments, we have used our measurements of the redshift to determine the Hubble parameter and thus that the universe is expanding — so the presentation here gets things in reverse, but is easier to follow mathematically.

It is helpful to define the *redshift parameter*:

$$z = \frac{\lambda_0 - \lambda_1}{\lambda_1} \Rightarrow z = \frac{1}{a(t_1)} - 1 \quad (1.21)$$

The universe today has a parameter $z = 0$; at $z = 1$ the universe had half the size it has today, and so on. And since z is related to the redshift (or blueshift) of spectral lines, we finally have a direct connection to measurement.

The redshift parameter has many important uses. The most crucial one for our purposes is that it allows us to divide the universe into a range of epochs that neatly classify how its evolution has changed with time. On the Big Bang model, the universe starts out being dominated by *radiation* $\rho_r \propto a^{-4}$. But with time, radiation scales slower than *matter* $\rho_m \propto a^{-3}$; matter will thus come to dominate at some point. This occurs at $z_{\text{eq}} \sim 3400$ ([Velten et al. 2014](#)).

Note also that as $a_0 = 1 > a(t_1) > 0$, we can see that $a(t_1)$ decreases as we look further back in time. (Conversely, z *increases* as we wind time backwards since it is zero today.) Since $a(t_1)$ appears in the denominator for the redshift parameter, we have a singularity at hand. This is *not* merely a coordinate singularity, but a real one; this is the Big Bang. We label it by $a(t_{BB}) = 0$.

We can estimate the size of the observable universe by making use of t_{BB} , since it implies, along with the fixed speed of light, that there is a fixed distance that light can have traveled since the early days:

$$c \int_{t_{BB}}^t \frac{dt'}{a(t')} = \int_0^{r_{\max}(t)} \frac{dr}{\sqrt{1 + \frac{kr^2}{R^2}}} \quad (1.22)$$

such that

$$l_{\text{horizon}} = a(t) \int_0^{r_{\max}(t)} \frac{dr}{\sqrt{1 + \frac{kr^2}{R^2}}} = ca(t) \int_0^t \frac{dt'}{a(t')} \quad (1.23)$$

where l_{horizon} is the *particle horizon*. Nothing outside this horizon can influence our bit of the universe today. But since the horizon depends on $a(t)$, the range of possible causal influences evolves with time.

1.2 Statistical Mechanics of Cosmology

The following section is not quite faithful to the true mathematical underpinnings of its own results, but provides an intuitive path to some equations and facts we need for later. We will make use of thermodynamic and classical intuitions rather than strictly statistical mechanical and general relativistic derivations. For more detailed derivations see [Lim \(2018\)](#), [Tong \(2018b\)](#), [Tong \(2018a\)](#), and [Lisanti \(2018\)](#).

Equations of state in cosmology are given by

$$P = w\rho \quad (1.24)$$

where $P(\rho)$ is the pressure and $\rho(t)$ the energy density of a perfect fluid that

we imagine to populate the whole universe. We use this fluid to model the large scale structure of the universe. Think of the fluid as our implementation of the cosmological principle to energy: it is our model of a roughly homogeneous and isotropic distribution of energy in the universe. We ignore the lumps and clumps of galaxies and stars, and pretend everything averages out to uniformity on the very largest scales.

As the universe expands, the energy density of the fluid should dilute. Note that the distances in one spatial direction scale by $a(t)$, so the volume of a bit of space increases by $V(t) = a^3(t)V_1$, where V_1 is the initial volume. Thus we have

$$\frac{dV}{dt} = 3\dot{a}a^2V_0 \quad (1.25)$$

by the chain rule. Since ρ is energy density, $E = V\rho$ and thus

$$E = \rho a^3 V_0 \quad (1.26)$$

and thus

$$\frac{dE}{dt} = \dot{\rho}a^3V_0 + 3\rho\dot{a}a^2V_0 \quad (1.27)$$

Now suppose the fluid in question can be treated classically, i.e. without going into the general relativistic derivation, and suppose we consider its time evolution to be reversible and adiabatic. Then we can apply the *first law of thermodynamics*: $dE = -PdV$. Taking the time derivative of both sides, we get

$$\begin{aligned} \dot{\rho}a^3V_0 + 3\rho\dot{a}a^2V_0 &= -3P\dot{a}a^2V_0 \\ \Rightarrow \dot{\rho} + 3\frac{\dot{a}}{a}(\rho + P) &= 0 \\ \Rightarrow \dot{\rho} + 3H(\rho + P) &= 0 \end{aligned} \quad (1.28)$$

This is the expression of energy-conservation in a cosmological setting: it is called the *continuity equation*. We will derive it more faithfully in chapter 6, though for full details see [Lim \(2019\)](#).

Using $P = w\rho$, we can integrate to get

$$\rho(t) = \rho_0 a^{-3(1+w)} \quad (1.29)$$

given $a_0 = 1$ and $\rho_0 = \rho(t_0)$. Recall that relativistic energy is related to momentum by

$$E^2 = p^2c^2 + m^2c^4 \quad (1.30)$$

and that

$$\begin{cases} \text{Non-Relativistic Limit: } pc \ll mc^2 \text{ entails } E \approx mc^2 \text{ and } \mathbf{v} \approx \mathbf{p}/m \\ \text{Relativistic Limit: } pc \gg mc^2 \text{ entails } E \approx pc \text{ and } |\mathbf{v}| \approx c \end{cases}$$

Suppose now that there are N atoms in V volume of the fluid, and let $n(p)$ be the number density of atoms in the fluid. Why as a function of p ? Because n cannot depend on position: the fluid is isotropic by assumption, so it can only depend on momentum. Different densities per momentum makes sense, but different densities per position breaks isotropy. Thus:

$$\frac{N}{V} = \int_0^\infty dp n(p) \quad (1.31)$$

and taking the flux of momentum across a surface area to be the pressure of the fluid, and noting that $\mathbf{v} \parallel \mathbf{p}$, we get

$$P = \frac{1}{3} \int_0^\infty dp vp n(p) \quad (1.32)$$

For a **non-relativistic gas** we have $p \approx mv$ and

$$P_{\text{non-rel}} = \frac{1}{3} \int_0^\infty dp mv^2 n(p) = \frac{1}{3} \frac{N}{V} m \langle v^2 \rangle \quad (1.33)$$

and noting that $E \approx mc^2 + \dots$ we get

$$P_{\text{non-rel}} = \frac{1}{3} \frac{NE}{V} \frac{\langle v^2 \rangle}{c^2} \quad (1.34)$$

But we know that in the non-relativistic domain the latter ratio is $\frac{\langle v^2 \rangle}{c^2} \approx 0$, so $P_{\text{non-rel}} \approx 0$. That is, $w \approx 0$. Approximately zero? Yes, but relative to the pressure of the expansion of the *universe*, not relative to the pressure of the earth's atmosphere and so on (relative to these it is a significant positive pressure).

Non-relativistic gases of this sort are called *dust* and are composed of matter of all kinds, including things like dark matter, galaxies, and planets. Of course, we are mostly interested in dark matter here, but we'll get back to that in later chapters.

For a **relativistic gas** we have $v \approx c$ and $E \approx pc$ such that

$$P_{\text{rel}} = \frac{1}{3} \int_0^\infty dp E n(p) = \frac{N \langle E \rangle}{3V} \quad (1.35)$$

and since the energy density is $\rho = N \langle E \rangle / V$ we simply have

$$P_{\text{rel}} = \frac{1}{3}\rho \quad (1.36)$$

which is simply our equation of state with $w = 1/3$. Relativistic gases of these sorts are called *radiation* and represent gases of photons and neutrinos and so on.

An important restriction on the equation of state is found in special relativity: $c_s \leq c$; the speed of sound in a gas must be less than or equal to the speed of light. That is,

$$c_s^2 = c^2 \frac{dP}{d\rho} \quad (1.37)$$

and thus $w \leq 1$. Recall now that $\rho(t) = \rho_0 a^{-3(1+w)}$. It follows from the above that

$$\begin{cases} \text{Matter: } w \approx 0 & \text{entails } \rho \sim \frac{1}{a^3} \\ \text{Radiation: } w = \frac{1}{3} & \text{entails } \rho \sim \frac{1}{a^4} \end{cases}$$

That is, the energy density of matter decreases by a^{-3} as the volume expands by a^3 ; the energy density of radiation, meanwhile, decreases by a^{-4} , with the added power of a^{-1} owed to the redshift. These differences in dilution-rates are crucial to our understanding of cosmological history.

A last thing to note here is that there is nothing special about matter and radiation. w can take on many more values, and these may or may not match up with other types of substances, structures, or processes. We will come back to this soon.

1.3 The Friedmann Equation

The final ingredient to the basic cosmological picture is the Friedmann equation. To derive this, we need some general relativity. We will not go into a lot of detail here, but will simply sketch the path for those interested in looking it up in more detail.

We will let $c \equiv 1$ and $K \equiv k/R^2$ in what follows, to simplify the calculations. The stress-tensor for a perfect fluid is

$$T_{\mu\nu} = (\rho + P) u_\mu u_\nu + P g_{\mu\nu} \quad (1.38)$$

where u^μ is the fluid four-velocity. Since we are working with the cosmological principle we get to assume isotropy, and thus we can simply choose a coordinate basis in which $u^i = 0$ and $u^0 = 1$. Thus

$$T_{\mu\nu} = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & & & \\ 0 & & a^2 \gamma_{ij} P & \\ 0 & & & \end{pmatrix} \quad (1.39)$$

Why the $a^2 \gamma_{ij}$? Because we are working with the FRW metric $g_{\mu\nu}$, and $\gamma_{ij} dx^i dx^j = dr^2/(1 - |k|r^2) + r^2 d\Omega^2$. The Einstein equation governs the evolution of the matter content of the universe:

$$G_{\mu\nu} = 8\pi G T_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R \quad (1.40)$$

where $\nabla_\mu T_{\mu\nu} = 0$ is the equation for conservation of energy. (This, incidentally, is the faithful starting point for those who want to properly derive the continuity equation from above.) Now take the time-time components to get

$$\begin{aligned} G_{00} &= R_{00} - \frac{1}{2} g_{00} R \\ &= 3 \left(\frac{\dot{a}}{a} \right)^2 + \frac{3K}{a^2} \end{aligned} \quad (1.41)$$

It's left as an exercise to those who are interested to do the full derivation. Note that $T_{00} = \rho$ and thus that

$$\begin{aligned} 3 \left(\frac{\dot{a}}{a} \right)^2 + \frac{3K}{a^2} &= 8\pi G \rho \\ \Rightarrow \quad \left(\frac{\dot{a}}{a} \right)^2 &= \frac{8\pi G \rho}{3} - \frac{K}{a^2} \end{aligned} \quad (1.42)$$

Recovering c and R and recalling $H \equiv \dot{a}/a$ we get

$$H^2 = \frac{8\pi G \rho}{3c^2} - \frac{kc^2}{R^2 a^2} \quad (1.43)$$

The evolution of the universe is thus governed by the Friedmann equation¹, the continuity equation, and the equation of state:

$$\begin{aligned} H^2 &= \frac{8\pi G \rho}{3c^2} - \frac{kc^2}{R^2 a^2}, \\ \dot{\rho} &= -3H(\rho + P), \\ P &= w\rho, \end{aligned} \quad (1.44)$$

¹We will see in chapter 6 that this is not entirely accurate. The Friedmann equation as we have seen it so far is *not* the equation of motion (it is not a second derivative in time, after all) and thus operates as a constraint equation rather than a dynamical equation; but this will not matter too much for our purposes.

where k determines the curvature and w determines the type of fluid we are considering. Since the equation of state is linear, it is possible to combine several types of fluids so long as each is a solution in its own right. But the equation for a that results from a linear combination of different solutions is nonlinear and very hard to solve; recall that $\rho_w = \rho_{w,0}a^{-3(1+w)}$.

Thankfully, we know (Collaboration et al. 2013) from observation that our universe is very close to flat, so close that for our purposes it will be fine to approximate $k \approx 0$, in which case we get

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G \rho_{w,0}}{3c^2} a^{-3(1+w)} \quad (1.45)$$

to which the solution is simply

$$a(t) = \left(\frac{t}{t_0}\right)^{2/(3+3w)} \quad (1.46)$$

for $t_0 = \left(3/2(1+w)\sqrt{8\pi G \rho_{w,0}/3c^2}\right)^{-1}$ such that we get $a(t_0) = 1 = a_0$, and such that integration constants are set to zero and $a(t_{BB}) = a(0) = 0$.

There is one last step to set us up. We have seen that the choice of k and w determine the curvature and content of the model of the universe that we are studying. We have seen that $k = 0$ is a good approximation for our flat universe, and that $w = 0$ yields a matter fluid; while $w = 1/3$ yields a radiation fluid.

But recall that the only restriction on w is that $w \leq 1$. That leaves open the possibility that $w < 0$. And indeed crucial possibility for this course is the case where $w = -1$:

$$P = -\rho_\Lambda \quad (1.47)$$

where Λ denotes this special fluid. Note that the density of Λ remains constant over time (homework): $\rho_\Lambda \sim a^0$.

This means that the energy density of Λ does not decrease with the expansion of the universe, which seems to disobey conservation of energy; and any universe with such a fluid will be dominated by it as any other fluid will be inversely proportionality to a . ρ_Λ is defined as

$$\rho_\Lambda = \frac{\Lambda c^2}{8\pi G} \quad (1.48)$$

where Λ [time] $^{-2}$ is the famous *cosmological constant*, and ρ_Λ is its energy density. It is also commonly referred to as the *vacuum energy*, as it is analogous to an overall linear shift in energy. The Friedmann equation with the cosmological constant

becomes

$$H^2 = \frac{8\pi G\rho}{3c^2} + \frac{\Lambda}{3} - \frac{kc^2}{R^2a^2} \quad (1.49)$$

The crucial thing for our purposes is that we can treat the constant as a fluid, and thus track its evolution and interaction with the other fluids in question, like matter and radiation. The cosmological constant, whose value we can measure, can thus be treated as a fluid that we have come to call *dark energy*. More on this in chapter 6.

1.4 The Point of All This Laborious Stuff

In approximately flat universes with fluids $\rho = \sum_w \rho_w$, the Friedmann equation is solved by $\rho_w = \rho_{w,0}a^{-3(1+w)}$ for each fluid. As with the cosmological constant, we can choose to track the curvature itself in this way, thus allowing for a dynamical structure that is modeled as a fluid ρ_k . Thus, we know that (homework):

$$\begin{aligned} \rho_k &\sim a^{-2} \\ \rho_m &\sim a^{-3} \\ \rho_r &\sim a^{-4} \\ \rho_\Lambda &\sim a^0 \end{aligned} \quad (1.50)$$

Note that we can assimilate all the components of the Friedmann equation into a linear combination:

$$H_0^2 = \frac{8\pi G\rho_0}{3c^2} + \frac{\Lambda}{3} - \frac{kc^2}{R^2a^2} \equiv \frac{8\pi G\rho_0^{\text{total}}}{3c^2} \quad (1.51)$$

We call this total energy density of today's universe the *critical energy density*, and label it

$$\rho_{\text{crit},0} \equiv \rho_0^{\text{total}} = \frac{3c^2}{8\pi G} H_0^2 \quad (1.52)$$

With it, we define the *density parameters* of each fluid (for today's universe, omitting the 0 subscript for convenience):

$$\Omega_{w,0} \equiv \frac{\rho_{w,0}}{\rho_{\text{crit},0}} \quad (1.53)$$

Conveniently, these density parameters must sum to

$$\sum_w \Omega_{w,0} = 1 + \frac{kc^2}{R^2 H_0^2} \quad (1.54)$$

We can use this to put the energy density of the curvature on equal footing with the rest, by letting $\rho_k = -3kc^4/(8\pi GR^2 a^2)$ and thus $\Omega_{k,0} = -kc^2/(R^2 H_0^2)$. And with these definitions we can write the Friedmann equation entirely in terms of density parameters:

$$\begin{aligned} H^2 &= \frac{8\pi G}{3c^2} \sum_w \rho_w - \frac{kc^2}{R^2 a^2}, \\ \left(\frac{H(a)}{H_0} \right)^2 &= \frac{\Omega_{r,0}}{a^4} + \frac{\Omega_{m,0}}{a^3} + \frac{\Omega_{k,0}}{a^2} + \Omega_{\Lambda,0}, \\ \left(\frac{H(z)}{H_0} \right)^2 &= \Omega_{r,0}(1+z)^4 + \Omega_{m,0}(1+z)^3 + \Omega_{k,0}(1+z)^2 + \Omega_{\Lambda,0} \end{aligned} \quad (1.55)$$

Now that is quite pretty. Each fluid's energy density scales differently over time with the expansion of the universe, and thus we can see how the universe will experience epochs of domination by each fluid at different stages; as well as how we can compare these epochs with the redshift parameter.

Observation and experiment have led to precise measurements of these density parameters, and we now know that ([Collaboration et al. 2013](#))

$$\begin{aligned} \Omega_{\Lambda,0} &\approx 0.69 \\ \Omega_{m,0} &\approx 0.31 \\ \Omega_{b,0} &\approx 0.05 \\ \Omega_{k,0} &\approx 0.004 \end{aligned} \quad (1.56)$$

Here, $\Omega_{b,0}$ is the density parameter for *baryonic* mass the mass in the Standard Model; the mass we understand. The point of this course is to explore the origin and the meaning of these numbers. Our questions about them are sometimes quite hard to state, but as a start we can just ask: what is all the non-baryonic and non-radiation stuff that we find? What are $\Omega_{CDM} = \Omega_m - \Omega_b$ and Ω_Λ ?

Chapter 2

Astrophysical Evidence for Dark Matter

[

All evidence in favor of particle [Dark Matter] thus far comes from observations of its gravitational effects on baryonic matter.

Mariangela Lisanti, (2018)

The evidence for dark matter is obtained by finding a mismatch between our normal-matter models and the experiments that attend these models. In general, the model will result in some density parameter Ω_b for baryonic matter, while our experiments will yield some total matter density parameter Ω_m such that we have a discrepancy $\Omega_m - \Omega_b = \Omega_{CDM} \neq 0$. The simplest explanation for that discrepancy is the existence of non-relativistic matter which doesn't interact with electromagnetic radiation at all, and so appears to be invisible, or "dark". We call this type of matter *cold dark matter*, or CDM.

The method thus requires an ideal model that yields a *prediction* Ω_b , and a separate experimental method that yields the *result* Ω_m . The following sections will explore a range of different models and experiments that suggest that the discrepancy is real: it is not a mere mistake of modelling or experiment that $\Omega_m > \Omega_b$; it is an important fact about cosmology.

The scientific method is not as clear cut as it seems. The fact that our models and experiments disagree does not in and of itself tell us which one to trust. Perhaps we have made a mistake in the experiment, or perhaps the experiment assumes the very bits of our theory that we ought to be questioning: the fact that things

look classical in most experiments, for instance, is not evidence against quantum mechanics; classical experiments yield classical results.

How then can we tell what to trust? Making another model or doing another experiment merely postpones the problem. It's important to think about this, because such choices need to be made in physics all the time, and we better get good at weighing up the evidence. The study of dark matter and dark energy provides an excellent testing ground for these kinds of methodological problems. We believe in dark matter for many disparate reasons, as with dark energy. But a lot of the individual pieces of evidence are unconvincing, or incomplete, if taken by themselves. And in the case of dark energy, as we will see in chapter 6, it is not even clear that there is a problem in the first place.

2.1 Methods of Indirect Detection

Before we get into the evidence for dark matter, we need to take a detour into the methods of observation. As we have seen, the broad idea is to estimate the density of baryonic matter on the one hand, and density of matter-in-general on the other, and to find a discrepancy that amounts to dark matter. The problem, as noted, is that dark matter is, well, dark: we cannot observe it directly with electromagnetic radiation. The detection of specific candidates for dark matter will be covered in chapter 5. Here we will sketch the methods for astrophysical detection. These rely, for the most part, on estimating the expected dynamics of galaxies due to baryonic gravitational effects, and detecting a very different dynamical behaviour that depends on a much stronger gravitational effect.

Recall that co-moving distance is $R\chi$ and physical distance is $l(t) = a(t)R\chi$. A problem: $l(t)$ is the distance traveled at time t , and we don't know what t is — we don't know, *a priori*, how long any given photon has been traveling. At closer distances we can get around this problem by using parallax effects to figure out distances. But this does not hold for distances larger than an order of 1000 pc. Thus we need to resort to indirect methods for determining distance.

The best indirect method we have is **luminosity distance**: look at the brightness of an object in the sky, and figure out the relation of the brightness to its distance; the two should correlate since light intensity dissipates by r^{-2} . The immediate problem is that we cannot know by looking whether something lacks intensity because it is far away or because it is small. So now we need another indirect method to determine *that*, i.e. the luminosity itself.

Our method for determining luminosity is to measure objects whose intrinsic

brightness can be known by other means; that is, whose intrinsic brightness has a known and fixed relation to other measurable properties. We call these objects *standard candles*. Cepheids pulsate with a periodicity that is related to their brightness; type Ia supernovae have a brightness determined by the Chandrasekhar limit; and gravitational waves (also called *standard sirens*) have a waveform related to the distance of their origin.

We call the *intrinsic luminosity* of an object L [energy][time] $^{-1}$. What we need for experimental utility is the *apparent luminosity* ℓ [energy][time] $^{-1}$ [area] $^{-1}$. In flat space this is easy, since energy spreads over the area of a sphere:

$$\ell = \frac{L}{4\pi r^2} \quad (2.1)$$

But in a generic FRW universe, we have to take account of the scale factor and the curvature of spacetime as well. The area of a sphere in the FRW metric is $4\pi R^2 S_k^2(\chi)$, which for $k = 0$ is identical to eq. (2.1).

The effect due to the expansion of space can be translated, as we have seen, into redshift effects on the photons we are measuring. We know that

$$\begin{aligned} \nu_0 &= \frac{2\pi c}{\lambda_0} = \frac{\nu_1}{1+z}, \\ \Rightarrow E_0 &= \hbar\nu_0 \frac{E_1}{1+z} \end{aligned} \quad (2.2)$$

and thus we have

$$\ell = \frac{L}{4\pi R^2 S_k^2(\chi)(1+z)^2} \quad (2.3)$$

for FRW universes. From this we can define the *luminosity distance* $d\ell = RS_k(\chi)(1+z)$, which is something we can measure with the luminosity L of a standard candle, the redshift parameter z , and the effect of the FRW metric $RS_k^2(\chi)$ which is just $R\chi = r$ in $k = 0$ flat space.

But in order to get back in touch with our cosmological models from chapter 1, we need to be able to measure H_0 as well. We begin by expanding the scale factor:

$$a(t) = 1 + H_0(t - t_0) - \frac{1}{2}q_0 H_0(t - t_0)^2 + \dots \quad (2.4)$$

where $q_0 \equiv -\ddot{a}/H_0$ is called the *deceleration parameter*¹.

The next step is to integrate a light path (exercise)

¹Which is obviously a stupid name given that our universe is doing no such thing. But we don't pick the names.

$$\begin{aligned}
R\chi &= c \int_{t_1}^{t_0} \frac{dt}{a(t)} \\
&= c \int_{t_1}^{t_0} dt [1 - H_0(t - t_0) + \dots] \\
&= c(t_0 - t_1) [1 + H_0(t_0 - t_1) + \dots]
\end{aligned} \tag{2.5}$$

where we call $t_0 - t_1$ the look-back time. We want this in terms of the redshift parameter (exercise):

$$z = 1 - \frac{1}{a(t_1)} \approx H_0(t_0 - t_1) + \frac{1}{2} (q_0 + 2) H_0^2(t_0 - t_1)^2 + \dots \tag{2.6}$$

which we get by inversion:

$$H_0(t_0 - t_1) = z - \frac{1}{2} (q_0 + 2) z^2 + \dots \tag{2.7}$$

from which we can infer that

$$\frac{H_0 R\chi}{c} = z - \frac{1}{2} (q_0 + 2) z^2 + \dots \tag{2.8}$$

which in flat space (eq. (2.3)) yields

$$\ell = \frac{c}{H_0} \left(z - \frac{1}{2} (q_0 + 2) z^2 + \dots \right) \tag{2.9}$$

for $z \ll 1$, which of course is no issue since $z_0 \approx 0$. Since we can measure both ℓ and z , we can use them to fix H_0 and q_0 . When all this is said and done we have a method of establishing, from standard candles and the maths in this section, the distance to galaxies, as well as the current rate of acceleration H_0 . This is key to the observations and evidence that follow in this chapter.

2.2 Galactic Clusters and the Virial Theorem

The Virial Theorem gives us a theoretical prediction for the mass of galaxy clusters which disagrees with the mass we can observe in detectable stars and gas. This section will discuss the Virial Theorem and the methods used for counting, as well as the discrepancies in their results.

The Virial Theorem should be familiar from classical mechanics and many other

areas of physics. You may have seen it in the following form:

$$\langle T \rangle = -\frac{1}{2} \langle V \rangle , \quad (2.10)$$

where $\langle X \rangle$ stands for the time average of a quantity X , T is the kinetic energy, and V the potential. It takes some steps to show why this is a general fact that we may borrow freely in cosmology.

We will use spherical coordinates to exploit the symmetry of gravitational potential which will allow us to ignore two degrees of freedom (θ and ϕ).

The gravitational potential between masses m_i and m_j is denoted by V_{ij} , and given by

$$V_{ij} = -\frac{Gm_i m_j}{|\mathbf{x}_j - \mathbf{x}_i|} , \quad (2.11)$$

$$= -\frac{Gm_i m_j}{|\mathbf{r}_{ij}|} , \quad (2.12)$$

where G is the gravitational constant, \mathbf{x}_i is the location of the i^{th} particle in some arbitrary coordinate system, and \mathbf{r}_{ij} is the separation of the i^{th} and j^{th} particle. Then we can say that the force \mathbf{F}_i on the i^{th} particle is given by²

$$\mathbf{F}_i = -\sum_{j \neq i}^N \nabla_i V_{ij} , \quad (2.13)$$

$$= \dot{\mathbf{p}}_i . \quad (2.14)$$

Due to spherical symmetry we can ignore all but the radial component of the grad³, thus getting the rather neat result

$$-\nabla_i V_{ij} = -\partial_r V_{ij} \hat{\mathbf{r}} , \quad (2.15)$$

$$= -\frac{Gm_i m_j}{|\mathbf{x}_i - \mathbf{x}_j|^2} \frac{\mathbf{r}_{ij}}{|\mathbf{r}_{ij}|} , \quad (2.16)$$

$$\equiv -\frac{Gm_i m_j}{|\mathbf{r}_{ij}|^2} \hat{\mathbf{r}} . \quad (2.17)$$

²Where we take the leftmost letter in below the summation sign to be the one summed over, j in this case.

³Recall that $\nabla = \partial_x + \partial_y + \partial_z$ in Cartesian coordinates and $\nabla = \partial_x \hat{\mathbf{x}} + \partial_y \hat{\mathbf{y}} + \partial_z \hat{\mathbf{z}}$ and $\nabla = \frac{\partial}{\partial r} \hat{\mathbf{r}} + \frac{1}{r} \frac{\partial}{\partial \theta} \hat{\theta} + \frac{1}{r \sin \theta} \frac{\partial}{\partial \varphi} \hat{\varphi}$ in spherical coordinates; thus, with spherical symmetry and only radial variation, we have $\nabla = \frac{\partial}{\partial r} \hat{\mathbf{r}} \equiv \partial_r \hat{\mathbf{r}}$.

The total potential is given by:

$$\sum_i^N \sum_{j \neq i}^N V_{ij} = V \quad (2.18)$$

Now recall that the total kinetic energy T of N masses m_i is written as

$$T = \sum_i^N \frac{1}{2} m_i \dot{\mathbf{x}}_i^2 = \sum_i^N \frac{1}{2} m_i (\dot{\mathbf{x}}_i \cdot \dot{\mathbf{x}}_i) \quad (2.19)$$

We now define *the virial* \dot{I} indirectly via I :

$$I = \frac{1}{2} \sum_i^N m_i (\mathbf{x}_i \cdot \mathbf{x}_i) \quad (2.20)$$

From which it follows that

$$\begin{aligned} \dot{I} &= \frac{1}{2} \sum_i^N m_i (\dot{\mathbf{x}}_i \cdot \mathbf{x}_i) + \frac{1}{2} \sum_i^N m_i (\mathbf{x}_i \cdot \dot{\mathbf{x}}_i) \\ &= \sum_i^N m_i (\dot{\mathbf{x}}_i \cdot \mathbf{x}_i) \\ &= \sum_i^N \mathbf{p}_i \cdot \mathbf{x}_i \end{aligned} \quad (2.21)$$

and thus

$$\begin{aligned} \ddot{I} &= \sum_i^N \dot{\mathbf{p}}_i \cdot \mathbf{x}_i + \sum_i^N \mathbf{p}_i \cdot \dot{\mathbf{x}}_i \\ &= \sum_i^N \mathbf{F}_i \cdot \mathbf{x}_i + \sum_i^N m_i (\dot{\mathbf{x}}_i \cdot \dot{\mathbf{x}}_i) \end{aligned} \quad (2.22)$$

which is the same as

$$\ddot{I} = \left(\sum_i \mathbf{F}_i \cdot \mathbf{x}_i \right) + 2T \quad (2.23)$$

But we know \mathbf{F}_i , so we get

$$\sum_{i \neq j}^N \mathbf{F}_i \cdot \mathbf{x}_i = - \sum_{i \neq j}^N \left(\left(\sum_{j \neq i}^N \nabla_i V_{ij} \right) \cdot \mathbf{x}_i \right) \quad (2.24)$$

Note that summing over all $i \neq j$ is equivalent to summing over all $i < j$ and all $i > j$. Let $\mathbf{f}_{ij} \equiv -\nabla_i V_{ij}$ and recall that Newton's 3rd law states that $\mathbf{f}_{ij} = -\mathbf{f}_{ji}$:

$$\begin{aligned} \sum_{i \neq j}^N \mathbf{F}_i \cdot \mathbf{x}_i &= \sum_{i < j}^N \left(\left(- \sum_{j \neq i}^N \nabla_i V_{ij} \right) \cdot \mathbf{x}_i \right) + \sum_{j < i}^N \left(\left(- \sum_{j \neq i}^N \nabla_i V_{ij} \right) \cdot \mathbf{x}_i \right) \\ &\equiv \sum_{i < j}^N \left(\sum_{j \neq i}^N \mathbf{f}_{ij} \cdot \mathbf{x}_i \right) + \sum_{j < i}^N \left(\sum_{j \neq i}^N \mathbf{f}_{ij} \cdot \mathbf{x}_i \right) \\ &= \sum_{i < j}^N \left(\sum_{j \neq i}^N \mathbf{f}_{ij} \cdot \mathbf{x}_i \right) - \sum_{j < i}^N \left(\sum_{j \neq i}^N \mathbf{f}_{ji} \cdot \mathbf{x}_i \right) \\ &= \sum_{i < j}^N \left(\sum_{j \neq i}^N \mathbf{f}_{ij} \cdot \mathbf{x}_i \right) - \sum_{j < k}^N \left(\sum_{j \neq k}^N \mathbf{f}_{jk} \cdot \mathbf{x}_k \right) \\ &= \sum_{i < k}^N \left(\sum_{k \neq i}^N \mathbf{f}_{ik} \cdot \mathbf{x}_i \right) - \sum_{i < k}^N \left(\sum_{i \neq k}^N \mathbf{f}_{ik} \cdot \mathbf{x}_k \right) \\ &= \sum_{i < k}^N \left(\sum_{k \neq i}^N \mathbf{f}_{ik} \right) \cdot (\mathbf{x}_i - \mathbf{x}_k) \\ &= \sum_{i < k}^N \left(- \sum_{k \neq i}^N \nabla_i V_{ik} \right) \cdot (\mathbf{x}_i - \mathbf{x}_k) \end{aligned} \quad (2.25)$$

And recalling eq. (2.15):

$$\begin{aligned}
\sum_{i \neq j}^N \mathbf{F}_i \cdot \mathbf{x}_i &= \sum_{i < j}^N \left(- \sum_{j \neq i}^N \nabla_i V_{ij} \right) \cdot (\mathbf{x}_i - \mathbf{x}_j) \\
&= \sum_{i < j}^N \left(- \sum_{j \neq i}^N \frac{Gm_i m_j}{|\mathbf{x}_i - \mathbf{x}_j|^2} \frac{\mathbf{r}_{ij}}{|\mathbf{r}_{ij}|} \right) \cdot (\mathbf{x}_i - \mathbf{x}_j) \\
&= - \sum_{i < j}^N \sum_{j \neq i}^N \frac{Gm_i m_j}{|\mathbf{x}_i - \mathbf{x}_j|^2} \frac{\mathbf{r}_{ij}}{|\mathbf{r}_{ij}|} \cdot \mathbf{r}_{ij} \\
&= - \sum_{i < j}^N \sum_{j \neq i}^N \frac{Gm_i m_j}{|\mathbf{x}_i - \mathbf{x}_j|} \\
&= \sum_{i < j}^N \left(\sum_{j \neq i}^N - \frac{Gm_i m_j}{|\mathbf{r}_{ij}|} \right) \equiv \sum_i^N \left(\sum_{j \neq i}^N V_{ij} \right) \\
&= V
\end{aligned} \tag{2.26}$$

And thus at long last we are approaching the theorem itself:

$$\ddot{I} = \left(\sum_i \mathbf{F}_i \cdot \mathbf{x}_i \right) + 2T = V + 2T \tag{2.27}$$

All that is left is to take the time average of \ddot{I} . The time average of a quantity $X(t)$ is given by

$$\bar{X}(\tau) \equiv \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_0^\tau X(t) dt \tag{2.28}$$

and thus (applying the Fundamental Theorem of Calculus)

$$\bar{\ddot{I}} = \lim_{\tau \rightarrow \infty} \frac{\dot{I}(\tau) - \dot{I}(0)}{\tau} = 0 \tag{2.29}$$

iff t grows more rapidly than $\dot{I}(t)$. This condition holds if we take the system to be in equilibrium and to be bound. As such, we have

$$0 = \overline{V + 2T} = \bar{V} + 2\bar{T}$$

$$\Rightarrow \bar{T} = -\frac{1}{2}\bar{V} \quad (2.30)$$

since the time average is linear.⁴ Thus, the Virial Theorem for time averages is settled.

