

Graduate Cosmology

Oliver H. E. Philcox

About these Notes

These notes provide a graduate-level introduction to cosmology. They focus on the physics of the late Universe, with a little early Universe physics thrown in for completeness, exploring the dynamics, structure, and evolution of the Universe starting from the Big Bang and ending several billion years later. The treatment is advanced, so we'll get deep into the mathematics behind key observations that have been instrumental in developing the standard cosmological model, such as galaxy clustering and the cosmic microwave background. By the end, you should be ready to perform cutting-edge research in cosmology!

Conventions: Throughout these notes, we'll assume $k_B = c = \hbar = 1$, and use the mostly-positive metric signature.

Acknowledgements: These notes are largely based on Daniel Baumann's *Cosmology* textbook (and my own dim and distant memory of accompanying Cambridge course), with additional input from Dragan Huterer's *A Course in Cosmology* and Colin Hill's lecture notes. Additional resources include the excellent textbooks by Dodelson & Schmidt and Weinberg, as well as my own lecture notes on the Effective Field Theory of Large-Scale Structure.

Contents

1	A Brief History of the Universe	3
2	The Cosmological Background	3
2.1	Cosmological Principle	3
2.2	Distances & Metrics	4
2.3	The FLRW Metric	4
2.4	Hubble Parameter and Hubble's Law	6
2.5	Geodesics	7
2.6	Cosmological Distances	8
2.7	Curvature & the Einstein Equation	10
2.8	Stress-Energy Tensor	11
2.9	Continuity Equation	12
2.10	Friedmann Equations	12
2.11	The Evolution of Species	13
2.12	The Evolution of the Universe	14

3	Thermal History of the Early Universe	15
3.1	A Primer on Statistical Mechanics	16
3.2	Equilibrium Thermodynamics	17
3.3	Non-Equilibrium Thermodynamics	23
3.4	Thermal Relics and Dark Matter Freeze-Out	25
3.5	Recombination and the Cosmic Microwave Background	27
3.6	Big Bang Nucleosynthesis	29
4	Inflation	34
4.1	Problems with the Hot Big Bang	34
4.2	The Inflationary Solution	37
4.3	Conditions for Inflation	40
4.4	Scalar Field Inflation	40
4.5	Slow-Roll Solution	42
4.6	Summary of Inflation	43
4.7	Reheating	43
4.8	Inflationary Fluctuations	45
5	Relativistic Perturbation Theory	50
5.1	Introduction to Perturbation Theory	51
5.2	Metric Perturbations	51
5.3	Matter Perturbations	55
5.4	Conservation Equations	57
5.5	Einstein Equations	59
5.6	Initial Conditions	61
5.7	Evolution of the Gravitational Potential	64
5.8	Evolution of Dark Matter Perturbations	65
5.9	The Matter Power Spectrum	67
5.10	Evolution of the Photons and Baryons	69
5.11	Summary	74
6	The Cosmic Microwave Background	74
6.1	CMB Statistics	75
6.2	The Temperature Anisotropy: Sachs-Wolfe and Beyond	75
6.3	The Sachs-Wolfe Plateau	76
6.4	Beyond the Sachs-Wolfe Limit	77
6.5	Diffusion and the Damping Tail	82
6.6	Summary So Far	83
6.7	Reionization	83
6.8	The Impact of Λ CDM Parameters	84
6.9	Additional Topics	84
7	Large-Scale Structure	89
7.1	Newtonian Fluid Equations	90
7.2	Linear Solution	90
7.3	Standard Perturbation Theory	91
7.4	The One-Loop Power Spectrum	92
7.5	Galaxy Bias	95

1 A Brief History of the Universe

Under the standard Hot Big Bang scenario, the early Universe was dense and hot. As the Universe expands and cools, it passes through a series of key epochs, triggered when the temperature drops below a characteristic energy scale. These correspond to changes in the particle content of the Universe, e.g., at the QCD phase transition quarks and leptons are replaced with hadrons, and Hydrogen replaces electrons and protons at recombination.

Here are the key epochs (very approximately):

Event	Energy	Time
Inflation	$\sim 10^{16}$ GeV	$\sim 10^{-36}$ – 10^{-32} s
Reheating	$\lesssim 10^{15}$ GeV (model-dependent)	$\lesssim 10^{-18}$ s
Electroweak Phase Transition	100 GeV	10^{-10} s
QCD Phase Transition	150 MeV	10^{-4} s
ν decoupling	1 MeV	1 s
BBN	0.1 MeV	10^2 s
Matter–radiation equality	0.8 eV	5×10^4 yr
Recombination	0.26 eV	3.8×10^5 yr
First structures form	10^{-3} eV	10^8 yr
First galaxies form	10^{-3} eV	5×10^8 yr
Matter–dark energy equality	10^{-3} eV	9×10^9 yr
Today	2.35×10^{-4} eV	13.8 Gyr

Note that the temperature of the Universe is approximately given by $T/1\text{MeV} \sim (t/1\text{s})^{-1/2}$. We will provide more detail on (most of) these epochs below.

2 The Cosmological Background

2.1 Cosmological Principle

The cosmological principle states that, on sufficiently large scales, the Universe is **homogeneous** and **isotropic**. Homogeneity implies that the Universe looks the same from any position (*i.e.* invariance under $\vec{x} \rightarrow \vec{x} + \vec{a}$), isotropy implies that the Universe looks the same in any direction (*i.e.* invariance under rotations).

This is experimentally supported (though still argued about). The primary evidence for isotropy comes from the CMB: the temperature anisotropy and (the inferred density fluctuations) are tiny, $\delta T/T \sim \delta \rho/\rho \sim 10^{-5}$. The distribution of galaxies imply homogeneity, at least above scales of around 100 Mpc.¹

¹1 Mpc = 3.086×10^{22} m. This is an annoying, but common, choice of units.

2.2 Distances & Metrics

The metric plays a central role in describing the geometry of the Universe. Fundamentally, this relates a set of coordinates ($\{x^1, x^2, \dots\}$, which are human constructs) to distances (which are observables). A particularly simple, but useful, set of metrics describe flat three-dimensional Euclidean space. Assuming the Cartesian coordinates $x^i = \{x, y, z\}$, we have the line element:

$$d\ell^2 \equiv \sum_{ij} \gamma_{ij} dx^i dx^j = dx^2 + dy^2 + dz^2. \quad (1)$$

implying that $\gamma_{ij} = \delta_{ij}$, for Kronecker delta δ_{ij} . A more interesting metric is obtained by using polar coordinates: $x^i = \{r, \theta, \phi\}$, with

$$d\ell^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2 = dr^2 + r^2 d\Omega^2 \quad (2)$$

(where $d\Omega^2 = d\theta^2 + \sin^2 \theta d\phi^2$ is the metric on the unit 2-sphere). Here, the metric is given by $\gamma_{ij} = \text{diag}(1, r^2, r^2 \sin^2 \theta)$. Importantly, ds^2 is invariant under a coordinate transform.

To describe physics in an expanding Universe, we need a *spacetime* metric (3+1 dimensions), which includes time. Defining the coordinates $x^\mu = \{x^0, x^1, x^2, x^3\}$ where x^0 describes time and x^i describes space,² we have the infinitesimal spacetime interval:

$$ds^2 = g_{\mu\nu}(x) dx^\mu dx^\nu, \quad (3)$$

where g is the metric. The simplest metric is Minkowski, which corresponds to flat space (the rest-frame in special relativity):

$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2, \quad g_{\mu\nu} = \text{diag}(-1, +1, +1, +1) \equiv \eta_{\mu\nu}. \quad (4)$$

$$ds^2 = -dt^2 + dr^2 + r^2 d\Omega^2, \quad g_{\mu\nu} = \text{diag}(-1, 1, r^2, r^2 \sin^2 \theta). \quad (5)$$

for $x^\mu = \{t, x, y, z\}$ and $x^\mu = \{t, r, \theta, \phi\}$ respectively. Timelike intervals ($-ds^2 \equiv d\tau^2 > 0$) correspond to the paths of massive particles; null intervals ($ds^2 = 0$) correspond to light rays.

In General Relativity, the metric becomes a dynamical, coordinate-dependent field, sourced by matter and energy (with the famous quip: matter tells space how to curve; space tells matter how to move). The relationship between matter content and geometry is governed by Einstein's field equations, which we'll use below.

2.3 The FLRW Metric

What is the metric of our Universe? On large-scales, the metric is heavily restricted by the assumptions of homogeneity and isotropy. We can start from the most general form (splitting time and space):

$$ds^2 = -A(t, \vec{x}) dt^2 + B_i(t, \vec{x}) dt dx^i + C_{ij}(t, \vec{x}) dx^i dx^j, \quad (6)$$

for unknown A, B, C . We then apply our symmetry requirements:

- a) Homogeneity: A, B, C cannot depend on $\vec{x}$.

²Here and throughout we will use the standard convention that $\mu \in \{0, 1, 2, 3\}$ describes spacetime indices and $i \in \{1, 2, 3\}$ describes spatial indices. We will also use Einstein summation convention: $g_{\mu\nu} dx^\mu dx^\nu \equiv \sum_{\mu\nu} g_{\mu\nu} dx^\mu dx^\nu$ et cetera.

- b) **Isotropy**: There are no preferred directions. Therefore, B_i must vanish (there is no special vector it could point along), and C_{ij} must be proportional to the spatial-metric, γ_{ij} (which we will define below).

Finally, we can redefine the time-coordinate such that $A(t)dt^2 \rightarrow dt^2$. Combining everything together, we find

$$\boxed{ds^2 = -dt^2 + a^2(t)\gamma_{ij}dx^i dx^j.} \quad (7)$$

Here, $a(t)$ is the **scale-factor**, which describes how physical distances change with time. We typically normalize it such that $a(t_0) = 1$, where t_0 is the time today.

The final ingredient in the metric is the spatial metric γ_{ij} , which describes the three-dimensional geometry of the Universe. Via homogeneity, this must be a space of constant spatial curvature. There are three options, parametrized by the curvature constant κ :

1. Flat ($\kappa = 0$): ordinary Euclidean 3-space, with

$$\gamma_{ij}dx^i dx^j = d\chi^2 + \chi^2(d\theta^2 + \sin^2\theta d\phi^2) = d\chi^2 + \chi^2 d\Omega^2. \quad (8)$$

Here, χ is a radial coordinate (known as the **comoving radial coordinate**), and θ, ϕ are the usual polar coordinates.

2. Spherical ($k = +1$, positive curvature). This is the 3D analogue of the 2-sphere $\mathbb{S}^2$, and can be thought of as coordinates living on the “surface” of a 4-dimensional sphere ($x_1^2 + x_2^2 + x_3^2 + x_4^2 = R_0^2$):

$$\gamma_{ij}dx^i dx^j = d\chi^2 + R_0^2 \sin^2(\chi/R_0) d\Omega^2. \quad (9)$$

This describes a closed Universe, with finite (but unbounded) spatial extent (like the surface of a sphere, but one dimension higher) and radius R_0 .

3. Hyperbolic ($k = -1$, negative curvature). This is the 3D analogue of the 2-hyperboloid $\mathbb{H}^2$, and can be thought of as coordinates living on the “surface” of a 4-dimensional hyperboloid ($x_1^2 + x_2^2 + x_3^2 - x_4^2 = -R^2$).

$$\gamma_{ij}dx^i dx^j = d\chi^2 + R_0^2 \sinh^2(\chi/R_0) d\Omega^2. \quad (10)$$

This describes an open Universe, with infinite extent defined by the radial distance R_0 (like a higher-dimensional saddle).

All three cases can be written in the unified form $\gamma_{ij}dx^i dx^j = d\chi^2 + S_k^2(\chi) d\Omega^2$, where

$$S_k(\chi) = \begin{cases} \chi & k = 0 \text{ (flat)} \\ R_0 \sin(\chi/R_0) & k = +1 \text{ (spherical)} \\ R_0 \sinh(\chi/R_0) & k = -1 \text{ (hyperbolic)}. \end{cases} \quad (11)$$

Combining with the time part, we obtain the full **Friedmann–Lemaître–Robertson–Walker (FLRW) metric**.³

$$\boxed{ds^2 = -dt^2 + a^2(t)[d\chi^2 + S_k^2(\chi) d\Omega^2]} \quad (12)$$

³Also known as Friedmann-Robertson-Walker, Robertson-Walker, or really any combination of F, L, R, W.