Now, this isn't actually all that helpful, since we obviously can't measure \bar{V} or \bar{T} , as there is no sense to measuring a quantity as it approaches an infinite lifetime (we sure won't be around by then). So we need to put the theorem in a more empirically manageable form. That is, we need to relate it to an the average values of a sampleable ensemble. This is achieved by the *ergodic hypothesis*, which states that all microstates of a system are equiprobable over infinite time-evolution of the system. As a consequence:

$$\lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_0^\tau X(t) dt \approx \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x(i) \quad (2.31)$$

This may look like a quantization of a continuous quantity, but that's not the right way to read it, as it holds for classical, relativistic, and quantum systems. Rather, it states that we can take the integrated average $X(t)$ of the system *over time* to be approximately equal to the sum *over particle-number* N of the microstates $x(i)$ at an instant.

For our purposes, the ergodic hypothesis simply allows us to go from the time-average to the ensemble-average:

$$\bar{T} \approx \langle T \rangle = \frac{1}{N} \sum_{i=1}^N m_i \dot{\mathbf{x}}_i^2, \quad (2.32)$$

Thus we arrive at the version of the Virial Theorem we need to estimate the masses of galaxy clusters. Suppose finally that $\forall i \in \{N\} (m_i \approx m)$, such that

$$\langle T \rangle \approx \frac{1}{2}m \langle \dot{\mathbf{x}}^2 \rangle \quad (2.33)$$

⁴Integration is linear because it is a sum. Limits are linear by definition. Division is linear because $\forall a, b, x, y, z (\frac{ax+by}{z} = a\frac{x}{z} + b\frac{y}{z})$. And any operation composed of only linear operations must itself be linear.

and so

$$\langle V \rangle \approx -m \langle \dot{\mathbf{x}}^2 \rangle \quad (2.34)$$

But we already know that

$$\langle V \rangle \approx -\frac{Gm^2}{\langle r \rangle} N \quad (2.35)$$

Therefore, we have

$$\langle \dot{\mathbf{x}}^2 \rangle \approx Gm^2 \left\langle \frac{1}{r} \right\rangle N \quad (2.36)$$

or

$$Nm \approx \frac{2 \langle \dot{\mathbf{x}}^2 \rangle}{G \langle r^{-1} \rangle} \quad (2.37)$$

This is the equation we need to make predictions. Measuring the average of the square-velocities of galaxies in clusters by measuring the redshift of their light spectra and assuming spherical symmetry, as well as measuring their average radii, allows us to obtain an estimate for the total mass $M_V = Nm$.

However, when we count the number of stars by measuring luminosity, we get a much smaller mass M_B . The Virial mass greatly exceeds the electromagnetically visible mass, in other words. The difference in mass $M_{CDM} = M_V - M_B$, we label as dark matter.

2.3 Modified Newtonian Dynamics

2.4 Gravitational Lensing

2.5 Bullet Cluster

The Bullet Cluster is our best evidence to support dark matter over modified theories of gravity.

The Bullet Cluster refers to the smaller of two galaxy clusters which underwent

a collision around 150 million years ago, although the name is also used frequently in reference to the pair of clusters; the descriptor refers to the smaller cluster having passed through the larger cluster with similar effects to a bullet.

The importance of the Bullet Cluster for dark matter lies in how different components of the cluster behave under collision.

Galaxy clusters are formed of gas and stars, and possibly dark matter.

Since stars within a cluster are extremely spaced out, the majority of them will pass straight through the other cluster unaffected, except for minor gravitational slowing. We observe this in the visible light spectrum.

The gases of each cluster, however, interact electromagnetically as well as gravitationally, and so are slowed to a much greater degree. This causes the gas, visible in the X-ray wavelengths, to lag behind the visible stars.

We can use gravitational lensing to observe where the majority of the mass of the clusters lies. If there is no dark matter, MOND predicts that this majority of mass would lie in the gas. Dark matter theories would instead place it in a similar location to the stars, since its theorised electromagnetic interaction cross-section is so small, and it would not have interacted with (and thus been slowed by) the gas; additionally, its self-interaction cross-section would have to be similarly small.

Indeed, analysis of the gravitational lensing of the two clusters places the majority of the mass outside the gas, more or less coincident with the stars. It has been confirmed to 8σ that this location cannot be explained via MOND alone⁵.

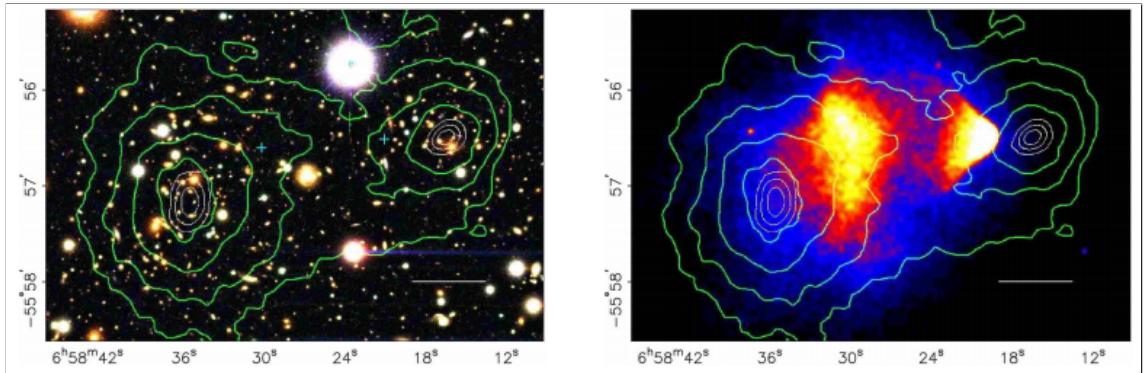


Figure 2.1: Left: Optical images from the Magellan telescope with over-plotted contours of spatial distribution of mass, from gravitational lensing. Right: The same contours over-plotted over Chandra x-ray data that traces hot gas in a galaxy. It can be seen that most of the matter resides in a location different from the gas.

The interpretation of this observation is that while the gas has collided with itself,

⁵Without MOND, the matter discrepancy is a factor of 10. With MOND, it is a factor of 2.

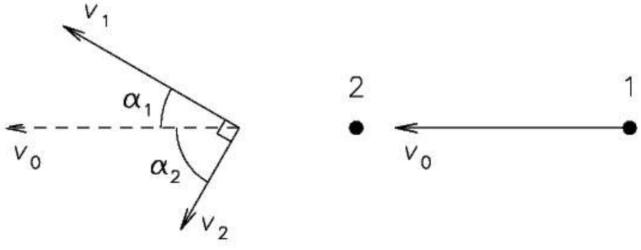


Figure 2.2: The Collision of two dark matter particles in the rest frame of the smaller cluster.

the galaxies haven't, and there is an additional component of mass which also hasn't collided with itself and is no longer in the same place as the gas. This is often cited as strong evidence against MOND, but also allows us to place a constraint on the self-interaction cross-section of dark matter.

Lensing suggests that the core of the smaller cluster has a radius of 70 kpc and a density of around $\rho = 1.3 \times 10^{-25} \text{ g cm}^{-3}$. It has an escape velocity of around 1200 km s^{-1} and observations of the gas dynamics suggest it crashed through the larger cluster at a speed of 4800 km s^{-1} .

There are various ways of using these observations to place a constraint on the self interaction cross section of dark matter, all of which give answers which are similar to within an order of magnitude or so. I am going to present what I think is the simplest one, which involves the fact that the smaller cluster survives the collision and still exists on the other side, having punched through the larger cluster.

Let's look at the collision of two dark matter particles of equal mass in the rest frame of the subcluster as shown in section 2.5 . In the smaller cluster's reference frame, particle 2 is at rest and particle 1, in the incoming flow, collides with it with a velocity

$$v_0 \approx 4800 \text{ km s}^{-1}, \quad (2.38)$$

where v_0 is the smaller cluster's velocity. Particle 1 scatters at an angle $\alpha_1 > 0$ with a velocity v_1 , while particle 2 acquires a velocity v_2 at an angle $\alpha_2 > 0$. From energy and momentum conservation, $\alpha_1 + \alpha_2 = \pi/2$ and we can see that the velocities after scattering will be given by

$$v_1 = v_0 \cos \alpha_1, \quad v_2 = v_0 \sin \alpha_1. \quad (2.39)$$

Now if the collision results in one of the particles escaping and the other remaining in the smaller cluster then this interaction will not reduce the probability of the smaller cluster surviving. We would like to look at situations where both dark matter particles are given enough velocity to escape from the smaller cluster, consequently

$$\begin{aligned}\frac{v_1^2}{v_0^2} &= \cos^2(\alpha_1) = 1 - \sin^2(\alpha_1) > \frac{v_{esc}^2}{v_0^2} \\ \frac{v_2^2}{v_0^2} &= \sin^2(\alpha_1) > \frac{v_{esc}^2}{v_0^2}\end{aligned}\quad (2.40)$$

So that the criterion to have both particles obtaining a velocity greater than the escape velocity of the smaller cluster can be written

$$\frac{v_{esc}^2}{v_0^2} < \sin^2(\alpha_1) < 1 - \frac{v_{esc}^2}{v_0^2}\quad (2.41)$$

Now we go to the centre of mass frame of the particle collision. In this reference frame, particle 1 scatters by an angle $\theta = 2\alpha_1$ ($0 < \theta < \pi$) and particle 2 scatters by $\pi - \theta$, so long as we maintain the definition and sense of α_1 and α_2 as set out in section 2.5. The criterion set out in eq. (2.41) is now

$$\frac{v_{esc}^2}{v_0^2} < \sin^2(\theta/2) < 1 - \frac{v_{esc}^2}{v_0^2}\quad (2.42)$$

and in this frame, we expect the scattering to be isotropic. Because of this, we need to take into account the fact that there is more solid angle around $\theta = \pi/2$ than around $\theta = 0$ or $\theta = \pi$, so we need to do a weighted integral in the normal way. Then the probability of ejection of both particles per collision is given by

$$\chi = \frac{\int_{\theta_{min}}^{\theta_{max}} 2\pi \sin \theta d\theta}{\int_0^\pi 2\pi \sin \theta d\theta}\quad (2.43)$$

where θ_{min} and θ_{max} will be determined by eq. (2.42). We can solve this neatly

$$\begin{aligned}\int_{\theta_{min}}^{\theta_{max}} 2\pi \sin \theta d\theta &= 2\pi [-\cos \theta]_{\theta_{min}}^{\theta_{max}} = 2\pi [2 \sin^2(\theta/2) - 1]_{\theta_{min}}^{\theta_{max}} \\ &= 2\pi \left(2 - 2\frac{v_{esc}^2}{v_0^2} - 1 - 2\frac{v_{esc}^2}{v_0^2} + 1 \right) = 4\pi - 8\pi \frac{v_{esc}^2}{v_0^2}\end{aligned}\quad (2.44)$$

and since the bottom integral in χ is just 4π we have

$$\chi = 1 - 2\frac{v_{esc}^2}{v_0^2}\quad (2.45)$$

Now the optical depth τ of the oncoming dark matter wind for a single dark matter particle gives a measure of the probability of a dark matter particle in the smaller cluster colliding with one from the bigger cluster (the probability of not being hit will be $e^{-\tau}$) for τ given by

$$\tau = \frac{\sigma}{m} \Sigma \quad (2.46)$$

where m is the mass of the dark matter particle, σ is its self interaction cross section and Σ is the surface density of the larger cluster, around 0.3 g cm^{-2} .

We would like to ensure that the smaller cluster does not lose more than 30% of its mass, since the smaller cluster's overall mass to light ratio is about eight for that kind of object. We therefore require that $\chi\tau < 0.3$ which leads to

$$\frac{\sigma}{m} < 1 \text{ cm}^2 \text{ g}^{-1} \quad (2.47)$$

This is actually quite a large cross section, for example, it corresponds to about 10^{-24} cm^2 for a GeV mass dark matter particle. This is very similar to the cross section corresponding to the nuclear force between atoms, so for example, neutrons would scatter off each other with this cross section⁶.

Note this assumes that the cross section is short range, in other words that the particles are a bit like little ping-pong balls. Technically, this corresponds to a Yukawa, short range interaction when the force carrying boson has a mass which sets the cross section. In the case where one considers long range forces, for example endowing the dark matter with a dark electric charge, the analysis is very much more complicated.

2.6 Rotation Curves

If we make the simplifying assumption that matter in galaxies is distributed with spherical symmetry, we can estimate the rotation velocity of matter at r distance from the center by

$$v_c = \sqrt{\frac{GM}{r}} \quad (2.48)$$

since Newton's 2nd law gives $mv_c^2/r = GMm/r^2$ for rotational motion. Thus

$$v_c \sim r^{-1/2} \quad (2.49)$$

But when we observe the rotation speeds we find that

⁶protons have electric charge so there would be a long range force.

$$\begin{aligned} M(r) &\sim r \\ \Rightarrow v_c &\sim \text{constant} \end{aligned} \tag{2.50}$$

for sufficiently large $r \gtrsim R_{\text{disk}}$. This is key evidence for dark matter. Baryonic matter alone simply do not result in rotational velocities that “flatten out” as the radius increases. Baryonic matter “slows down” by the inverse square law.

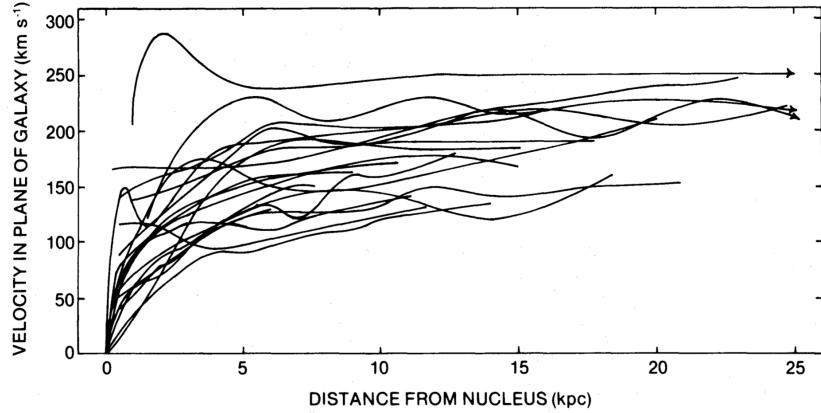


Figure 2.3: Rotation curves of spiral galaxies, measured by [Rubin et al. \(1980\)](#); figure via [Lisanti \(2018\)](#).

2.6.1 The Isothermal Sphere

Qualitatively, we expect dark matter to be distributed in a spherically symmetric halo about the galaxy-center; since is not dissipative, it therefore doesn’t collapse into a disk the way run-of-the-mill baryons do. It doesn’t quite float free, since it feels the gravitational pull of surrounding dark matter and baryonic matter, but it doesn’t collapse either. The isothermal sphere⁷ is a model of this which leads to flat rotation curves.

We need $v_c \sim r^{-1/2} \approx \text{constant}$. Therefore we need $M \propto r$:

$$m(r') = \int_0^{r'} 4\pi\rho(r)r^2 dr \tag{2.51}$$

Note that we are assume a flat rotation curve right from the centre of the galaxy ($r = 0$). This is not quite accurate, but is fine for our purposes.

If we say that

$$\rho(r) = \rho_s \left(\frac{r}{r_s} \right)^\gamma \tag{2.52}$$

⁷You will establish why this name is appropriate in a homework set.

where ρ_s is the value for the density at some radius r_s . This equation leads us to

$$M(r) = \frac{4\pi}{3+\gamma} \rho_s \frac{r^{3+\gamma}}{r_s^\gamma} \quad (2.53)$$

We can see that $\gamma = -2$ gives us the correct rotation curve. Thus, $\rho(r) = \rho_s(r_s/r)^2$ is the equation which will be useful to us.

The mass of the dark matter halo can be estimated as

$$M_{\text{halo}} \sim 10^{12} M_\odot \quad (2.54)$$

and the density value for dark matter today is estimated at roughly

$$\rho_{CDM,0} \sim 0.3 \text{ GeV cm}^{-3} \quad (2.55)$$

such that

$$M_{\text{halo}} \sim \int_0^{R_{\text{halo}}} dr r^2 \rho(r) \quad (2.56)$$

$$\Rightarrow R_{\text{halo}} \sim 100 \text{ kpc}$$

Average dark matter velocity in halo is given by the Virial Theorem:

$$\langle v \rangle \sim \sqrt{\frac{GM_{\text{halo}}}{R_{\text{halo}}}} \sim 200 \text{ km s}^{-1} \quad (2.57)$$

Thus, the dark matter in the halo is non-relativistic. Rotation curves thus provide both evidence for dark matter and a constraint on its behavior.

2.7 Freeze-Out

Freeze-out occurs for a *particle species*⁸ χ when $H \gtrsim \Gamma$; that is, when the rate of expansion of spacetime starts to approach and exceed the rate of interactions between χ s.

You can think of freeze-out as the point at which different subsystems stop communicating with each other: space expands at such a rate that we cannot treat their interactions with statistical averages any longer, because the different subsystems are not able to send signals back and forth at a rate high enough to affect each other's internal dynamics — they become isolated.

⁸Not for a *particle*. That is nonsense, since freeze-out is a statistical concept.

Why should we care about freeze-out? A combination of methods from statistical mechanics and particle physics allows us to estimate, given reasonable assumptions and experimental constraints, when exactly a species *froze out* and how densely it should populate the universe at any given past, present, and future time. The thought is simple, the method is hard.

This section contains a lot of mathematics that seems simple of the surface level, but only because it hides nine levels of complexity that you need to understand to really see what is going on. See the suggested readings for more, and use Google, Wikipedia, and StackExchange liberally, and help each other.

The full description of a system of N particles is given by its evolution in $6N$ -dimensional phase space. The probability of finding the *system* (*not* a particle by itself, so this is not yet a measurable quantity) in a $6N$ -dimensional region $[Q_k, Q_k + dQ_k]^{6N}$ for $Q_k = x_i, p_j$ and $i, j, k < N \in \mathbb{N}$ at time t is given by

$$\rho(x_1, \dots, x_N; p_i, \dots, p_N; t) \prod_i^{3N} dx_i \prod_j^{3N} dp_j \quad (2.58)$$

where $dx_i \prod_j^{3N} dp_j$ is the volume element and $\rho(x_1, \dots, x_N; p_i, \dots, p_N; t)$ is the phase space density distribution for the system. In practice, this cannot be calculated by human beings or modern computers for an interesting system. We need a simplification.

Boltzmann found one: an average density distribution $f(\mathbf{x}, \mathbf{p}, t)$ defined over 6×1 -dimensional phase-space for *one* particle gives $f(\mathbf{x}, \mathbf{p}, t) d^3\mathbf{x} d^3\mathbf{p}$ chance of finding the system in a given $d^3\mathbf{x} d^3\mathbf{p}$ volume element. We haven't said anything about the nature of $f(\mathbf{x}, \mathbf{p}, t)$. More on that later⁹. First, we constrain it by normalizing $\frac{f(\mathbf{x}, \mathbf{p}, t)}{N}$:

$$\int \int d^3\mathbf{x} d^3\mathbf{p} \frac{f(\mathbf{x}, \mathbf{p}, t)}{N} \equiv 1 \quad (2.59)$$

and note that it must scale as the full integration of the horrible (but specific) density above:

$$\frac{1}{N} f(\mathbf{x}, \mathbf{p}, t) \sim \int \cdots \int \rho(x_1, \dots, x_N; p_i, \dots, p_N; t) \prod_i^{3N} dx_i \prod_j^{3N} dp_j \quad (2.60)$$

The state of a system at an instance is hard to figure out precisely, as we have seen,

⁹See recommended readings for applications of $f(\mathbf{x}, \mathbf{p}, t)$.

so it should come as no surprise that the time-evolution is even harder. Again we turn to $f(\mathbf{x}, \mathbf{p}, t)$, and take its time-evolution to be given by the Liouville equation:

$$\frac{df}{dt} = 0 \quad (2.61)$$

where df/dt is a total derivative along a phase space trajectory $(\mathbf{x}(t), \mathbf{p}(t))$, such that

$$\frac{\partial f}{\partial t} + \dot{\mathbf{x}} \cdot \frac{\partial f}{\partial \mathbf{x}} + \mathbf{F} \cdot \frac{\partial f}{\partial \mathbf{p}} = 0 \quad (2.62)$$

This assumes that we are considering an ideal gas, where particles do not collide with each other, but are rather evolving according to individual and non-interacting Hamiltonians. If that is the case, then the Boltzmann distribution for the gas will evolve according to (2.62) — it is called the collisionless Boltzmann equation.

You are encouraged to find out the deeper facts about the Boltzmann equation and how it fares in more complex situations, where we allow for collisions and non-ideal gases (the right hand side is non-zero in these cases, and nontrivial to derive; see [Landau et al. \(1981, vol 10\)](#)). The complex version actually looks a lot prettier once we use Einstein summation, so I will put it here for those who are curious:

$$\underbrace{\mathbf{L}[f]}_{\text{Liouville operator}} = \underbrace{\mathbf{C}[f]}_{\text{Collision operator}} = p^\alpha \frac{\partial f}{\partial x^\alpha} - \Gamma_{\beta\gamma}^\alpha p^\beta p^\gamma \frac{\partial f}{\partial p^\alpha} \quad (2.63)$$

Equation 2.63 expresses the most general equation for the time-evolution of a generic Boltzmann distribution f . But since the Dark Matter we observe in the Milky Way appears to be approximately collisionless and definitely non-relativistic (see section 2.6), we can stick to eq. (2.62).

Assume now that we are dealing with a steady state, i.e. $\partial f / \partial t = 0 \Rightarrow f(\mathbf{x}, \mathbf{p}, t) = f(\mathbf{x}, \mathbf{p})$. It follows from the Jean's Theorem ([Binney and Tremaine 2011](#), p. 283) that steady-state solutions to (2.62) must be functions of phase-space coordinates through an integral of motion $I(\mathbf{x}, \mathbf{p})$

$$\frac{d}{dt} I(\mathbf{x}(t), \mathbf{p}(t)) = 0 \quad (2.64)$$

and that the phase-space distribution can thus always be a function solely of energy E

$$f(\mathbf{x}, \mathbf{p}) = f(E) \quad (2.65)$$

for $E = \Psi - \frac{1}{2}m\dot{\mathbf{x}}^2$, where Ψ is the gravitational potential. (This is not the *only*

solution: there are many different and interesting ways of cashing in f , but for our purposes $f(E)$ will do just fine.) Suppose now that $f(E) \propto e^E$ such that

$$\rho \propto \int_0^\infty dv v^2 f(v) = \int_0^\infty dv v^2 \exp\left(\frac{\Psi - v^2/2}{\sigma^2}\right) \propto \exp\left(\frac{\Psi}{\sigma^2}\right) \quad (2.66)$$

where σ is the velocity dispersion — how stretched or squeezed the distribution is about its mean value.

The Poisson equation is as usual

$$\nabla^2 \Psi = -4\pi G \rho \quad (2.67)$$

and if we take $\Psi \propto \ln \rho$ then

$$\rho(r) = \frac{\sigma}{2\pi G r^2} \propto \frac{1}{r^2} \quad (2.68)$$

for $f(v) \propto e^{-v^2/\sigma}$.

Consider the interaction $\chi\bar{\chi} \leftrightarrow x_i\bar{x}_i$. In chemical equilibrium, χ get replenished continuously via inelastic collisions. But as the universe expands, swaths of dark matter decouple and can no longer be treated as though they are in chemical equilibrium; this is freeze-out. The density reached at freeze-out is final. It occurs when

$$\Gamma_{inelastic} = n_\chi \langle \sigma v \rangle \sim H = \frac{\dot{a}}{a} \quad (2.69)$$

where n_χ is the number density of Dark Matter and $\langle \sigma v \rangle$ is the *thermally* averaged (or averaged over velocity) cross-section, which you can think of as a distribution of reaction-probabilities indexed by velocity, so that we can track how likely a reaction is at any given velocity *and* what the distribution of such probabilities looks like. *Cold* Dark Matter (CDM) is non-relativistic at freeze-out, and its $n_\chi \sim T^{3/2} e^{m_\chi/T}$ — where T is the temperature of χ s; *Hot* Dark Matter is relativistic with $n_\chi \sim T^3$; Warm Dark Matter falls in between.

After freeze-out, chemical equilibrium is as discussed not possible, but thermal equilibrium via elastic collisions with components x_i of surrounding systems *is*. But this interaction as well will freeze out once the remaining systems decouple, at

$$\Gamma_{elastic} = n_{x_i} \langle \sigma v \rangle \quad (2.70)$$

for n_{x_i} the number density of x_i , which is taken to be relativistic (since it must survive the inelastic freeze-out which occurs for lower than relativistic velocities) and

scales like T^3 .

The rest of this section will only consider Cold Dark Matter (CDM), since this is what we appear to find in our universe. To calculate its current density we need the Boltzmann equation, which from our simplifications above and letting $f(E, t)$ evolve over time:

$$\mathbf{L}[f] = E \frac{\partial f}{\partial t} - \frac{\dot{a}}{a} |\mathbf{p}|^2 \frac{\partial f}{\partial E} \quad (2.71)$$

The phase-space number-density n_χ of a species χ is given by

$$n_\chi = g \int f(E, t) \frac{d^3 p}{(2\pi)^3} \quad (2.72)$$

where g is the spin degrees of freedom for χ . (This is pretty basic statistical quantum mechanics, so you will need to look this up if you don't understand it already.) The next step is to simply make a substitution:

$$g \int \mathbf{L}[f] \frac{d^3 p}{(2\pi)^3} = \frac{1}{a^3} \frac{d}{dt} (na^3) = \frac{dn}{dt} + 3Hn \quad (2.73)$$

Recall that $\mathbf{L}[f] = \mathbf{C}[f]$, so the above equation becomes rather complicated if the collision term is too general:

$$\begin{aligned} g_1 \int \mathbf{C}[f_1] \frac{d^3 p_1}{(2\pi)^3} &= - \sum_{\text{spins}} [f_1 f_2 (1 \pm f_3) (1 \pm f_4) |\mathcal{M}_{12 \rightarrow 34}|^2 - f_3 f_4 (1 \pm f_1) (1 \pm f_2) |\mathcal{M}_{34 \rightarrow 12}|^2] \\ &\times (2\pi)^4 \delta^4(p_1 + p_2 - p_3 - p_4) \prod_{i=1}^4 \frac{d^3 p_i}{(2\pi)^3 2E_i} \end{aligned} \quad (2.74)$$

Which is as ugly as it looks: g_i is spin degrees of freedoms for particle i and f_i is the phase space density for i ; $\mathcal{M}_{x \rightarrow y}$ is the matrix element for the reaction $x \rightarrow y$; $1 \pm f_i$ represents Pauli blocking and Bose enhancement, with $+, -$ representing bosons and fermions respectively, as usual. All this is a complicated mathematical way of spelling out the following fact: boson-boson reactions are preferred over boson-fermion reactions and fermion-fermion reactions are avoided.

We need some simplifying assumptions to get out of this mess if we are to stand any chance of getting useful cosmological facts out of it. First, we assume there is kinetic equilibrium, so that we can use Fermi-Dirac and/or Bose-Einstein statistics for f . Second, we assume that each species i has a temperature $T_i \ll E_i - \mu_i$, such that they can be treated with the Maxwell-Boltzmann distribution and $(1 \pm f_i) \sim 1$.

Third, the Standard Model particles in the interaction are in thermal equilibrium with the photon bath, such that we can use known information about photons to eliminate unknowns.

The matrix element relates to the cross section by:

$$\begin{aligned} \sum_{\text{spins}} \int |\mathcal{M}_{ij \rightarrow kl}|^2 \times (2\pi)^4 \delta^4(p_1 + p_2 - p_3 - p_4) \frac{d^3 p_k}{(2\pi)^3 2E_k} \frac{d^3 p_l}{(2\pi)^3 2E_l} \\ = 4g_i g_j \sigma_{ij} \sqrt{(p_i \cdot p_j)^2 - (m_i m_j)^2} \end{aligned} \quad (2.75)$$

where σ_{ij} is the cross-section for the scattering process. Substitution yields:

$$g_1 \int C[f_1] \frac{d^3 p_1}{(2\pi)^3} = - \int \{(\sigma v_{\text{Møller}})_{12} dn_1 dn_2 - (\sigma v_{\text{Møller}})_{34} dn_3 dn_4\} \quad (2.76)$$

with

$$(v_{\text{Møller}})_{ij} \equiv \frac{\sqrt{(p_i \cdot p_j)^2 - (m_i m_j)^2}}{E_i E_j} \quad (2.77)$$

Note now that $\sigma v_{\text{Møller}}$ varies slowly with the number density. Thus, we can factor it out of the integrand:

$$\frac{dn_1}{dt} + 3Hn_1 = -\langle \sigma v_{\text{Møller}} \rangle_{12} n_1 n_2 - \langle \sigma v_{\text{Møller}} \rangle_{34} n_3 n_4 \quad (2.78)$$

From now on let $v_{\text{Møller}} \equiv v$ to simplify the notation. Recall that we are interested in a situation where particles of type 1 and 2 are of the same type (cold dark matter) and where particles 3 and 4 are in equilibrium with the photon bath

$$\langle \sigma v \rangle_{12} n_{eq}^2 = \langle \sigma v \rangle_{34} n_3^{eq} n_4^{eq} \quad (2.79)$$

and thus we have

$$\dot{n} + 3Hn = \langle \sigma v \rangle (n_{eq}^2 - n^2) \quad (2.80)$$

We now introduce a new quantity Y , defined as the dark matter number density n over the total entropy density s of the universe:

$$Y = \frac{n}{s} \quad (2.81)$$

Since entropy increases with time, we can expect Y to decrease with time so long as n either does not increase or increases less rapidly than s . By substitution and

rearranging (homework) into eq. (2.80):

$$\frac{dY}{dt} = \langle \sigma v \rangle s(Y_{eq}^2 - Y^2) \quad (2.82)$$

and

$$\frac{dY}{dx} = -\frac{x \langle \sigma v \rangle}{H(m)} (Y^2 - Y_{eq}^2) \quad (2.83)$$

where we have defined the rescaled time variable as $x = m/T$ where m is the mass of the dark matter. Thus we have $dx/dt = H(x)x$, since $T \propto 1/a$.

We have arrived at an expression for the evolution of Y as the universe cools. That's why we have rescaled time in terms of temperature, in turn in terms of the expansion rate of the universe. We went through entropy to do so, but for the purposes of this course you can count that as a convenient trick. The trick has the important effect of allowing us to measure the number density of dark matter without getting distracted by the effect of the expansion rate — that's why n alone is not good enough when we are considering freeze-out.

The final step is to spell out the thermally averaged cross section. We will not do that here, as that is a task for a statistical physics course. Suffice it to say that after suitable redefinitions and algebraic gymnastics, we get

$$\langle \sigma v \rangle = \frac{T}{8m^4 T K_2^2 m} \int_{4m^2}^{\infty} \sigma(\tilde{s} - 4m^2) \sqrt{\tilde{s}} K_1(\sqrt{\tilde{s}/T}) ds \xrightarrow{\text{non-rel}} b_0 + \frac{3}{2} b_1 x^{-1} + \dots \quad (2.84)$$

K_i are i^{th} order Bessel functions (appendix A.1) and $\tilde{s} = 2m^2 + 2E_1 E_2 - 2\mathbf{p}_1 \cdot \mathbf{p}_2$; we have expanded in the non-relativistic limit with coefficients $b_{0,1}$.

There are no analytic methods for solving this equation for the evolution of Y . But there are interesting limiting cases that give us a good feeling for how Y works. That means it's finally time to tie this all up — to return to our very first claim for this section, namely that freeze-out occurs when $\Gamma \lesssim H$.

Recall that when $\Gamma \gg H$, then the dark matter and photon bath interact with enough frequency to maintain equilibrium; conversely, when $\Gamma \ll H$, then the dark matter cannot interact without enough frequency to maintain equilibrium — it freezes out. Formally:

$$\begin{aligned} Y(x \lesssim x_f) &\simeq Y_{eq}(x) \\ Y(x \gtrsim x_f) &\simeq Y_{eq}(x_f) \propto \text{constant} \end{aligned} \quad (2.85)$$

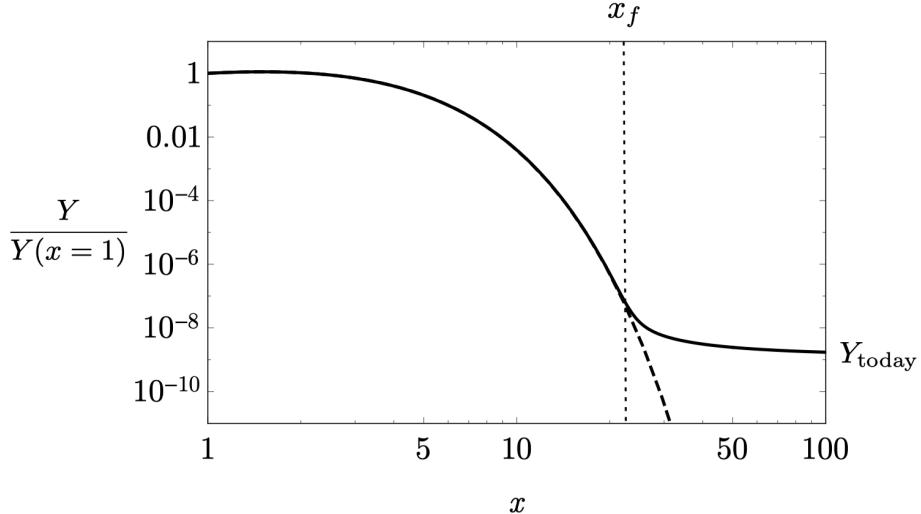


Figure 2.4: The time-evolution of Y , rescaled to screen off the expansion rate. For $x < x_f$ we see the solid and dashed lines as one, indicating that before freeze-out Y (solid) tracks Y_{eq} (dashed). Freeze-out occurs at x_f , after which Y is approximately constant over time. Figure courtesy of Lee et al. (2015), via Lisanti (2018).

Now suppose we approximate the thermal average cross-section with the non-relativistic expansion in (2.84), factor out dependence on x , and define a constant λ , such that we get (homework):

$$Y = -\frac{\lambda}{x^{n+2}} \text{ where} \\ \lambda = \frac{\langle \sigma v \rangle_0 s_0}{H(m)} \quad (2.86)$$

and in which $\langle \sigma v \rangle = \langle \sigma v \rangle_0 x^n$ and $s = s_0 x^{-3}$.¹⁰ If we suppose $n_0 = 0$, then we get (homework)

$$Y_{today} \simeq \frac{x_f}{\lambda} \quad (2.87)$$

on the assumption that the value of Y at freeze-out is greater than its value today. If the thermally averaged cross-section depends on x , the calculation and the results become trickier. Whether it does depends on the details of the particle physics model of the dark matter in question; Y_{today} will have higher powers of x_f if $n > 0$.

We know the fraction of today's critical density contributed by dark matter, given by the formula

¹⁰This assumes that one or both of the first terms in the expansion dominates, which is often approximately true.

$$\Omega_\chi = \frac{m s_{today} Y_{today}}{\rho_{cr}} \quad (2.88)$$

and thus we can approximate

$$\Omega_\chi h^2 = \left[\int_{x_f}^{\infty} \frac{\langle \sigma v \rangle}{x^2} dx \right]^{-1} \sim \frac{1 \times 10^{-26} \text{ cm s}^{-1}}{\langle \sigma v \rangle} \simeq 0.1 \left(\frac{0.01}{\alpha} \right)^2 \left(\frac{m}{100 \text{ GeV}} \right)^2 \quad (2.89)$$

on the assumption that $x_f = 10$ and $\langle \sigma v \rangle \sim (\alpha/m)^2$. If we assume weakly interacting dark matter, we get the correct abundance as measured by Planck and WMAP ([Collaboration et al. 2013](#)) for $\alpha \sim 0.01$ and $m_\chi \sim 100 \text{ GeV}$.

This is quite a striking coincidence, and has been made to sound even more striking by its nickname: the “WIMP miracle”; where WIMP stands for Weakly Interacting Massive Particles. Models with the above parameters are WIMP models of dark matter. We will discuss WIMPs in chapter 3, and how we are trying to detect them in chapter 5.

But in fact there is not really much of a coincidence. The constraint comes from $\langle \sigma v \rangle \sim (\alpha/m)^2$, and not from the specific choices of α or m_χ . So long as we keep $\alpha \ll 1$ and fix the ratio, there is significant wiggle-room for m_χ and α that allow for a correct estimate of today’s abundance. Models with these parameteres are called WIMPless dark matter models.

The key take-away from freeze-out is that the ratio $\langle \sigma v \rangle \sim (\alpha/m)^2$ must be satisfied for the correct abundance to fall out of the equations. There are countably infinite ways to do this; not all of which make much physical sense, not all of which are testable at present. But this ratio provides a powerful constraint on the models we construct, and can help to rule out otherwise plausible-sounding suggestions.

2.8 N-body Simulations

In order to simulate the distribution of dark matter, people generally use N-body simulations. If these simulations only contain dark matter and one assumes they are collisionless, this is simply a case of using Newton’s inverse law of gravitation to update the positions and forces on all particles.

There are around 10^{11} stars in a galaxy like the Milky Way, and if the mass of the dark matter is around 100 GeV for example then there are about 10^{67} particles. Updating Newton’s Law for every particle is a process which goes like $\text{cputime} \propto N^2$

so this would take a very long time. While it is possible to get this down to cputime $\propto N \ln N$, one still has to simulate a lot of mass with each particle, for example, the biggest simulations of the Milky Way have only around a billion particles in them, so each particle would represent a mass of 10^6 solar masses. That's a quarter of the mass of the central supermassive black hole.

If the dark matter starts without baryons then the smallest structures form first and the larger ones later. Typically in dark matter only simulations, the dark matter forms profiles with the functional form something like this

$$\rho(r) = \frac{\rho_s}{\left(\frac{r}{r_s}\right)\left(1 - \frac{r}{r_s}\right)^2} \quad (2.90)$$

which is called the Navarro-Frenk-White profile, normally known as the NFW profile. The parameter ρ_s is the density at the characteristic scale radius r_s where the density drops like r^{-2} . The density falls much less rapidly in the inner part of the halo compared to the outer part. Flat rotation curves, where $M(r) \sim \rho(r)r^3 \propto r$ must therefore occur close to the scale radius r_s . So we know that the scale radius must correspond to the very extent of the visible galaxy where the rotation curves are measured, showing that the dark matter halo must extend far beyond the baryonic part of the Galaxy.

You can define a virial radius, depending upon

$$\frac{3M(r_{vir})}{4\pi r_{vir}^3} = \rho_{av}(r < r_{vir}) = \Delta\rho_{crit} \quad (2.91)$$

where Δ is some overdensity that you have decided on, around 200 given the discussion above. It therefore marks the sphere within which the average density of the halo is bigger than the critical density by the factor Δ . You can also define a virial mass which is simply $M_{vir} = M(r_{vir})$.

The concentration parameter $c = r_{vir}/r_s$ is bigger for smaller haloes, reflecting the fact that the inner parts of halos retain some memory of the density of the Universe at the point they left the Hubble flow and collapsed. For more details, see [Lisanti \(2018\)](#).

2.9 Nucleosynthesis

Big Bang Nucleosynthesis is the process by which light elements were created in the early universe. It began at around 10s after the Big Bang, and ended around 20 minutes later.

The relevance of nucleosynthesis for our consideration of dark matter is not immediately obvious - there is no mention of dark matter in the theory of nucleosynthesis. However, the theory works extremely well. We deduce from this that dark matter particles must exist at a much lower temperature than the rest of the thermal bath, and thus would not be interacting, nor would be created in the same manner as the particles we are about to discuss¹¹.

A number of different parameters go into the calculations of nucleosynthesis, and a fully comprehensive calculation is too involved to cover here. However, we will sketch out the basic theory.

2.9.1 Neutrons and Protons

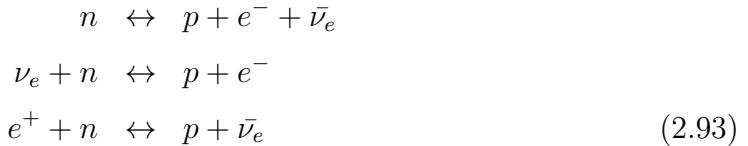
Even before nucleosynthesis began, other processes were in action which lead to important constraints.

A useful equation for the following sections will be the approximate relation between time and temperature:

$$\frac{t}{\text{1 second}} \approx \frac{2.4}{\sqrt{g_*}} \left(\frac{1\text{MeV}}{k_B T} \right)^2 \quad (2.92)$$

where $g_* \approx 3.4$ is the number of relativistic species at this time.

At very early times, $t \ll 1\text{s}$, in the hottest temperatures, $k_B T \gg 1\text{MeV}$, the reactions



occur constantly, without suppression. Since the mass of the electron is $\approx 0.5\text{ MeV}$, the universe is full of relativistic electron-positron pairs; photons and neutrinos are also relativistic, and the protons and neutrons of the universe are non-relativistic. These reactions maintain equilibrium through the weak nuclear force.

However, when the temperature begins to fall, these reactions cannot continue to proceed in both directions; the mass difference between protons and neutrons is

$$\Delta m = m_n - m_p = 939.6 - 938.3 = 1.29\text{MeV} \quad (2.94)$$

¹¹We will discuss later one possible process for the creation of dark matter particles, WIMPs in particular.

The Boltzmann equation for the ratio of neutrons to protons is given by

$$\frac{n_n}{n_p} = \left(\frac{m_n}{m_p} \right)^{\frac{3}{2}} e^{-\frac{\Delta m}{T}} \approx e^{-\frac{1.3 \text{ MeV}}{T}} \quad (2.95)$$

So at temperatures above $k_B T \gg 1.3$ MeV, there are equal numbers of protons and neutrons; as the temperature cools, the number of neutrons will fall exponentially, until they reach freeze out temperature.

The freeze-out temperature for neutrons is $k_B T(\Gamma \approx H) \approx 0.8$ MeV. Using eq. (2.92) this dates neutron decoupling to $t \approx 2$ seconds after the Big Bang. At this point, the neutron to proton ratio is

$$e^{-\frac{1.3}{0.8}} \approx \frac{1}{5} \quad (2.96)$$

i.e 5 protons to every neutron. Here, all weak force reactions cease, except the beta decay of neutrons to protons: neutron decay has a half life of around 880 seconds, so the number density of neutrons will decay as

$$n_n(t) \approx \frac{1}{5} n_p(t \approx 2 \text{ seconds}) e^{-\frac{t}{880 \text{ seconds}}} \quad (2.97)$$

2.9.2 Deuterium, Helium, and Heavier Nuclei

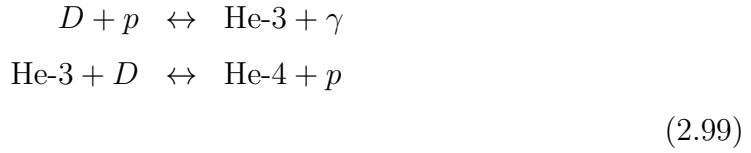
The first element which can be created after hydrogen is deuterium, D ; this forms through the reaction



The binding energy of deuterium is therefore $E_{\text{bind}} = m_n + m_p - m_D \approx 2.2$ MeV; deuterium forms when $k_B T \approx 0.06$ MeV (you will show this in a homework set), which is at approximately $t \approx 360$ seconds.

Heavier nuclei have higher binding energies; there is not sufficient time for heavier nuclei to be created from three or more separate nucleons, so heavier nuclei are formed from deuterium. That these heavier nuclei must wait to be created until deuterium has been formed is referred to as the ‘Deuterium Bottleneck’.

Once deuterium has been formed, helium forms almost instantly via the reactions



Equation 2.97 tells us that the ratio of neutrons to protons at this time is

$$\frac{n_n}{n_p} = \frac{1}{5} e^{-\frac{360}{880}} \approx 0.13 \tag{2.100}$$

Since each He-4 nucleus contains 2 protons and 2 neutrons, the ratio of helium to hydrogen is

$$\frac{n_{\text{He}}}{n_H} = \frac{n_n/2}{n_p - n_n} \approx 0.07 \tag{2.101}$$

(ie the maximum number of helium which could have been created from the available neutrons, over the number of remaining protons)

Since the mass of a helium atom is roughly 4x that of a hydrogen atom, roughly 25% of the combined mass of the two is helium, and 75% hydrogen. This is close to the observed ratio.

Indeed, more comprehensive calculations and consideration of the creation of lithium and beryllium are all in very good agreement with observed abundances.

2.9.3 Cosmological Parameters

One important parameter which is involved in these calculations is g_* . If there were more relativistic species, the Deuterium Bottleneck is passed much earlier on; this would generate a higher proportion of helium and heavier elements.

Since we do not observe these, we conclude that any dark matter particles must be non-relativistic. This places constraints on how early dark matter particles must have fallen out of thermal equilibrium.

Chapter 3

Candidates for Dark Matter

3.1 WIMPS and Thermal Relics

Weakly Interacting Massive Particles (WIMPS) are a genre of theorised candidate particles for cold dark matter.

WIMPS interact primarily through the weak and gravitational forces, and their interactions through other forces cannot be stronger than the weak force; this means that they do not significantly interact electromagnetically, hence ‘dark’ matter. They are significantly heavier than standard model particles; this would lead them to have low velocities, hence ‘cold’ dark matter. WIMPs are labelled ‘Thermal Relics’ since they would have been produced via thermal particle production in the early ages of the universe.

Thermal Particle Production can create particles which have very low interaction rates. In general, a hot thermal bath (a system in equilibrium) can create particles which are *not* in equilibrium with it through decay or scattering of the equilibrium-thermal-bath particles. The created particle, since it essentially does not interact with the thermal bath, can then escape from the bath with negligible effect on the bath itself. One example of this is the creation of photons in the early universe quark-gluon plasma; another might be the production of WIMPs shortly after reheating.

By considering the particle production rate, and using the Boltzmann distribution, one can find an equation for the particle number density over time. Since we have estimates for the current number density of WIMPs, if they do constitute dark matter, we can infer various other features of them. For a detailed discussion of Thermal Particle Production, see <https://arxiv.org/pdf/1903.03026.pdf>.

3.1.1 Cross-Sections

Massless Bosons

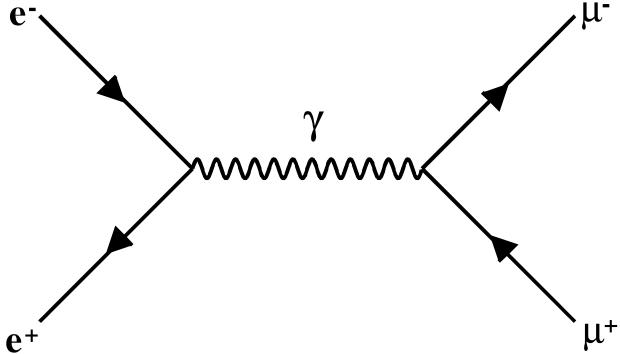


Figure 3.1: Electron-positron annihilation to form a muon and an anti-muon via a virtual photon.

If we look at fig. 3.1 we can see that we have two incoming electrons - one electron and one positron, annihilating to form two muons. The first thing of note is that, since the mass of an electron is less than the mass of a muon, this can only take place if they have some high initial momentum. The probability of this process occurring depends upon the coupling between photons and electrons, which is given by $\sqrt{\alpha_{em}} = e/\sqrt{4\pi}$, where e is the charge of the electron. It also depends upon the coupling between the photon and muons, which is also given by $\sqrt{\alpha_{em}}$.

There is another thing we need to worry about though: in the centre of momentum frame, once the electrons annihilate, the total energy of the four vector of the system will have the form $p^\mu = (\sqrt{s}, 0, 0, 0)$. Here \sqrt{s} is the centre of mass energy $s = 2|\mathbf{p}|^2 + 2m_e^2$ where \mathbf{p} is the initial momentum of the electrons, which will be equal and opposite by definition in the centre of momentum frame.

The energy momentum at the moment that the electrons annihilate, $p^\mu = (\sqrt{s}, 0, 0, 0)$, is not a good momentum to be carried by photons. It is the momentum of a stationary particle with mass \sqrt{s} : stationary such that it is on a time-like trajectory through space-time. Photons have no mass which means they cannot stay still, you cannot get into a Lorentz frame where their trajectory is time-like. Because of this, we know the photon γ will be a *virtual* particle, it will have to “borrow” momentum-energy from the vacuum and then give it back within a short time $\Delta t \sim \hbar/\Delta E$. This becomes more difficult as \sqrt{s} increases. The overall difficulty needs to be reflected in the probability of the process occurring, so we have a suppression factor called the *propagator* (since it is associated with the probability of the photon

propagating from the left to the right).

The propagator for bosons has the form $P(s) = 1/(s - m_{bos}^2)$ where the mass of the boson m_{bos} is zero for a massless photon. We then, going from left to right, obtain the amplitude $\sqrt{\alpha} \times (1/s) \times \sqrt{\alpha}$ - ie a cross section which is proportional to the amplitude squared

$$\sigma = const \times \phi \frac{\alpha^2}{s^2} \quad (3.1)$$

where ϕ is some factor related to the phase space available for the process to take place. If we were doing the calculation properly we would obtain the full expression for the phase space, but we can often do quite well just by looking at the dimension and knowing that we expect the cross section have dimension [length]².¹

So we end up with

$$\sigma_{e^+ e^- \rightarrow \mu^+ \mu^-} \sim \frac{\alpha^2}{s} \quad (3.2)$$

The correct answer is $4\pi/3 \times \alpha^2/(s)$, so our estimate is pretty good.

Massive Bosons

If we have a massive propagator then we need to include that too, as it will change when the boson is happy to carry away the energy momentum of the collision, and when it isn't. So for example for the diagram in fig. 3.2 we have a cross-section

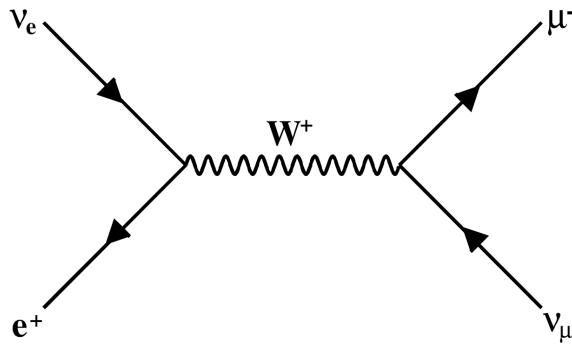


Figure 3.2: Positron-neutrino annihilation to form an anti-muon and via a W^+ boson which can be virtual or not depending upon \sqrt{s} .

$$\sigma_{e^+ \nu_e \rightarrow \mu^+ \nu_\mu} \sim \frac{\alpha_w^2 s}{(s - M_W^2)^2} \quad (3.3)$$

which rises at low energies, becomes very large when $\sqrt{s} \sim M_W$ and then dies down again at high energies. Note here we have a slightly different value of α as it is

¹These kind of naive estimations can go badly wrong if helicities of the particles create problems. However this is not what is happening here.

related to α_{em} via the weak mixing angle, but it is roughly the same. At $\sqrt{s} = M_W$, it looks like the cross-section becomes singular, but actually what happens is that there is no virtuality about the particle at all; it is created ‘on shell’ - it has precisely the energy and momentum that it needs to have given its mass M_W and the centre of mass \sqrt{s} such that it exists happily as a classical particle. What then determines what happens is how long it takes for the particle to decay. If you think about the classical wave equation for a massive particle, adding an imaginary term to the mass will make the amplitude decrease exponentially over time, which when we quantize looks like a decay. In fact Γ is the decay rate of the particle.

So we add a width to the propagator $M_W \rightarrow M_W - i\Gamma_W/2$, and then

$$(s - M_W^2) \rightarrow [s - (M_W - i\Gamma_W/2)^2] = s - M_W^2 + iM_W\Gamma_W + \Gamma_W^2/4 \sim s - M_W^2 + iM_W\Gamma_W \quad (3.4)$$

where we have dropped the $(1/4)\Gamma_W^2$ term since we assume $\Gamma \ll M_Z$ in order to have a well defined resonance. We can then write

$$\sigma_{e^+\nu_e \rightarrow \mu^+\nu_\mu} \sim \left| \frac{\alpha_w \sqrt{s}}{(s - M_W^2 + i\Gamma_W M_W)} \right|^2 = \frac{\alpha_w^2 s}{(s - M_W^2)^2 + M_W^2 \Gamma^2} \quad (3.5)$$

where Γ_W is the decay rate of the W-boson into other particles. The width Γ is related to the uncertainty in the mass of unstable particles such as the W-boson - because it is shortlived, there is uncertainty in its energy and this makes it easier to produce from a variety of \sqrt{s} around its actual mass.

Lets imagine that dark matter is a heavy neutrino, then the annihilation rate would be given by

$$\sigma_{\nu\bar{\nu} \rightarrow Z \rightarrow ALL} \sim \left| \frac{\alpha \sqrt{s}}{(s - M_Z^2 + i\Gamma_Z M_Z)} \right|^2 = \frac{\alpha_w^2 4 m_{\nu DM}^2}{M_Z^4} \sim 10^{-37} \text{cm}^2 \left(\frac{m_{\nu DM}}{1 \text{GeV}} \right)^2 \quad (3.6)$$

So we can see that $m_{\nu DM} \sim 3 \text{ GeV}$ gives us about the right number density of relic neutrinos. However, there is a problem here since the Z-boson also decays into pairs of quarks and leptons.

$$\Gamma_{Z total} = \Gamma_{Z \rightarrow e^+e^-} + \Gamma_{Z \rightarrow \mu^+\mu^-} + \Gamma_{Z \rightarrow u\bar{u}} + \dots \quad (3.7)$$

so the total lifetime is much shorter depending upon how many species Z can decay into. Z particles are produced at the LHC but because the LHC is smashing protons together things are a bit messy. The Z particle has also been produced at LEP and

its cross section measured carefully as a function of energy. Because we know that

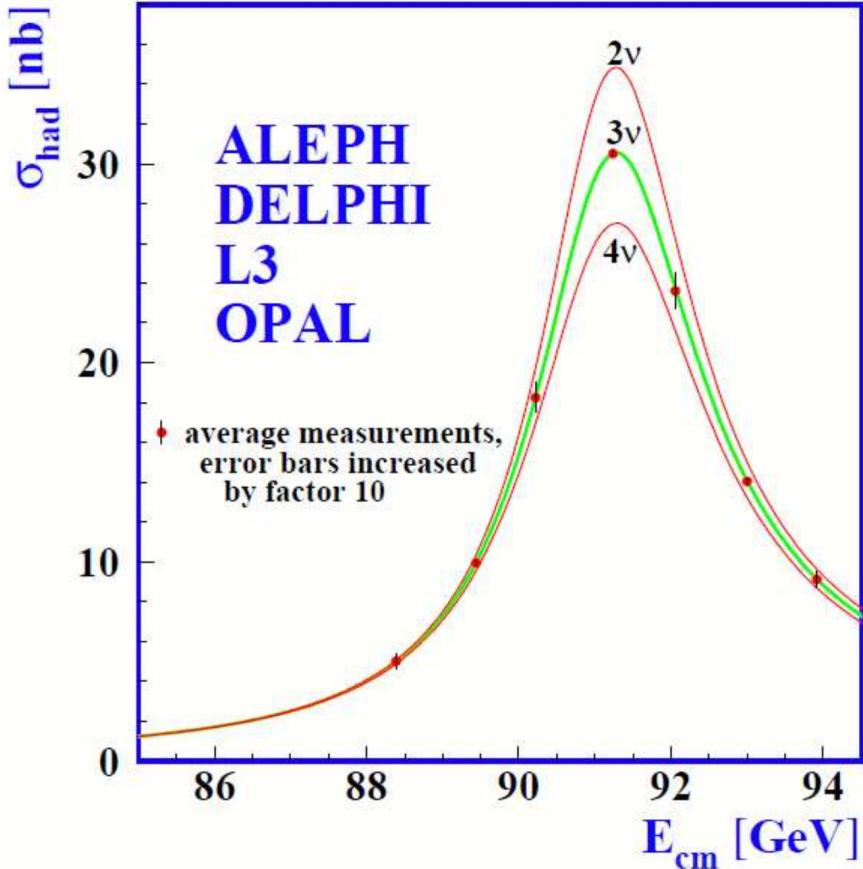


Figure 3.3: The cross section for the production of the Z-boson was measured in great detail at LEP. It does not allow for a 4th generation of neutrinos which are kinematically accessible when the Z decays.

the width of the resonance for Z-boson production is set by the decay channels of the Z-boson, we can tell that there cannot be another neutrino with a mass less than half the Z-boson mass, i.e. less than 45 GeV. So it does not appear that heavy neutrinos could be our thermal relics.

3.1.2 A model of dark matter

We would like to consider a thermal relic particle which does give the correct relic abundance without messing up the precision electroweak physics observed at LEP, Tevatron, LHC and other colliders. We have seen that a heavy neutrino does not fit the bill. Let us consider another process in fig. 3.4.

Here we introduce another dark matter particle χ which couples to the standard model particles via a new Z-boson which we call Z' . Let us assume that it only

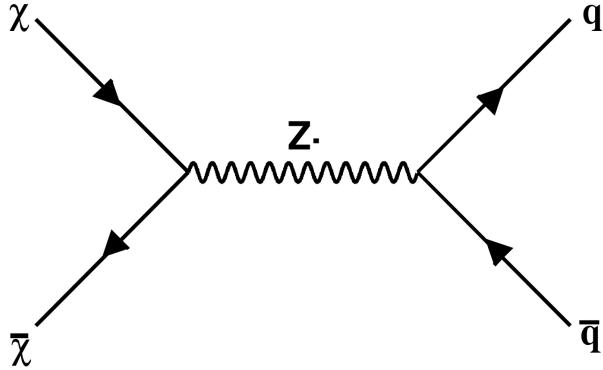


Figure 3.4: The coupling of dark matter χ with a new hypothetical Z -boson Z' which also couples to quarks .

couples to quarks. The annihilation cross section is then given by

$$\sigma_{\chi\bar{\chi} \rightarrow Z' \rightarrow q\bar{q}} \sim \frac{4g_q^2 g_\chi^2 s}{3\pi (s - M_{Z'}^2)^2 + M_{Z'}^2 \Gamma_{Z'}^2} \quad (3.8)$$

and then by setting the correct values of g_q , g_χ and $M_{Z'}$ we will be able to get the correct relic abundance.

3.1.3 Self-Annihilating Dark Matter

Straight away we can see that if dark matter in N-body simulations concentrates inside the centre of galactic halos then this is the place where we expect to see the most annihilation. The equation for the number of annihilations will be given by

$$\frac{dn}{dt} = n^2 \langle \sigma_A |\vec{v}| \rangle \quad (3.9)$$

which will give us the annihilation in units of $s^{-1}m^{-3}$. More generally, if we assume the inner region of some dark matter halo from $r = 0$ to $r = r_{max}$ then we expect the annihilation to be

$$\frac{dE}{dt} = 2m_{dm} \int_{r=0}^{r=r_{max}} \left(\frac{\rho}{m_{dm}} \right)^2 \langle \sigma_A |\vec{v}| \rangle 4\pi r^2 dr \quad (3.10)$$

and the precise channel that this manifests itself in depends upon the nature of the dark matter in question. This could be (very rarely) directly into gamma rays. More often it is into quarks which then hadronise producing gamma rays and anti-protons.

There is currently a signal apparently coming from the galactic centre observed

by the Fermi gamma rays telescope which could be consistent with gamma rays from dark matter annihilation. However it could also be consistent with unresolved astrophysical sources.

3.2 Alternatives to Thermal Relics

3.2.1 Axions

Axions (A^0) are proposed chargeless and spinless elementary particles of a low mass (between 10^{-5} and 10^{-3} eV). They are postulated primarily as a solution to the strong CP problem, but their supposed small cross-sections for strong and weak force interactions make them a candidate for dark matter. As we will see later, there are a number of different theorised axions, which have different properties.

The strong CP problem relates to why QCD seemingly conserves Charge-Parity symmetry; that is, if all particles in a reaction are switched for their anti-particles (which carry opposite charge), and the parity is switched (considering the reaction from right to left), nothing would change. Experiments have given every suggestion that this is the case.

The strong CP problem then arises, because there are terms in the QCD-Langrangian which could break the CP-symmetry: there is a fundamental parameter $\bar{\theta}^2$ which could take any value between 0 and 2π . There is no a priori suggestion of its value, hence it must be determined experimentally.

The best test for the value of $\bar{\theta}$ comes from measuring the electric dipole moment (EDM) of the neutron. It should be of the order

$$|d_n| \sim 10^{-1} |\bar{\theta}| e \text{ m}, \quad (3.11)$$

The experimental upper bound for this measurement is roughly 10^{-36} ; essentially 0. This means that $|\bar{\theta}| \approx 0$. There is no theorised reason for its value being so close to zero.

One of the ways people solve this fine-tuning problem is by promoting the dimensionless parameter $\bar{\theta}$ to a pseudo-scalar field a which we call the axion. We cannot directly replace $\bar{\theta}$ with a directly for dimensional reasons, but we can replace

²The QCD Lagrangian is $\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{n_f g^2 \theta}{32\pi^2} F_{\mu\nu}\bar{F}^{\mu\nu} + \bar{\psi}(i\gamma^\mu D_\mu - me^{i\theta' \gamma_5})\psi$. θ' is the chiral quark mass phase, and can be changed to make the contribution to the total effective angle $\bar{\theta}$.

$\bar{\theta}$ with a/f_a , where f_a is a mass scale.

Furthermore, when a increases from zero, it gives rise to complicated vacuum effects which give it an effective potential of the form

$$V = V_0 \left(1 - \cos\left(\frac{a}{f_a}\right) \right) \quad (3.12)$$

As always, the field tends towards the minimum of its potential, which occurs when $a = 0$. So by promoting θ to a field a you gain a dynamic solution to explain its value.

A variety of different axions fields have been theorised. In general, axions which solve the strong CP problem have a relationship between $m_a f_a \sim m_\pi f_\pi$ where m_π is the pion mass and f_π is the pion decay constant. Both of these parameters have masses around 100 MeV.³

We must do some background on scalar fields before we can assess the axion's candidacy for dark matter.

The equation of motion for a scalar field is given by

$$\frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)} \right) = 0 \quad (3.13)$$

from which we get

$$m^2 \phi - \partial_\mu \partial^\mu \phi = 0 \quad (3.14)$$

for flat space $\nu^{\mu\nu} = -1, 1, 1, 1$. Or, more generally

$$\frac{dV(\phi)}{d\phi} - \partial_\mu \partial^\mu \phi = 0 \quad (3.15)$$

for when $V(\phi) \neq (1/2)m^2\phi^2$. If we then assume that there are no spatial gradients for the field, in other words that it is homogeneous throughout space but can oscillate over time then we get

$$\mu^{\mu\nu} \partial_\mu \partial_\nu \phi = \mu^{00} \partial_0 \partial_0 \phi = -\frac{\partial^2 \phi}{\partial t^2} = m^2 \phi \quad (3.16)$$

³In theories such as string theory there are sometimes a great number of axions which come from the way that the extra dimensions are compactified. For these axions, the same relationship between m_a and m_π does not exist, and one can treat them to some degree as free parameters (depending upon which string theorist you talk to). In this situation, or any situation where one is agnostic about the precise value of the couplings, these particles are called "axion-like particles" or ALPs.

which has a very simple solution

$$\phi(t) = A \sin(mt) \quad (3.17)$$

however the pressure P and density ρ of the scalar field are given by

$$\begin{aligned} \rho &= \frac{1}{2}\dot{\phi}^2 + V(\phi) \\ P &= \frac{1}{2}\dot{\phi}^2 - V(\phi) \end{aligned} \quad (3.18)$$

Then the equation of state

$$w = \frac{P}{\rho} = \frac{\dot{\phi}^2 - 2V}{\dot{\phi}^2 + 2V} \quad (3.19)$$

will also vary over time. For the potential $V(\phi) = (1/2)m^2\phi^2$ we would like to find out the average equation of state as the field oscillates, in other words $\langle w \rangle$.

$$\begin{aligned} w &= \frac{\dot{\phi}^2 - 2V}{\dot{\phi}^2 + 2V} = \frac{A^2m^2 \cos^2(mt) - A^2m^2 \sin^2(mt)}{A^2m^2 \cos^2(mt) + A^2m^2 \sin^2(mt)} \\ &= \frac{\cos^2(mt) - \sin^2(mt)}{\cos^2(mt) + \sin^2(mt)} = \cos^2(mt) - \sin^2(mt) \end{aligned} \quad (3.20)$$

and the time average of this quantity

$$\langle w \rangle = \frac{\int_0^{t=2\pi/m} \{\cos^2(mt') - \sin^2(mt')\} dt'}{\int_0^{t=2\pi/m} dt''} = \left[\frac{m \sin(2mt)}{4\pi} \right]_0^{t=2\pi/m} = 0 \quad (3.21)$$

So we can see that this oscillating scalar field has the same equation of state as matter.