This is the most general spacetime metric consistent with homogeneity and isotropy, and will form a central role in these lectures, as the background metric for our Universe. It also defines the *proper time* coordinate, via $d\tau^2 \equiv -ds^2$ (encoding the time difference at fixed spatial coordinates).

The FLRW metric is often written in two alternative choices of coordinates. First, we can use the standard spherical polar coordinates: r, θ, ϕ . Defining $r \equiv S_k(\chi)$ (such that the angular piece takes the canonical form: $r^2 d\Omega^2$) and doing some mild trigonometry, this gives

$$ds^2 = -dt^2 + a^2(t) \left[\frac{dr^2}{1 - kr^2/R_0^2} + r^2 d\Omega^2 \right]. \quad (13)$$

where it is clear that $r \leq R_0$ for a closed Universe. Secondly, one can define **conformal time** via $d\eta \equiv dt/a$, such that the scale factor multiplies the entire metric:

$$ds^2 = a^2(\eta) [-d\eta^2 + d\chi^2 + S_k^2(\chi) d\Omega^2]. \quad (14)$$

This is useful because the causal structure becomes transparent. In particular, photons travel on null curves satisfying $ds = 0$: a radial photon thus has $d\eta = d\chi$, implying that it propagates on straight lines at 45° in the (η, χ) plane, just as in Minkowski space.

2.4 Hubble Parameter and Hubble's Law

With the FLRW metric in hand, we can define several important physical quantities. The **scale factor** $a(t)$ describes how physical distances stretch with the expansion. For an object at fixed comoving distance χ from the observer, the **physical distance** is $d_{\text{phys}}(t) = a(t)\chi$ – this describes the true distance between the observer and object if we could ‘freeze’ the expansion of the Universe and lay down a (rather long) ruler.

The **Hubble parameter** is the logarithmic time derivative of the scale-factor:

$$H(t) \equiv \frac{\dot{a}}{a}, \quad (15)$$

where $\dot{X} \equiv dX/dt$. This describes the expansion rate of the Universe. The aforementioned object at comoving distance χ is transported away from the observer as the Universe expands: this occurs at a physical velocity:⁴

$$V_{\text{phys}}(t) = \dot{a}(t) \chi = \frac{\dot{a}}{a} \cdot a\chi = H(t) d_{\text{phys}}(t). \quad (16)$$

This is **Hubble's law**: the recession velocity is proportional to the physical distance:

$$V = H d. \quad (17)$$

The present-day value of Hubble is $H_0 \equiv H(t_0) \approx 100 h \text{ km/s/Mpc}$ with $h \approx 0.67$.⁵ Hubble's law is a direct consequence of the homogeneous expansion of space and holds for any pair of comoving observers.

⁴We ignore peculiar motions, which introduce $V_{\text{phys}}(t) \supset a(t)\dot{\chi}(t) \equiv V_{\text{pec}}(t)$.

⁵The value of h is hotly debated. This is (in part) the reason for the appearance of h in many cosmological units.

2.5 Geodesics

Above, we have specified the background geometry of the Universe. How do particles move in it? As we learnt in General Relativity, free particles follow geodesics, which are the relativistic analog of straight lines in Minkowski space. The geodesic is defined as the ‘shortest path’ between two points (or formally, that that extremizes the spacetime interval). To define this, we note that the distance between points A and B can be naturally extracted from the metric:

$$S_{AB} = \int_A^B ds = \int_{\lambda_A}^{\lambda_B} \sqrt{-g_{\mu\nu} \frac{dx^\mu}{d\lambda} \frac{dx^\nu}{d\lambda}} d\lambda \quad (18)$$

introducing some affine parameter λ labelling how far we are along the curve $x^\mu(\lambda)$. Extremizing with respect to $x^\mu(\lambda)$ (via Euler-Lagrange) gives the **geodesic equation**:

$$\frac{d^2 x^\mu}{d\lambda^2} + \Gamma_{\alpha\beta}^\mu \frac{dx^\alpha}{d\lambda} \frac{dx^\beta}{d\lambda} = 0, \quad (19)$$

which depends on the **Christoffel symbols** (connection coefficients).⁶

$$\Gamma_{\alpha\beta}^\mu = \frac{1}{2} g^{\mu\rho} (\partial_\alpha g_{\rho\beta} + \partial_\beta g_{\rho\alpha} - \partial_\rho g_{\alpha\beta}). \quad (20)$$

Note the presence of the inverse metric: $g^{\mu\nu}$. This satisfies $g^{\mu\nu} g_{\nu\lambda} = \delta_\lambda^\mu$, with $g^{\mu\nu} \neq g_{\mu\nu}$ in general. A simple example of geodesic equations is provided by Newtonian gravity: this has the metric $ds^2 = -(1 + 2\Phi)dt^2 + (1 - 2\Phi)d\vec{x}^2$ (collapsing to Minkowski space if $\Phi \rightarrow 0$), which leads to the geodesic equations $d^2 x^i / dt^2 = -\partial_i \Phi$, matching Newton.

For the FLRW metric, the only non-vanishing Christoffel symbols are

$$\Gamma_{ij}^0 = a\dot{a}\gamma_{ij}, \quad \Gamma_{0j}^i = \frac{\dot{a}}{a}\delta_j^i, \quad \Gamma_{jk}^i = \frac{1}{2}\gamma^{il} (\partial_j \gamma_{kl} + \partial_k \gamma_{jl} - \partial_l \gamma_{jk}). \quad (21)$$

For a flat Universe in t, x, y, z coordinates, Γ_{jk}^i vanishes, since $\gamma_{ij} = \delta_{ij}$ is coordinate-independent.

We can use the geodesic equation to understand the propagation of photons. First, we note that the 4-momentum of photons is given by $P^\mu = \{E, p^i\}$, where E is the energy and p^i is its 3-momentum. This can be explicitly related to the photon trajectory via $P^\mu \equiv \frac{dx^\mu}{d\lambda}$, where $\lambda(t)$ parametrizes the time along the curve.⁷ From the geodesic equation, we find

$$\frac{dP^\mu}{d\lambda} + \Gamma_{\alpha\beta}^\mu P^\alpha P^\beta = 0. \quad (22)$$

Writing $d/d\lambda = (dx^0/d\lambda)d/dx^0 = P^0 d/dx^0$ we find

$$P^0 \frac{dP^\mu}{dx^0} + \Gamma_{\alpha\beta}^\mu P^\alpha P^\beta = 0. \quad (23)$$

Further setting $\mu = 0$ (with $P^\mu \equiv E$) and inserting the Christoffel symbols for the FLRW metric, we arrive at

$$E \frac{dE}{dt} = -\Gamma_{ij}^0 p^i p^j = -a\dot{a} p^i p^j \delta_{ij} \quad (24)$$

⁶If you haven’t seen these before: fear not! They are just matrices describing how the metric changes with respect to coordinates.

⁷For massive particles, we have a similar relation: $P^\mu = m dx^\mu / d\tau$ for proper time τ .

Finally, we note that, since photons propagate on null geodesics, $g_{\mu\nu}P^\mu P^\nu = 0$ thus $E^2 = a^2 \vec{p} \cdot \vec{p}$. This implies that

$$\frac{1}{E} \frac{dE}{dt} = -\frac{1}{a} \frac{da}{dt} \Rightarrow E \propto a^{-1}. \quad (25)$$

This proves that the energy of a photon decays with the expansion of the Universe.

Consider a photon emitted at some scale-factor a_{em} with frequency ν_{em} (with $E \propto \nu$). Due to the expansion of the Universe, this will be observed to have frequency $\nu_{\text{obs}} < \nu_{\text{em}}$ at some later-time with scale-factor a_{obs} . Via the above argument

$$\frac{\nu_{\text{obs}}}{\nu_{\text{em}}} = \frac{a_{\text{em}}}{a_{\text{obs}}}. \quad (26)$$

Defining the *redshift* as the fractional decrement in the photon wavelength $\lambda \propto \nu^{-1}$ ($z \equiv \lambda_{\text{obs}}/\lambda_{\text{em}} - 1$), and setting $a_{\text{obs}} = 1$ (today), we find

$$\boxed{a_{\text{em}} = (1 + z_{\text{em}})^{-1}}. \quad (27)$$

This implies that the scale-factor is uniquely determined by the cosmological redshift, which itself encodes the expansion-induced change in energy of a photon traversing from that time until today. We'll use z , t and a as time-coordinates interchangeably going forward.

2.6 Cosmological Distances

The above discussion raises an important question: how do we define ‘distance’ in cosmology? This is a bit subtle: we can’t simply lay down rulers and measure since the Universe (and potentially our rulers) is expanding. There’s a proliferation of different definitions for distance that crop up in cosmology: here are the five most important ones:

1. **Comoving distance:** This is the (comoving) distance a photon travels from a source at time t to the observer at time t_0 (today). Since photons travel on null geodesics ($ds^2 = 0$, giving $d\chi = dt/a = d\eta$), the comoving distance can be computed as

$$\chi = \int_{\chi(t)}^{\chi(t_0)} d\chi = \int_{t_{\text{em}}}^{t_0} \frac{dt}{a(t)}. \quad (28)$$

Using $H(t) = d \log a / dt$ and $a = (1 + z)^{-1}$, this can be rewritten purely in terms of the expansion history:

$$\chi(z) = \int_0^z \frac{dz'}{H(z')}. \quad (29)$$

This is a key result: given $H(z)$, we can compute the comoving distance to any redshift.

2. **Physical (proper) distance:** $d_{\text{phys}}(t) = a(t) \chi$, where χ is the **comoving distance**. This is the ‘true’ distance between two objects (defined from the spatial metric: $d\ell = a d\chi$), *i.e.* the quantity you’d measure if you froze the expansion of the Universe and laid down some rulers.
3. **Luminosity distance:** A classical way to measure distances involves ‘standard candles’. If I know the intrinsic luminosity of an object, L , and I measure the flux, F , I can obtain the

distance d by using the inverse-square law: $F = L/(4\pi d_L^2)$. In an FLRW Universe, this is modified to

$$F = \frac{L}{4\pi a^2(t_0) S_k^2(\chi[z])} \cdot \frac{1}{1+z} \cdot \frac{1}{1+z} \quad (30)$$

where $4\pi a^2(t_0) S_k^2(\chi)$ is the proper surface area of the sphere of photons from comoving distance $\chi(z)$ at time t_0 (from the angular part of the metric), and the two $1+z = a(t_0)/a(t_{\text{em}})$ factors accounts for (1) the reduction in the arrival rate of photons due to expansion; (2) the loss of energy of photons due to expansion (as derived above). Comparing to the classical result, we find the luminosity distance

$$d_L = (1+z) S_k(\chi), \quad (31)$$

(since $a(t_0) = 1$), which reduces to $(1+z)\chi$ if the Universe is flat.

4. **Angular diameter distance:** We can also measure distances using ‘standard rulers’. In Euclidean space, an object with known physical size D and distance d_A subtends an angle $\delta\theta \approx D/d_A$ in the sky; measuring $\delta\theta$ at known D thus gives us an estimate for d_A . For the FLRW equivalent, we note that, upon emission at time t_{em} , the physical distance between two photons at opposite sides of a source is given by $a(t_{\text{em}}) S_k(\chi) \delta\theta$ (from the metric), which must be equal to D . Comparing to the Euclidean case, we find

$$\boxed{d_A = a(t_{\text{em}}) S_k(\chi[z]) = S_k(\chi[z])/(1+z)} \quad (32)$$

Note that the luminosity and angular diameter distances are connected by the **distance duality relation** (Etherington reciprocity):

$$d_L = (1+z)^2 d_A; \quad (33)$$

this is a surprisingly general result (assuming only that photons travel on null geodesics and that their number density is conserved).

A few more distance measures will also prove useful.

1. **Hubble radius:** $d_H = 1/H$. This is the characteristic distance scale of the expansion: from the definition of physical velocity, objects at $d \gtrsim d_H$ are receding superluminally.⁸ We will use this quantity a lot when discussing inflation.
2. **Particle horizon:** This is the maximum comoving distance from which a photon could have reached us since the Big Bang ($t = 0$), *i.e.* the size of the observable Universe at time t . From the metric:

$$\chi_{\text{hor}}(t) = \int_0^t \frac{dt'}{a(t')} = \eta(t), \quad (34)$$

which is equal to the conformal time. The physical horizon distance is $d_{\text{hor}}(t) = a(t) \chi_{\text{hor}}(t) = a(t) \eta(t)$. Today, this is approximately equal to $\chi_{\text{hor}}(t_0) \approx 14\,000$ Mpc, but it is growing as the Universe expands and more of it becomes causally connected (*i.e.* able to interact via light rays).