Since this field behaves like matter, it can in fact play the role of dark matter, although strictly speaking it is more like a homogeneous bose-Einstein condensate throughout space. It does cluster like dark matter though. The differences are not obvious and even experienced physicists argue over interpretations.

Since we are in an expanding space-time and we understand General relativity, we should use covariant derivatives

$$\rightarrow g^{\mu\nu} \frac{\nabla}{\partial x^\mu} \frac{\partial}{\partial x^\nu} \phi - \frac{dV}{d\phi} = \frac{\partial^2 \phi}{\partial x^\mu \partial x^\nu} - \Gamma_{\mu\nu}^\alpha \frac{\partial \phi}{\partial x^\alpha} - \frac{dV}{d\phi} \quad (3.22)$$

And assuming spatial isotropy ($\partial\phi/\partial x^i = 0$ where $i = 1, 2, 3$) this gives us a slightly

different euqation of motion

$$\ddot{\phi} + 3H\dot{\phi} + \frac{dV}{d\phi} = 0 \quad (3.23)$$

which takes into account the redshifting of the energy. This is exactly like a damped harmonic oscillator, but the damping comes from the expansion of the Universe. So as the universe expands, the energy density redshifts and the amplitude decreases ($A \rightarrow A(t)$).

The Axion as Dark Matter

If we assume that the axion is a scalar field, then

$$\ddot{a} + 3H\dot{a} + \frac{dV}{da} = 0 \quad (3.24)$$

It is easy to see that as long as H is large, the field will not be able to move in its potential - any positive \dot{a} will lead to a negative \ddot{a} , preventing any acceleration of the field away from its value. In fact only when $H \ll m$ (where m^2 is the second derivative of the potential with respect to a) will the field be able to move. This creates a damped harmonic oscillator; whether it is critically damped or not depends upon $H^2 - 4m^2$.

Recall that

$$\cos \theta = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots \quad (3.25)$$

so that close to the origin, the potential for the axion $\propto (1 - \cos \theta)$ is exactly like a quadratic mass potential.

So if the axion is the substance of dark matter, it must obtain some value very early on, perhaps during inflation, but only begins to fluctuate once the density of the Universe drops to the point where $H \sim m$. At this point its value will start to move, oscillating as a damped harmonic oscillator and forming dark matter.

Indeed this fits with the unsatisfactory nature of older axion theories, specifically ‘invisible axions’; these axions were lighter than $10^{-11} e$ (the critical mass) and would have been abundantly produced in the early universe. This would lead to ‘hot’ axions as they interacted with the high energy particles, and so would have been observed by now. That newer theories of axions prevent an abundance of axions in the early universe makes a more compelling argument for them.

3.2.2 Primordial Black Holes

Primordial black holes are a form of massive compact halo objects (MACHOs⁴⁵) and are thus a candidate for dark matter.⁶ They are a type of black holes, theoretically created in the early Universe (< 1s after the Big Bang), under the same conditions as those required for black hole formation now.

Primordial black holes are created in the early Universe where a region has an overdensity which collapses into a black hole *before* the Universe becomes matter dominated. Typically this happens within the first 1s of the universe. Unlike modern black holes, which are produced from stellar collapse and hence must have a mass of a considerable size - the smallest black holes are around 10 solar masses, primordial black holes can be as light as 10^{-8} kg.⁷

The criterion for a spherical region of radius R to be a black hole to exist is:

$$R = \frac{2GM}{c^2} = \frac{2G\rho R^3}{c^2} \quad (3.26)$$

Hence regions such that $G\rho R^2 \sim 1$ ⁸ give rise to collapsing regions which will lead to black holes. This requires over-densities of $\frac{\delta\rho}{\rho} \sim 0.1$, which are not very regularly seen in the modern universe. However several theories can produce such over-densities in abundance - certain inflationary models and cosmological phase transitions, for example.

3.2.3 Sterile neutrinos

In the standard model all neutrinos have left-handed helicities⁹, and all anti-neutrinos right-handed.

⁴Named by Griest, intended as a humorous alternatives to WIMPs.

⁵Other MACHOs include neutron stars and brown dwarfs. The categorization of primordial black holes as a MACHO is contentious, as they are non-baryonic, while all other MACHOs are baryonic.

⁶It should be noted though that even if MACHOs account for an amount of dark matter, constraints from the CMB set the ratio of baryonic matter to total matter, and as we will see later, there is a large amount of non-baryonic matter in the universe. As the previous footnote notes, though, this does not affect primordial black holes. However, various constraints from other cosmological observations deny the possibility that primordial black holes could constitute the majority of dark matter.

⁷It should be noted though that primordial black holes of that size would no longer exist, due to Hawking radiation. The smallest primordial black holes which would still be present today would be 10^{11} kg.

⁸In N.U.

⁹A particle has a right-handed helicity if its spin is parallel to its direction of motion, and a left-handed helicity if it is antiparallel.

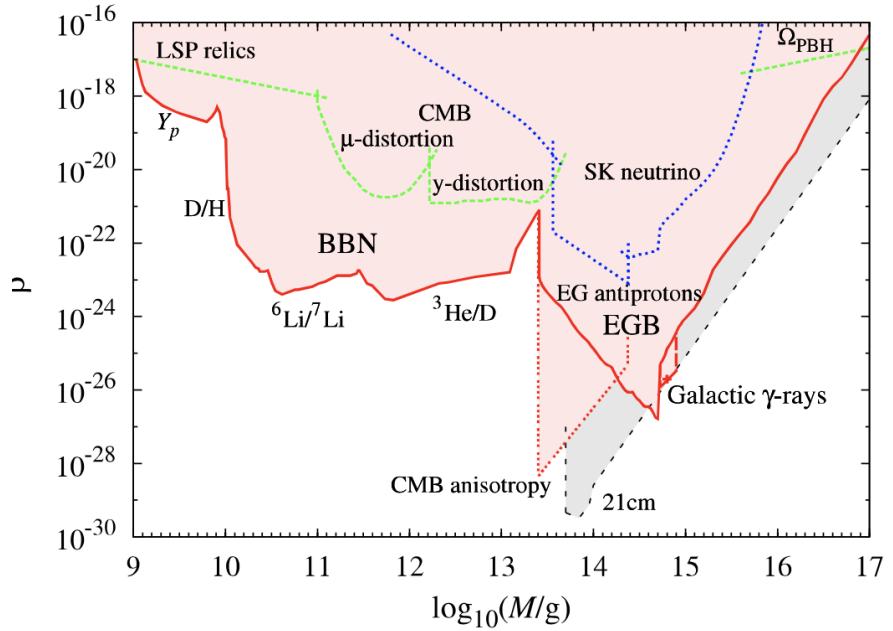


Figure 3.5: Constraints on different masses of primordial black holes due to different constraints

Since we now know that neutrinos are not massless, there is a distinction between their chirality and their helicity; since they have mass, they do not travel at the speed of light, and it is possible for an observer to travel faster than them to overtake them, and change reference frame. The neutrinos would then appear to be spinning in reverse, and its helicity would be reversed. Their chirality, on the other hand, is relativistically invariant. Thus, we would expect to observe right-handed neutrinos, and left-handed antineutrinos.

However, we have only ever observed neutrinos with a left-handed *chirality*, and anti-neutrinos with a right-handed *chirality*; these are referred to as ‘active neutrinos’. Despite this, the standard model could permit the existence of right-handed neutrinos, and left-handed anti-neutrinos, referred to as ‘sterile neutrinos’.

Sterile neutrinos, so named because they do not interact with the rest of the standard model at low energy, interact only via the gravitational force; hence, if they are massive enough and there are enough of them, they could constitute dark matter. Their theorised mass is anywhere between < 1 eV and 10^{15} GeV. However, to be a candidate for dark matter this region shrinks to between 100 eV to 10 keV.

Chapter 4

The CMB Constraints

The observation of the Cosmic Microwave Background (CMB) was the conclusive evidence for an expanding universe. Our understanding of dark matter is essential in constructing a theory to explain it.

In particular, the CMB is evidence for the Big Bang. Under the big bang theory, the universe began in a hot, compressed, dense state, and would thus leave observable thermal radiation. This radiation would form a blackbody spectrum, and be uniform in every direction. Initially we would expect this radiation to have a very high temperature, but by the time of our observation, it should be redshifted due to expansion of the universe. The photons from this era form the ‘Surface of Last Scattering’ (see section 4.2.1 below), a surface beyond which we cannot ever see.

Our measurements of the CMB fit these criteria; data from COBE shows that the present-day CMB almost perfectly fits a blackbody spectrum at a temperature of $T = 2.725\text{K} \pm 0.002\text{K}$. It is not completely isotropic, but this fits with our understanding of Baryonic Acoustic Oscillations (BAO) (see section 4.3.3); it has a primary anisotropy at a scale of 1/100,000 and subsequent smaller anisotropies at higher scales. These appear to us as temperature fluctuations.

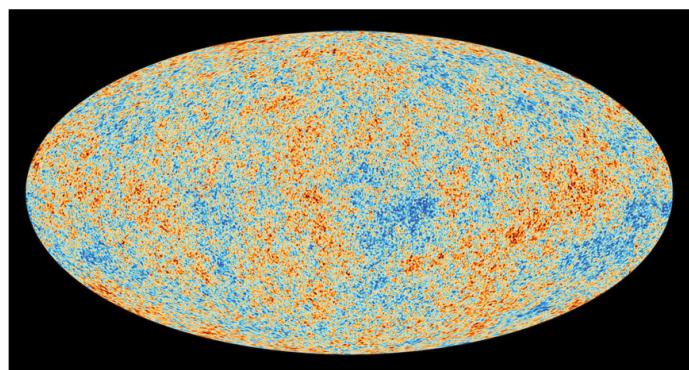


Figure 4.1: The temperature map of the CMB generated by the Planck spacecraft.

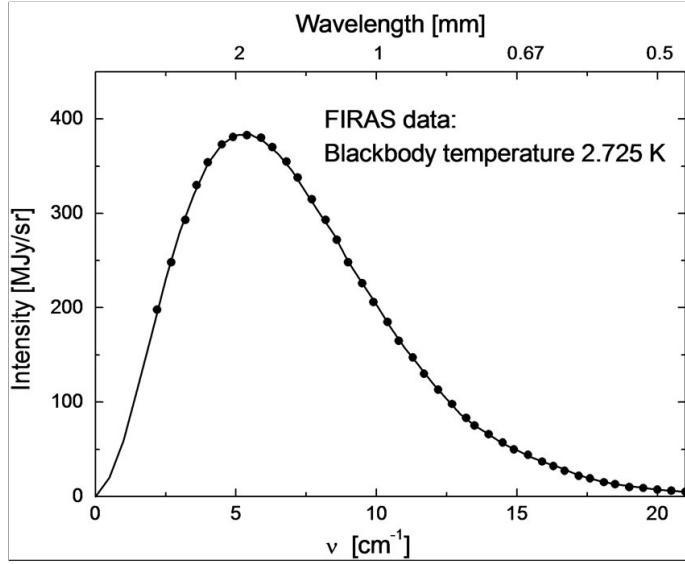


Figure 4.2: A graph of the spectral radiance of the CMB; the crosses are data points from COBE. The label ‘intensity’ is an old term for (spectral) radiance. The unit MJy is the mega-janksy, where $1 \text{ Jy} = 10^{-26} \text{ W m}^{-2} \text{ Hz}^{-1}$. The line on the graph is the theoretical blackbody spectrum for a temperature of 2.725K.

These temperature fluctuations are linked to fluctuations in baryon density in local regions of the photon-baryon plasma 380,000 years after the big bang; by considering the temperatures of the CMB at various points in the sky now, we can learn about the state of the early photon-baryon plasma.

These early density fluctuations are the seeds of the grand density fluctuations we observe now in both matter and dark matter (stars, solar systems, galaxies amongst the vacuum of interstellar space); we will look at this briefly in section 4.6.

4.1 The Angular Power Spectrum

Since we observe the LSS as a spherical shell surrounding the Earth, it is convenient to record data in spherical polar co-ordinates. We can then construct an angular power spectrum with which to analyse the temperature fluctuations.

We define the temperature fluctuation at a point as

$$\Theta(\hat{\mathbf{n}}) = \frac{T(\hat{\mathbf{n}}) - \langle T \rangle}{\langle T \rangle} \quad (4.1)$$

where $\hat{\mathbf{n}} \equiv (\theta, \phi)$ is an angle in the sky, $T(\hat{\mathbf{n}})$ is the temperature measured at that angle, and $\langle T \rangle$ is the average temperature across the sky ($\langle T \rangle = 2.725 \text{ K}$).

4.1.1 Spherical Harmonics

We use spherical harmonics to represent the data points.

Spherical harmonics are defined as

$$Y_{l,m} \equiv \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} P_l^m(\cos(\theta)) e^{im\phi} \quad (4.2)$$

where $l = 0, 1, \dots, \infty$ is the multipole moment, m is an integer such that $-l \leq m \leq l$, and P_l^m are Legendre polynomials.

The spherical harmonics form an orthonormal and complete set over the unit sphere.

Analogously to the Fourier series, one can compile a spherical harmonic series which represents some function over the sphere; this is used to break down the temperature fluctuation function over the SLS. It is given by

$$f(\hat{\mathbf{n}}) = \sum_{l=0}^{\infty} \sum_{m=-l}^l a_{l,m} Y_{l,m}(\hat{\mathbf{n}}) \quad (4.3)$$

where $a_{l,m}$ gives the amplitude of a given spherical harmonic.

Thus our temperature fluctuations can be expanded using

$$\Theta(\hat{\mathbf{n}}) = \sum_{l=0}^{\infty} \sum_{m=-l}^l a_{l,m} Y_{l,m}(\hat{\mathbf{n}}) \quad (4.4)$$

Due to the mutual orthogonality of the spherical harmonics, the amplitudes $a_{l,m}$ are given by

$$a_{l,m} = \int_0^{2\pi} \int_0^\pi Y_{l,m}^*(\hat{\mathbf{n}}) \Theta(\hat{\mathbf{n}}) d\Omega \quad (4.5)$$

where $d\Omega = \sin(\theta)d\theta d\phi$.

Power Spectrum

Again, analogously to a Fourier series, we can create a power spectrum¹. We define the power spectrum coefficients C_l , as the variance of the spherical harmonics'

¹This formulation is called a ‘power spectrum’, since classically the power of a wave is proportional to the square of its amplitude.

amplitudes $a_{l,m}$ for a fixed l .

$$C_l = \langle |a_{l,m}|^2 \rangle = \frac{1}{2l+1} \sum_{m=-l}^l a_{l',m'}^* a_{l,m} \quad (4.6)$$

A simple dimensional analysis gives us that $a_{l,m}$ is units of temperature (since spherical harmonics are dimensionless), and C_l has units of [temperature]².

In order to evaluate the C_l , we will need a different formula; our method of observation will involve comparing Θ at two different angles.

Substituting the equation for $a_{l,m}$ into the one for C_l given above gives us:

$$C_l = \frac{1}{2l+1} \int_{\Omega} \int_{\Omega'} \left(\sum_{m=-l}^l Y_{l,m}(\hat{\mathbf{n}}') Y_{l,m}^*(\hat{\mathbf{n}}) \right) T(\hat{\mathbf{n}}') T(\hat{\mathbf{n}}) d\Omega' d\Omega \quad (4.7)$$

From the above equations for C_l , it can be shown that C_l is a *correlation function*². That is, if $\hat{\mathbf{n}}'$ and $\hat{\mathbf{n}}$ are peaks in the CMB temperature function, and if $Y_{l,m}$ (or alternatively, if P_l) also has peaks at $\hat{\mathbf{n}}'$ and $\hat{\mathbf{n}}$, then C_l will have a larger value. More simply: the more points with temperature peaks at the l -harmonic angular separation, the greater C_l .

Importantly, the power spectrum will show peaks at l values where CMB temperature peaks are spaced at angles of $\theta = \frac{\pi}{l}$ apart.

4.2 The Theory

4.2.1 The Surface of Last Scattering

The Surface of Last Scattering (SLS) plays an important role in understanding the CMB; it will be useful to spend some time here to properly explain it.

There are two important aspects of the SLS to visualise. The first relates to why we describe the SLS as being as far back into the universe as we can see: recall that the photon-baryon plasma is opaque, and that, after Recombination, the universe becomes transparent. We can make an analogy of a cloud - from the ground, we can only see the bottom surface of a cloud, which is where photons were last scattered. In the same way, we can see the opaque plasma-fog only as it was when photons were last scattered.

²A correlation function gives the statistical correlations between two variables. Here, between l and Θ . We will see another correlation function shortly, the *Galaxy 2 Point Correlation Function* in section 4.6.

The second important property of the SLS is its shape and distance from us. The SLS is discussed as being a spherical shell surrounding us, where its distance from us is equal to the distance a photon has travelled from it, to us.

The boundary between the opaque and the transparent is what forms the SLS. Since photons travel only at the speed of light, at the point of recombination, only the space close to an observer would appear transparent. The finite speed of light would entail that the SLS would travel outwards from an observer at the speed of light³

A range of cosmological phenomena has generated the anisotropies in the CMB; we split them into two categories:

- **Primary Causes:** those which occur up to and during Recombination (ie the creation of the CMB)
- **Secondary Causes:** those which occur after Recombination (ie when the CMB is travelling to us)

4.3 Primary Causes of Anisotropies in the CMB

In the Big Bang model of cosmology, the very first moments of the universe are still unclear. The **Planck epoch**⁴, which lasts for the first 10^{-43} s after the Big Bang⁵, is not covered in any current theory. The general holding is that this era is dominated by quantum gravitational effects, as gravity separates from the three other fundamental forces at this time.

Following this, we enter the **Grand Unification epoch** (from 10^{-43} s to 10^{-36} s after the Big Bang), during which the strong force also splits off. It is suggested that this causes a violation of the conservation of baryon number, which leads to our universe having more matter than anti-matter. It is also suggested that this is the trigger for inflation, which occurs next.

At these very small time scales, however, nothing is particularly certain. Fortunately, the state of the universe at this time is largely irrelevant, as inflation washes out almost all irregularities.

³In fact, the SLS is currently receding at approximately twice the speed of light, due to the expansion of the universe.

⁴It is called the ‘Planck’ epoch, as it lasts for one Planck unit of time (approximately 10^{-43} s).

⁵Inflationary and non-inflationary models disagree slightly on the exact endpoint of the Planck era.

The **Inflationary Epoch** is estimated to have started at 10^{-36} s after the Big Bang, and to have ended somewhere between 10^{-33} s and 10^{-32} s after the Big Bang. During this time, the volume of the universe expanded by a factor of (at least) 10^{78} .

The driving mechanism behind inflation is still not agreed upon (and neither is the theory itself, although it does form part of the standard cosmological model), but it is posited to be a scalar field called the Inflaton. Under this theory of inflation, the decay of the inflaton particle (into those particles we see today) is short-lived, and inflation ends following its decay.

Inflation theory predicts that this expansion was very nearly exponential. Inflation acts to separate two neighbouring sections of space at an accelerating rate. The energy density of the inflaton field will remain constant, but the energy density of radiation, matter, and curvature drop dramatically; thus inflation smooths out initial fluctuations of density, and flattens the curvature of space. The only remaining anisotropies will be quantum fluctuations.

After inflation, the universe goes through a **quark epoch**, a **hadron epoch**, **neutrino decoupling**⁶, a lepton epoch, and big bang nucleosynthesis (which we discussed in some detail in section 2.9), before entering its photon epoch.

4.3.1 The Photon-Baryon Plasma

Around 10 s after the Big Bang, we enter the **Photon Epoch**⁷. A very small number of atomic nuclei are formed in the first 500s or so, but the vast majority of the universe at this time is single photons, baryons, electrons,⁸ and dark matter. These cosmological constituents form an extremely hot and dense ‘photon-baryon’ plasma.

This plasma is opaque, as photons are affected by both Thomson and Compton scattering; since they are scattered so frequently, the photons would look like an opaque ‘fog’. Photons have such a short mean free path that they are essentially locked into local regions of the plasma.

The initial quantum fluctuations persist through to this era because Inflation is theorised to be exponential growth (and thus at every stage, the universe would be an exact enlargement of itself). At this stage of growth, the quantum fluctuations have become spatial variations in the matter and dark matter distributions inside

⁶It is this neutrino decoupling which causes the cosmic neutrino background which was observed by WMAP. I will not go into detail about it here, but [?](#) covers the initial discovery, and some useful background on the physics of this phenomena.

⁷So-named, because the energy of the universe is mainly that of photons.

⁸It should be noted that electrons have a negligible mass for the processes discussed later, and are therefore frequently neglected from calculations.

the plasma.

4.3.2 Perturbations in the Photon-Baryon Plasma

The variations in the dark matter distribution of the plasma lead to areas with deeper and shallower gravitational wells. Those with deeper gravitational wells draw in not only more dark matter, but also more ordinary matter. Local regions that end up with a high baryon density continue to draw in more baryons and photons through the gravitational force. As this happens, the pressure in the area increases, until the pressure has built enough that the overall force acts repulsively. The high density of photons set up a pressure gradient, down which the electrons move. They, in turn, drag protons away from the over-densities, through the electromagnetic force⁹. The repelled particles set up a pressure wave which propagates as a spherical wave through the plasma. This process repeats, with local regions moving between states of maximum density, and maximum rarefaction.

We model the propagating baryon pressure wave as an acoustic wave; that is, as an acoustic wave travelling through an ideal gas¹⁰, where gas molecules act in the same way as the baryons in the plasma.

The speed of sound in a fluid is given by

$$c_s = c \sqrt{\frac{\partial P}{\partial \rho}} \quad (4.8)$$

where $\rho = mn$ is the energy density in natural units¹¹. c_s is around 1.7×10^8 ms⁻¹ for our plasma.

General Fluid Mechanics Equations

We are now going to consider the mathematics of these perturbations of the plasma.

⁹The dark matter is not repelled because it does not interact with photons.

¹⁰Although it is not always appropriate to approximate a plasma to an ideal gas, in this case we can. The important difference between a plasma and an ideal gas relates to the permitted interactions: an ideal gas permits only interactions through direct elastic collisions of particles, while in a plasma there are also long-range interactions via the electromagnetic (and gravitational) force. However, in this early photon-baryon plasma, the densities are sufficiently high to neglect long-range interactions, while keeping short-range gravitational attraction - such a high density in the plasma leads to a large number of collisions, which should act to shield these long-range interactions.

¹¹It is $\rho = mnc^2$ in SI units.

There are three equations from fluid mechanics which we will need to use; in their most simple forms they are:

1. Continuity Equation

$$\frac{\partial n}{\partial t} + \nabla(n\mathbf{v}) = 0 \quad (4.9)$$

2. Euler Equation

$$mn \left(\frac{\partial}{\partial t} + (\mathbf{v} \cdot \nabla) \right) \mathbf{v} = -\nabla P \quad (4.10)$$

3. Poisson Equation

$$\nabla^2 \Phi = 4\pi G \rho \quad (4.11)$$

where n is number density, \mathbf{v} is velocity, P is pressure, and Φ is gravitational potential. The Continuity Equation is the condition for conservation of number of particles; we will see it later in a different form as the conservation of energy. The Euler Equation is the fluid mechanics counterpart of Newton's Second Law: the left hand side here is equivalent to ma , and the right hand side to F . The Poisson Equation here is that for a gravitational potential.

We will begin by considering the case of a density perturbations in a static fluid (ie $\vec{v} = 0$), with no gravity. Here we will only use equations 4.24 and 4.25 ; in addition, we need an equation of state of the form $P = P(n, T)$.

To the first order, perturbations will be of the form:

$$\begin{aligned} n &= \bar{n} + \delta n \\ P &= \bar{P} + \delta P \end{aligned} \quad (4.12)$$

where \bar{n} is the average number density, and \bar{P} the average pressure in the fluid; δn and δP are then the perturbations.

Invoking the definition of the speed of sound in a fluid: $c_s = \sqrt{\partial P / m \partial n}$ leads to the solution for these density perturbations

$$\left(\frac{\partial^2}{\partial t^2} - c_s^{-2} \nabla^2 \right) \delta n = 0 \quad (4.13)$$

The ansatz for this equation is

$$\delta n = C(\mathbf{k}) e^{i(\omega t - \mathbf{k} \cdot \mathbf{x})} \quad (4.14)$$

where $\mathbf{k} = 2\pi/\lambda$ is the wavevector.¹² It is related to the frequency by $\omega = c_s |\mathbf{k}|$.

Density Perturbations in a Static Spacetime

Now we will consider the same perturbations with the inclusion of gravity.

We use equations 4.24 and 4.26 in the same form, but eq. (4.25) gains an extra term:

$$mn \left(\frac{\partial}{\partial t} + (\vec{v} \cdot \nabla) \right) \vec{v} = -\nabla P - mn\nabla\Phi \quad (4.15)$$

We consider again 1st order perturbations:

$$\begin{aligned} n &= \bar{n} + \delta n \\ P &= \bar{P} + \delta P \\ \Phi &= \bar{\Phi} + \delta\Phi \end{aligned} \quad (4.16)$$

$$(4.17)$$

with constant \bar{n} , \bar{P} , and $\nabla\Phi = 0$.¹³ We require that the perturbed potential obeys $\nabla^2\delta\Phi = 4\pi Gm\delta n$.

This gives a new wave equation, counterpart to eq. (4.13):

$$\left(\frac{\partial^2}{\partial t^2} - c_s^{-2} \nabla^2 \right) \delta n = 0 \quad (4.18)$$

which has the same ansatz as before (eq. (4.14)), except now with the condition

$$\omega^2 = c_s^{-2} k^2 - 4\pi Gm\bar{n} \quad (4.19)$$

$$= c_s^{-2} (k^2 - k_J^2) \quad (4.20)$$

where the Jeans' wavenumber is defined as

$$k_J \equiv \sqrt{\frac{4\pi Gm\bar{n}}{c_s^{-2}}} \quad (4.21)$$

From this, we define the Jeans' length as

¹²This is not quite correct, since the perturbation is real; it is simply more convenient to write it in terms of an exponential than sin and cos; we use only the real part.

¹³This approach is known as Jeans' Swindle: $\nabla\Phi = 0$ does not actually obey the Poisson Equation. This makes sense physically, as there is no infinite static self-gravitating field. However, the maths will lead us to the next section where we consider perturbations in expanding spacetime, so we will turn a blind eye for now.

$$\lambda_J \equiv \frac{2\pi}{k_J} = c_s \sqrt{\frac{\pi}{Gm\bar{n}}} \quad (4.22)$$

The qualitative understanding of the solutions depends on whether the wavelength of the solution is larger or smaller than the Jeans' wavelength.

For $\lambda < \lambda_J$, the solutions oscillate as before (recall: as damped harmonic oscillators).

For $\lambda > \lambda_J$, the frequency is imaginary; this suggests that the perturbations grow or decay exponentially:

$$\delta n \propto e^{i\omega t} \quad (4.23)$$

Thus any perturbation with a sufficiently large wavelength will grow exponentially due to gravity. Physically we can understand this by considering the balance between the pressure and gravitational terms in the Euler equation: a region collapses under gravity; this collapse is opposed by the pressure; neither of these incidents are instantaneous - in a perturbation with a large wavelength, the time it would take for the pressure to build sufficiently to oppose the gravitational collapse is always smaller than the time it takes for gravitational collapse. Hence there are no oscillations and the region undergoes exponential collapse.

Density Perturbations in an Expanding Spacetime

We will now consider what effect an expanding, flat, spacetime has on density perturbations. The initial and final equations are listed here; you will show the steps from one to the other in a homework set.

We will make use of both physical and co-moving co-ordinates (see section 1.1). In replicating the previous section, we should consider the equations in terms of physical co-ordinates; however, the subsequent calculations are more elegant in co-moving terms.

As you will show in a homework set, the three equations become:¹⁴

1. Continuity Equation

$$\frac{\partial n}{\partial t} + 3Hn + \frac{1}{a}\nabla(n\mathbf{v}) = 0 \quad (4.24)$$

¹⁴For convenience, we have dropped the subscripts on the derivatives; here, ∇ is shorthand for $\nabla_{\mathbf{x}}$, and $\frac{\partial}{\partial t}$ for $\frac{\partial}{\partial t}|_{\mathbf{x}}$.

2. Euler Equation

$$mn \left(\frac{\partial}{\partial t} + \left(\frac{\mathbf{v}}{a} \cdot \nabla \right) \right) \mathbf{u} = -\nabla P - mn\nabla\Phi \quad (4.25)$$

3. Poisson Equation

$$\nabla^2\Phi = 4\pi G m n a^2 \quad (4.26)$$

As before, we next perturb the density:

$$n(\mathbf{x}, t) = \bar{n}(t) + \delta n(\mathbf{x}, t) \quad (4.27)$$

$$= \bar{n}(t)[1 + \delta(\mathbf{x}, t)] \quad (4.28)$$

where, for a non-relativistic fluid, $\bar{n}(t) \propto a^{-3}$, and we have defined $\delta = \delta n / \bar{n} = \delta\rho / \bar{\rho}$.

Substituting this into the continuity equation leads to the perturbed equation

$$\dot{\delta} = -\frac{1}{a} \nabla \cdot \mathbf{v} \quad (4.29)$$

In co-moving co-ordinates, for $\mathbf{v} = 0$, $\nabla \bar{P} = 0$, and $\Phi = \bar{\Phi}$, the Euler and Poisson equations combine¹⁵ to give

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3} m \bar{n} \quad (4.30)$$

Again perturbing with $P = \bar{P} + \delta P$, $\Phi = \bar{\Phi} + \delta\Phi$, the Euler and Poisson equations read respectively:

$$m \bar{n} a (\dot{\mathbf{v}} + H \mathbf{v}) = -\nabla \delta P - m \bar{n} \nabla \delta \Phi \quad (4.31)$$

$$\nabla^2 \delta \Phi = 4\pi G m \bar{n} a^2 \delta \quad (4.32)$$

Now to combine the three equations: the time derivative of Equation 4.29 combines with the gradient of Equation 4.31 and the Poisson equation to give

$$\ddot{\delta} + 2H\dot{\delta} - c_s^2 \left(\frac{1}{a^2} \nabla^2 + k_J^2 \right) \delta = 0 \quad (4.33)$$

which tells us the evolution of the density perturbation δ in an expanding space-time.

¹⁵Note that the two are no longer incompatible; a spatially homogenous fluid under the influence of gravity *can* exist in an expanding universe.

The growth of these perturbations is slightly more complex than that for a static change; we need to consider not only the Jeans' length $\lambda_J = 2\pi/k_J$, but also the expansion of the universe.

Again, only perturbations with wavelength $\lambda > \lambda_J$ can grow. Additionally, only perturbations are constrained by $\lambda < d_H$, where $d_H \approx c/H$ is the apparent horizon, since systems cannot interact at separations above that scale.

4.3.3 Baryonic Acoustic Oscillations

So now we understand how the perturbations themselves arise. We now consider the effect that these perturbations have on the plasma around them; they set up Baryonic Acoustic Oscillations (BAO), or acoustic waves which propagate through the plasma. Different initial densities of local regions will set up acoustic waves with different wavelengths; the plasma therefore has a whole spectrum of waves propagating through it in all directions.

These various waves all interfere with each other. As with all propagating waves, waves with the same wavelength λ interfere to form a standing wave. Since there are so many waves propagating in all directions, waves with different λ will interfere such that they cancel out their amplitudes. The waves in the plasma, then, interfere such that we are left with only a series of standing waves, with various different wavelengths λ_n .

It is important to note that the standing waves cause changes in density as they oscillate. At the moment of recombination, it is these different densities which will cause the primary anisotropies in the observed CMB (see fig. 4.3).