⁸This does not violate special relativity, since the recession is due to the expansion of space itself, not motion *through* space.

3. **Event horizon:** This is the maximum comoving distance from which a photon emitted at time t will reach us in the future. From the metric

$$\chi_{\text{eh}}(t) = \int_t^\infty \frac{dt'}{a(t')} = \eta(\infty) - \eta(t). \quad (35)$$

Note the difference between the particle and event horizons: anything outside the *particle* horizon we don't know about *today* (photons haven't reached us), but anything outside the *event* horizon we will *never* know about (photons will never reach us). In the standard cosmological model, $\chi_{\text{eh}}(t_0) \approx 5000$ Mpc.

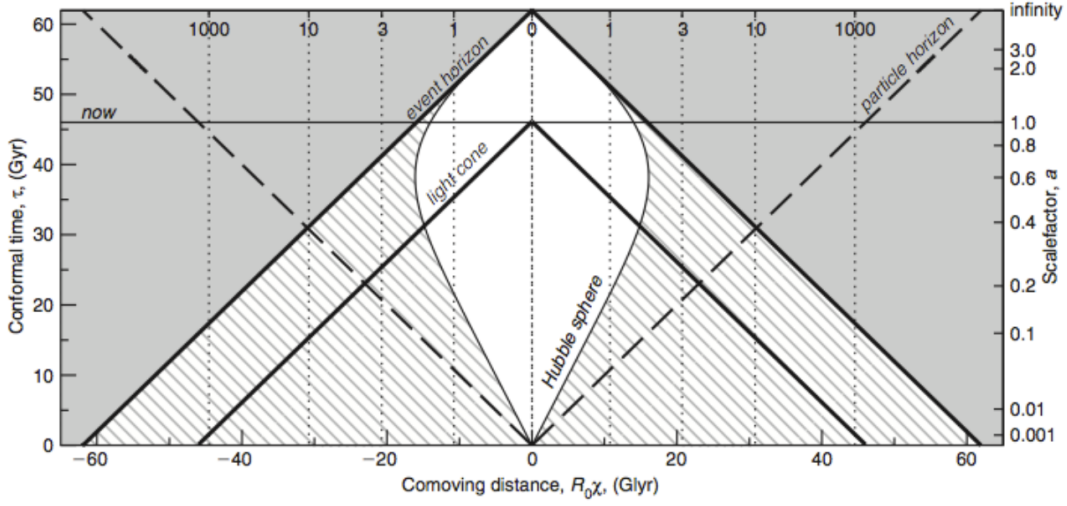


Figure 1: Space-time diagram (reproduced from Davis & Lineweaver 2004). This shows the interplay of the comoving distance and conformal times (here labelled η) for an observer at $\chi = 0$, where $\eta = 0$ is the end of inflation, and $\eta \approx 60$ Gly is the asymptotic future. Regions outside the particle horizon have never been in contact with us in the past; regions outside the event horizon will never be in contact with us in the future.

2.7 Curvature & the Einstein Equation

We've now described the background geometry of the Universe and understood how particles and photons traverse it. But there's still a key missing ingredient: the expansion history, $H(t)$ (and thus the scale factor $a(t)$). To determine the dynamics of the expansion, we can use the Einstein equation, which states

$$G_{\mu\nu} = 8\pi G T_{\mu\nu}. \quad (36)$$

The left-hand side is the **Einstein tensor**, which describes the curvature of space. The right-hand side is the **stress-energy** (also energy-momentum) tensor, which we'll define below.⁹ To solve for $a(t)$ we'll need to evaluate both sides using the FLRW metric.

⁹If you've not seen this before, don't panic. It's just a set of 10 equations (one for each μ, ν , with $G_{\mu\nu} = G_{\nu\mu}$) linking derivatives of the metric (left hand side) with the density and pressure of the Universe (right hand side).

Let's start with the Einstein tensor. This is defined in terms of the Ricci tensor, $R_{\mu\nu}$ (which is itself a contraction of the Riemann tensor: $R^\lambda_{\mu\lambda\nu}$), as

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R \quad (37)$$

where $R \equiv g^{\mu\nu} R_{\mu\nu}$. In terms of Christoffel symbols:

$$R_{\mu\nu} \equiv \partial_\lambda \Gamma^\lambda_{\mu\nu} - \partial_\nu \Gamma^\lambda_{\mu\lambda} + \Gamma^\lambda_{\lambda\rho} \Gamma^\rho_{\mu\nu} - \Gamma^\rho_{\mu\lambda} \Gamma^\lambda_{\nu\rho} : \quad (38)$$

this uniquely defines $G_{\mu\nu}$ in terms of the metric and its derivatives. For the FLRW metric, there are only two non-vanishing contributions:

$$R_{00} = -3 \frac{\ddot{a}}{a}, \quad R_{ij} = \left(a\ddot{a} + 2\dot{a}^2 + \frac{2k}{R_0^2} \right) \gamma_{ij} \quad \Rightarrow \quad R = 6 \left[\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a} \right)^2 + \frac{k}{a^2 R_0^2} \right]. \quad (39)$$

where γ_{ij} is the spatial metric.¹⁰ From this it is straightforward to obtain the Einstein tensor, $G_{\mu\nu}$.

2.8 Stress-Energy Tensor

Next, we must specify $T_{\mu\nu}$, which encodes the particle content of the Universe. Typically, this is done by assuming that (on sufficiently large scales) the Universe can be described by a fluid. $T_{\mu\nu}$ can be written in terms of the hydrodynamic properties of this fluid:¹¹

$$T_{\mu\nu} = \begin{pmatrix} T_{00} & T_{0i} \\ T_{i0} & T_{ij} \end{pmatrix} = \begin{pmatrix} \text{mass-energy density} & \text{energy flux / momentum density} \\ \text{energy flux / momentum density} & \text{stress-tensor} \end{pmatrix} \quad (40)$$

where the ij components (the stress-tensor) encode the forces between fluid elements including pressure and viscosity.

The assumptions of homogeneity and isotropy greatly restrict the form of $T_{\mu\nu}$. In a frame comoving with the fluid, homogeneity implies that $T_{\mu\nu}$ must be invariant under translations thus $T_{\mu\nu}(t, \vec{x}) = T_{\mu\nu}(t)$. By definition, T_{00} describes the energy density, thus $T_{00} = \rho(t)$. By isotropy, T_{i0} and T_{0j} must vanish (there are no preferred directions, since we are in a rest frame) and T_{ij} must be proportional to the spatial metric: $T_{ij} = P(t)g_{ij}$, where $P(t)$ is the fluid pressure (this is the isotropic force per unit area exerted by a fluid). In combination, we find that $T_{\mu\nu} = \text{diag}(\rho, Pa^2, Pa^2, Pa^2)$ (in t, x, y, z coordinates).

To use the Einstein equation, we require a form for $T_{\mu\nu}$ that is valid in all frames, not just the fluid's rest frame. The general covariant form is given by

$$T^{\mu\nu} = (\rho + P) U^\mu U^\nu + P g^{\mu\nu}, \quad (41)$$

where U^μ is the 4-velocity of the fluid relative to the observer. This satisfies homogeneity and isotropy explicitly (noting that U^μ and $g^{\mu\nu}$ are the only vectors and tensors in the problem), and reduces to the above form if $U^\mu = (1, 0, 0, 0)$ (as appropriate for a comoving observer).

¹⁰A simpler scenario is the weak-field Newtonian metric: $ds^2 = -(1 + 2\Phi)dt^2 + (1 - 2\Phi)dx^2$. This has $G_{00} = 2\nabla^2\Phi$; noting that $T_{00} = \rho$, this leads to the Poisson equation: $\nabla^2\Phi = 4\pi G\rho$.

¹¹The stress-energy tensor can also be rigorously defined from variational principles, given an action $S[g_{\mu\nu}]$.

2.9 Continuity Equation

In General Relativity, the stress-energy tensor is conserved: $\nabla_\mu T^\mu_\nu = 0$.¹² Explicitly inserting the definition of ∇_μ :

$$\nabla_\mu T^\mu_\nu = \partial_\mu T^\mu_\nu + \Gamma^\mu_{\mu\lambda} T^\lambda_\nu - \Gamma^\lambda_{\mu\nu} T^\mu_\lambda = 0, \quad (42)$$

which gives a set of four equations (for $\nu = 0, 1, 2, 3$). These describe conservation of energy and momentum, analogously to the continuity and Euler equations in fluid mechanics. Noting that $T^i_0 = 0$, the $\nu = 0$ component simplifies to

$$\partial_0 T^0_0 + \Gamma^\mu_{\mu 0} T^0_0 - \Gamma^\lambda_{\mu 0} T^\mu_\lambda = 0, \quad (43)$$

Inserting the Christoffel symbols for the FLRW metric, alongside the comoving-observer definition $T^\mu_\nu = \text{diag}(-\rho, P, P, P)$, we obtain the **continuity equation**:

$$\boxed{\dot{\rho} + 3H(\rho + P) = 0.} \quad (44)$$

This encodes conservation of energy in an expanding background and plays a crucial role in cosmology. Physically, the $3H\rho$ term represents the dilution of energy density from the expanding volume ($V \propto a^3$, so $\dot{V}/V = 3H$), and the $3HP$ term captures the thermodynamic PdV work done by the pressure.

2.10 Friedmann Equations

Let's return to the Einstein equation armed with our knowledge of the energy-momentum tensor. The $(0, 0)$ component of the Einstein equation (in an (t, x, y, z) coordinate system) gives

$$G_{00} = R_{00} + \frac{1}{2}R = 3\left(\left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2 R_0^2}\right) = 8\pi G \rho, \quad (45)$$

noting that $g_{00} = -1$ and $T_{00} = \rho$. Inserting the FLRW definitions of the Ricci tensor, we find the **first Friedmann equation**:

$$\boxed{H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2 R_0^2}.} \quad (46)$$

This relates the expansion rate H to the total energy density ρ and the spatial curvature k/R_0^2 .¹³ This is a first-order differential equation in time: by combining with the continuity equation (which defines $\rho(t)$), we can solve for $a(t)$.

The (i, j) component of the Einstein equation gives $R_{ij} - \frac{1}{2}g_{ij}R = 8\pi G P g_{ij}$, leading to

$$2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2 R_0^2} = -8\pi G P \quad (47)$$

Combining with the first Friedmann equation, we find the **second Friedmann equation**:

$$\boxed{\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P),} \quad (48)$$

¹²This involves the covariant derivatives, which generalize the usual partial derivatives (∂_μ) to arbitrary coordinate systems. Also note that $T^\mu_\nu = g^{\mu\lambda} T_{\lambda\nu}$.

¹³This can also be derived without assuming GR. See Baumann or Huterer for a nice derivation.

which defines the acceleration of the Universe. Notably, the expansion *decelerates* ($\ddot{a} < 0$) when $\rho + 3P > 0$, which holds for ordinary matter and radiation (and indeed any species obeying the strong equivalence principle). The expansion *accelerates* ($\ddot{a} > 0$) when $\rho + 3P < 0$. Notably, only two of the three evolution equations are independent: differentiating the first Friedmann equation with respect to time and inserting the continuity equation yields the second Friedmann equation.