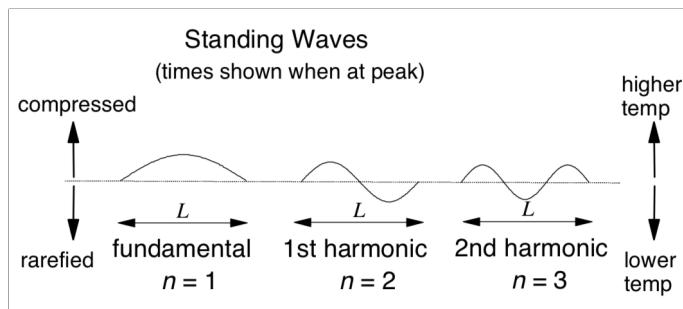


Figure 4.3: The first three standing waves, simplified to 1D. Note that the x -axis is position, and the y -axis is density (where ‘compressed’ denotes high density, ‘rarefied’ denotes low density, and $y = 0$ denotes equilibrium).

Gaussianity of Waves

It should be noted that, although the figures here depict the standing waves as being sinusoidal, this is not actually the case. In fact, the waves should be Gaussian.

This can be understood intuitively, if one considers that the initial quantum fluctuations have Gaussian distributions. Since these are what cause the matter fluctuations, we can expect those to be Gaussian too, as well as the pressure waves which result.

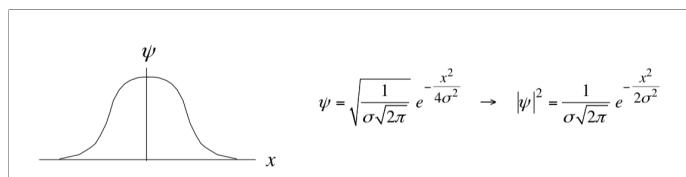


Figure 4.4: The probability distribution of a quantum particle; its wavefunction and probability density are on the right and far right, respectively.

There are now new theories which propose non-Gaussianities¹⁶, but Gaussianity is an important assumption for the basic mathematical understanding of the CMB. As such, we will continue to assume Gaussianity for this module.

4.3.4 Recombination

378,000 years after the Big Bang, the universe has expanded such that its temperature has now fallen enough that atoms can form¹⁷.

Since atoms can only be excited to discrete energies, much fewer photons are now absorbed. This gives photons a much longer mean free path, releasing them from the local regions in which they were trapped.¹⁸

At the point of Recombination, photon frequencies relate to the conditions of the local regions they were previously locked into: a higher local density ρ leads to higher

¹⁶See ? for more information.

¹⁷Recombination is dated through analysis of when electrons and protons could form hydrogen atoms, but a small number of helium and deuterium formed at a similar timescale too. For a full explanation of how recombination can be dated, see ?

¹⁸This is a simplification. In actuality, recombination and photon decoupling (when photons are initially freed from local regions) are separate events; after recombination, the sudden drop in the numbers of scattering electrons means that the mean free path of a photon increases to greater than the Hubble length (This is the radius of a Hubble sphere; a Hubble sphere is a sphere surrounding an observer, beyond the surface of which is space no longer causally connected to the observer, as it is receding at more than the speed of light. The Hubble radius is given by c/H_0 , where H_0 is the Hubble constant.). However, it is the photon-baryon ratio, and matter density of our universe that causes recombination and decoupling to happen almost simultaneously; in another universe, the two processes might happen at completely distinct times, and we would not have a CMB to analyse in the same way.

baryon pressure P ; this generates a higher local temperature T (this can be deduced from the Ideal Gas Law: $PV = nRT$), which itself leads to photons having higher energies, and thus also higher frequencies ν . So the photons, now free to travel throughout the universe, carry with them information about the initial conditions of the photon-baryon plasma through their frequencies.

We can now place parameters on the standing waves in the plasma: the standing waves must all have a length which is equal to furthest distance an acoustic wave could have travelled in the 378,000 years - we call this length the Sound Horizon, r_s . The sound horizon is calculated by considering the distance a sound wave could have travelled, from $t \approx 0$ at the Big Bang¹⁹, to $t = t_*$ at Recombination²⁰, and is therefore approximately: $r_s \approx \int_0^{t_*} c_s dt$.²¹

From this, we see that the standing waves will have wavelengths of $\lambda_n = 2r_s/n$. We assume that all waves have random amplitudes and directions.

Since we also assume that all wave formation begins at $t = 0$, all waves of the *same mode* will be in phase; they will all peak at the same time. But *different modes* of standing waves have different oscillatory frequencies, and so will peak at different times.

We use the standard set of equations for standing waves²²:

$$\lambda_n = \frac{2L}{n} \quad (4.34)$$

where L is the width of the standing waves, r_s in our case, and n is the mode of the standing waves. And

$$f_n = \frac{c}{\lambda_n} = nf_1 \quad (4.35)$$

gives the frequencies of the different modes, where f_1 is the frequency of the fundamental mode.

We assume that the fundamental mode is at a point of maximum compression when we reach recombination²³. This gives us a value for f_1 , from which we can

¹⁹Although there was no photon-baryon plasma directly at $t = 0$, the time between the Big Bang and when pressure waves could viably have started to propagate is negligible compared to the duration over which the waves propagate.

²⁰Note that parameters with the underscore $*$ denote their value at recombination.

²¹You will recognise the structure of the equation from eq. (1.23). A more accurate calculation of the sound horizon will be considered in Section 4.5.2, when the analysis of the power spectrum graph is discussed.

²²These carry the assumption that the waves all propagate with the same speed. We can make this assumption because the fluctuations in the density of the plasma are negligible when considering the speed of the pressure waves through it.

²³As will become clear later in discussion of the Angular Power Spectrum graph, we know that

determine not only the frequencies of the other modes, but also, using eq. (4.35), at what point of their oscillation they will be at recombination. Figure 4.5 illustrates this.

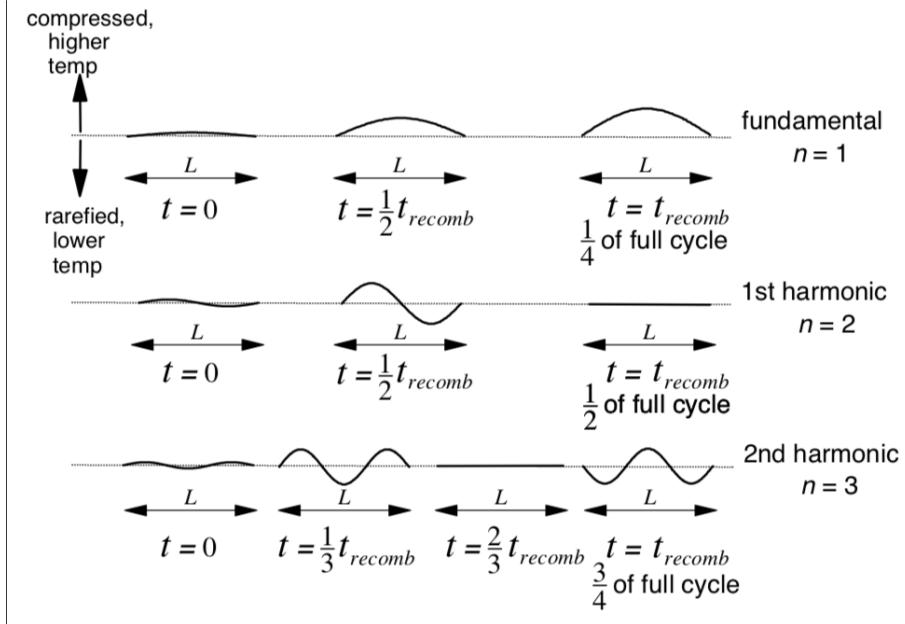


Figure 4.5: This shows at what point in their cycle the first three modes will be, at recombination.

Figure 4.5 shows the different modes at recombination. Note that the fundamental mode is overall a compressed state, while the $n = 2$ mode is at equilibrium, and the $n = 3$ mode is overall rarefied. Recall that a dense (or compressed) region will produce photons of a higher temperature and frequency (these will relate to modes $n = 1, 5, 9\dots$). By the same logic, a rarefied region will produce photons of a lower temperature ($n = 3, 7, 11\dots$), and one in equilibrium will produce photons of average temperature ($n = 2, 4, 6, 8\dots$). Thus odd-numbered modes will give rise to detectable anisotropies in the CMB, while even-numbered ones will not. This will be discussed further in discussion of the graph.

Diffusion Damping

The other primary anisotropy comes from diffusion (or ‘Silk’²⁴) damping, where photons diffuse from higher temperature space to lower temperature space. This happens during photon-decoupling (recall that this is coeval with recombination).

it is neither a point of maximum rarefaction, nor equilibrium, else the peaks in the graph would appear differently.

²⁴After Joseph Silk who first described the phenomenon, and after whom the Silk Scale on which diffusion damping operates is named; this scale corresponds to the sizes of galaxies of the present day.

It is a consequence of the fact that recombination and photon-decoupling do not happen instantaneously (this leads to the Surface of Last Scattering having a finite depth²⁵).

The diffusing photons take with them electrons, which in turn attract the protons²⁶; this acts to exponentially reduce the density fluctuations, and to make the matter more isotropic at scales of $l \geq 800$.

4.4 Expansion, and Secondary Causes of Anisotropies in the CMB

In the time between Recombination and the present day, the universe has been undergoing expansion. This leads to a cosmological redshifting of the CMB photons, increasing their wavelengths. In general, cosmological redshift, z is given by:

$$z = \frac{a_{\text{now}}}{a_{\text{then}}} - 1 = \frac{\lambda_{\text{observed}}}{\lambda_{\text{emitted}}} - 1 \quad (4.36)$$

where a is the time-dependent cosmic scale factor²⁷:

$$\begin{aligned} a(t) &= \frac{\mathbf{r}(t)}{\mathbf{r}(t_0)} \\ &= \frac{\mathbf{r}(t)}{\mathbf{r}_0} \end{aligned} \quad (4.37)$$

where $\mathbf{r}(t)$ is the distance between two points in space. It should then be noted that the scale factor represents the expansion of space.

Since $z_* \approx 1100$, the observed wavelengths of the CMB will be

$$\lambda_{\text{observed}} \approx (1100 \times \lambda_{\text{emitted}}) \quad (4.38)$$

In addition to this cosmic expansion, there are other effects which have altered the CMB photons, causing Secondary Anisotropies. The three most relevant are the ionized state of matter in the universe, the Sachs-Wolfe effect, and the Sunyaev-Zel'dovich effect.

²⁵In cases where we ignore the previous assumption of instantaneous recombination, the SLS takes on the property of finite depth; we can consider it now a spherical shell with a finite thickness.

²⁶For this reason, the density fluctuations are adiabatically damped (the ratio of photons to baryons is constant).

²⁷It is less a time-dependency than a state-of-the-universe-dependency - see Table 4.1 for further details.

Table 4.1: Scale Factor and Recessional Velocity dependency in different eras. Note that $v \propto \dot{a}(t)/a(t)$.

Domination Era	Time after Big Bang	$a(t)$	$\dot{a}(t)$	v
Radiation	10^{-32} s - 47,000 years	$a \propto t^{\frac{1}{2}}$	$\dot{a} \propto \frac{1}{2}t^{-\frac{1}{2}}$	$v \propto \frac{1}{t}$
Matter	47,000 years - 9.8 billion years	$a \propto t^{\frac{2}{3}}$	$\dot{a} \propto \frac{2}{3}t^{-\frac{1}{3}}$	$v \propto \frac{1}{t}$
Dark Energy	9.8 billion years - present	$a \propto e^{\sqrt{\Lambda/3} \times t}$	$\dot{a} \propto \sqrt{\frac{\Lambda}{3}} e^{\sqrt{\Lambda/3} t}$	$v \propto \sqrt{\Lambda/3}$

4.4.1 Ionization

After recombination, the universe remains un-ionized for 149.5 million years²⁸. But between 150 million years and 1 billion years after the Big Bang, re-ionization occurs. During this time, hydrogen atoms are ionized by energetic objects; the universe reverts to another ionized plasma. We are still not certain what these energetic objects are (although they must be objects which produce ultraviolet radiation, as this is the wavelength which corresponds to the energy required to ionize neutral hydrogen); it is suggested that they might be quasars and early stars.

This second ionized plasma, which makes up the majority of the intergalactic medium, scatters photons through Thomson scattering. This leads to an erasure of temperature anisotropies at a very small angular scale, while also introducing polarization anisotropies on large angular scales.

4.4.2 Sachs-Wolfe Effect

The Sachs-Wolfe effect causes additional fluctuations in the observed temperatures of the CMB photons, through gravitational red- and blue-shifting. It is the primary source of anisotropies at angular scales of over 10° .

The Sachs-Wolfe effect (SWE) can be split into two components: the Non-Integrated SWE, and the Integrated SWE.

The Non-Integrated Sachs-Wolfe Effect is caused by the gravitational redshifting of the CMB photons which occurs at the point of recombination. As such, it is actually a primary anisotropy. However, its similarity to the Integrated Sachs-Wolfe Effect (below) is such that it makes more sense to discuss both here.

Recall that the dark matter distribution across the plasma is uneven, but that its density does not directly affect the frequencies of the CMB photons. However, it does create uneven gravitational potential wells; when the photons first begin to

²⁸An era which is ‘humorously’ referred to as the Dark Ages, as no stars form in this time.

travel, they first have to exit these potential wells, and are redshifted as a result. The unequal nature of these potential wells means that CMB photons in different locations lose different amount of energy, thus adding additional anisotropies.

It occurs on a large angular scale as a result of the dark matter over-densities being large, in comparison to the photon-baryon over-densities.²⁹

The Integrated Sachs-Wolfe Effect (ISW) is caused by the gravitational red-shifting which occurs between recombination, and our observation of the photons. It occurs only when the universe is not significantly matter-dominated.

Through different phenomena, gravitational wells and voids are shallowed. At a high enough rate, this can occur quickly enough that a photon entering a gravitational potential well/void ends up keeping/losing an amount of the potential energy. This results in photons which have a higher/lower energy (and frequency) than would otherwise be expected (see fig. 4.6).

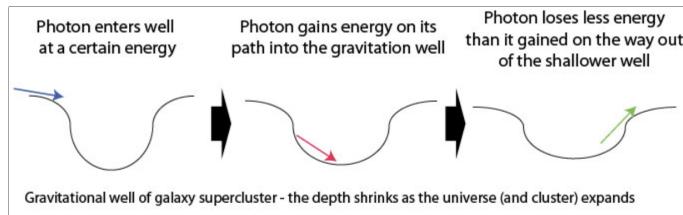


Figure 4.6: The Integrated Sachs-Wolfe Effect.

The ISW is split into early-time and late-time ISW.

Early-Time ISW happens soon after recombination, when, although the universe is technically matter-dominated, there is a sudden excess of radiation, which can also cause decay to the potential wells themselves; when photons leave the potential wells, the density will decrease, and so too will the gravitational potential.

Local regions of the photon-baryon plasma which have a higher density produce photons of a higher energy. However, they also have a higher gravitational potential, so escaping photons lose a greater amount of energy; you can think of the two effects battling to cancel each other out.

Late-Time ISW happens when dark energy begins to affect the cosmological constant.

That Late-Time ISW does not occur during matter-dominated times is related to how the scale factor changes during different eras. From table 4.1, it can be seen

²⁹Recall that the cross-sections of dark matter for both self-interaction, and interaction with matter, are small.

that the recessional velocity of two points in space (here they might be points on either edge of a gravitational potential well) is proportional to $1/t$ in both radiation- and matter-dominated eras, and thus the expansion of a potential well would not be fast enough to affect a photon via the Sachs-Wolfe effect. However, during radiation-domination, the dispersion of photons from a potential well also affects the decay rate of the well. This will be discussed further shortly.

The universe is considered to be dark-energy-dominated at around 9.8 billion years after the Big Bang. During this era, the expansion of space is happening fast enough that a photon entering a gravitational potential will exit a potential which has been shallowed. At this point, cosmological expansion is accelerating, and this is when the Sachs-Wolfe effect has the largest effect.

The overall effect is that CMB photons which pass through superclusters end up warmer than they would otherwise be, and those which pass through supervoids are cooler. Superposition of observed superclusters and supervoids over the CMB corroborate this.

4.5 What Does the CMB Tell Us About the Universe?

4.5.1 The Angular Power Spectrum Graph

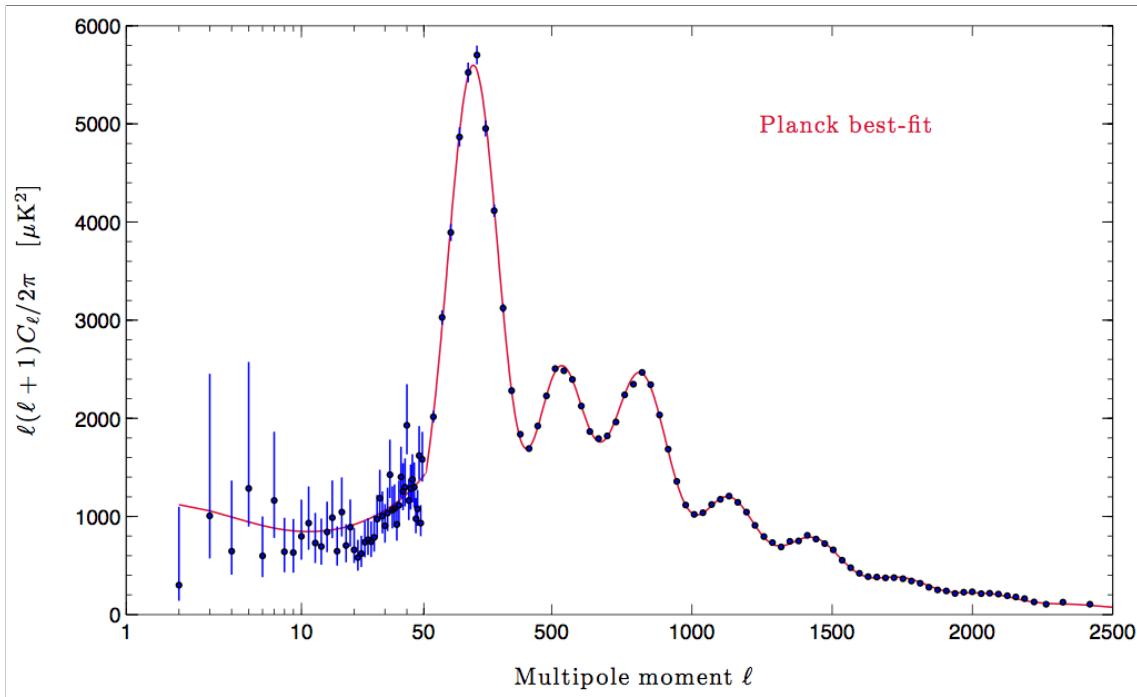


Figure 4.7: The angular power spectrum, plotted from data from Planck 2018. Note the log scale on the x -axis up to $l = 50$. Frequently, angular power spectrum graphs have a log scale for the whole x -axis, but it is useful here to see the characteristic angular scale between the peaks.

Axes

The axes on fig. 4.7 are not straightforward; note the log-scale on the x -axis³⁰, and the unusual value plotted on the y -axis: $l(l + 1)C_l/2\pi$.

The log-scaling is typically used because the first few peaks are the most important; a log-scale gives more graph space to the lower values of l - due to the damping, higher values of l do not contribute significant peaks, and can thus often be given less space.

To explain the value plotted on the y -axis, we must consider the relation between C_l and the Fourier amplitude $A(k)$.

³⁰This graph has a log-scale up to $l = 50$, and is linear after that. This is an unusual format but it will be useful to show the Characteristic Angular Scale (discussed in section 4.5.2).

Inflation posits a uniform distribution of standing waves across wavenumber k . This would be easy to plot on a Fourier analysis graph, as each Fourier amplitude $A(k)$ is independent of the other amplitudes. However, the spherical harmonic power spectrum coefficients C_L are all a mixture of every Fourier amplitude:

$$C_l = \frac{2}{\pi} \int \langle |A(k)|^2 \rangle j_l^2(kr) k^2 dk \quad (4.39)$$

where j_l are spherical Bessel functions.

On a Fourier power spectrum graph, we would plot the power per unit k :

$$P(k) = \frac{k^2}{2\pi^2} \langle |A(k)|^2 \rangle \quad (4.40)$$

Solving eq. (4.40) for $\langle |A(K)|^2 \rangle$, and substituting this into eq. (4.39) gives C_l in terms of power per unit k :

$$C_l = 4\pi \int P(k) j_l^2(kr) dk \quad (4.41)$$

But, since we have chosen a log-scale for the x-axis, we want an equivalent to eq. (4.41) with $d(\log(k))$ - power per unit $\log(k)$. We rewrite eq. (4.41) as:

$$\begin{aligned} C_l &= 4\pi \int P(k) j_l^2(kr) k \frac{dk}{k} \\ &= 4\pi \int P(k) j_l^2(kr) k d(\log(k)) \\ &= 4\pi \int j_l^2(kr) \Delta^2(k) d(\log(k)) \end{aligned} \quad (4.42)$$

where we have defined $\Delta^2(k) \equiv kP(k)$ as the power per unit $\log(k)$.

A property of spherical Bessel functions of order l , $j_l(z)$ is that they peak only when $z \approx l$; it is a very narrow peak, and j_l is zero everywhere else³¹. For our purposes, this means that we need only consider when $j_l(kr) \approx j_l(l)$, or when $kr \approx l$. Over this narrow range, $\Delta^2(k) \approx \Delta^2(l/r) = (l/r)P(l/r)$ is approximately constant. Thus we get:

$$C_l \approx 4\pi \Delta^2(k = l/r) \int j_l^2(kr = l) d(\log k) \quad (4.43)$$

Using a standard integral,

³¹See appendix A.1 for more information.

$$C_l \approx \frac{2\pi}{l(l+1)} \Delta^2(k) \quad (4.44)$$

$$\Delta^2(k) \approx \frac{l(l+1)}{2\pi} C_l \quad (4.45)$$

which gives us the power per unit $\log(k)$ in terms of C_l - this is what we wanted to plot.

Small values of l

You will notice that there are no values plotted for $l = 0$ and $l = 1$ on the graph. This is because the Y_0 spherical harmonic is constant over the whole sphere (a monopole), and Y_1 is a dipole.

The temperature value at the $l = 0$ angular scale is the average temperature, and the value at the $l = 1$ scale is the dipole which gives us information about the Earth's velocity relative to the CMB³².

4.5.2 First Peak - The Curvature of the Universe

The first peak in fig. 4.7 is at approximately $l = 221$, or at an angular separation of around 1° . This location tells us about the curvature of the universe.

Although we observe the CMB photons close to Earth, we can consider that they occur at the Surface of Last Scattering (SLS) (see fig. 4.8 below). Since the first peak is caused by the fundamental standing wave, we also have a width at the SLS, r_s ³³. We can determine the distance to the SLS, and thus by measuring the angle subtended by the fundamental mode, we can determine the curvature of space; we are essentially constructing a triangle, which will have different properties in different curvatures: in flat space, the angles in a triangle sum to 180° (a flat universe), in hyperbolic space, to $< 180^\circ$ (an open universe), and in elliptical space, to $> 180^\circ$ (a closed universe).

³²See appendix A.4 for further information.

³³Recall from earlier that the width of the standing waves must all be equal to the distance acoustic waves could have travelled by recombination - the sound horizon, r_s .

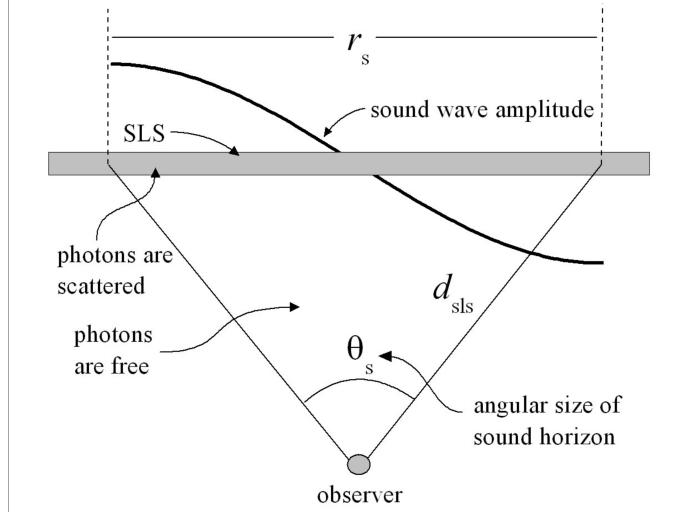


Figure 4.8: A simplified explanation of what we are trying to work out here.

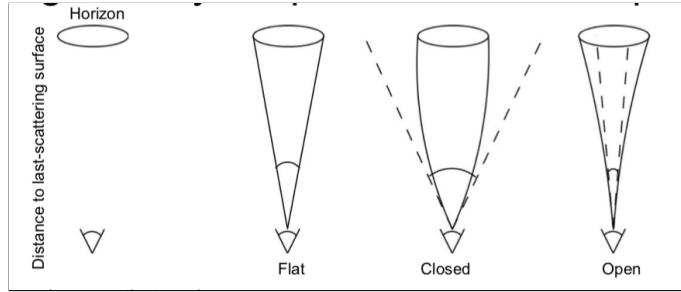


Figure 4.9: A 2D simplification of the angle subtended by the sound horizon, as it is observed now.

Earlier we discussed r_s as being approximately equal to $\int_0^{t_*} c_s dt$. Although this is not quite an equality, it approximates the result well enough.

The speed of the acoustic waves is given by

$$c_s \approx \frac{c}{\sqrt{3 \left(1 + \frac{3\Omega_b}{4\Omega_r}\right)}} \quad (4.46)$$

and dt is given by

$$dt = \frac{1}{H_0(1+z)} \times \frac{1}{\sqrt{(1+z)^2(1+\Omega_{m,D}z) + z(z+2)[(1+z)^2\Omega_{r,D} - \Omega_{\Lambda,D}]}} dz \quad (4.47)$$

We will use these later to determine the sound horizon r_s . Additionally, the distance to the SLS d_{SLS} is related to the angular radial distance r_{SLS} by:

$$d_{SLS} = \frac{r_{SLS}}{1 + z_*} \quad (4.48)$$

where r_{SLS} is given by:

$$r_{SLS} = \frac{c}{H_0} \int_0^{z_*} \frac{1}{\sqrt{\Omega_m(1+z)^3 + \Omega_\Lambda}} dz \quad (4.49)$$

It will suffice here to only consider the derivation for a flat universe, with the understanding that an observed angle larger than its prediction would indicate an open universe, and smaller than its prediction would indicate a closed one. We will also assume a matter dominated universe, as this is true for the majority of the time it has taken for photons to reach us.

With these assumptions, equations 4.46 and 4.47 simplify to:

$$c_s = \frac{c}{\sqrt{3}} \quad (4.50)$$

$$dt = \frac{1}{H_0 \sqrt{\Omega_m} (1+z)^{\frac{5}{2}}} dz \quad (4.51)$$

This gives us an equation for r_s :

$$\begin{aligned} r_s &= \frac{c_s}{H_0 \sqrt{\Omega_m}} \int_{z_*}^{\infty} \frac{1}{(1+z)^{\frac{5}{2}}} dz \\ &= \frac{2c}{3\sqrt{3}H_0\sqrt{\Omega_m}(1+z_*)^{\frac{3}{2}}} \end{aligned} \quad (4.52)$$

And using a Binomial expansion on eq. (4.49), neglecting the higher order terms, and substituting into eq. (4.48) gives:

$$d_{SLS} = \frac{2c(9 - 2\Omega_m^3)}{7H_0(1+z_*)\sqrt{\Omega_m}} \quad (4.53)$$

Since $l \approx d_{SLS}/r_s$, our estimate for the l -value of the first peak, in a flat universe is:

$$l \approx \frac{(9 - 2\Omega_m^3)\sqrt{1+z_*}}{\sqrt{3}} \approx 221 \quad (4.54)$$

So it turns out, our universe is approximately flat.

Ratio of Heights of Even and Odd Peaks - Baryon Density of the Universe

It can be seen from fig. 4.7 that the peaks do not all have the same amplitude; there is a trend for the amplitudes to decrease as l increases. Partially this is due to damping (which will be discussed in section 4.5.3), but there is also a trend for even-numbered peaks to be lower than odd-numbered peaks (see the difference in the heights of the 1st and 2nd peak, and similarity in the 2nd and 3rd peaks).

This is a consequence of odd-numbered peaks being compressive, and even-numbered peaks being rarefied. Recall from section 4.3.4 that some modes of standing wave are overall compressive in nature, some are rarefied, and some are in equilibrium. Those waves which are compressed give rise to the odd-numbered peaks in the graph, those which are rarefied to the even-numbered peaks, and those which are at equilibrium to the troughs³⁴.

We can treat the system which sets up the standing waves as a spring-mass system (see fig. 4.10). In such a system, maximum rarefaction (which corresponds to the initial state of the system) is the same for all masses, or all numbers of baryons. But for greater masses, or for larger numbers of baryons, maximum extension (compression in the plasma) will be greater. Hence we can determine from the ratio of the heights of the odd- to even-numbered peaks, the overall baryon density of the universe.

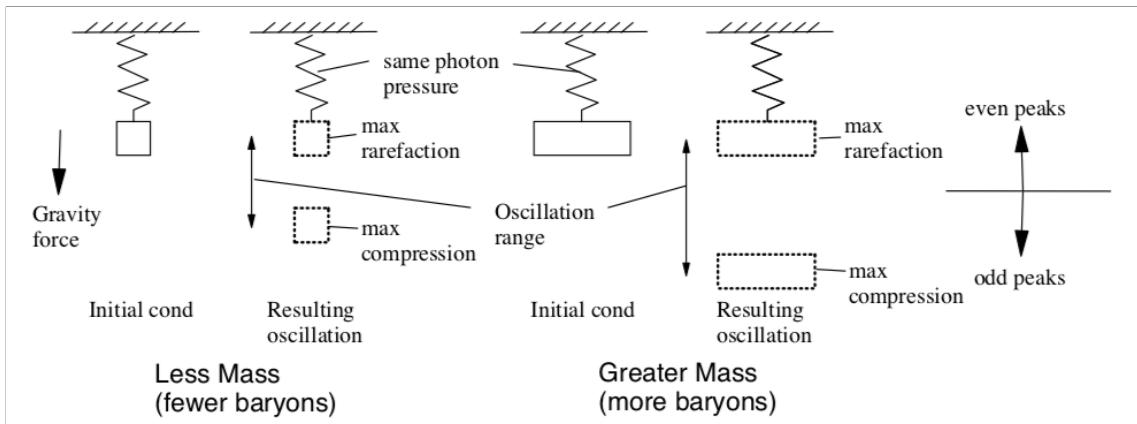


Figure 4.10: An analogy of a spring-mass system to the photon-baryon plasma pressure systems. Note that the mass relates to baryon mass, and the spring to pressure. In this analogy, a state of extension relates to a state of compression in the plasma.

Characteristic Angular Scale

As can be seen in fig. 4.7, there is a characteristic angular scale between the higher-order peaks; the peaks are spaced an average $\approx 310 l$ apart (from fig. 4.7).

An explanation of this can be given by consideration of the higher-order modes of the standing waves. In the same way that we used $r_s = 1/2\lambda_1$ to determine the angular scale of the first peak, so too can we approximate the angular scales of the higher-order peaks.

³⁴The axes on the graph are constructed such that a lack of anisotropies would generate a straight horizontal line; those waves which were at equilibrium would generate no anisotropies, so at their angular scale, we expect the graph to revert to that horizontal line.

Recall from section 4.5.1 that

$$l_{n=2m} \approx \frac{d_{SLS}}{1/2\lambda_{2m}} = \frac{d_{SLS}}{\frac{\lambda_1}{2m}} = \frac{2m \times d_{SLS}}{r_s} = 2ml_1 \quad (4.55)$$

where we have substituted $n = 2m$ as only even-numbered standing wave modes contribute peaks in the graphs, and used the standard standing wave equations from section 4.3.1. This gives an idea of the regularity between the l -values of the subsequent peaks. It does not give an accurate numerical approximation, it should be noted, as the further peaks are also dependent on other parameters (see section 4.5.4).

4.5.3 Damping

The damping of the peaks in the power spectrum has a few causes. Primarily, it is to do with the finite thickness of the Surface of Last Scattering; throughout the recombination age, there is a mixture of atoms, and single baryons and electrons. As we progress through recombination, the ratio of atoms to particles increases, and the amount of photon scattering decreases, but most photons will scatter at least once before being released across the universe.

When the mean free path of a photon is larger than a wavelength λ_n , continued scattering throughout recombination will damp the anisotropies at the angular scale l associated with that standing wave, thus reducing the height of its peak in the Angular power spectrum graph. Since shorter λ_n (which relate to high values of l) will begin to be damped first, their overall damping will be largest. This is what causes the exponential damping tail of fig. 4.7. ?

Other factors also affect the damping. A greater baryon density, for example, will result in more photon scattering, and thus in a greater degree of damping (see section 4.5.4 for more).