2.11 The Evolution of Species

To close the set of equations, we require a relation between ρ and P . Most cosmological species (matter, radiation, et cetera) can be described by a constant **equation-of-state**:

$$\boxed{P(t) = w\rho(t)} \quad (49)$$

From the continuity equation

$$\dot{\rho} + 3H(1+w)\rho = 0 \quad \Rightarrow \quad \frac{\dot{\rho}}{\rho} = -3(1+w)\frac{\dot{a}}{a} = 0 \quad \Rightarrow \quad \boxed{\rho \propto a^{-3(1+w)}}. \quad (50)$$

There are three interesting types of matter:

1. **Matter.** For non-relativistic matter (also known as ‘dust’), $P_m \ll \rho_m$, thus $w_m \approx 0$. The continuity equation implies that $\rho_m \propto a^{-3}$, *i.e.* the density dilutes with the expansion of the Universe (as $V \propto a^3$). Here, matter includes baryons, dark matter, and, if non-relativistic, neutrinos.
2. **Radiation.** For photons and relativistic matter, $P_r = \rho_r/3$, thus $w_r = 1/3$.¹⁴ From the continuity equation $\rho_r \propto a^{-4}$. In addition to dilution (a^{-3}), each photon loses energy due to redshifting ($\propto 1/a$), giving the extra power of a . In the early Universe, neutrinos are relativistic, thus act as radiation.
3. **Dark Energy:** This is a mysterious force invoked to explain the accelerating expansion of the Universe. In the standard model (featuring a cosmological constant, Λ), $P_\Lambda = -\rho_\Lambda$, thus $w_\Lambda = -1$, and $\rho_\Lambda \propto a^0$, implying that the density is constant in time (being a property of space itself). Given our lack of knowledge of what dark energy actually is, it has become popular (though not necessarily well-motivated) to test whether dark energy has a non-standard (and possibly varying) equation-of-state: $w_\Lambda \neq -1$.

Note that the Friedmann equations depend on the total density and pressure of matter: $\rho_{\text{tot}} = \sum_i \rho_i$ and $P_{\text{tot}} = \sum_i P_i$, whilst the continuity equation holds for each species individually.

There are two equivalent ways to include canonical dark energy (with $w = -1$) in the cosmological model. First, we can treat it as a physical source of stress-energy, including it in the density and pressure budget as above. Secondly, we can add it directly to the Einstein equations, via the modification:

$$G_{\mu\nu} \rightarrow G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}, \quad (51)$$

for constant Λ . The second approach is the one originally advocated for by Einstein (in his infamous ‘biggest blunder’); but the first approach is more common in the contemporary literature, so we’ll assume it henceforth.

¹⁴This is a standard result; it can be derived by noting that the pressure is equal to the momentum flux, thus $P = \frac{1}{3}n\langle pv \rangle$ for momentum p and velocity v , where $1/3$ arises from averaging over direction. Setting $v = c = 1$ and $p = E$, with $nE = \rho$, we find the desired result.

2.12 The Evolution of the Universe

Let's write the first Friedmann equation in terms of the densities of matter, radiation, and dark energy:

$$H^2 = \frac{8\pi G}{3} (\rho_{r,0}a^{-4} + \rho_{m,0}a^{-3} + \rho_{\Lambda,0}) - \frac{k}{a^2 R_0^2}, \quad (52)$$

where $\rho_{i,0}$ is the density of the i -th species today (with $a(t_0) = 1$). Since the right-hand side now depends only on a , it is straightforward to solve for $a(t)$. Before doing so, it is useful to rescale the species by the **critical density** $\rho_c(t) \equiv 3H(t)^2/(8\pi G)$. This is the energy required for a flat Universe ($k = 0$): if $\rho_{\text{tot}} < \rho_c$, the Universe will expand forever, whilst if $\rho_{\text{tot}} > \rho_c$ it will collapse. Defining the relative density parameters $\Omega_i \equiv \rho_{i,0}/\rho_c(t_0)$, we find

$$\boxed{\frac{H^2}{H_0^2} = \frac{\Omega_r}{a^4} + \frac{\Omega_m}{a^3} + \frac{\Omega_k}{a^2} + \Omega_\Lambda.} \quad (53)$$

where we have also defined an effective energy-density in curvature: $\Omega_k \equiv -\frac{k}{R_0^2 H_0^2}$. Note the negative sign: $\Omega_k < 0$ implies a positively curved closed Universe, whilst $\Omega_k > 0$ implies a negatively curved open Universe. By definition, the fractional energy densities sum to unity:

$$\Omega_m + \Omega_r + \Omega_\Lambda + \Omega_k = 1, \quad (54)$$

thus the Universe is flat if $\Omega_m + \Omega_r + \Omega_\Lambda = 1$. In the consensus cosmological model (mostly driven by CMB observations), we find $\Omega_m \approx 0.3$, $\Omega_\Lambda \approx 0.7$ and $\Omega_r \sim 10^{-4}$, with negligible curvature ($|\Omega_k| \lesssim 10^{-2}$).

Expansion History The Friedmann equation represents a balance between different cosmological species: at early times ($a \ll 1$), radiation dominates the expansion, before being taken over by matter, then curvature (if present), and finally Λ . Given this competition of scales it is difficult one must solve for $a(t)$ numerically: here, we will consider a few limiting regimes.

1. **Radiation domination:** At early times, the energy density is dominated by radiation, thus $\rho(a) \propto a^{-4}$ and $H = \dot{a}/a \propto a^{-2}$. This can be solved to find $a(t) \propto t^{1/2}$. In conformal time, with $d\eta = dt/a$ from the FLRW metric, we have $H = a^{-2}da/d\eta$ thus $a(\eta) \propto \eta$.
2. **Matter domination:** Here, $\rho(a) \propto a^{-3}$, leading to $a(t) \propto t^{2/3}$ and $a(\eta) \propto \eta^2$. This is the famous 'Einstein de-Sitter' solution.
3. **Dark energy domination:** At late times, $\rho(a)$ becomes constant, thus $a(t)$ grows exponentially with time and $a(\eta) \propto -1/\eta$. This is the 'de Sitter' solution, which will crop-up again when we consider inflation.

The full solution for $a(t)$ interpolates between these three regimes. The transition from radiation to matter domination occurs at **matter radiation equality**, with $\rho_m(a) = \rho_r(a)$. Solving for a (and thus z) we find $1 + z_{\text{mr}} = \Omega_m/\Omega_r \approx 3500$. Similarly, the transition from matter to dark energy domination occurs at $1 + z_{\Lambda m} = (\Omega_\Lambda/\Omega_m)^{1/3} \approx 1.3$.

Age Using $a(t)$, we can compute the age of the Universe. Writing $dt = da/(aH)$ and integrating from the Big Bang to today, we find:

$$t_0 = \int_0^1 \frac{da}{a H(a)} = \frac{1}{H_0} \int_0^1 \frac{da}{a \sqrt{\Omega_r a^{-4} + \Omega_m a^{-3} + \Omega_k a^{-2} + \Omega_\Lambda}}, \quad (55)$$

For the Einstein de-Sitter Universe, we find $t_0 = \frac{2}{3H_0} \approx 9 \text{ Gyr}$. This is substantially less than the age of the oldest globular clusters, which provided some early evidence for dark energy. Inserting the observed Λ CDM parameters and integrating numerically, we find $t_0 \approx 13.8 \text{ Gyr}$.

Deceleration We can quantify whether the expansion is speeding up or slowing down via the deceleration parameter:

$$q(t) \equiv -\frac{\ddot{a} a}{\dot{a}^2}. \quad (56)$$

A positive q means deceleration; a negative q means acceleration. For a single component universe with equation-of-state w : $q = (1 + 3w)/2$, implying that $w \leq -1/3$ is needed to drive acceleration (which is clear from the second Friedmann equation). Observations of Type Ia supernovae indicate that $q_0 < 0$ today, providing direct evidence for dark energy and an accelerating expansion.

The Future The solutions for $a(t)$ derived above predict qualitatively different expansion histories depending on the matter content and curvature. For example, a matter only Universe without curvature has $a \propto t^{2/3}$ and $q > 0$: this implies a decelerating but forever-expanding Universe. Open universes without dark energy ($\Omega_k > 0$) expand forever, whilst closed universes without dark energy ($\Omega_k > 0$) will expand to a maximum size, then recollapse via a ‘Big Crunch’.¹⁵ In the presence of a cosmological constant, the Universe will expand exponentially in the asymptotic future, regardless of curvature.

3 Thermal History of the Early Universe

We now turn to the thermal history of the early Universe: the sequence of events that unfolds as the Universe cools from extremely high temperatures ($T \gg 100 \text{ GeV}$) to around 0.1 eV , including recombination, nucleosynthesis and beyond. At these energy scales, the Universe is filled with a hot, dense bath of particles and radiation, interacting through a variety of chemical processes. To describe this, we will use a selection of tools from statistical mechanics, which we’ll recap below. The key epochs that we’ll describe in this section are the following:

Event	Energy	Time
EW Phase Transition	100 GeV	10^{-10} s
QCD Phase Transition	150 MeV	10^{-4} s
ν decoupling	1 MeV	1 s
$e^\pm$ annihilation	0.5 MeV	6 s
BBN	100 keV	3 min
Recombination	0.25 eV	300 kyr
Photon decoupling	0.25 eV	380 kyr

¹⁵To see this, note that $H = 0 \Rightarrow \Omega_m/a^3 + \Omega_k/a^2 = 0$, which has valid solutions only for $\Omega_k < 0$.

3.1 A Primer on Statistical Mechanics

Distribution Functions The key quantity characterizing the behavior of the various species present in the early Universe is the distribution function $f(\vec{x}, \vec{p}, t)$. This gives the probability of finding a particle at position $\vec{x}$ with momentum $\vec{p}$ at time t (per unit phase-space volume). By homogeneity and isotropy, this can only depend on $p = |\vec{p}|$ and t , though we'll often parametrize the time coordinate by T , the temperature of the photon bath.

To connect the distribution function to observables such as densities and pressures, we must specify the phase-space density of states (which tells us how to ‘sum up’ the possible values of momentum). Since we set $\hbar = 1$, this is given by $g/(2\pi)^3$, where g is the **degeneracy factor**, which counts the number of internal degrees of freedom (*e.g.* a massive spin- s fermion has $g = 2s + 1$ and a photon has $g = 2$, due to the two polarization states).¹⁶ Given the phase-space density, we can define three quantities of interest: the number density, energy density, and pressure:

$$\begin{aligned} n(t) &= \frac{g}{(2\pi)^3} \int d^3p f(p, t) \\ \rho(t) &= \frac{g}{(2\pi)^3} \int d^3p E(p) f(p) \\ P(t) &= \frac{g}{(2\pi)^3} \int d^3p \frac{p^2}{3E(p)} f(p, t), \end{aligned} \tag{57}$$

where $E(p) = \sqrt{p^2 + m^2}$ is the relativistic energy.¹⁷

Equilibrium Distributions In thermal equilibrium, the distribution function for some species i takes a simple form:

$$f_i(p, T_i) = \frac{1}{e^{(E_i(p) - \mu_i)/T_i} \pm 1}, \tag{58}$$

where $+$ is for fermions (the Fermi–Dirac distribution) and $-$ is for bosons (the Bose–Einstein distribution). Here T_i is the temperature and μ_i is the chemical potential, which we discuss more below. In the classical limit ($E_i - \mu \gg T_i$), both reduce to the Boltzmann distribution: $f_i \rightarrow e^{-(E_i - \mu_i)/T_i}$ with $E_i \approx m_i$. Coupled with (57), this uniquely sets the number density, energy density, and pressure of a given particle species. Notably, different species can have different temperatures, despite remaining in thermal (but not chemical) equilibrium.

Chemical Potential Thermodynamically, the chemical potential is defined as $\mu \equiv (\partial E / \partial N)_{S, V}$: the (Helmoltz) energy cost of adding one particle at fixed entropy and volume. This appears primarily to ‘balance the books’, ensuring that the relevant conservation laws are obeyed (*e.g.*, ensuring that the fractional excess of electrons over positrons is held fixed). This has a few useful properties:

¹⁶This can be justified as follows. Consider a cubic box of size L . For each internal degree of freedom, the allowed 3-momentum values are $\vec{p} = (h/L)(N_1\hat{x} + N_2\hat{y} + N_3\hat{z})$ for integer N_i . This gives a momentum-space density $(L/h)^3$, and thus a phase space density $1/(2\pi)^3$ in natural units.

¹⁷To derive the pressure formula, note that pressure is, by definition, the rate of change of momentum per unit area with respect to time, *i.e.* $\delta p / (A \delta t)$. The elastic scattering of a single particle moving in the $+x$ direction induces a momentum change $2p_x$; from the distribution function, the number of such events is $f(p, t) \delta V$, where $\delta V = A v_x \delta t = A p_x \delta t / E(p)$ for velocity v . Dividing by $A \delta t$, and averaging over the momentum direction, we obtain the above result.