4.5.4 Cosmological Parameters

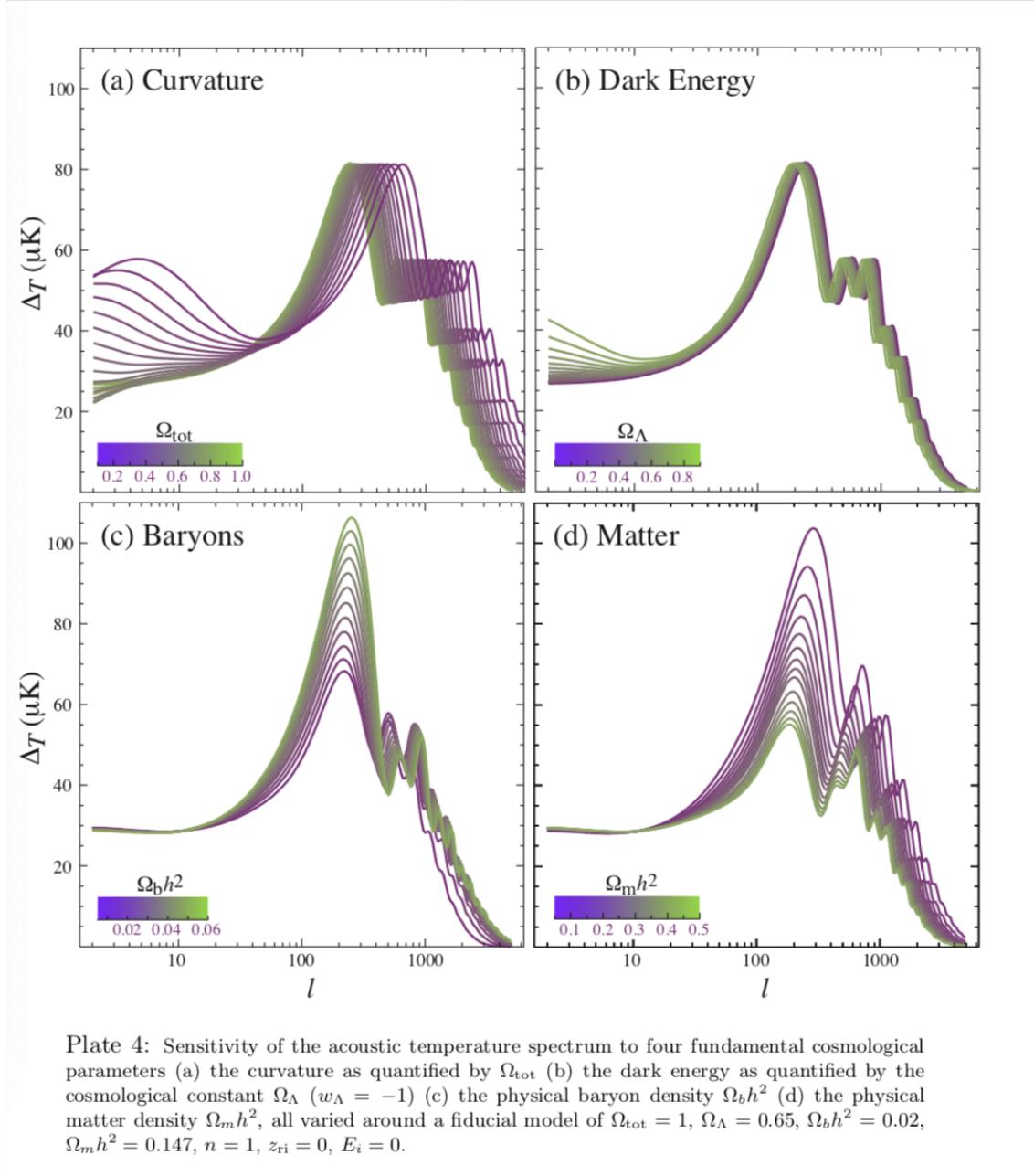


Figure 4.11: The theorised effects of cosmological parameters on the angular power spectrum graph.

Figure 4.11 gives a qualitative suggestion of how the angular power spectrum graph is affected by the variation of different cosmological parameters. From this, we can see that:

- The **Curvature** of space affects the l positions of the peaks, as well as the initial plateau. An open universe will have peaks at lower l , with little effect

on the plateau, while a closed one will have peaks at higher l , as well as a peak of sorts instead of the plateau.

- The proportion of **Dark Energy** has a small effect on the positions of the peaks; a greater dark energy percentage causes peaks to occur at lower l . A higher proportion of dark energy also increases the power at very low l , affecting the shape of the plateau.
- The **Baryon Density** $\Omega_b h^2$ has a very small effect on the l -positions of the peaks, but dramatically affects their heights³⁵; a greater baryon density will cause the odd-numbered peaks to be higher, and the even-numbered peaks to be lower. A low baryon density also increases the damping.
- The overall **Matter Density** $\Omega_b h^2$ again primarily affects the heights of the peaks, with some effect on their position. A higher matter density results in lower peaks, at slightly lower l values; lower matter density does the opposite, and has a more pronounced effect on the peak heights.

NB from the values the CMB data gives us for $\Omega_b h^2$ and $\Omega_b h^2$, we can deduce the **Dark Matter Density**: $\Omega_{\text{dark matter}} h^2 = \Omega_b h^2 - \Omega_b h^2$.

4.6 The Galaxy 2 Point Correlation Function

Galaxies in the universe are not evenly spaced out; they are clustered on and at the end of filaments and sheets surrounding large voids which are around 100 Mpc.

The Galaxy 2 Point Correlation Function is defined as a measure of the additional probability, relative to the Poisson distribution, of finding 2 galaxies, contained within the volume elements dV_1 and dV_2 , separated by \mathbf{r} :

$$dP_{1,2} = n^2[1 + \xi(\mathbf{r})]dV_1 dV_2 \quad (4.56)$$

where n is the mean number density over the whole sample volume. ξ measures an excess of clustering ($\xi > 0$) or a scarcity ($\xi < 0$), compared to the prediction of the Poisson distribution ($\xi = 0$).

³⁵Section 4.5.2 covers this in more detail.

From this we can deduce the equation for the conditional probability of finding a galaxy in dV at a distance r from a given galaxy:

$$dP = n[1 + \xi(r)]dV \quad (4.57)$$

The number of galaxies, on average, located between r and $r + dr$ from a given galaxy is:

$$n[1 + \xi(r)]4\pi r^2 \quad (4.58)$$

The function usually follows a power law which drops off quickly with radius $\xi(r) \propto (r/r_0)^{-\gamma}$ where typical values of r_0 are about 8 Mpc and γ is around 1.8 but this varies for different types of galaxies. This arises simply because of collisionless gravitational collapse. The BAOs mean that there is a slightly increased probability of finding an overdensity of baryons separated by the characteristic scale corresponding to the peaks in the CMB. It turns out that this results in a slightly increased probability of forming a galaxies which are separated by that scale. This has actually been detected in the two point function. It is a very small effect and very difficult to measure. However it does mean that we can observe the same comoving ruler at a totally different redshift ($z \sim 0.5$ rather than $z \sim 1100$). By performing the same analysis with the angular distance measurement at two different redshifts, we can break the degeneracy in the density content of the Universe.

Chapter 5

Detection

In contrast to the astrophysical evidence for dark matter described in chapter 2, this chapter will address the current undertakings into finding evidence for specific candidates for dark matter.

It should be noted that there has yet to be any strong claim of dark matter particle detection; this sets ever narrowing constraints on the mass of proposed dark matter particles, and their interaction cross-sections.

5.1 Detection of WIMPs

5.1.1 The Theory

There is a the possibility of WIMPs scattering off quarks. According to the diagram in fig. 5.1, this results in a scattering off nuclei. In the Milky Way, the dark matter is moving at around 200 km s^{-1} which corresponds to $v \sim 10^{-3}c$. If we look at the scattering of a dark matter particle off a nucleus at rest we get

$$m_\chi v_{\chi 1} = m_\chi v_{\chi 2} + m_A v_{A2} \quad (5.1)$$

$$\frac{1}{2} m_\chi v_{\chi 1}^2 = \frac{1}{2} m_\chi v_{\chi 2}^2 + \frac{1}{2} m_A v_{A2}^2 \quad (5.2)$$

And we get a typical recoil energy

$$E_{rec} = \frac{2m_A m_\chi^2}{(m_\chi + m_A)^2} v_\chi^2 = 2 \frac{\mu_{A\chi}^2}{m_A} v_\chi^2 \quad (5.3)$$

where m_A is the mass of the nucleus which is struck by the dark matter, m_χ is the mass of the dark matter particle, and $\mu_{A\chi} = m_\chi m_A / (m_\chi + m_A)$ is the reduced mass.

If we approximate the Milky Way as being an isothermal sphere then we calculate

the circular velocity $v^2 = GM/r = 2\sigma^2$ (where here σ is the velocity dispersion of the isothermal sphere rather than a cross section). The value of the circular velocity of the Sun is around 220-230 km s⁻¹. In this way we see that $\sigma \sim 160$ km s⁻¹. However since we are moving through the dark matter halo, we might expect the typical velocity with which we hit dark matter to be around 200 km s⁻¹; in dimensionless units, $v \sim 10^{-3}$.

Quickly considering some examples where the dark matter has the same mass as the target nucleus: 100 GeV of dark matter and Xenon (roughly 130 nucleons) gives $E_{rec} \sim m_\chi v^2 \sim 25 - 50$ keV. Lighter Dark matter, eg for 10 GeV dark matter particle, gives $E_{rec} \sim 1$ keV.

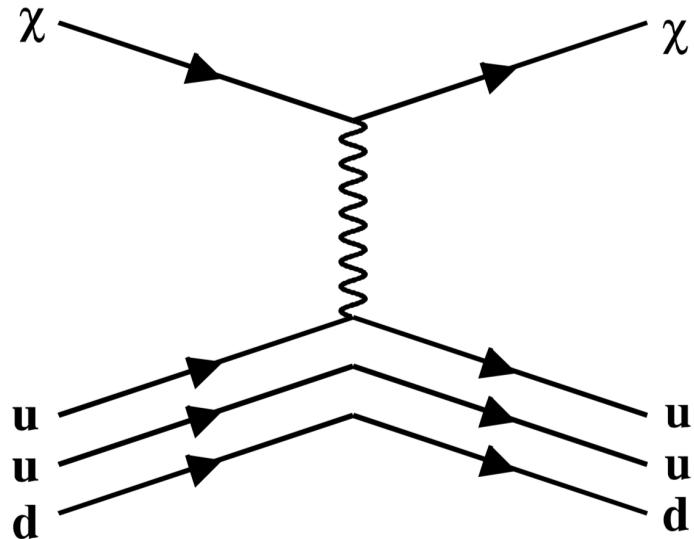


Figure 5.1: The scattering of dark matter χ of a nucleon via a new hypothetical Z-boson Z' .

The cross section for this happening is given by

$$\sigma_{\chi N}^{SI} = \frac{\mu_{\chi N}^2}{M_{Z'}^4} f_N^2 \quad (5.4)$$

where the superscript SI means spin independent, so this is when the interaction of the boson with the nucleon does not depend upon the overall spin of the nucleus.

The parameter $\nu_{\chi N}$ is the reduced mass of the nucleon and

$$f_p = g_\chi (2g_u + g_d) \quad (5.5)$$

$$f_n = g_\chi (g_u + 2g_d) \quad (5.6)$$

The deBroglie Wavelength of the dark matter particles will be $\lambda \sim h/(mv)$ and for a 10 GeV dark matter particle, this will correspond to around $\lambda \approx 10^{-3} \text{ MeV}^{-1}$, which is $2 \times 10^{-14} \text{ m}$. And since the De Broglie Wavelength of the dark matter $\lambda \sim 1/(mv)$ is larger than the size of the nucleus, the matrix element contains a scatter off all A nuclei *coherently*, which means that the cross section is enhanced by A^2 .

This is only the case for spin independent cross sections, i.e. cross sections which do not depend upon the orientation of the dark matter spin and the nucleon spin. Those cross sections which do depend on this will not receive this A^2 enhancement, since the spin of most of the nuclei are paired, and only nuclei with unpaired nucleons will be able to detect such interactions.

For more massive or more quickly moving dark matter, this A^2 starts to be suppressed because the de Broglie wavelength starts to become comparable or smaller than the size of the nucleus, hence the interaction contains a *form factor* which takes this into account.

So now we imagine that $g_u = g_d = g_\chi = 0.1$ and that $M_{Z'} = 500 \text{ GeV}$ and $m_\chi = 50 \text{ GeV}$. Let us also imagine we are scattering off a particle with 100 nuclei in (approximating the mass of a nucleon to 1 GeV) then we have

$$\sigma_{\chi A}^{SI} = A^2 \sigma_{\chi N} = 100^2 \frac{\mu_{\chi N}^2}{M_{Z'}^4} f_N^2 = 100^2 \frac{(100 \text{ GeV} \times 1 \text{ GeV})^2}{(101 \text{ GeV})^2 \times (500 \text{ GeV})^4} 0.1^4 = 10^{-39} \text{ cm}^2 = 10^{-3} \text{ pb} \quad (5.7)$$

where here pb means pico-barns or 10^{-12} barns. A "barn" is a unit of cross section used by particle physicists corresponding to 10^{-24} cm^2 and it a relatively large cross section by particle physics standards (roughly the cross section for two neutrons to scatter off each other).

Now how big a detector do we need to see such an object? The local dark matter density is $\rho_{DM} = 0.3 \text{ GeV cm}^{-3}$ so for a particle moving relative to dark matter at rest at 200 km s^{-1} , the number of times it will be hit by a dark matter particle every second is

$$\frac{\rho_{dm}}{m_{dm}} \sigma_{\chi A} v = \frac{0.3 \text{ GeV cm}^{-3}}{50 \text{ GeV}} 10^{-39} \text{ cm}^2 \times 2 \times 10^7 \text{ cm s}^{-1} \sim 10^{-34} \text{ s}^{-1} \quad (5.8)$$

In practise we need to take into account the Maxwell-Boltzmann distribution of dark matter in the Milky Way, including the cut-off at the escape velocity. We also need to take into account the motion of the Earth around the Sun which changes the velocity distribution in the rest frame of the earth around the year - annual modulation.

There are a variety of different methods currently employed in the hope of direct detection of WIMPs.

5.1.2 Cryogenic Dark Matter Search (CDMS)

CDMS is a series of cryogenic crystal detectors, which utilise large ($\approx 10\text{cm}$) germanium and silicon crystals cooled to the milli-Kelvin range, with tungsten strips at superconducting temperature at either end of the crystals. The crystals are cooled to this extent in order to limit background interactions; the detectors are also set up in remote locations for the same reason.

The detectors are set up to measure the ionization of every interaction, as well as the phonons produced by it. The ratio of ionization to phonon production is different for particle interactions with atomic electrons, and atomic nuclei. It is expected that dark matter is the only particle interaction which would involve the atomic nuclei inside the crystal.

Phonon detection is measured by considering the tungsten transition edge sensors (TESs): if a particle interacts with an atom inside a crystal, this will cause the crystal to heat up; this affects the conductivity of the tungsten, which can be detected by a change in its resistance.

The ionization is measured with field-effect transistors (FETs), which control the flow of current through the crystals by manipulating an electric field.

The CDMS experiments are currently in their third detector: CDMS I and II have been run to completion, as has SuperCDMS; SuperCDMS SNOLAB started construction in 2018, and will continue the experiments when it begins operation in the next few years.

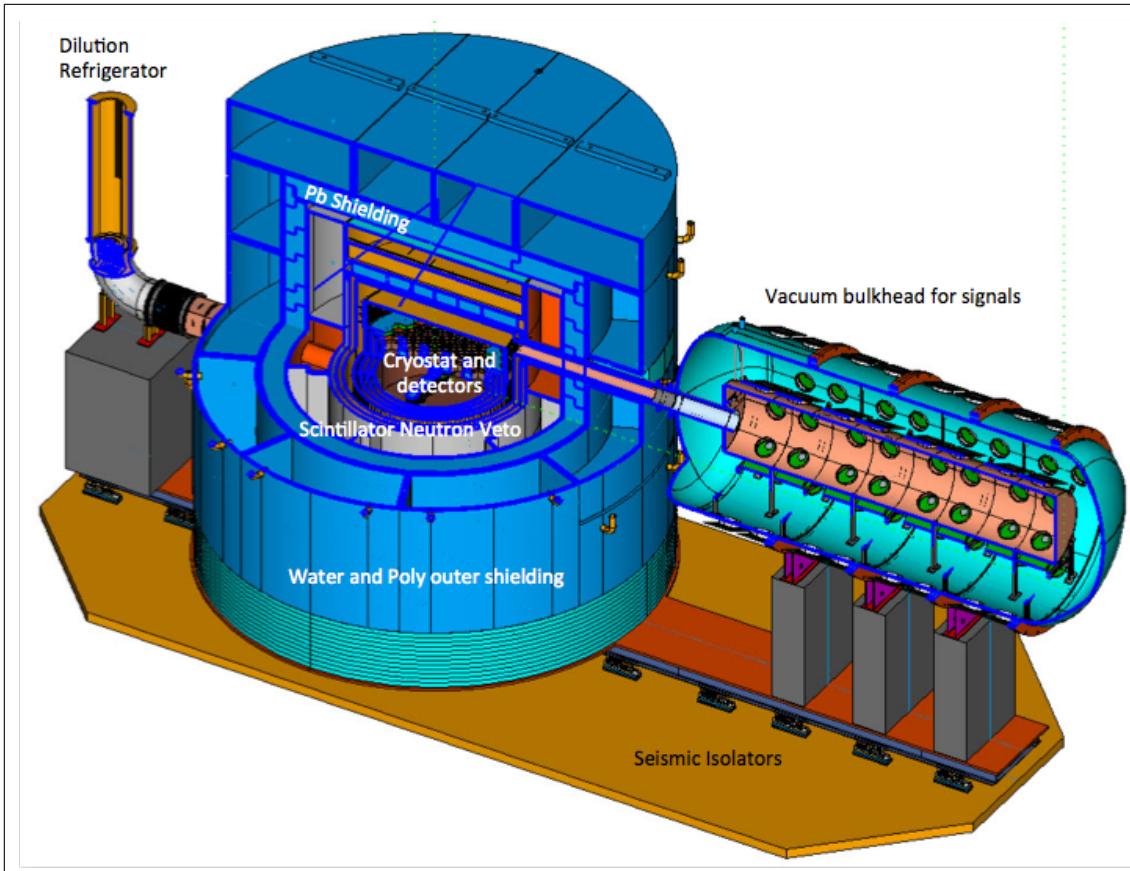


Figure 5.2: The newest CDMS detector, SuperCDMS. Currently in construction.

Although there has been no ‘conclusive’ evidence from these experiments to show that dark matter exists and is formed of WIMPs, there has been conclusive detection of interactions.

In 2013, CDMS II published 3 WIMP detections in silicone crystal, reported at the 3σ level (99.8% confidence). Most recently, the results published in 2014 from SuperCDMS found 11 separate events, for WIMP masses under the 30 GeV range.

5.1.3 Noble Gas and Crystal Scintillators

The general mechanism of a scintillator is to use a ‘scintillating’ material, one which luminesces when excited by ionising radiation, to observe interactions.

Noble gas scintillators consist of enormous containers of an inert gas, most commonly liquid argon or liquid xenon. Crystal scintillators instead use a scintillating crystal, often NaI.

The Dark matter Experiment using Argo Pulse-shape discrimination (DEAP) is currently in its second phase. DEAP-1 used 7kg of liquid argon, but DEAP-3600 has 3600kg. The additional sensitivity allows it to detect scattering cross-sections to

the limit to 10^{-46} cm 2 , for a dark matter particle of mass 100GeV. It is currently collecting data. In collaborations with other argon scintillators, a new detector with over 20 tonnes of argon is planned. An additional detector with capacity for multiple hundred tonnes of argon is in development.

The benefits of using argon as the scintillating material are that it is transparent to its own scintillation light, and that it allows pulse shape discrimination (PSD). The former allows easier detection of ionization events, as the produced photons are not reabsorbed almost immediately. The latter is very important - since the elastic scatter of a WIMP with an argon nucleus is a very low energy reaction (in the range of keV), the detection threshold is very low. This leads to an enormous number of background events being detected. In order to differentiate between dark matter interactions and background events, the timing signature of the produced light is analysed - interaction between various particles and argon creates different proportions of excitation states of argon, which have different lifetimes.

DAMA/LIBRA is a crystal scintillation experiment, which uses NaI(Tl)¹ crystals in a 5 by 5 matrix. Each of these crystals is connected to photomultipliers which record incident photons. The detection matrix is surrounded by nitrogen, sealed in a copper box, and 1m of concrete. All operative parameters are recorded and attempted to be kept constant. Again, we expect different photons to result from dark matter striking a crystal to normal matter.

The experiment ran for 14 years; it confirmed the existence of an annual modulation effect in the data range of 2-6 keV. This would satisfy most dark matter theories, but did not conclusively point to any one theory.

A separate experiment, COSINE-100 attempted to replicate these results in 2018; it concluded that the annual modulation detected by DAMA could not have been WIMP-nucleon interactions. It did not suggest an alternative source.

5.1.4 Bubble Detectors

Bubble detectors function similarly to the bubble chambers you will have seen before; the distinction is that a bubble detector suspends small drops of superheated liquid in a gel, rather than simply having a chamber full of superheated liquid. This allows the detector to stay active for much longer periods of time, as droplets undergo phase transitions (turning into a gas bubble when enough energy has been given to it) independently of each other.

When a droplet becomes a gas bubble, it creates an acoustic wave which can be

¹Thallium-doped Sodium Iodide.

picked up by sensors. Bubble detectors have the unusual benefit of being almost entirely unaffected by background radiation.

The PICASSO experiment is a bubble detector, with the liquid involved being Freon (chlorofluorocarbon, ‘CFC’); in 2013 it merged with COUPP which was using trifluoroiodomethane (CF_3I) to form PICO. Neither PICASSO nor COUPP ever detected any dark matter. Thus far, neither has PICO.

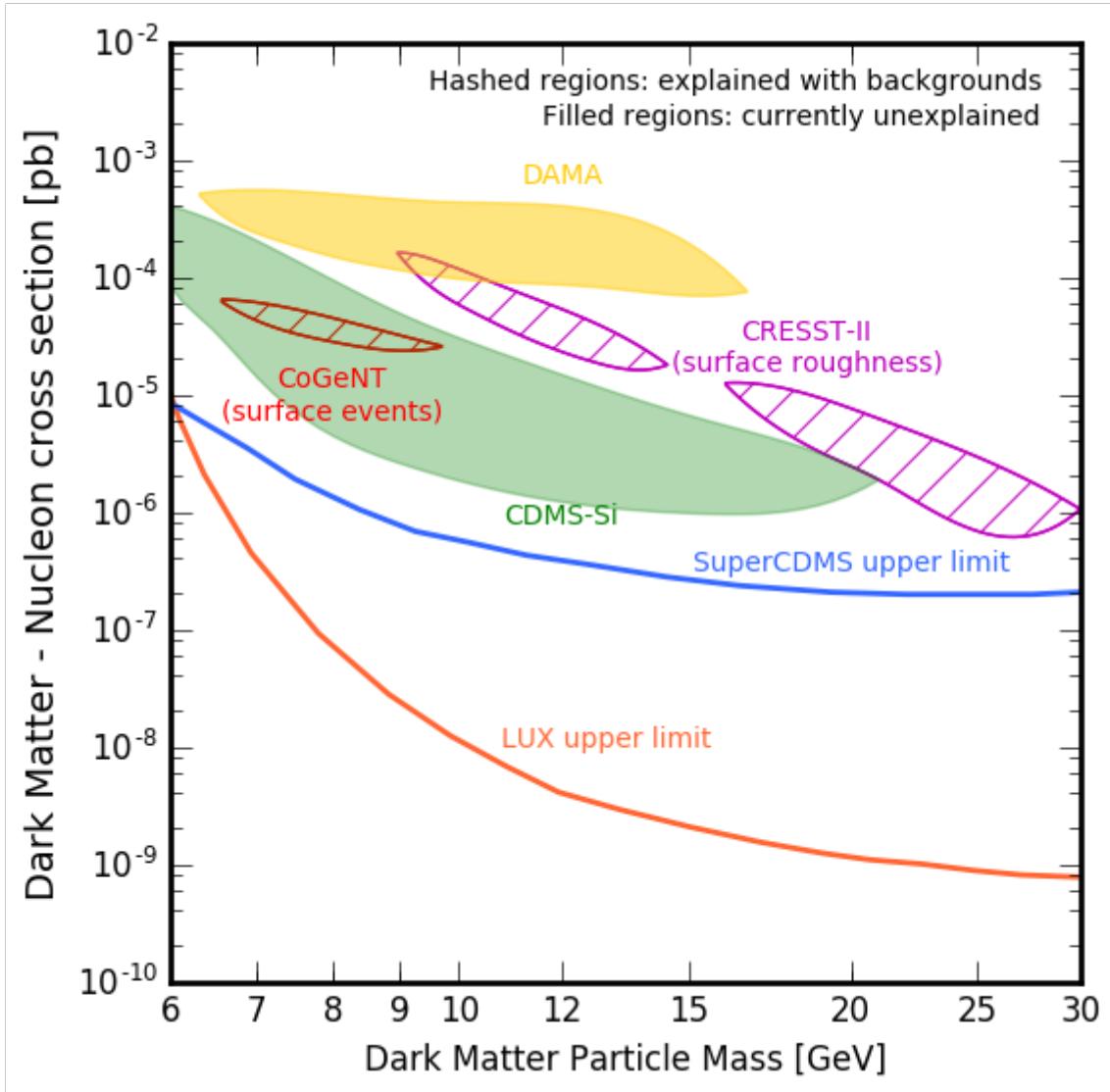


Figure 5.3: The constraints set by various experiments.

5.2 Detection of Axions

The Theory

One way to detect axions is to use the coupling to the electromagnetic field strength which appears in the QCD Lagrangian. For simplicity we will write only the relevant

part of the Lagrangian:

$$\mathcal{L} = \frac{1}{2}(\partial^\mu a \partial_\mu a - m^2 a^2) - \frac{1}{4} \frac{a}{M} F_{\mu\nu} \tilde{F}^{\mu\nu} - \frac{1}{4} F_{\mu\nu} F^{\mu\nu},$$

where $F_{\mu\nu}$ is the electromagnetic stress tensor and $\tilde{F}_{\mu\nu} = \epsilon_{\mu\nu\rho\lambda} F_{\rho\lambda}$ is its dual, a denotes the axion, m is the axion mass, and M is the inverse axion-photon coupling. Because of the $F_{\mu\nu} \tilde{F}^{\mu\nu}$ term, there is a finite probability for the photon to mix with the axion in the presence of a magnetic field. Mixing also occurs between photon components with different polarizations.

Technically, the mixing may be described as follows. We represent the photon field $A(t, x)$ as a superposition of fixed-energy components $A(x)e^{-i\omega t}$. If the magnetic field does not change significantly on the photon wavelength scale and the index of refraction of the medium $|n - 1| \ll 1$, one can decompose the operators in the field equations as (for a photon moving in the z direction) $\omega^2 + \partial_z^2 \rightarrow 2\omega(\omega - i\partial_z)$, so that the field equations become Schrodinger-like,

$$i\partial_z \Psi = -(\omega + \mathcal{M}) \Psi \quad ; \quad \Psi = \begin{pmatrix} A_x \\ A_y \\ a \end{pmatrix}, \quad (5.9)$$

where

$$\mathcal{M} \equiv \begin{pmatrix} \Delta_p & 0 & \Delta_{Mx} \\ 0 & \Delta_p & \Delta_{My} \\ \Delta_{Mx} & \Delta_{My} & \Delta_m \end{pmatrix}.$$

The mixing is determined by the refraction parameter Δ_p , the axion-mass parameter Δ_m , and the mixing parameter Δ_M . These three parameters are equal to

$$\begin{aligned} \Delta_{Mi} &= \frac{B_i}{2M} = 540 \left(\frac{B_i}{1 \text{ G}} \right) \left(\frac{10^{10} \text{ GeV}}{M} \right) \text{ pc}^{-1}, \\ \Delta_m &= \frac{m^2}{2\omega} = 7.8 \times 10^{-11} \left(\frac{m}{10^{-7} \text{ eV}} \right)^2 \left(\frac{10^{19} \text{ eV}}{\omega} \right) \text{ pc}^{-1}, \\ \Delta_p &= \frac{\omega_p^2}{2\omega} = 1.1 \times 10^{-6} \left(\frac{n_e}{10^{11} \text{ cm}^{-3}} \right) \left(\frac{10^{19} \text{ eV}}{\omega} \right) \text{ pc}^{-1}, \end{aligned}$$

respectively. Here $\omega_p^2 = 4\pi\alpha n_e/m_e$ is the plasma frequency squared (effective photon mass squared), n_e is the electron density, B_i , $i = x, y$ are the components of the magnetic field B , m_e is the electron mass, α is the fine-structure constant and ω is the photon (axion) energy.

For constant magnetic field and electron density, the conversion probability is

$$P = \frac{4\Delta_M^2}{(\Delta_p - \Delta_m)^2 + 4\Delta_M^2} \sin^2 \left(\frac{1}{2} L \Delta_{\text{osc}} \right),$$

where

$$\Delta_{\text{osc}}^2 = (\Delta_p - \Delta_m)^2 + 4\Delta_M^2$$

and we assumed that imaginary parts of all ‘ Δ ’s can be neglected. If B and n_e change spatially, the probability can be found by a numerical solution of (eq. (5.9)). The condition for strong mixing is

$$4\Delta_M^2 \gg (\Delta_p - \Delta_m)^2. \quad (5.10)$$

Other maximal-mixing conditions, which also must be met, are

$$\Delta_m \ll 2\Delta_M,$$

and

$$\Delta_p \ll 2\Delta_M,$$

which are equivalent to

$$\omega \gg 70 \text{ eV} \left(\frac{m}{10^{-9} \text{ eV}} \right)^2 \left(\frac{B}{\text{G}} \right)^{-1} \left(\frac{M}{10^{10} \text{ GeV}} \right), \quad (5.11)$$

$$n_e \ll 10^{20} \text{ cm}^{-3} \left(\frac{\omega}{10^{19} \text{ eV}} \right) \left(\frac{B}{\text{G}} \right). \quad (5.12)$$

In addition, to have large mixing one should require that the size L of the region in which the conditions in equations 5.11 and 5.12 . are fulfilled should exceed the oscillation length,

$$L \gtrsim \frac{\pi}{\Delta_{\text{osc}}},$$

that is

$$L \gtrsim 5.8 \times 10^{-3} \text{ pc} \left(\frac{B}{\text{G}} \right)^{-1} \left(\frac{M}{10^{10} \text{ GeV}} \right). \quad (5.13)$$

So that axions in a magnetic field can actually mix together with photons. This means that in principle, one can convert photons into axions in a magnetic field, leading to the concept of shining light through walls experiments. This is where one (tries) to convert photons into axions in a magnetic field, then send the beam of photons and axions into a wall. Only the axions will penetrate the wall, then you put a magnetic field on the other side of the wall and try to convert the axions back

into photons on the other side.

Axions can also affect many astrophysical systems, they can carry away too much energy during supernova explosions etc. We expect axions to be produced in the Sun, in which case, one can detect them on Earth using experiments like CAST - the CERN axion Solar Telescope.

One can also look for dark matter axions converting back into microwaves in resonant cavities but this has to be extremely resonant, so that one can only look for one dark matter mass at a time.

The Experiment

The Axion Dark Matter eXperiment² (ADMX) is an ongoing experiment searching for dark matter axions in the local galactic dark matter halo. It is one of the only dark matter detectors which is not located deep underground, or in a remote location³. It is a supersized version of the axion haloscope invented in 1983. ADMX has been operational since 1995, moved location in 2010, and is currently having upgrades to increase its sensitivity across a broader range of masses.

The method is based on dark matter axions being converted into photons under a strong magnetic field. The conversion has a very low energy, and result in photons of the microwave wavelength. The microwave cavity of the experiment is designed to mimic this; if the cavity's resonant frequency is tuned to the axion mass, the interaction between axions in the galactic halo should be enhanced, which will create a very tiny amount of power in the cavity. A microwave receiver then records the axion signal. The setup is cooled to around 4K, and the receiver can detect powers at the 10^{-26} Watt scale.

The two theorised mass ranges for dark matter axions are 1.9-3.53 μeV and 2-20 μeV . ADMX has eliminated the former range.

5.3 Detection of Primordial Black Holes

Since primordial black holes are MACHOs, their detection is vastly different from that of other dark matter candidates. Additionally, we are theoretically able not only to detect them, but also to determine their masses.

Hawking radiation is our primary method of detecting primordial black holes. Since Hawking radiation is inversely proportional to mass, primordial black holes (which,

²Not a typo; annoyingly the entire published data uses ‘eXperiment’ throughout.