1. For a reaction in equilibrium, denoted $X + Y \rightleftharpoons Z + W$, we have $\mu_X + \mu_Y = \mu_Z + \mu_W$.
2. For photons, $\mu_\gamma = 0$, since photon number is not conserved (since reactions like $e^- + \gamma \rightleftharpoons e^- + 2\gamma$ are allowed).
3. For a particle–antiparticle pair $X, \bar{X}$, the chemical potential satisfies, $\mu_X + \mu_{\bar{X}} = 0$ (since $X + \bar{X} \rightleftharpoons 2\gamma$, with $\mu_\gamma = 0$).

In the below, we'll generally ignore the chemical potential $\mu \ll T$, which is an excellent approximation for most species in the early Universe.¹⁸

Reaction Rates In this section, we'll be discussing a variety of chemical reactions, with the most important being annihilations, *i.e.* $X + \bar{X} \rightleftharpoons 2\gamma$. The importance of such reactions is set by an interplay of two factors: the interaction rate Γ and the expansion rate $H \equiv d \log a / dt$. The former is given by $\Gamma = n\sigma v$, where σ is the cross-section (set by the microphysics) and v is the velocity (usually 1 if relativistic). If $\Gamma \gg H$, we attain chemical equilibrium, whilst if $\Gamma \ll H$, the reactions are inefficient (since the particles move apart too fast to interact) and the species *decouples* from the bath of photons. After decoupling at some time t_{dec} , the distribution function of the particles is frozen, with $f(p_i, t) = f(p_{\text{dec}} a(t) / a(t_{\text{dec}}), t_{\text{dec}})$, recalling that three-momentum satisfies $p \propto 1/a$.

3.2 Equilibrium Thermodynamics

3.2.1 Number Densities and Energy Densities

At sufficiently high energies ($T > 100 \text{ GeV}$), all standard model particles are in thermal equilibrium, with distribution functions described by (58). Explicitly, their number densities and energy densities satisfy

$$\begin{aligned} n_i(T_i) &= \frac{g_i}{2\pi^2} \int_0^\infty dp \frac{p^2}{e^{\sqrt{p^2 + m_i^2}/T_i} \pm 1} \\ \rho_i(T_i) &= \frac{g_i}{2\pi^2} \int_0^\infty dp \frac{p^2 \sqrt{p^2 + m_i^2}}{e^{\sqrt{p^2 + m_i^2}/T_i} \pm 1}, \end{aligned} \quad (59)$$

which depend only on their temperature T_i and mass m_i (assuming $|\mu_i| \ll T_i$, as above). In general, these integrals are difficult to solve numerically; however, they take simple forms in the ultra-relativistic $T_i \gg m_i$ and non-relativistic $T_i \ll m_i$ limits. To see this, it is useful to first change variables to $x = m_i/T_i$, $\xi = p/T_i$, such that the integrals become

$$n_i(T_i) = \frac{g_i T_i^3}{2\pi^2} \int_0^\infty d\xi \frac{\xi^2}{e^{\sqrt{\xi^2 + x^2}} \pm 1}, \quad \rho_i(T_i) = \frac{g_i T_i^4}{2\pi^2} \int_0^\infty d\xi \frac{\xi^2 \sqrt{\xi^2 + x^2}}{e^{\sqrt{\xi^2 + x^2}} \pm 1}, \quad (60)$$

¹⁸For relativistic fermions, it is straightforward to show that the excess density of fermions over antifermions satisfies $(n - \bar{n})/n \sim \mu/T \sim 10^{-9}$.

Relativistic Limit For $x \ll 1$ (*i.e.* $T_i \gg m_i$), we have¹⁹

$$\begin{aligned} \int_0^\infty d\xi \frac{\xi^2}{e^{\sqrt{\xi^2+x^2}} \pm 1} &\approx \int_0^\infty d\xi \xi^2 e^{-\xi} \left(1 \pm e^{-\xi}\right)^{-1} = \int_0^\infty d\xi \xi^2 \sum_{j=1}^\infty (\mp 1)^{j-1} e^{-j\xi} \\ &= \sum_{j=1}^\infty (\mp 1)^{j-1} \frac{2}{j^3} = 2 \left[\frac{1}{1^3} \mp \frac{1}{2^3} + \frac{1}{3^3} \mp \frac{1}{4^3} + \dots \right] \end{aligned} \quad (61)$$

For bosons (+ sign), the expression in the brackets is the Riemann-Zeta function $\zeta(3) \approx 1.202$. For fermions (− sign), we can rewrite the bracket as the difference of two sums:

$$\left[\frac{1}{1^3} - \frac{1}{2^3} + \frac{1}{3^3} - \frac{1}{4^3} + \dots \right] = \left[\frac{1}{1^3} + \frac{1}{2^3} + \frac{1}{3^3} + \frac{1}{4^3} + \dots \right] - \frac{2}{2^3} \left[\frac{1}{1^3} + \frac{1}{2^3} + \dots \right] \quad (62)$$

which gives $\frac{3}{4}\zeta(3)$. Thus we obtain the relativistic limit of the particle number density:

$$n_i(T_i) = \frac{\zeta(3)}{\pi^2} g_i T_i^3 \times \begin{cases} 1 & (\text{bosons}) \\ 3/4 & (\text{fermions}) \end{cases} \quad (63)$$

The energy density can be derived similarly, giving

$$\rho_i(T_i) = \frac{\pi^2}{30} g_i T_i^4 \times \begin{cases} 1 & (\text{bosons}) \\ 7/8 & (\text{fermions}) \end{cases}, \quad (64)$$

noting that $\zeta(4) = \pi^4/90$. The pressure can be derived also, and satisfies $P_i = \rho_i/3$, as expected.

Using this result, we can compute the energy density of photons in the Universe today (since, being massless, they are relativistic at all times). Inserting the current temperature of the CMB, $T_\gamma \approx 2.73$ K, we find $n_{\gamma,0} = 411 \text{ cm}^{-3}$, and a fractional energy density (see above) $\Omega_\gamma h^2 = 2.5 \times 10^{-5}$.

Non-Relativistic Limit For $x \gg 1$ (*i.e.* $T_i \ll m_i$), the integral controlling the number density becomes

$$\int_0^\infty d\xi \frac{\xi^2}{e^{\sqrt{\xi^2+x^2}} \pm 1} \approx \int_0^\infty d\xi e^{-\sqrt{\xi^2+x^2}} \approx e^{-x} \int_0^\infty d\xi e^{-\xi^2/(2x)}, \quad (65)$$

which we recognize as a Gaussian integral. Solving, we find the non-relativistic number density

$$n_i(T_i) = g_i \left(\frac{m_i T_i}{2\pi} \right)^{3/2} e^{-m_i/T_i}, \quad (66)$$

(where $e^{-m_i/T_i} \rightarrow e^{-(m_i-\mu_i)/T_i}$ if we keep the chemical potential). Since $E_i \approx m_i + p^2/(2m_i)$ in the non-relativistic limit, it is easy to show that the density and pressure are given by

$$\rho_i(T_i) \simeq m_i n_i(T_i), \quad P_i(T_i) = n_i(T_i) T_i. \quad (67)$$

Notably, all quantities contain the exponential factor e^{-m_i/T_i} (‘Boltzmann suppression’), implying that species with $m_i \gg T_i$ are extremely rare: their abundance is depleted by annihilation processes ($X + \bar{X} \rightleftharpoons 2\gamma$), which proceed efficiently as long as the species is in equilibrium. Moreover, the ratio of pressure to density is small: $P_i(T_i)/\rho_i(T_i) \approx m_i/T_i \ll 1$, justifying our previous assumptions about the equation of state of non-relativistic particles.

¹⁹Whilst we could have just stated these results, they’re more fun to prove, since Riemann-Zeta functions are rare to see in the wild.

3.2.2 Energy Density in the Primordial Plasma

An important lesson from the above calculations is that the energy density of the early Universe is dominated by *relativistic* species (those with $T_i \gg m_i$), since the abundance and densities of non-relativistic species are Boltzmann-suppressed. To obtain the total energy density of the early Universe, we can sum up the contributions from each relativistic species, *i.e.*

$$\rho(T) = \sum_i \rho_i(T_i), \quad (68)$$

where we note that different species can have their own temperatures, T_i (as we will see for neutrinos below). It is convenient to parameterize this in terms of the photon temperature $T \equiv T_\gamma$:

$$\boxed{\rho(T) = \frac{\pi^2}{30} g_*(T) T^4}, \quad (69)$$

where $g_*(T)$ is the **effective number of relativistic degrees of freedom**. Comparing expressions, this is given by

$$g_*(T) = \sum_{i: \text{bosons}} g_i \left(\frac{T_i}{T} \right)^4 + \frac{7}{8} \sum_{i: \text{fermions}} g_i \left(\frac{T_i}{T} \right)^4. \quad (70)$$

When all species share a common temperature ($T_i = T$), this simplifies to $g_* = \sum_{\text{bosons}} g_i + \frac{7}{8} \sum_{\text{fermions}} g_i$, where the factor of 7/8 comes from the different distribution functions for fermions and bosons.

At very high temperatures ($T > 100$ GeV), the entire Standard Model is in thermal equilibrium and relativistic.²⁰ The combined particle content is:

Particle	Type	g
Quarks (u, d, c, s, t, b)	Fermion	$6 \times 2 \times 3 \times 2 = 72$
Leptons ($e, \mu, \tau, \nu_e, \nu_\mu, \nu_\tau$)	Fermion	$3 \times 2 \times 2 + 3 \times 2 \times 1 = 18$
$W^\pm, Z^0$	Boson	$3 \times 3 = 9$
Photon (γ)	Boson	2
Gluons	Boson	$8 \times 2 = 16$
Higgs	Boson	1

where the factors account for spin, color, and particle/antiparticle degrees of freedom (remembering that each massive spin- s species contributes $(2s + 1)$ degrees of freedom). Note that neutrinos contribute only one degree-of-freedom for each of the six particle and antiparticles, since they are left-handed (*i.e.* they have only one helicity state). Summing these contributions:

$$g_*(T > 100 \text{ GeV}) = (9 + 2 + 16 + 1) + \frac{7}{8}(72 + 18) = 106.75, \quad (71)$$

which defines the energy density in the $T > 100$ GeV universe, assuming no additional species attained thermal equilibrium.

²⁰Since this is above the electroweak phase transition, particles have not yet acquired mass via the Higgs mechanism.

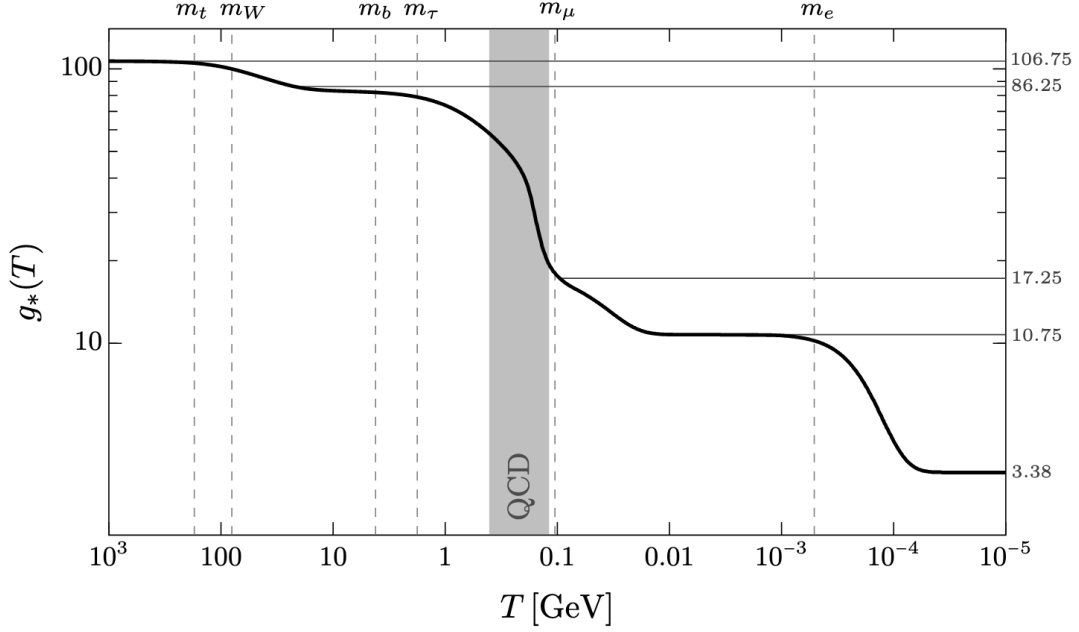


Figure 2: Evolution of the effective number of relativistic degrees of freedom in the early Universe (and thus the energy density). Reproduced from D. Baumann, *Cosmology* (Cambridge University Press, 2022).

As the Universe cools, species with $T < m_i$ become non-relativistic, annihilate, and decouple from the thermal bath (since the forwards reaction $X + \bar{X} \rightarrow 2\gamma$ becomes energetically favored over the reverse), reducing g_* in a step-wise fashion (assuming each transition happens quickly compared to the expansion rate of the Universe, which is usually true). This is shown in Fig. 2. For example, the top quark annihilates around 30 GeV, reducing g_* by $\frac{7}{8} \times 12$ units. Next, the Higgs and $W^\pm, Z^0$ gauge bosons bow out (giving $g_* = 86.25$), before the bottom, charm, and tau leptons. At the QCD phase transition ($T \sim 150$ MeV), quarks and gluons confine into hadrons, and g_* rapidly drops to 10.75 (following a shortlived period dominated by muons and pions, with $g_* = 17.25$). What happens to all of the energy contained in the formerly relativistic species? It is transferred to the remaining relativistic species that are still in equilibrium, reheating them. To understand this quantitatively (and the remainder of the Universe's thermal history), we require conservation laws, which will be provided by thermodynamics.

3.2.3 Entropy

The first law of thermodynamics gives a relationship between the entropy of the Universe, S , the internal energy, U , and the pressure P :

$$TdS = dU + PdV = d(\rho V) + PdV, \quad (72)$$

again ignoring the chemical potential. Introducing the **entropy density** $s = S/V$ and rearranging, we find

$$(\rho + P - Ts)\frac{dV}{V} = Tds - d\rho = \left(T\frac{ds}{dT} - \frac{d\rho}{dT}\right)dT, \quad (73)$$

since s, ρ can depend only on T . Since this must hold for independent variations in T and V , we find

$$s(T) = \frac{\rho(T) + P(T)}{T} \quad \text{and} \quad \left(T \frac{ds}{dT} - \frac{d\rho}{dT} \right) = 0, \quad (74)$$

where the first equation defines the entropy density. To simplify the second equation, we insert the continuity equation, which states that $\dot{\rho} = -3H(\rho + P)$, thus

$$\frac{ds}{dT} = \frac{1}{T} \frac{d\rho}{dT} = -\frac{3}{a} \frac{\rho + P}{T} \frac{da}{dT} = -\frac{s}{a^3} \frac{d(a^3)}{dT} \Rightarrow d(a^3 s) = 0 \quad (75)$$

thus $a^3 s \propto S$ is conserved.

In the early Universe, the entropy density is dominated by relativistic species. Since $P = \rho/3$ for radiation, we find $s_i(T_i) = 4\rho_i(T_i)/(3T_i)$, which gives the total entropy

$$s(T) = \sum_i \frac{\rho_i(T_i) + P_i(T_i)}{T_i} = \frac{2\pi^2}{45} g_{*S}(T) T^3, \quad (76)$$

where we have defined the effective number of degrees of freedom in entropy:

$$g_{*S}(T) = \sum_{\text{bosons}} g_i \left(\frac{T_i}{T} \right)^3 + \frac{7}{8} \sum_{\text{fermions}} g_i \left(\frac{T_i}{T} \right)^3. \quad (77)$$

Note that $g_{*S}(T) \approx g_*(T)$ if all relativistic species in the Universe share a common temperature, which is true for $T \gtrsim 1 \text{ MeV}$. Conservation of entropy then implies:

$$\boxed{g_{*S}(T) T^3 a^3 = \text{const.}} \quad (78)$$

This is an important result. When g_{*S} is roughly constant (away from phase transitions), we recover the familiar $T \propto 1/a$. When a species annihilates and g_{*S} decreases, the temperature of the coupled species increases relative to $1/a$, with the annihilation energy redistributed among the remaining equilibrium species. Crucially, species that have already *decoupled* do not share in this reheating.

Finally we note that conservation of entropy allows us to compute the time-dependence of a and T in the early Universe. From the first Friedmann equation, we have

$$H^2(t) = \frac{8\pi G}{3} \rho(t) = \frac{4\pi^3 G}{45} g_*(T) T^4 \quad (79)$$

(neglecting curvature). Since $g_{*S} T^3 a^3 = s_{\text{eq}}$ is conserved (and equal to some equilibrium entropy today), we find

$$\frac{1}{a} \frac{da}{dt} = \left(\frac{8\pi G}{3} \rho(t) \right)^{1/2} = \left(\frac{4\pi^3 G s_{\text{eq}}^{4/3}}{45} \right)^{1/2} \left(\frac{g_*(T)}{g_{*S}(T)} \right)^{1/2} g_{*S}^{-1/6}(T) a^{-2}. \quad (80)$$

In periods of constant g_* , g_{*S} , we find $\dot{a}(t) \propto a^{-1}$, thus $a(t) \propto t^{1/2}$ and $T(t) \propto t^{-1/2}$, with bumps when particular species annihilate.

3.2.4 Neutrino Decoupling

Armed with our knowledge of entropy conservation, we can understand the physical processes occurring near the right edge of Fig. 2. At temperatures $T \gtrsim 1$ MeV, the Universe is a thermal bath filled with relativistic electrons, positrons, photons, neutrinos and antineutrinos (with all other particles decoupled and non-relativistic). These are kept in equilibrium by weak interactions such as $\nu_e + \bar{\nu}_e \rightleftharpoons e^+ + e^-$ and $\nu_e + e^- \rightleftharpoons \nu_e + e^-$.²¹ The rate of such reactions is given by $\Gamma = n\sigma v \approx n\sigma$, since all species are relativistic (noting that $m_e \approx 500$ keV). By dimensional analysis, the cross-section scales as $\sigma \sim G_F^2 T^2$ (where G_F is the Fermi constant). Given that $n \propto T^3$ for relativistic particles, we find

$$\Gamma \sim G_F^2 T^5. \quad (81)$$

If $\Gamma \gtrsim H$, the weak interactions keep the neutrinos in chemical equilibrium with the photons, electrons and positrons. On the other hand, when $\Gamma \lesssim H$, the neutrinos decouple and are no longer able to exchange energy with the photon and electron bath. During radiation domination, the expansion rate is $H \sim T^2/M_{\text{Pl}}$, for (reduced) Planck mass $M_{\text{Pl}} = (8\pi G)^{-1/2}$. Combining equations, the temperature of neutrino decoupling, T_{dec} , is given by

$$G_F^2 T_{\text{dec}}^5 \sim \frac{T_{\text{dec}}^2}{M_{\text{Pl}}} \Rightarrow T_{\text{dec}} \sim \left(\frac{1}{G_F^2 M_{\text{Pl}}} \right)^{1/3} \sim 1 \text{ MeV}. \quad (82)$$

After decoupling, neutrinos free-stream forming the *Cosmic Neutrino Background* (CνB). Noting that the momentum of neutrinos redshifts as $p \propto 1/a$, the distribution function at some time t after decoupling is given by

$$f(p, t > t_{\text{dec}}) = f\left(\frac{pa(t)}{a(t_{\text{dec}})}, t_{\text{dec}}\right) = \frac{1}{e^{pa(t)/T_{\text{dec}}} + 1} \equiv \frac{1}{e^{p/T(t)} + 1} \quad (83)$$

where the first expression relates the momentum p to its value at decoupling, noting that the distribution function is frozen at t_{dec} . This is just a thermal distribution with time-dependent temperature $T_\nu(a) = T_{\text{dec}}/a$.²² Notably, this differs from the evolution of the photon temperature, which satisfies $g_{*S}(T_\gamma)a^3T_\gamma^3 = \text{const.}$

3.2.5 Electron–Positron Annihilation

Shortly after neutrino decoupling, the temperature drops below $m_e \simeq 0.5$ MeV, such that the annihilation reaction between electrons and positrons ($e^+ + e^- \rightarrow 2\gamma$) is energetically preferred and the electrons and positrons drop out of equilibrium. This transfers energy to the photon sector raising T_γ ; since the neutrinos have already decoupled, their energy is not affected by this process, leading to the imbalance $T_\gamma > T_\nu$.

To quantify this reheating, we invoke conservation of entropy. Before annihilation the plasma has an effective number of degrees of freedom in entropy of

$$g_{*S}^{\text{before}} = 2 + \frac{7}{8} \times (2 + 2) = \frac{11}{2}, \quad (84)$$

²¹Note that $\nu + \bar{\nu} \rightleftharpoons \gamma + \gamma$ is loop-suppressed, since neutrinos interact only via the weak force. This explains why our treatment of neutrinos differs from that of other species.

²²This analysis assumes that the neutrinos are massless, which is not quite true. In practice, the distribution function will distort at late-times when the neutrinos become non-relativistic.

counting $g = 2$ for photons and $g = 2$ each for e^- and e^+ (ignoring the decoupled neutrinos). After annihilation, only photons remain in the bath, thus

$$g_{*S}^{\text{after}} = 2. \quad (85)$$

Since $g_{*S} T^3 a^3$ is conserved and the annihilation is fast (*i.e.* $a_{\text{before}} \approx a_{\text{after}}$) we find

$$\frac{T_{\gamma}^{\text{after}}}{T_{\gamma}^{\text{before}}} = \left(\frac{g_{*S}^{\text{before}}}{g_{*S}^{\text{after}}} \right)^{1/3} = \left(\frac{11}{4} \right)^{1/3}. \quad (86)$$

Before $e^{\pm}$ annihilation, the photon and neutrino sectors had the same temperature; after annihilation, the temperature of both sectors scales as a^{-1} , since g_{*S} is constant (there are no more species to annihilate). In combination, we find the important result:

$$\boxed{\frac{T_{\gamma}}{T_{\nu}} = \left(\frac{11}{4} \right)^{1/3} \approx 1.40.} \quad (87)$$

which holds for all times after $e^{\pm}$ annihilation. Today, we find $T_{\gamma} = 2.73$ K, implying that the cosmic neutrino background has a temperature $T_{\nu} = 1.95$ K.

The annihilation of electrons and positrons represents the last reheating event in the thermal history of the Universe. After this point, the effective number of degrees of freedom becomes constant, with

$$g_* = 2 + \frac{7}{8}(2N_{\text{eff}}) \left(\frac{4}{11} \right)^{4/3} \simeq 3.38, \quad g_{*S} = 2 + \frac{7}{8}(2N_{\text{eff}}) \left(\frac{4}{11} \right) \simeq 3.94, \quad (88)$$

where $N_{\text{eff}} \approx 3.044$ is the effective number of neutrino species.²³ Using the previous expression for the number density of a relativistic species, we can compute the total number density of each flavor of neutrinos today²⁴

$$n_{\nu,0} \simeq \frac{3}{4} \times \frac{\zeta(3)}{\pi^2} T_{\nu}^3 = \frac{3}{4} \times \frac{4}{11} n_{\gamma,0} \approx 112 \text{ cm}^{-3} \quad (89)$$

Furthermore, if the neutrinos are massless, their current energy density is given by

$$\rho_{\nu,0} = \frac{7}{8} N_{\text{eff}} \left(\frac{4}{11} \right)^{4/3} \rho_{\gamma,0}, \quad (90)$$

which yields $\Omega_{\nu,0} h^2 \approx 1.7 \times 10^{-5}$.

3.3 Non-Equilibrium Thermodynamics

So far, our treatment of the early Universe has assumed thermal equilibrium, such that the distribution function of each species is given by the Fermi–Dirac or Bose–Einstein form. However, some of the most important processes in cosmology, including recombination (sourcing the Cosmic Microwave Background), the formation of light elements, and (potentially) dark matter production, occur precisely when species fall out of equilibrium. To describe these phenomena, we need a new tool: the **Boltzmann equation**.

²³ N_{eff} is somewhat larger than 3 since the decoupling of neutrinos has not quite finished by the time electrons annihilate, leading to some leakage of energy into the neutrino sector.

²⁴ This remains true if the neutrinos become non-relativistic, since their number is locally conserved.