³It is based in Washington.

recall, can be much lighter than ordinary black holes) would radiate intense bursts of gamma radiation towards their end.

Small, massive objects can affect visible object through gravitational lensing; gamma-rays passing by primordial black holes should have an altered luminosity.

NASA's Fermi Gamma-Ray Burst Monitor has been observing gamma-ray bursts, and has placed constraints on primordial black holes: they cannot contribute importantly to dark matter in the mass range 10^{14} to 10^{17} kg.

We can further place constraints on primordial black holes by considering the abundance of neutron stars in globular clusters; primordial black holes with masses of between 10^{15} and 10^{22} are of the right size and age that they could be ‘captured’ by neutron stars. By considering the abundance of neutron stars, and comparing it to theory, we can estimate how many may have captured primordial black holes, and been destroyed.

The final way we can detect primordial black holes is through gravitational microlensing (the effect of something massive passing between a distant shining object and an observer, causing gravitational lensing of the light).

We have constrained the abundance of primordial black holes between the masses of 10^{23} and 10^{31} kg by considering the microlensing of stars in the Magellanic Clouds⁴. It should be noted though that this assumes a model of roughly equally spaced out primordial black holes, whereas models which involve halos of black holes would not heed these same constraints.

It should also be possible to consider the microlensing of supernovae due to primordial black holes (for example Type 1a supernovae, since they have a standard brightness), but this has yet to be done experimentally.

5.4 Detection of Sterile Neutrinos

There are many experiments ongoing which are attempting to find evidence of sterile neutrinos. So far, the strongest evidence for them comes from the MiniBooNE Experiment which is observing neutrino oscillations. In 2018, their findings were published, suggesting that some muon neutrinos switch to sterile neutrinos briefly, before switching to electron neutrinos.

The MiniBooNE experiment is an off-shoot of the BooNE experiment (Booster

⁴Look up EROS survey for further information.

Neutrino Experiment). BooNE's purpose is to establish the neutrino masses through neutrino oscillation. MiniBooNE in particular was designed to confirm or refute the results of the LSND (Liquid Scintillator Neutrino Detector) experiment which ran between 1993 and 1998.

The MiniBooNE detector consists of a muon neutrino beam, produced using 8 GeV protons from the Fermilab booster. The beam can also produce anti-neutrinos. The neutrinos pass through solid earth for \sim 500 m, before reaching the spherical 12m detector.

The detector records the number of electron neutrinos which arrive; this is then compared to the number which should arrive coincidentally from background sources, with the number that may have oscillated (via a sterile neutrino) from a muon neutrino from the beam.

In combination with data LSND, MiniBooNE stated the likelihood for the existence of sterile neutrinos as 6.1σ , or 99.999999% accuracy.

However, even if these results are accurate⁵, we still have a long way to go to understand this potential fourth neutrino. For example, the CMB places a constraint on the number of neutrino species around in the early universe - 3. A further experiment, MicroBooNE, is planned.

⁵ And we do have serious reasons to want further experiments; there have been a good number of experiments which took place between LSND and MiniBooNE which disagreed with their results.

Chapter 6

Dark Energy

6.1 Late Universe Acceleration

[Perlmutter et al. \(1997\)](#) and [Riess et al. \(1998\)](#) provided independent evidence, from studies of type Ia supernovae, that the expansion of the universe *today* is accelerating. We now call this phenomenon *late universe acceleration*, and the topic of this section is our attempt to explain it.

We expect that the large scale development of the universe will be dominated by gravity, as the other forces are weaker on this scale. And, in fact, from the naive Friedmann equation, using only matter and radiation and curvature, we expect a matter-dominated universe to *decelerate* its expansion due to these forces. But observations show that the universe's expansion is not decelerating, it is instead accelerating; $\ddot{a} > 0$. We are missing an ingredient in our equations, or General Relativity is simply wrong.

The earliest observations used type Ia supernovae as standardizable candles (see section 2.1). It was found that they were 0.25 mag dimmer than the expectation in a decelerating universe. This indicated that the universe had an accelerated expansion for the past 5 G yr. A summary of the results are given in fig. 6.1. There is a lot more detail about the evidence to be found in [Frieman \(2008\)](#), including recent developments, CMB and BAO contributions to the dark energy search, error measurements, and other corroborating evidence. These are too diverse and technical for our purposes.

In terms of the missing ingredients or GR failure mentioned above, our options are threefold. First, we may be missing some important *fact about geometry* that isn't captured by the Einstein equation at present, something that manifests only on cosmological scales. This way, we can retain the correctness of GR on the scales we have tested it so far, but nevertheless maintain that it is mistaken on the whole;

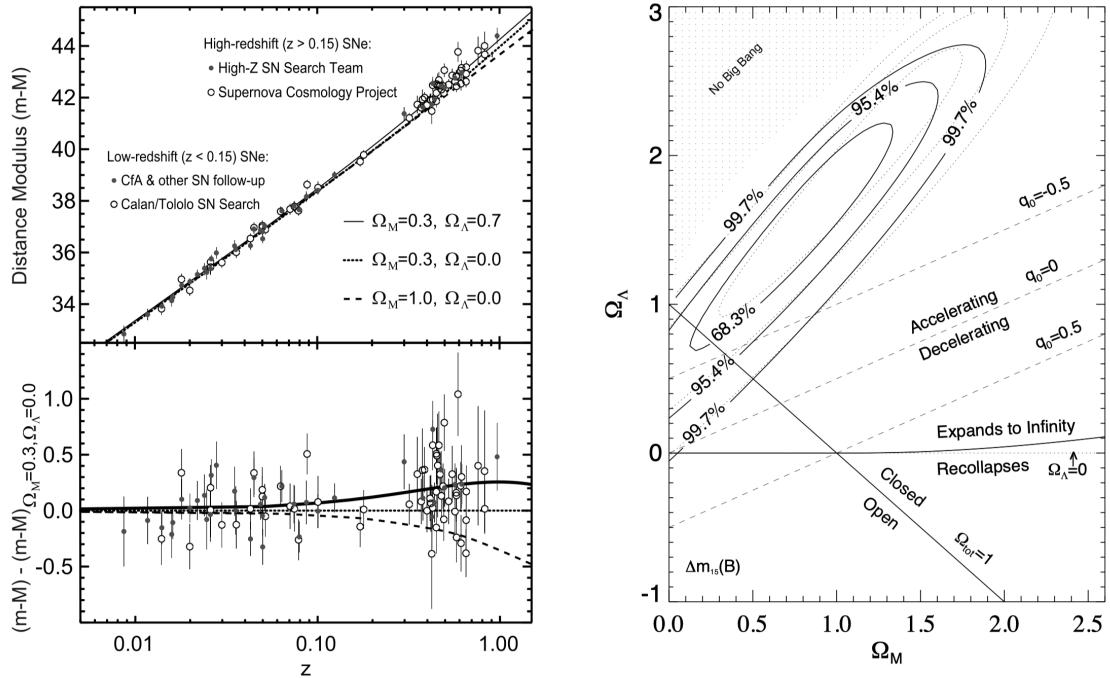


Figure 6.1: Figure and caption from [Frieman \(2008\)](#). **Left panel (a):** Discovery data: Hubble diagram of SNe Ia measured by the SCP and HZT. Bottom panel shows residuals in distance modulus relative to an open universe with $\Omega_0 = \Omega_m = 0.3$. **Right panel (b):** constraints on Ω_m and Ω_Λ from the HZT data

much like Newton’s theory is true on the scale of people and cars, but not on the scale of electrons or galaxies. This is quite far fetched given the enormous success of GR, but not entirely ruled out since we know that GR has shortcomings when it comes to describing microphysics and black holes.

Second, we may be missing some important *diversity* that our models have filtered out by their assumptions of homogeneity and isotropy, something about the large scale structure that causes the expansion. This is much harder to theorize than a geometric alteration of GR, and much harder to test, since it will rely on us finding enough new diversity to account for the expansion without violating the well-established evidence from the CMB that the universe is mostly isotropic. Such a solution would have to explain why the universe looks so damn isotropic and homogeneous when, on its view, it really isn’t.

Finally, we may be missing an important *component* that counter-acts the tendencies of gravity — a repulsive stress-energy that we call *dark energy* simply because we have no clear idea what it *is*. This is the simplest solution, so we will spend this chapter exploring it in more depth. Much of the chapter will be qualitative, as there is simply not enough space or time to go into much depth on the Quantum

Field Theoretic origins of much that will be said here. We can nonetheless get a good intuition about the problems that dark energy solves, as well as the problems it creates. The simplest place to start is with the cosmological constant from eq. (1.49).

6.2 The Cosmological Constant

Recall from section 1.4 that the density parameter for the cosmological constant is $\Omega_\Lambda \approx 0.7$. In this section we will explore the mystery of this number, which is quite hard to state, and even harder to solve.

We begin with the Einstein equation:

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi GT_{\mu\nu} \equiv G_{\mu\nu} \quad (6.1)$$

where we are working in 3+1 dimensions.¹ In General Relativity the mathematical expression of the conservation of energy is

$$\begin{aligned} \nabla^\mu T_{\mu\nu} &= 0 \\ \Rightarrow \nabla^\mu \left(R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R \right) &= 0 \end{aligned} \quad (6.2)$$

where the $T_{\mu\nu}$ is the stress-energy tensor, here taken to describe an ideal fluid in accordance with the cosmological principle; that is, a fluid that is homogeneous and isotropic. We model the fluid as

$$T_{\mu\nu} = (\rho + P)u_\mu u_\nu + P g_{\mu\nu} \quad (6.3)$$

where we assume that there is no position-dependence, since the fluid is isotropic. Thus we get (in co-moving coordinates) that

$$T_{\mu\nu} = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & & & \\ 0 & & a^2 \gamma_{ij} P & \\ 0 & & & \end{pmatrix} \quad (6.4)$$

since $u^\mu = (1, 0, 0, 0)$. The time-time component then yields

¹This is a good time to ask yourself if anything interesting happens by increasing the numbers of dimensions, analogously to how the quantum mechanical wavefunction lives in a $3N$ dimensional configuration space, rather than in $3 + 1$ dimensional space.

$$\begin{aligned} G_{00} &= 8\pi GT_{00} = 8\pi G\rho = R_{00} - \frac{1}{2}g_{00}R \\ &= 3\left(\frac{\dot{a}}{a}\right)^2 + \frac{3K}{a^2} \end{aligned} \quad (6.5)$$

and thus

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{K}{a^2} \quad (6.6)$$

which is, of course, just the good old Friedmann equation, with $c \equiv 1$ and $K = k/R^2$. The diagonal space-space components each give the same equation (homework):

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P) \quad (6.7)$$

This is the equation of motion for the scale factor, since it is a second derivative in time; the original Friedmann equation actually acts more like a constraint on this equation (though this will not matter much for our purposes).

The continuity equation is

$$\frac{\dot{\rho}}{\rho} = -3(1+w)\frac{\dot{a}}{a} \quad (1.28 \text{ revisited})$$

which has its solution as

$$\rho \propto a^{-3(1+w)} \quad (6.8)$$

as we have seen in chapter 1. But what we saw earlier as a mere arbitrary energy shift in the Friedmann equation is now simply the natural integrating factor that results from eqs. (6.1) and (6.2):

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + \Lambda g_{\mu\nu} = 8\pi GT_{\mu\nu} = G_{\mu\nu} \quad (6.9)$$

This is our cosmological constant from chapter 1, this time derived (somewhat hastily) from general relativistic principles, rather than an ad-hoc quasi-classical addition to the Friedmann equation to fit our experimental data. In the Friedmann equation we know it, of course, as

$$H^2 = \frac{8\pi G}{3c^2}\rho + \frac{\Lambda}{3} - \frac{kc^2}{R^2a^2} \quad (1.49 \text{ revisited})$$

But last time we did not have the space-space equations (homework):

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + \frac{3P}{c^2}) + \frac{\Lambda c^2}{3} \quad (6.10)$$

Einstein originally introduced Λ to allow for a static universe, where there was neither expansion nor contraction of space as time passes; i.e. where $\dot{a} = 0$. Before we go on, we need to allow ourselves the use of natural units, otherwise the algebra is just going to be too annoying. So we let $c \equiv \hbar \equiv 1$, and for good measure let $R = 1$ as well. Let us also note that we are (for now) interested only in a matter-dominated universe (which ours is), in which $\rho_m = \rho$ and $P_m = 0$. In that case:

$$\begin{aligned} 0 &= \frac{8\pi G}{3}\rho_m + \frac{\Lambda}{3} - \frac{K}{a^2} \\ 0 &= -\frac{4\pi G}{3}\rho_m + \frac{\Lambda}{3} \\ \Rightarrow K &= 4\pi G\rho_m a^2 \\ \Rightarrow \Lambda &= 4\pi G\rho_m \end{aligned}$$

And thus there are static solutions for a , so long as we thereby fix the cosmological constant and the curvature to very specific numbers, determined by ρ_m . Of course, since we know that $|K| \sim 0.01$, this gives us very little space to breathe in terms of these two values. Indeed, even the slightest perturbations of ρ_m (which, recall, is determined by dynamical processes in the early universe) would result in vastly different values of Λ and K . And at any rate we now have evidence that $\dot{a} \neq 0$, so this was just wrong.

We have already seen in the introduction that we can treat the cosmological constant (and the curvature of space, for that matter) as a fluid in its own right, with equation of state

$$P_\Lambda = -\rho_\Lambda \quad ; \quad w_\Lambda = -1 \quad (1.24 \text{ revisited})$$

It is when we do this that we elevate it to dark energy: a cosmological constant on its own is just a constant; a cosmological constant that drives expansion is a dark energy candidate. But you might think that it's weird to treat a constant as a fluid (or a curvature as a fluid). Many other cosmologists do as well.

There are primarily two problems: first, the constant and its particular value seem too fine tuned to come out of nowhere, but all our candidates for its origin are miserably inadequate; second, the constant just so happens to be of the same order ($0.7 \sim 0.3$) as the matter-content of our universe, despite not diluting at all during the expansion ($\rho_\Lambda \sim \text{constant}; \rho_m \sim a^{-3}$), and this seems like too much of a perfect coincidence (had we only started our cosmological search a little earlier or a little

later, the orders would be wildly different). We will discuss these problems in turn.

6.2.1 The Constant Problem

Dark energy is really not mysterious at all when seen from the perspective of classical field theory, where the addition of a constant that leaves the equation of motion invariant is simply a mathematical trick born of the nature of the integral. The problem is that, under the regime of general relativity, “space,” as Wheeler said, “tells matter how to move [and] matter tells space how to curve”. That leaves out what kind of thing dark energy is, and how it fits into the interplay between space (or spacetime) and matter (or energy). To really understand dark energy we need an account of its origin. That is, we need to find out what the cosmological constant *is*.

We want to know what a constant fluid that is both isotropic and homogeneous might be. A decent start, since we don’t have much else to go from, is to use a similar concept from another domain: vacuum energy. [Zel’dovich et al. \(1968\)](#) first pointed out that the cosmological constant belongs to the stress-energy side of the Einstein equation, and not to the geometric side as we had it above:

$$G_{\mu\nu} - \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R \quad (6.11)$$

With this in order, recall that in Quantum Mechanics there is a nonzero floor to the lowest energy state of some systems, such as the harmonic oscillator; in Quantum Field Theory there is a constant background vacuum energy that permeates all of space. Perhaps this energy *is* dark energy. If it is, then we should be able to learn things about dark energy from QFT; or, at the very least, we should get similar values from QFT and GR. The *constant problem* is that we *don’t*.

In fact, we *really* don’t. The observed vacuum energy density (i.e. dark energy density) is

$$\rho_\Lambda = 3.7 \times 10^9 \text{ eV m}^{-3} \quad (6.12)$$

But it’s much more conveniently to set $c = \hbar = 1$ such that

$$\rho_\Lambda \approx (10^{-3} \text{ eV})^4 \quad (6.13)$$

These are *natural units*. In these, the value of the vacuum energy density is pretty close to zero, but nonetheless (and importantly) nonzero.

There are many QFT variants to compare with, but each has a ground state energy term which represents the vacuum energy of the field in question (weak, strong, electromagnetic). The major difference between QFT and QM lies that

integrals that in QM converge are generally divergent in QFT unless we *renormalize* them: we simply assume that we can cut off the integral at some energy-level at which we know for sure that the theory reproduces the correct results. Of course, this is quite circular (we need the method to prove the method), but it is incredibly predictively powerful, so we ignore those issues in our experimentation and fixing of constants and so on. QFT is an effective theory despite being quite obviously incomplete.

To calculate the vacuum energy of QFT we need to sum over the zero-point energies all the different fields of the Standard Model:

$$\rho_{\text{vac}} = \frac{1}{2} \sum_{\text{fields}} g_i \int_0^\infty \sqrt{k^2 + m^2} \frac{d^3 k}{(2\pi)^3} \simeq \sum_{\text{fields}} \frac{g_i k_{\max}^4}{16\pi^2}, \quad (6.14)$$

where g_i is the degree of freedom of the field, positive for bosons and negative for fermions; while the sum runs over all the fields — quarks, leptons, gauge fields, etc. In order to fit this with the expected (from cosmology) value, at most at the critical density, we need $k_{\max} < 0.01 \text{ eV}$.

But the most intuitive cut-offs are given either by the highest energies tested at the LHC: $M_{UV} = 1 \text{ TeV}$ in case we find evidence for a Supersymmetric model (SUSY)²; or, if we don't find such evidence, by the Planck scale: $10 \times 10^{19} \text{ GeV}$. The intuition for the LHC cut-off is that it represents the highest energy scale at which we have tested QFT without observing any contradictions, so it's safe to assume it still holds up to this limit, but it *might* break beyond it. The intuition for the Planck scale is that it seems reasonable that QFT will hold even further than we have tested, but that there is a limit due to quantization, most naturally assumed to lie with our familiar Planck constant. But with *these* as our cut-offs, the energy-density predicted by QFT is ([Velten et al. 2014](#))

$$\begin{aligned} \rho_{\text{Planck}} &= 10^{120} \rho_\Lambda \\ \rho_{\text{SUSY}} &= 10^{60} \rho_\Lambda \end{aligned} \quad (6.15)$$

which is, to put it very gently, quite far from the value we want. And it simply will not work to cut off QFT at a lower energy, since we *know* it works up to at least the LHC values. So whatever dark energy is, it is not the naive QFT vacuum energy — though it may yet turn out to be some modified QFT vacuum energy, in an extended Standard Model theory, for instance.

Perhaps the problem is overblown, though, by the fact that $\rho_{\text{vac}} \sim [\text{energy}]^4$. But if we translate to a mass scale ([Carroll 2001](#)), then our new estimates get only a

²We will not go into any details of SUSY models here.

slight improvement:

$$\begin{aligned} M_{\text{Planck}} &= 10^{30} M_\Lambda \\ M_{\text{SUSY}} &= 10^{15} M_\Lambda \end{aligned} \tag{6.16}$$

In other words, we still have a giant problem. It is still possible to fine-tune the constant by shifting the QFT prediction by an arbitrary constant — QFT measurements only care about *differences* in energy, after all. But that isn't very helpful, since the correction is plainly arbitrary, unless we have a better way of determining the number than by just *fixing* it according to $\rho_{\text{fine-tuning constant}} = \rho_\Lambda - \rho_{\text{QFT}}$. Just plugging this value in is emphatically not a solution to our problem (to find out what constitutes dark energy), since that merely postpones the problem to what in all the world the fine-tuning constant is.

6.2.2 The Coincidence Problem

From the above results we already know enough to see that

$$\begin{aligned} \frac{\rho_{\Lambda,0}}{\rho_{\text{dm},0}} &\sim \mathcal{O}(1), \\ z_{\Lambda,\text{eq}} &\sim 0.5 \end{aligned} \tag{6.17}$$

where the value of z_Λ is here a very rough estimate (for now we are only interested in its order). It appears to indicate that matter and the cosmological constant just recently entered a period of approximate equality. But *why now?* Matter's energy density decreases by a^{-3} while the cosmological constant is—well it's in the name, isn't it. Time-dependent variables (so long as they are monotonic) do not stay for long within an order of a constant.

We do not have any reason to think that there should be a matter- Λ equality right now or recently in the past (or, for that matter, soon in the future). Why do we live in times like these? At most other times in the evolution of the universe, the matter- Λ proportions were radically different; at most times in the future (if our math is right) they will become very different again.

The *easiest* answer is: well why *shouldn't* we? Coincidences occur. Why can't this be one too?

Well, here are some potential reasons to suspect that it's *too much* of a coincidence. First, it starkly contrasts to the much older epoch-changes we observe between other components of the Friedmann equation: last scattering at $z \sim 1000$, matter-radiation at $z \sim 5000$, and nucleosynthesis at $z \sim 10^{10}$. Matter- Λ is the only recent equality. Why does it stick out? But then again: why not?

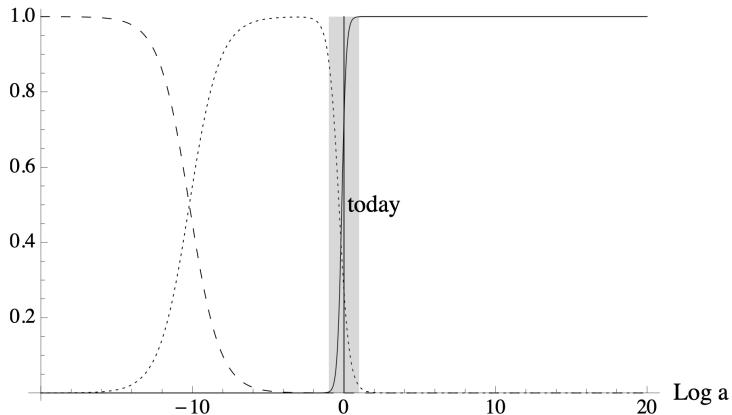


Figure 6.2: Dotted: Ω_m ; dashed: Ω_r ; solid: Ω_Λ . Figure from [Sivanandam \(2013\)](#).

Second, we should expect our time to be like all other times. Indeed, that is very much the ethos of the cosmological principle, which seems to be that there is nothing special about our perspective. Of course, the cosmological principle does not say anything about equality in time, so we need to be careful with our extrapolations. Nevertheless, it is strange that there should be something so special about *our* time: it indicates that we may be awaiting another Copernican twist in cosmology, that can reveal yet again how we are not special.

Let's take a step back and review what we actually mean by the term *coincidence*. Do we mean *unlikely*? Or maybe we mean that our era is somehow *special*? If so, we need to tread carefully. What notion of likelihood or specialness are we using?³

Let's first consider whether it is sensible to say that the order of the cosmological constant is *unlikely*. What could this mean? Likelihoods are usually understood either as objective or as subjective. If they are objective, we rely on a frequency or a disposition in order to produce expectations or predictions: we count the successes and divide them by the total number of trials to get a frequency or rate; or we derive a set of dispositions or predictions from the equation that governs the evolution of some object (the wavefunction-squared of a fermion, say). Neither can be right for our purposes. There is only one universe and thus no more than one trial, so the frequency view fails. Moreover, we do not know the precise initial conditions of the universe nor the equations that govern them (we do not have a Theory of Everything), so we cannot derive a disposition either. There is nothing *objectively unlikely* about the order of the cosmological constant.

If the likelihoods are subjective, we run into some interesting issues regarding Bayesian probabilities. Subjective probabilities are given by three things: first, a

³For a deeper dive into the nature of probability, and the problems it presents, see Laurence Sklar's second chapter of the excellent *Physics and Chance* (1995).

prior complete set of credences; second, an adherence to the fundamental axioms of probability (the Kolmogorov axioms); third, an updating of credences according to Bayes Theorem. (See [Sklar \(1995\)](#) for details.) The idea is that there is no way to beat the predictions of someone who follows these rules. As such, while they are subjective probabilities, they will come to track the way the world is given enough new and repeated exposure to the evidence; everyone should converge on the same credences for the same beliefs. The problem is that the idea of something being unlikely then comes to depend on subjective choices of priors. Someone might give the coincidence of similar parameter-orders a low prior, and they will be surprised to see what we now see. But someone who sets their priors high or simply to 50-50, should not really be very surprised at all. And we have no independent way of adjudicating which priors are better, since we have no other universes on which to base our priors.

It appears that we cannot argue convincingly that it is *unlikely* that the matter- Λ equality was recent. Can it still be a coincidence in some nefarious way? There is a chance that, while not unlikely, our time is still *special* in the evolutionary history of the universe up until today; this may call for an explanation of why we live in such special times, and not instead in more ordinary times. Note that the problem here is *not* that we *expect* to be normal (that would be a probabilistic reading of coincidence). It is instead a *curiosity with why* we are *not* normal (when, by definition, most of the co-moving hyperslices of our universe are). On *this* reading, there is a coincidence problem only for those who are curious. But we are phycisists, so we better be.

The literature is riddled with discussions about this issue. Almost none of it makes the distinctions that we have made, between the likelihood problem and the specialness problem. But it is important to note that the literature gets around this only by insisting on a very strange assumption, as best described by [Lineweaver and Egan \(2007\)](#): “Under the assumption that observers could exist at any time during the history of the Universe, this large number coincidence could not be explained in the standard cosmology”. Observers could not exist at any time during the history of the universe: the Big Bang was too hot for it, and for a long time after the conditions simply were not right. There is a pocket of time in which it is possible for observers to exist, and it includes the time it takes for observers to evolve all the way from clumps of stardust into semi-intelligent hominids with good enough eyesight to study the heavens. We should expect to live in a certain kind of era of the universe, to the exclusion of many, perhaps most, others. This argument, the one made in this paragraph, is called the *anthropic principle*, best summarized by [Vilenkin \(1995\)](#): “We should observe a universe that is typical amongst those containing observers”. Many scientists dismiss this kind of thinking because it sounds unscientific; it is not,

they say, grounded in testable evidence. You may be sympathetic, and there is room for argument here. Egan and Lineweaver (2007; 2008) have argued that we live in approximately the kind of period we should expect to live in; Sivanandam (2013) asses their argument. For a discussion of fine-tuning in general, see Friederich (2018).

Finally, we should note that even if we agree with the anthropic principle, we still need to answer the specialness problem regarding the timing of the matter- Λ equality: of all the times at which universe could contain observers, why do we find ourselves observing the matter- Λ equality so recently, when it isn't obviously connected to the possibility of there being observers? After all, we might as well be living in the distant future, separated by much greater distances from this epoch; life will still be possible then (we think). Why now, why not much later? This is a hard question.

Or maybe it's just an artefact of selection effects. Let's change the variables from redshift to time, and compare the epochs in that way instead. The result (fig. 6.3) is summarized by Velten et al. (2014):

The dark-energy dominated epoch started 3.5 Gyrs ago. This is about 1/4 of the Universe's lifetime. Therefore, having these numbers in mind, the dark-energy epoch did not start so recently as the redshift parametrization suggests.

The important thing to notice in fig. 6.3 is that the redshift parameter variation itself changes with time. That is, it is not a linear function of time at all, not even approximately. So to say that something is “recent” in redshift has a very different meaning from being recent in time, even on cosmological scales. Viewed from the perspective of time, our epoch seems far less special than when viewed from the redshift parameter.

Nevertheless, there is an interesting perspective left to consider. We can estimate (Velten et al. 2014) the “minimum” age of the universe by looking at known globular clusters. These are ~ 10 G yr old. An Einstein-de-Sitter universe, with $\Omega_\Lambda = 0$ has an age of ~ 9 G yr, and so is effectively ruled out by these observations. These are not very precise estimates, but they are effective at ruling out extreme values for Ω_Λ , as shown in fig. 6.4.

Moreover, we can ask about the relation between the cosmological constant and the possibility of structure formation in the universe. We can define a matter-density contrast $\Delta = \delta\rho_m/\rho_m$, where $\delta\rho_m$ is the inhomogeneous density perturbations and ρ_m is background density; in essence, Δ is a measure of how much inhomogeneity we observe in an otherwise mostly homogeneous space. It is a good way to estimate the inhomogeneity we need for structure-formation. For a small Δ with fluid speed of sound c_s and wavenumber k_n , we have

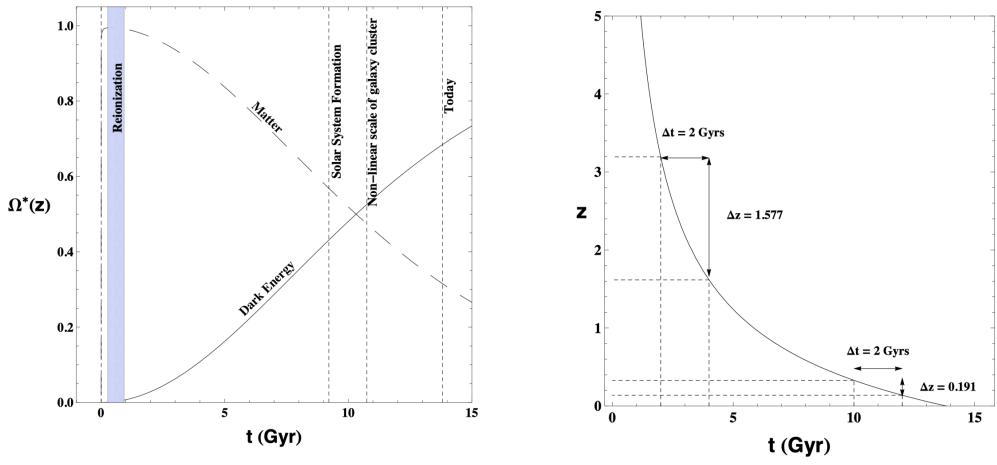


Figure 6.3: **Left:** Ω_s as a function of the cosmic time; **right:** redshift variation as a function of cosmic time. Figure from [Velten et al. \(2014\)](#).

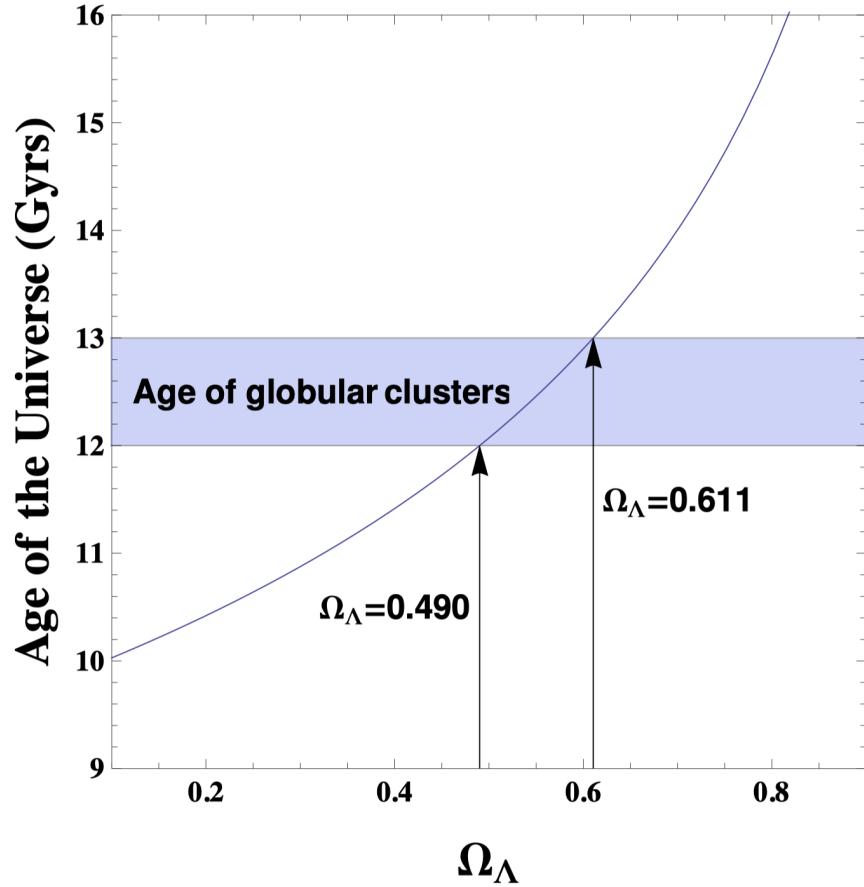


Figure 6.4: $\Omega_\Lambda(t)$ plotted against the age of known globular clusters.

$$\ddot{\Delta} + 2H\dot{\Delta} + \left(\frac{c_s^2 k_n^2}{a^2} - 4\pi G\rho \right) \Delta = 0 \quad (6.18)$$

where we suppose $k = 0$, i.e. the universe is flat. This breaks down when $\Delta \approx 1$, since we would then not be able to use a linear approximation. The problem is that the observed structures in the universe are highly *non-linear*. Thus we can do a little backflip reasoning to weed out some previously good-looking alternatives. We ask: what parameters for dark energy, baryonic mass, curvature, and so on can lead to non-linear solutions, i.e. those where $\Delta \approx 1$? Only these can be candidates for a model of our universe. Thus, from the innocent assumptions of flat curvature and 0.05 baryonic density today, we can get bounds for dark energy, *as a function of the scale factor*.