3.3.1 The Boltzmann Equation

The Boltzmann equation governs the non-equilibrium evolution of the distribution function $f_i(\vec{x}, \vec{p}, t)$, for some species i . This is built on the **Liouville theorem**, which states that, in the absence of collisions, the distribution function is conserved along particle trajectories, *i.e.* it behaves like an incompressible fluid in phase space.²⁵ Including collisions, the Boltzmann equation reads

$$\mathcal{L}f_i = C_i[\{f_j\}], \quad (91)$$

where $\mathcal{L}$ is the Liouville operator encoding the total derivative along the particle worldline and $C_i[\{f_j\}]$ is the **collision operator**, encoding all interactions, annihilations, and decays. Note that the right-hand-side depends on *any* species present in the Universe that interact with the particles of interest.

Given a particle worldline $\{x^\mu(\lambda), p^\mu(\lambda)\}$ for affine parameter λ , the Liouville operator is specified by²⁶

$$\mathcal{L} = \frac{dx^\mu}{d\lambda} \frac{\partial}{\partial x^\mu} + \frac{dp^\mu}{d\lambda} \frac{\partial}{\partial p^\mu} \quad (92)$$

Inserting the geodesic equation and defining the momentum via $dx^\mu/d\lambda = p^\mu$, this can be written

$$\mathcal{L} = p^\mu \frac{\partial}{\partial x^\mu} - \Gamma_{\nu\rho}^\mu p^\nu p^\rho \frac{\partial}{\partial p^\mu}. \quad (93)$$

Inserting the Christoffel symbols for the FLRW metric (and asserting homogeneity and isotropy, so that f_i depends only on $|\vec{p}|$ and t), this simplifies to

$$\boxed{E_i \frac{\partial f_i}{\partial t} - \frac{\dot{a}}{a} |\vec{p}|^2 \frac{\partial f_i}{\partial E_i} = C_i[\{f_j\}].} \quad (94)$$

where E_i is the energy of the species and $|\vec{p}|^2 = p_j p^j$, which we label p^2 . The second term on the left-hand side describes the effect of the cosmological expansion on the distribution function: in the absence of collisions ($C_i = 0$), it simply encodes the redshifting of momenta.

Integrated Boltzmann Equation For most cosmological applications, we do not need the full phase-space distribution, and can restrict to the number density $n_i \equiv \frac{g_i}{(2\pi)^3} \int d^3p f_i$. Dividing by E_i and integrating over 3-momenta, the equation simplifies to

$$\frac{dn_i}{dt} + 3 \frac{\dot{a}}{a} n_i = \frac{g_i}{2\pi^2} \int d^3p \frac{C_i[\{f_j\}]}{E_i} \quad (95)$$

where we have used integration-by-parts, noting that $\partial_{\log E_i} = \partial_{\log p}$ given the mass-shell condition. Defining the right-hand-side as a new collision term C' , the result is the **integrated Boltzmann equation**:

$$\boxed{\frac{1}{a^3} \frac{d}{dt} (a^3 n_i) = C'_i[\{n_j\}],} \quad (96)$$

where C' couples interacting species together. In the absence of interactions ($C'_i = 0$), this gives $n_i \propto a^{-3}$, implying that the number density simply dilutes with the expansion, as assumed previously.

²⁵This is a general result, which holds true as long as particle number is conserved and particles follow Hamiltonian equations of motion.

²⁶For a Newtonian universe, this simplifies to the standard form used in fluid dynamics: $\frac{\partial}{\partial t} + \vec{p} \cdot \nabla_{\vec{x}} - (\nabla_{\vec{x}} \Phi) \cdot \nabla_{\vec{p}}$, noting that $\dot{\vec{p}} = -\nabla_{\vec{x}} \Phi$ by Newton's second law.

Collision Terms To solve the Boltzmann equation, we must specify the collision term, C' . To understand this, let's take a simple two-body reaction specified by $1 + 2 \rightleftharpoons 3 + 4$ (which describes most interactions we will encounter in cosmology). The collision term for species 1 involves a competition between destruction and creation:

$$\frac{1}{a^3} \frac{d}{dt}(a^3 n_1) = - \underbrace{\alpha n_1 n_2}_{\text{destruction}} + \underbrace{\beta n_3 n_4}_{\text{creation}}, \quad (97)$$

where the rates (in units of particles per unit volume per unit time) are proportional to the densities on the left- and right-sides of the equation respectively. By definition, $\alpha = \langle \sigma v \rangle$, where $\langle \sigma v \rangle$ is the thermally-averaged cross-section times velocity. In equilibrium, the left-hand side vanishes (such that the comoving number of particles is conserved), which fixes the creation rate coefficient:

$$\beta = \left(\frac{n_1 n_2}{n_3 n_4} \right)_{\text{eq}} \langle \sigma v \rangle. \quad (98)$$

This is known as the **detailed balance** criterion, and involves the equilibrium abundances, which follow the Fermi-Dirac and Bose-Einstein distributions derived previously. Substituting back, the Boltzmann equation becomes²⁷

$$\frac{1}{a^3} \frac{d}{dt}(a^3 n_1) = - \langle \sigma v \rangle \left[n_1 n_2 - \left(\frac{n_1 n_2}{n_3 n_4} \right)_{\text{eq}} n_3 n_4 \right]. \quad (99)$$

It is useful to rewrite the above expression in terms of the comoving abundance of particles $N_i \equiv n_i/s \propto n_i a^3$ (using entropy conservation), as well as the interaction rate $\Gamma_i = n_2 \langle \sigma v \rangle$. We find

$$\frac{d \ln N_1}{d \ln a} = - \frac{\Gamma_1}{H} \left[1 - \left(\frac{N_1 N_2}{N_3 N_4} \right)_{\text{eq}} \frac{N_3 N_4}{N_1 N_2} \right]. \quad (100)$$

Here, the prefactor Γ/H measures the efficiency of the interaction relative to the expansion rate, while the bracketed term measures the departure from equilibrium. If $\Gamma \gg H$, the system is driven rapidly toward equilibrium (regardless of the sign of the bracket); if $\Gamma \lesssim H$, the scattering becomes inefficient, leading to the comoving abundances ‘freezing out’ ($N_i \rightarrow \text{const}$).

3.4 Thermal Relics and Dark Matter Freeze-Out

An important application of the Boltzmann equation is to the production of **thermal relics**: massive species that were once in thermal equilibrium but froze out as the Universe expanded and cooled. Consider a massive species X (e.g., a massive dark matter particle) that annihilates into light Standard Model particles ℓ via $X + \bar{X} \rightleftharpoons \ell + \bar{\ell}$. For temperatures above the particle mass ($T \gtrsim m_X$) the massive species is in equilibrium, with abundance set by the Bose-Einstein or Fermi-Dirac distribution. When T falls below m_X , the particles become non-relativistic and the massive species decouples. The integrated Boltzmann equation gives

$$\frac{1}{a^3} \frac{d}{dt}(a^3 n_X) = - \langle \sigma v \rangle [n_X^2 - (n_X^{\text{eq}})^2]. \quad (101)$$

where we have noted that $n_\ell = n_\ell^{\text{eq}}$ since the light particles remain in thermal equilibrium (assuming that they are tightly coupled to the photon bath).²⁸ Defining $Y_X \equiv n_X/T^3$ (proportional to the

²⁷This expression can also be written in terms of scattering amplitudes, defined in the usual particle-physics sense. Explicitly, $C'[n] = \int_{p_1 \dots p_4} (2\pi)^4 \delta_D(p^1 + p^2 - p^3 - p^4) |\mathcal{M}|^2 (f_3 f_4 - f_1 f_2)$, where $\int_p = \frac{g_i}{(2\pi)^3} \int \frac{d^3 p}{E(p)}$, the Dirac delta ensures conservation of 4-momentum, and $|\mathcal{M}|^2$ is a scattering amplitude. Note that integrating over $\int \frac{d^3 p}{E(p)}$ is equivalent to integrating over 4-momentum given the mass-shell constraint.

²⁸We also assume that $n_X \approx n_{\bar{X}}$.

number of particles per comoving) and using $x \equiv m_X/T$ as the time variable (with $x \sim 1$ at decoupling), this can be rewritten

$$\frac{1}{a^3} \frac{d}{dt} (a^3 T^3 Y_X) = T^3 \frac{dY_X}{dt} = T^3 \frac{dx}{dt} \frac{dY_X}{dx} = -T^6 \langle \sigma v \rangle [Y_X^2 - (Y_X^{\text{eq}})^2]. \quad (102)$$

where we have assumed that $T \propto 1/a$ (which holds if there are no reheating events in the time period of interest, *i.e.* if g_{*S} is constant). Noting that $dx/dt = -(m_X/T^2)dT/dt = (m_X/aT)da/dt = xH(T)$ and $H(T) = H(m_X)(T/m_X)^2$ in radiation domination, this simplifies to

$$\boxed{\frac{dY_X}{dx} = -\frac{\lambda}{x^2} [Y_X^2 - (Y_X^{\text{eq}})^2]}, \quad (103)$$

where $\lambda \equiv m_X^3 \langle \sigma v \rangle / H(m_X) = \Gamma(m_X)/H(m_X)$ is approximately constant if $\langle \sigma v \rangle$ does not depend on temperature. This is a **Riccati equation** and describes the abundance of X through both equilibrium and freeze-out regimes. The solution for $Y_X(x)$ is shown in Fig. 3.

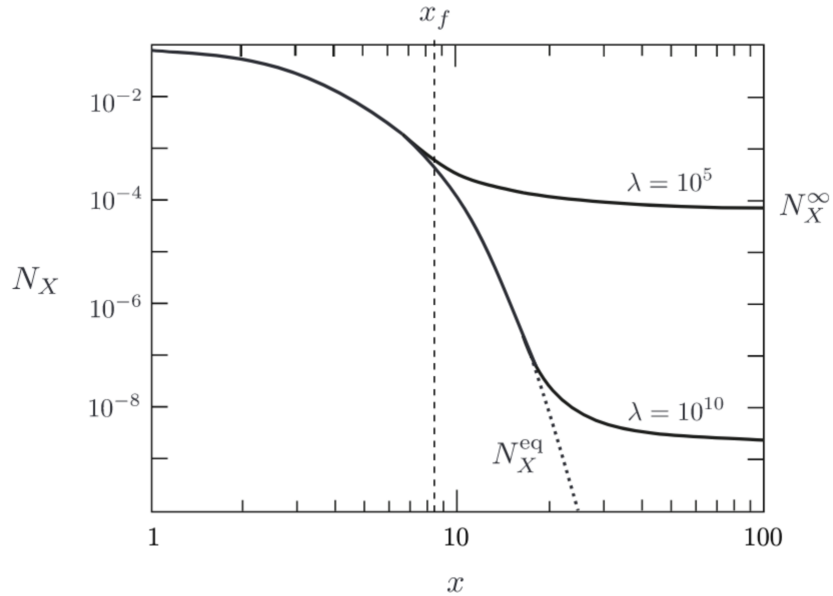


Figure 3: The abundance of some massive species X (denoted by Y_X in the text) as a function of $x = m_X/T$. For $x \ll 1$ the species is in equilibrium, whilst for $x \gg 1$ it freezes-out leaving a relic-abundance Y_X^∞ (which is much larger than the equilibrium value, since annihilations become inefficient). Reproduced from D. Baumann's *Cosmology* lecture notes.