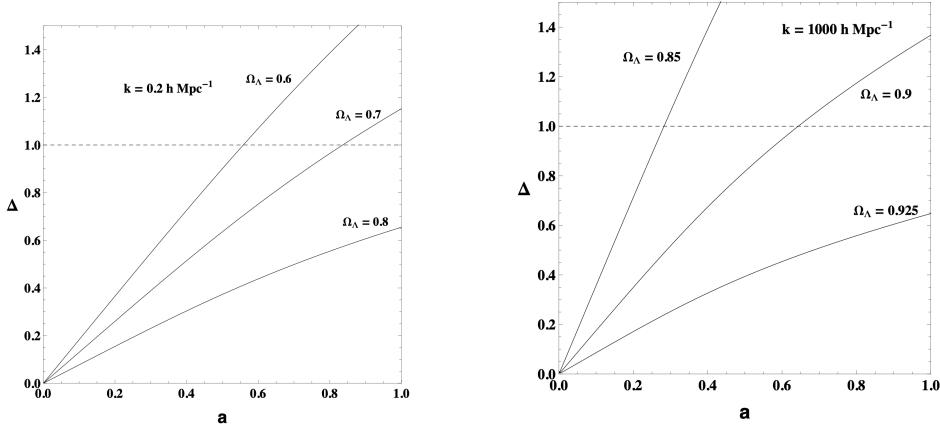


Figure 6.5: Dark matter perturbation growth as a function of the scale factor, in proto-galaxies and galaxy-clusters respectively. The dashed line represents $\Delta \approx 1$, i.e. the nonlinear regime (Velten et al. 2014).

From these considerations we can see that there are upper bounds on Ω_Λ as well: if it is too large, it will not lead to sufficient structure formation. Therefore, we ought to expect a cosmological constant squashed in between the considerations of structure formation and those of globular clusters from above.

Is there a coincidence problem? Cosmologists do not agree. The opinions expressed in these notes are controversial, but right. (If we thought they were wrong, there wouldn't be much point in stating them.) Our time is not as special as it at first seems. A little, maybe, and as curious physicists we should try to find out if there is a connection between the value of the cosmological constant and the possibility of life evolving to the point of being able to observe such intricate star-stuff. But there is no point at any stage to haggle over likelihoods or expectations. The coincidence problem is interesting even if the anthropic principle is right.

The simplest explanation of late universe acceleration is to treat the cosmological constant as the stress-energy of a fluid, dark energy. But once we do, we end up with two puzzling problems: why is the value of the cosmological constant so small,

but nonzero? why does it not even remotely match our predictions from QFT? why do we live in early aftermath of the matter- Λ equality? These are puzzling questions. But perhaps most pressing is the even simpler worry: what is this strange fluid made of? how does it fit into the set of ingredients to our fundamental theories? why does it evolve so differently from them? We have no clue.

6.3 Other Alternatives

Given the mystery remaining with even the simplest solution to late universe acceleration, it is no wonder that we have seen many attempts at more complex but nuanced solutions. These require detours into recesses mathematics that are too finicky for us to deal with here, so we will instead treat them qualitatively, with only a little peek at the relevant mathematics.

6.3.1 Light Scalar Fields and Quintessence

An alternative solution to the cosmological constant as dark energy is the idea that the universe is not currently in its ground state: the ground-state vacuum energy is $\Lambda = 0$, we are simply not there (yet/now). Temporary vacuum energy might exist if there exists a field that takes a long time, on a cosmological scale, to reach its ground state. Incidentally, this matches the reasoning behind early cosmological inflation models: these postulate an epoch of rapid acceleration in the (very) early universe.

Such a model requires (see [Frieman \(2008\)](#) and references therein for details) an extremely light scalar field ϕ such that its mass is $m_\phi < H \sim 10^{-33}$ eV, which is minuscule by particle physics standards, which allows it to spread smoothly without gravitational clustering — though it allows some anisotropies. A scalar field with Lagrangian density

$$\mathcal{L} = \frac{1}{2}\partial^\mu\phi\partial_\mu\phi - V(\phi) \tag{6.19}$$

has stress-energy in the form of a perfect fluid (see sections 1.3 and 6.1) with

$$\begin{aligned} \rho_{\text{light}} &= \frac{\dot{\phi}^2}{2} + V(\phi), \\ p_{\text{light}} &= \frac{\dot{\phi}^2}{2} - V(\phi) \end{aligned} \tag{6.20}$$

where $\phi(\mathbf{x}, t) = \phi(t)$, $\dot{\phi}^2/2$ is the kinetic energy, and $V(\phi)$ is the potential energy, as usual. (If field theory is foreign, get a copy of [Susskind and Friedman \(2017\)](#). No

other resource explains field theory quite as well or quite as concisely.) The equation of motion for the field is

$$\ddot{\phi} + 3H\dot{\phi} + \frac{dV(\phi)}{d\phi} = 0 \quad (6.21)$$

Scalar-field dark energy can thus be described by the *non-constant* equation of state parameter

$$w_\phi = \frac{\dot{\phi}^2/2V - 1}{\dot{\phi}^2/2V + 1} \quad (6.22)$$

If $m_\phi = \sqrt{dV(\phi)/d\phi} \ll H(t)$, then $w_\phi \approx -1 = w_\Lambda$. This scalar field will behave like a slowly varying vacuum energy with $\rho_{\text{vac}} \simeq V[\phi(t)]$. But even slight deviations from this will throw the equation of state parameter off course, making it vary in time more than we want.

Many models satisfy this criterion. The simplest one has a free scalar field $V(\phi) = m_\phi^2 \phi^2/2$. In order for this field to have negative w_ϕ and energy-density magnitude, the field amplitude must be very large: $\phi \sim 10^{28} \text{ eV} \sim M_{\text{Planck}}$. This, in turn, implies that the field is very flat and weakly self-coupled, which makes it hard to fit with particle physics models. These kinds of solutions to late universe acceleration are usually called Quintessence models. Because of the problems we have noted, they are not expected to be the right way to go. See [Frieman \(2008\)](#) and references therein for more details and examples of concrete models.

It's important to note that ([Carroll 2001](#)) Quintessence models do not solve the coincidence or constant problems at all: we still need to fix its parameters manually to fit our present values. They are thus harder to implement and fit with the data and equally arbitrary as cosmological constant models; they are less elegant and less parsimonious. Is this a good reason to reject them?

6.3.2 Modified Gravity

A different approach is to pin late universe acceleration on gravitational effects rather than the repulsive force of dark energy. That is, instead of adding the cosmological constant to the right-hand, stress-energy side of the Einstein equation, we modify the left-hand, geometric side:

$$G_{\mu\nu} \rightarrow G_{\mu\nu} + f(g_{\mu\nu}) \quad (6.23)$$

where $f(g_{\mu\nu})$ represents our modification. There are many examples of such theories which we will not go into here. In general, they will take the form of

modifying the GR action from

$$S = \frac{M_{\text{Planck}}^2}{2} \int d^4x \sqrt{-g} R + \int d^4x \sqrt{-g} \mathcal{L}_M \quad (6.24)$$

to

$$S = \frac{M_{\text{Planck}}^2}{2} \int d^4x \sqrt{-g} \left(R - \frac{\mu^4}{R} \right) + \int d^4x \sqrt{-g} \mathcal{L}_M \quad (6.25)$$

where μ is an empirical mass scale parameter, which comes into play only for small R . This is a common edit to make in higher-dimensional theories such as String Theory. Unfortunately, these models have a very hard time accounting for smaller structures like solar systems as we understand them through standard GR. Modified gravity models of this kind need very high values of curvature for quantum effects to become important enough to drive late acceleration, $R \sim M_{\text{Planck}}^2$, but these usually prevent us from making sense of smaller, more mundane stuff.

However, this comes with a neat flipside: we can use the predictions of modified gravity models to test GR. GR with the cosmological constant and dark matter, the ΛCDM -model, yields an equation for the growth of large-scale matter-density perturbations of the form we have seen in eq. (6.18). A modification of GR in the form considered here would lead to additional terms in this equation, and thus affect predictions for the growth rate of structures. If these fit observations better than GR, we might have reason to believe in a modified gravity solution on independent grounds.

These models do stand the chance of explaining the coincidence and constant problems because the background theory is more complex. String Theory, at least in principle, can explain the values we observe as emergent from a deeper structure. But that is at the cost of a much more complex machinery that is much more abstract and much more distant from experimental practices than our current theories.

6.4 Conclusion: Expanded Anthropic Reasoning; or, Why Physics Needs Philosophy

Given that we have seen the use of the anthropic principle, it's hard to resist the urge to apply it to everything all at once: if things weren't the way they are, we wouldn't be here to see them; thus, they are the way they are — QED.

This is a bad argument, because it can prove everything and predict nothing; it's the opposite of what we are trying to do in physics. And yet, we have said that it works (or if you are more skeptical: that it *might* work) in the case of the coincidence

problem. Why does it not work everywhere?

Physical theories can be idealized to a set of axioms and theorems. Electromagnetism, QM, SR, and GR are all good examples of physical theories in this sense: they have a set of foundational assumptions, which are true in some domains of the world (QM holds for single particles in non-relativistic settings; GR holds on scales where gravity is the chief dynamical factor), from which we can derive facts and predictions and laws of nature *given a set of initial conditions*. The problem is that such theories can never hope to explain the initial conditions. For that, we would need a bigger theory, with more axioms, from which we could derive the smaller theory as well as its initial conditions.

To make this more concrete, imagine that String Theorists construct a model that entails the verified predictions of the Standard Model and the verified predictions of GR, but that nevertheless is unified and without contradictions or breakdowns. This is the bigger theory, SM and GR are the smaller theories. Such a String Theory might even predict the present value of the cosmological constant or the mass of the Higgs boson without needing post hoc experimental fixing. Such a String Theory would count as a grand victory for physics: we would have explained *why* nature is the way QFT and GR tell us it is, at least approximately.

The problem is that we would now want to know why such a String Theory itself is true, and why *its* initial conditions obtain. That, alas, would require an even bigger theory, with even more (or: stronger) axioms. And so on and so forth.

But can't a theory prove its own initial conditions, perhaps? The answer, unfortunately, is that a consistent and true theory cannot possibly do so: there can be no true Theory of Everything.

In 1931, [Gödel \(1992\)](#) proved two monumental theorems in formal logic: (G1) any consistent and sufficiently complex theory T will contain a *true* sentence G which is *neither provable nor disprovable* in T; and (G2) any consistent and sufficiently complex theory T cannot prove its own consistency.

(G1) says that our vocabulary for String Theory (or any sophisticated enough theory, for that matter) will be able to produce true sentences which the axioms of String Theory cannot prove or disprove. How do we know they are true? Well, because we can check experimentally (using different theories as our aids) or because we can construct an even more complex theory that *can* prove it. In physics, this may not be so much of a problem, since we can always avail ourselves of a rich ensemble of theories, observations, and experiments to help us decide, test, or probe the new sentences we come up with. Nevertheless, insofar as we are in search of a true Theory of Everything, (G1) is ruinous: it is impossible to construct a theory that decides everything (unless it is so simple and uninformative that we cannot

derive any predictions from it).

(G2) ruins whatever we have left of an aspiration to a true Theory of Everything. Sure, because of (G1) we can't prove certain sentences in the theory. But who said these would be important? Maybe we can still prove all we want. Alas. (G2) tells us that we can never prove whether our theory is consistent. And any theory that hopes to be *true* must at the very least be consistent. So a Theory of Everything could at most be a Theory of Everything apart from its own fit with the world. In short: we could never *complete* a physical theory.

What does this all mean? At a minimum, it means that we cannot truly have a Theory of Everything. That is, if a theory attempts to account for initialy conditions, for the laws of nature, *and* for itself, then it must be inconsistent at some point — and thus false. Consistent theories need experimentally fixed parameters. As such, they need us to take some things for granted; such as the consistency and truth of the theory, which we cannot possibly prove. Of course, that's what we want the very most: to know if our theory is *right*.

We've taken a circuitous route to a problem we find in a series of ancient questions: how do we get something from nothing? why is there anything at all? and why, given that there is something, is there *this* and not something utterly (or marginally) different? The world could have been many ways, but it is only one way: the way we see it when we look at it and perform experiments on it. We can only discover what the initial conditions must have been in order for the world to be *this way*, not also why it is there in the first place or why it isn't otherwise. Physics must stop at some point: some things simply cannot be explained without cheats.

The problem we are forced to face, then, is *where* we must stop. This is a hard question worth dealing with, and our investigations into dark matter and dark energy and cosmological models in general bring this issue to the foreground. The simple versions of the scientific method that we all know are not enough to tell us when anthropic principles do or do not apply. That problem cannot be settled here. But a lesson can and must be drawn: we need to engage in philosophy of science, whose aim is to understand fully what makes for good science, and what doesn't. Insofar as we are for or against antropic reasoning, we are already doing philosophy. Insofar as we are deciding how to interpret the constants in our equations, we are already doing metaphysics. So we better get those disciplines right, too, if we want to keep doing physics. As Einstein put it ([Howard and Giovanelli 2019](#)):

I fully agree with you about the significance and educational value of methodology as well as history and philosophy of science. So many people today—and even professional scientists—seem to me like somebody who

has seen thousands of trees but has never seen a forest. A knowledge of the historic and philosophical background gives that kind of independence from prejudices of his generation from which most scientists are suffering. This independence created by philosophical insight is—in my opinion—the mark of distinction between a mere artisan or specialist and a real seeker after truth. (Einstein to Thornton, 7 December 1944, EA 61–574)

Appendix A

Appendices

A.1 Spherical Bessel Functions

Spherical Bessel functions are the solutions to Bessel's differential equation in spherical coordinates:

$$x^2 \frac{d^2y}{dx^2} + 2x \frac{dy}{dx} + (x^2 - n(n+1))y = 0 \quad (\text{A.1})$$

where n is an integer.

There are two different kinds of spherical Bessel function, to correspond to the two independent solutions to the above equation:

$$\begin{aligned} j_n(x) &= (-x)^n \left(\frac{1}{x} \frac{d}{dx} \right)^n \frac{\sin x}{x} \\ y_n(x) &= (-x)^n \left(\frac{1}{x} \frac{d}{dx} \right)^n \frac{\cos x}{x} \end{aligned} \quad (\text{A.2})$$

j_n are the important kind for our purposes; they will be used in our plane wave expansion (see below), and in our derivation of the axes in section 4.5.1.

Standard Integral

For a spherical Bessel function, j_l , there exists the following standard integral:

$$\int j_l^2(z) d(\log(z)) = \frac{1}{2l(l+1)}$$

In section 4.5.1 we want to integrate

$$\int j_l^2(kr) d(\log(k))$$

which is similar, but not identical.

To evaluate this integral, we start with

$$\begin{aligned}
\frac{1}{2l(l+1)} &= \int j_l^2(kr) \, d(\log(kr)) \\
&= \int j_l^2(kr) \frac{d(kr)}{kr} \\
&= \int j_l^2(kr) \frac{rd(k)}{kr} \quad (\text{since } r \text{ is a constant}) \\
&= \int j_l^2(kr) \frac{d(k)}{k} \\
&= \int j_l^2(kr) \, d(\log(k))
\end{aligned}$$

Hence the standard integral still applies in our case:

$$\int j_l^2(kr) \, d(\log(k)) = \frac{1}{2l(l+1)}$$

A.2 Spherical Harmonics

Legendre Polynomials

Legendre polynomials P_l , form a power series expansion, which cuts off after l terms. They have the form:

$$P_l(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n \quad (\text{A.3})$$

where the coefficients a_i are normalized such that $\sum_0^n a_i = 1$.

There are many formulations of the definition for Legendre polynomials; their differences are not important, so I shall give only one.

$$P_l(x) = \frac{1}{2^n} \sum_{k=0}^n \binom{n}{k}^2 (x-1)^{n-k} (x+1)^k \quad (\text{A.4})$$

In the coming definition for spherical harmonics, we will use the Associated Legendre Polynomials. These are related to ordinary Legendre Polynomials by:

$$P_{l,m}(\cos \theta) = (1 - \cos^2 \theta)^{\frac{m}{2}} \frac{d^m}{d \cos^m \theta} P_l(\cos \theta) \quad (\text{A.5})$$

Spherical Harmonics

The spherical harmonics are the solutions to the Laplace equation¹ in spherical coordinates:

$$\Delta^2 \psi = \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial \psi}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial \psi}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 \psi}{\partial \phi^2} = 0 \quad (\text{A.6})$$

The solutions have the form:

$$\psi = \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} P_{l,m}(\cos \theta) e^{im\phi} \equiv Y_{l,m}(\theta, \phi) \quad (\text{A.7})$$

where $P_{l,m}$ are the Associated Legendre Polynomials, and the value in the square root is a normalization constant.

Spherical harmonics are orthonormal, and complete over the 2D surface of a sphere. These properties follow from properties of the Legendre polynomials. I will not derive them here, but the full derivations for both Legendre polynomials and spherical harmonics as solutions to the Laplace equation, and their properties, can be found readily online.

That they are orthonormal gives: $\int Y_{l,m}(\theta, \phi) Y_{l',m'}(\theta, \phi) d\Omega = \delta_{ll'} \delta_{mm'}$

and that they are complete gives: $f(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^l a_{l,m} Y_{l,m}(\theta, \phi)$

Visualising the Spherical Harmonics

To visualise the spherical harmonics, we first split $Y_{l,m}$ into its real and imaginary parts. Since the associated Legendre polynomials are all real, the only imaginary component will come from $e^{im\phi}$.

Using Euler's formula, $e^{i\theta} = \cos \theta + i \sin \theta$,

$$\begin{aligned} Y_{l,m} &= \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} P_{l,m}(\cos \theta) \times (\cos(m\phi) + i \sin(m\phi)) \\ \text{Re}[Y_{l,m}] &= \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} P_{l,m}(\cos \theta) (\cos(m\phi)) \\ \text{Im}[Y_{l,m}] &= \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} P_{l,m}(\cos \theta) \times (i \sin(m\phi)) \end{aligned} \quad (\text{A.8})$$

¹They are also the angular solutions to the Helmholtz equation; we will consider the Laplace equation as it is a simpler derivation.

We will now consider when the real part is equal to zero². This will form a set of points equal to zero, which are called nodal lines.

There will be l nodal lines in total: the number of longitudinal nodal lines is equal to $|m|$; the number of latitudinal nodal lines is equal to $l - |m|$. The segments between these nodal lines must then alternate between being positive and negative (see fig. A.1).

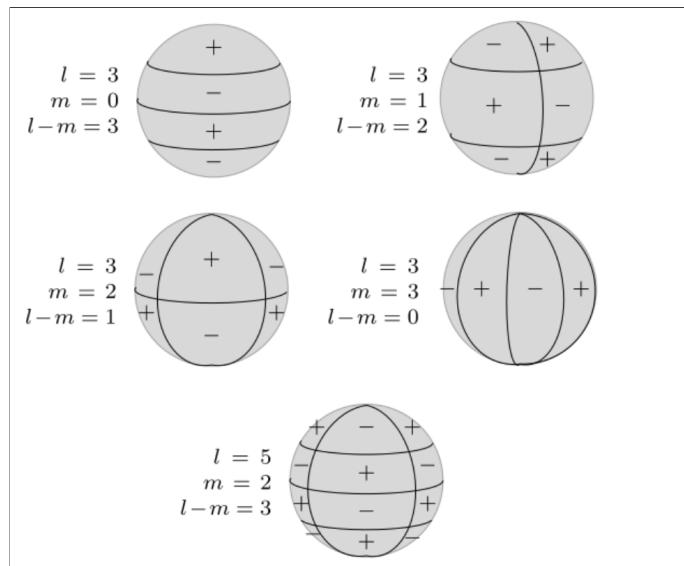


Figure A.1: The nodal lines of the $l = 3$ spherical harmonics (and $l = 5$, $m = 2$ harmonics, for extra visualisation). The solid lines represent the nodal lines. Note how the regions alternate between being positive and negative.

It is more complicated to ascertain the values at each point on the sphere, but Figure A.2 gives a qualitative idea.

²It makes no difference which of the real and imaginary parts we choose.

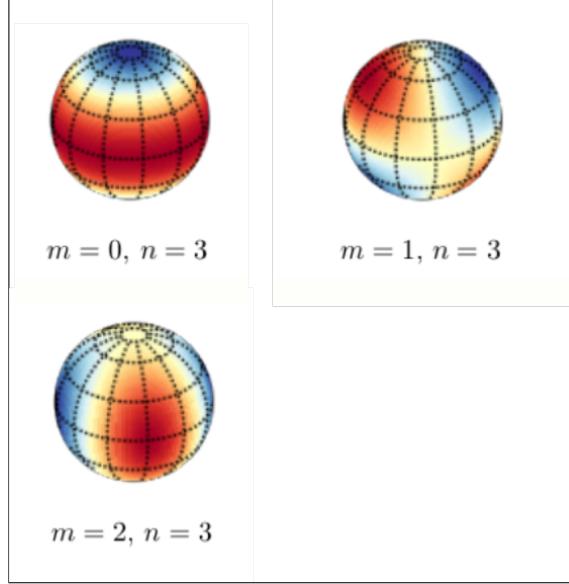


Figure A.2: A colour representation of the three spherical harmonics in the upper left corner of Figure A.1. Red shows areas with a high positive value, yellow with a zero value, and blue with a negative value. Note that in this figure, $n = l$. Figure adapted from ?.

A.3 Angular Power Spectrum Derivation

Since spherical harmonics are complete over a sphere (see appendix A.2), we can write any well-behaved function as:

$$f(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^l a_{l,m} Y_{l,m}(\theta, \phi) \quad (\text{A.9})$$

That is, as a sum of different spherical harmonics.

We can then obtain the amplitudes of each harmonics, using the completeness relation, and multiplying each side by $Y_{l,m}^*$, and integrating over Θ :

$$\int f(\theta, \phi) Y_{l,m}^* d\Omega = \int \sum_{l=0}^{\infty} \sum_{m=-l}^l a_{l,m} Y_{l,m}(\theta, \phi) Y_{l,m}^*(\theta, \phi) d\Omega \quad (\text{A.10})$$

Then using the orthogonality relation:

$$\begin{aligned} \int f(\theta, \phi) Y_{l,m}^* d\Omega &= \sum_{l=0}^{\infty} \sum_{m=-l}^l a_{l,m} \delta_{l,l'} \delta_{m,m'} \\ &= a_{l',m'} \end{aligned} \quad (\text{A.11})$$

Dropping the primes:

$$a_{l,m} = \int f(\theta, \phi) Y_{l,m}^*(\theta, \phi) d\Omega \quad (\text{A.12})$$

Variance Calculated by Two Angles

It was mentioned in chapter 4 that we need an equation for C_l which involves comparing the temperature fluctuation at two different points on the SLS.

We can devise an equation which gives C_l in terms of $T(\theta, \phi)$ and $T(\theta', \phi')$, integrated over the whole surface of the sphere. We begin with the definition of C_l , then substitute the equation we have just derived for the amplitudes:

$$\begin{aligned} C_l &= \frac{1}{2l+1} \sum_{m=-l}^l a_{l,m}^* a_{l,m} \\ &= \frac{1}{2l+1} \sum_{m=-l}^l \left(\int_0^{2\pi} \int_0^\pi Y_{l,m}(\theta', \phi') T(\theta', \phi') d\Omega' \right) \left(\int_0^{2\pi} \int_0^\pi Y_{l,m}^*(\theta, \phi) T(\theta, \phi) d\Omega \right) \\ &= \frac{1}{2l+1} \int_0^{2\pi} \int_0^\pi \int_0^{2\pi} \int_0^\pi \left(\sum_{m=-l}^l Y_{l,m}(\theta', \phi') Y_{l,m}^*(\theta, \phi) \right) T(\theta', \phi') T(\theta, \phi) d\Omega' d\Omega \end{aligned} \quad (\text{A.13})$$

if we use \mathbf{n} as the unit vector in the θ, ϕ direction:

$$C_l = \frac{1}{2l+1} \int_\Omega \int'_\Omega \left(\sum_{m=-l}^l Y_{l,m}(\mathbf{n}') Y_{l,m}^*(\mathbf{n}) \right) T(\mathbf{n}') T(\mathbf{n}) d\Omega' d\Omega \quad (\text{A.14})$$

We can use this equation for our analysis, as the data we use comes in the form of a set of data points $T(\mathbf{n})$.

A.4 The CMB Dipole

The map of the CMB in chapter 4 has had the CMB dipole removed; A map of the CMB without its removal is shown in Figure A.3. It is removed from most maps, because its intensity effectively drowns out the small anisotropies which we are trying to analyse; its variation is around 100 times larger than the anisotropies.

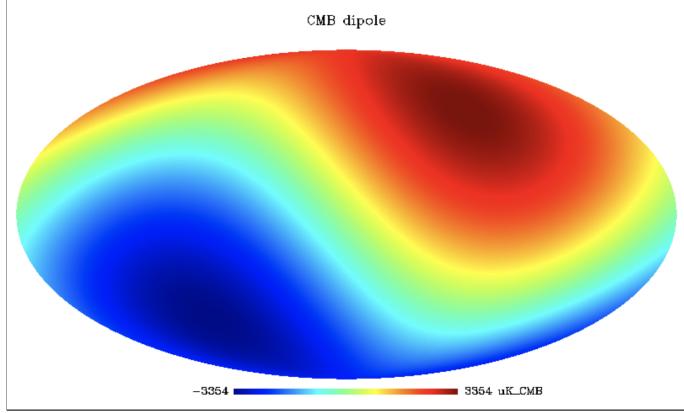


Figure A.3: The CMB dipole.

The dipole is caused by the Doppler effect, as a result of the Earth's motion relative to the CMB rest frame. We are moving towards the upper right of the map in fig. A.3; the CMB is hotter in that direction, so it must be being blue-shifted.

Recall from chapter 1 that we typically consider the universe homogenous on a cosmic scale. However, on very small scale compared to this, the space surrounding us is not homogenous. The Earth moves relative to the Sun, which in turn is moving relative to the Milky Way. We can analyse the CMB dipole to give us a velocity relative to the CMB.

From the appearance on the dipole over the SLS, we can tell that we are not moving fast, compared to the speed of light; if we were, the relativistic Doppler effect would be taking place, and the dipole would look different. Thus we can use the non-relativistic Doppler effect equations³.

In general, the Doppler frequency is given by:

$$\nu = \frac{\nu_0 \sqrt{1 - \frac{v_s^2}{c^2}}}{1 - \frac{v_s}{c}} \quad (\text{A.15})$$

where ν is the observed frequency, ν_0 is the source frequency, v_s is the source velocity, and c is the speed of light.

We expand the square root in eq. (A.15) using a binomial expansion, and can neglect all but the first two terms, since our motion is not relativistic. This gives:

$$\begin{aligned} \nu &\approx \nu_0 \frac{1 - \frac{v_s^2}{2c^2}}{1 - \frac{v_s}{c}} \\ \frac{\nu - \nu_0}{\nu_0} &= \frac{\Delta\nu}{\nu_0} \approx \frac{v_s}{c} \end{aligned} \quad (\text{A.16})$$

Since temperature is proportional to frequency,

³This is also useful, as the non-relativistic Doppler effect is a simple superposition over the CMB anisotropies, so it is easy to remove.

$$\frac{\Delta T}{T} \approx \frac{v_s}{c} \quad (\text{A.17})$$

Using eq. (A.17), the velocity of the Earth with respect to the CMB is 368 ± 2 km s^{-1} .

Bibliography

- Binney, J. and Tremaine, S. (2011). *Galactic dynamics*, volume 20. Princeton university press.
- Boolos, G. S., Burgess, J. P., and Jeffrey, R. C. (2002). *Computability and logic*. Cambridge university press.
- Carroll, S. M. (2001). Dark energy and the preposterous universe. *arXiv preprint astro-ph/0107571*.
- Collaboration, P., Ade, P., and Aghanim, N. (2013). Planck 2015 results. *XVI. Cosmological parameters*, 1303:v1.
- Egan, C. A. and Lineweaver, C. H. (2008). Dark-energy dynamics required to solve the cosmic coincidence. *Physical Review D*, 78(8):083528.
- Friederich, S. (2018). Fine-tuning. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2018 edition.
- Frieman, J. A. (2008). Lectures on dark energy and cosmic acceleration. In *AIP Conference Proceedings*, volume 1057, pages 87–124. AIP.
- Gödel, K. (1992). *On formally undecidable propositions of Principia Mathematica and related systems*. Courier Corporation.
- Gourgoulhon, E. (2012). *3+ 1 formalism in general relativity: bases of numerical relativity*, volume 846. Springer Science & Business Media.
- Howard, D. A. and Giovanelli, M. (2019). Einstein’s philosophy of science. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall 2019 edition.
- Koleva, M., Prugniel, P., Ocvirk, P., Le Borgne, D., and Soubiran, C. (2008). Spectroscopic ages and metallicities of stellar populations: validation of full spectrum fitting. *Monthly Notices of the Royal Astronomical Society*, 385(4):1998–2010.

- Krauss, L. M. and Chaboyer, B. (2003). Age estimates of globular clusters in the milky way: constraints on cosmology. *Science*, 299(5603):65–69.
- Landau, L., Lifshitz, E., and Pitaevskij, L. (1981). *Course of theoretical physics. vol. 10: Physical kinetics*. Oxford.
- Lee, S. K., Lisanti, M., Mishra-Sharma, S., and Safdi, B. R. (2015). Modulation effects in dark matter-electron scattering experiments. *Physical Review D*, 92(8):083517.
- Lim, E. (2018). Statistical mechanics.
- Lim, E. (2019). General relativity.
- Lineweaver, C. H. and Egan, C. A. (2007). The cosmic coincidence as a temporal selection effect produced by the age distribution of terrestrial planets in the universe. *The Astrophysical Journal*, 671(1):853.
- Lisanti, M. (2018). Lectures on dark matter physics (2016). *arXiv preprint arXiv:1603.03797*.
- Nagel, E. and Newman, J. R. (2012). *Godel's proof*. Routledge.
- Perlmutter, S., Gabi, S., Goldhaber, G., Goobar, A., Groom, D., Hook, I., Kim, A., Kim, M., Lee, J., Pain, R., et al. (1997). Measurements* of the cosmological parameters ω and λ from the first seven supernovae at $z \geq 0.35$. *The astrophysical journal*, 483(2):565.
- Persic, M., Salucci, P., and Stel, F. (1996). The universal rotation curve of spiral galaxies—i. the dark matter connection. *Monthly Notices of the Royal Astronomical Society*, 281(1):27–47.
- Pitaevskii, L. and Lifshitz, E. (2012). *Physical kinetics*, volume 10. Butterworth-Heinemann.
- Riess, A. G., Filippenko, A. V., Challis, P., Clocchiatti, A., Diercks, A., Garnavich, P. M., Gilliland, R. L., Hogan, C. J., Jha, S., Kirshner, R. P., et al. (1998). Observational evidence from supernovae for an accelerating universe and a cosmological constant. *The Astronomical Journal*, 116(3):1009.
- Rubin, V. C., Ford Jr, W. K., and Thonnard, N. (1980). Rotational properties of 21 sc galaxies with a large range of luminosities and radii, from ngc 4605/r= 4kpc/to ugc 2885/r= 122 kpc. *The Astrophysical Journal*, 238:471–487.

- Schmidt, B. P., Suntzeff, N. B., Phillips, M., Schommer, R. A., Clocchiatti, A., Kirshner, R. P., Garnavich, P., Challis, P., Leibundgut, B., Spyromilio, J., et al. (1998). The high-z supernova search: measuring cosmic deceleration and global curvature of the universe using type ia supernovae. *The Astrophysical Journal*, 507(1):46.
- Sivanandam, N. (2013). Is the cosmological coincidence a problem? *Physical Review D*, 87(8):083514.
- Sklar, L. (1995). *Physics and chance: Philosophical issues in the foundations of statistical mechanics*. Cambridge University Press.
- Susskind, L. and Friedman, A. (2017). *Special Relativity and Classical Field Theory*. Penguin UK.
- Tong, D. (2018a). Lectures on cosmology.
- Tong, D. (2018b). Lectures on statistical physics.
- Velten, H., Vom Marttens, R., and Zimdahl, W. (2014). Aspects of the cosmological “coincidence problem”. *The European Physical Journal C*, 74(11):3160.
- Vilenkin, A. (1995). Predictions from quantum cosmology. *Physical Review Letters*, 74(6):846.
- Wallace, D. (2012). *The emergent multiverse: Quantum theory according to the Everett interpretation*. Oxford University Press.
- Zel'dovich, Y. B., Zeldovich, Y. B., and Krasinski, A. (1968). The cosmological constant and the theory of elementary particles. *Gen. Rel. Grav.*, 11:209.