At early times ($x \ll 1$, *i.e.* $T \gg m_X$), the interaction rate greatly exceeds the expansion rate, and Y_X tracks its equilibrium value, *i.e.*, $Y_X(x) \approx Y_X^{\text{eq}}(x) = \frac{3\zeta(3)}{4\pi^2} \approx 0.1$ (for fermionic X). As the Universe cools below $T \sim m_X$, the massive particle decouples and the equilibrium abundance becomes Boltzmann-suppressed ($Y_X^{\text{eq}}(x) \propto x^{3/2} e^{-x}$, since non-relativistic). Eventually, the annihilation rate can no longer keep up with the expansion, and Y_X freezes out at some constant value Y_X^∞ (which is much larger than the equilibrium value). This can be obtained by noting that at late-times ($x \rightarrow \infty$), $Y_X^{\text{eq}} \rightarrow 0$ and the Boltzmann equation reduces to $dY_X/dx \approx -\lambda Y_X^2/x^2$. Integrating from the freeze-out time ($x_f \approx 10$, defined via $H(x_f) \approx \Gamma(x_f)$) to infinity we find the **relic abundance**

$$Y_X^\infty \simeq \frac{x_f}{\lambda}. \quad (104)$$

The WIMP Miracle After freeze-out, the number density of the massive particles decay as a^{-3} following the expansion of the Universe. Invoking conservation of entropy (and thus of $g_*sa^3T^3$), we find the current density

$$\rho_{X,0} = m_X n_{X,0}, \quad n_{X,0} = Y_X^\infty T_1^3 \left(\frac{a(T_1)}{a(T_0)} \right)^3 = Y_X^\infty T_0^3 \frac{g_{*S}(T_0)}{g_{*S}(T_1)} \quad (105)$$

where T_0 is the photon temperature today and $T_1 \gtrsim m_X$ is some time close to x_f , where the species has reached its relic abundance. Converting to a dimensionless density parameter, one finds

$$\Omega_{X,0} h^2 \sim 0.1 \left(\frac{x_f}{10} \right) \sqrt{\frac{100}{g_*(m_X)}} \frac{10^{-8} \text{ GeV}^{-2}}{\langle \sigma v \rangle}, \quad (106)$$

(setting $T_1 \approx m_X$ and inserting the definition of λ), which depends on the annihilation cross-section but *not* on the mass m_X (at least to leading order). Let's assume that the X particles couple to the standard model via weak interactions. In this case,

$$\langle \sigma v \rangle \sim G_F^2 m_X^2 \sim 10^{-6} \text{ GeV}^{-2} (m_X/100 \text{ GeV})^2 \quad (107)$$

for Fermi constant G_F . Fixing $m_X \sim 100 \text{ GeV}$ (and keeping the various numerical factors), we recover a late-time density that is remarkably close to the observed density of dark matter ($\Omega_{\text{cdm}} h^2 \approx 0.12$). This is the **WIMP miracle**: the freeze-out of a weakly-coupled massive particle induces a relic abundance similar to that of dark matter. This coincidence is known as the “WIMP miracle”, and has motivated decades of experimental searches for Weakly Interacting Massive Particles (WIMPs), though no direct detection has yet been made (and indeed, the couplings seem a little unlikely given the lack of new physics detections from the Large Hadron Collider).

3.5 Recombination and the Cosmic Microwave Background

The last great thermal event in the history of the Universe is the sourcing of the Cosmic Microwave Background (CMB), triggered by the (re)combination of ionized species. This occurs at temperatures around 0.3 eV, when the Universe's energy density (and thus expansion rate) becomes dominated by non-relativistic matter (noting that matter-radiation equality occurs at $T_{\text{eq}} \approx 0.8 \text{ eV}$).

The formation of the CMB occurs via two processes. First, electrons and protons form neutral hydrogen via the reaction $e^- + p \rightarrow H + \gamma$, which is energetically favored for $T \lesssim 0.3 \text{ eV}$. This rapidly dilutes the number density of free electrons, and thus Compton scattering processes, which ensure equilibrium between photons and electrons. Once the density drops sufficiently, the photons decouple and free-stream, generating the CMB.

3.5.1 Hydrogen Recombination

Around $T \sim 1 \text{ eV}$, the Universe was filled with a plasma of ionized hydrogen, ionized helium, free electrons, and photons (with neutrinos and dark matter having decoupled much earlier). Two sets of processes kept the plasma in thermal equilibrium: photons and electrons were tightly coupled via Compton scattering ($\gamma + e^- \rightleftharpoons \gamma + e^-$), whilst electrons and protons were interacting via Coulomb scattering ($H^+ + e^- \rightleftharpoons H^+ + e^-$). Neutral hydrogen forms via the reaction $e^- + p \rightleftharpoons H + \gamma$, whose equilibrium depends on the ionization energy $E_I = m_p + m_e - m_H = 13.6 \text{ eV}$. For $T \gg E_I$, the ionized states are energetically preferred, whilst for $T \ll E_I$, the equilibrium shifts toward neutral hydrogen, which leads to a suppression of the electron energy density.

To compute the electron density, n_e , we can use the above formulae for the equilibrium abundances of non-relativistic species (since $T \ll m_e \approx 511$ keV). Since $\mu_e + \mu_p = \mu_H$ in chemical equilibrium:

$$\left(\frac{n_H}{n_e n_p} \right)_{\text{eq}} = \frac{g_H}{g_e g_p} \left(\frac{m_H}{m_e m_p} \right)^{3/2} \left(\frac{2\pi}{T} \right)^{3/2} e^{E_I/T}, \quad (108)$$

with $g_H = 4$ (from one singlet and three triplet states, due to electron-proton spin interactions), and $g_e = g_p = 2$. Since $m_H \approx m_p$, this simplifies to

$$\left(\frac{n_H}{n_e^2} \right)_{\text{eq}} = \left(\frac{2\pi}{m_e T} \right)^{3/2} e^{E_I/T}, \quad (109)$$

where we have asserted charge neutrality such that $n_e = n_p$.

It is useful to define the **free-electron fraction** $X_e \equiv n_e/n_b$, where $n_b = n_p + n_H$ is the total number of baryons (ignoring Helium), which can be related to the number of photons via $n_b = \eta n_\gamma$, where η is the baryon-to-photon ratio. If $X_e = 1$, the plasma is fully ionized, with $X_e = 0$ indicating a fully neutral medium. Noting that $(1 - X_e)/X_e^2 = n_H n_b/n_e^2$, we obtain the **Saha equation**:²⁹

$$\left[\frac{1 - X_e}{X_e^2} \right]_{\text{eq}} = \frac{2\zeta(3)}{\pi^2} \eta \left(\frac{2\pi T}{m_e} \right)^{3/2} e^{E_I/T}, \quad (110)$$

where we have inserted the definition of $n_\gamma = 2\zeta(3)T^3/\pi^2$. We define the recombination temperature, T_{rec} via $X_e(T_{\text{rec}}) = 0.5$; this evaluates to 0.3 eV, corresponding to ≈ 3760 K or $z \approx 1300$, which is deep in the matter-dominated era. The recombination temperature is much lower than the ionization energy (13.6 eV), due to the large excess of photons over baryons, and the high-energy tail of the photon distribution (analogous to the formation of deuterium in BBN).

3.5.2 Photon Decoupling

As discussed above, equilibrium between photons and electrons is maintained by Compton scattering ($\gamma + e^- \rightleftharpoons \gamma + e^-$), which has the photon interaction rate

$$\Gamma_\gamma(T) = n_e(T) \sigma_T, \quad (111)$$

where $\sigma_T = 8\pi\alpha^2/(3m_e^2)$ is the Thomson cross-section. When this falls below the expansion rate, $H(T)$, the photons decouple from the plasma. To compute the decoupling temperature, T_{dec} , we note that $H(T) = H_0\sqrt{\Omega_m}(T/T_0)^{3/2}$ in the matter dominated era (where $T_0 \approx 2.73$ K is the CMB temperature today), thus $H(T_{\text{dec}}) = \Gamma_\gamma(T_{\text{dec}})$ requires

$$H_0\sqrt{\Omega_m} \left(\frac{T_{\text{dec}}}{T_0} \right)^{3/2} = \frac{2\zeta(3)}{\pi^2} \eta T_{\text{dec}}^3 X_e(T_{\text{dec}}) \sigma_T \quad (112)$$

setting $n_e = X_e n_b = X_e \eta n_\gamma$. Using the Saha equation for $X_e(T)$, we find $T_{\text{dec}} \approx 0.25$ eV, which corresponds to a decoupling redshift of $z_{\text{dec}} \approx 1090$ and a time of $t_{\text{dec}} \approx 370,000$ yr. By this temperature, the free-electron fraction has dropped to $X_e(T_{\text{dec}}) \sim 10^{-3}$, thus the Universe is almost entirely neutral.³⁰

²⁹A more careful treatment involves the Boltzmann equation, which accounts for out-of-equilibrium effects that become important for $X_e \rightarrow 0$.

³⁰A small residual population of free electrons remains after decoupling due to freeze-out effects (which occur since the recombination rate falls below H before all electrons are captured by protons). Solving the appropriate Riccati equation for X_e , we find $X_e^\infty \approx 10^{-4}$.

The decoupling of photons from the primordial plasma is associated with the cessation of electron-photon scatterings. In general, the probability of a photon scattering between $[t, t + \delta t]$ is given by $\Gamma_\gamma(t)\delta t$, which motivates defining the **optical depth**:

$$\tau(t) = \int_t^{t_0} \Gamma_\gamma(t') dt', \quad (113)$$

which gives the expected number of scatterings of a photon between time t and today. The **last scattering surface** is the locus of points for which $\tau(t_*) = 1$ (*i.e.* for which we expect just one scattering event), which defines a last-scattering time $t_* \approx t_{\text{dec}}$. More precisely, last-scattering occurs at $z_* \simeq 1080$ (accounting for helium recombination, multi-level hydrogen atoms, and other corrections), with a width $\Delta z \simeq 80$ reflecting the finite duration of decoupling. The photons released at this epoch free-stream to us today as the **cosmic microwave background** (CMB).

3.5.3 Baryon Drag

After photon decoupling, the photons and baryons remain tightly coupled due to residual Compton scattering ($\gamma + e^- \rightleftharpoons \gamma + e^-$) which continues to transfer momentum between photons and baryons. Whilst photon decoupling is defined by the condition $\Gamma_\gamma \equiv n_e(T)\sigma_T = H(T)$, the relevant criterion for the decoupling of baryons is the suppression of the photon-induced drag force in the baryon Euler equation (which we'll discuss more when we consider perturbations). This introduces the effective drag rate

$$\Gamma_{\text{drag}}(T) = \frac{4\rho_\gamma(T)}{3\rho_b(T)} n_e(T)\sigma_T \equiv \frac{\Gamma_\gamma(T)}{R(T)}, \quad (114)$$

where $R \equiv 3\rho_b/(4\rho_\gamma)$ is the baryon loading parameter, which quantifies the inertia of baryons relative to the photon pressure. We define the redshift of *baryon drag* by setting $\Gamma_{\text{drag}}(T_{\text{drag}})$ equal to $H(T_{\text{drag}})$, which gives $z_{\text{drag}} \approx 1060$. Notably, baryon drag occurs a little after photon decoupling, since the photon pressure is enhanced by the smallness of the baryon loading parameter (with $R \propto a$). The finite duration of this process leads to a small separation between the photon last-scattering surface at $z_* \simeq 1090$ and the baryon drag epoch, which sets the sound horizon relevant for large-scale structure and baryon acoustic oscillations.

3.6 Big Bang Nucleosynthesis

One of the great successes of the Hot Big Bang model is its prediction of the primordial abundances of light elements, including deuterium (D), helium-3 (^3He), helium-4 (^4He), and lithium-7 (^7Li). This process is known as **Big Bang Nucleosynthesis** (BBN), and can be described using the non-equilibrium thermodynamics tools introduced above.

In this section, we'll trace the evolution of baryons from $T \sim 1$ MeV onwards, picking up the story of thermal history from where we left-off (neutrino decoupling). Our main goal will be to predict the primordial Helium mass fraction from first principles, and compare to the observed result: $Y_p \approx 0.25$. As we'll show, the baryon sector involves a wide variety of reactions, representing both chemical and alchemical processes. These include:

- $n + \nu_e \rightleftharpoons p + e^-$ (β -decay)
- $n + e^+ \rightleftharpoons p + \bar{\nu}_e$ (inverse β -decay)
- $n + p \rightleftharpoons D + \gamma$ (deuterium formation)

- $D + p \rightleftharpoons {}^3\text{He} + \gamma$ (${}^3\text{He}$ formation)
- ${}^3\text{He} + D \rightleftharpoons {}^4\text{He} + p$ (${}^4\text{He}$ formation)

The first two reactions occur in equilibrium for $T \gtrsim 1$ MeV, but freeze-out after neutrinos decouple, which fixes the number of neutrons (up to decay processes). Gradually, the free neutrons combine with protons to form deuterium, which then forms ${}^3\text{He}$ and ${}^4\text{He}$ (as well as trace quantities of other light elements, including ${}^7\text{Li}$). Below, we will discuss each process in this hierarchy in detail.

Phase 1: Neutron–Proton Equilibrium At $T \gtrsim 1$ MeV, neutrons and protons are kept in thermal equilibrium via the weak interactions listed above ($n + \nu_e \rightleftharpoons p + e^-$ and $n + e^+ \rightleftharpoons p + \bar{\nu}_e$). As such, we can compute their number densities using the Fermi-Dirac distribution. Assuming $T \ll m_{p,n} \approx 940$ MeV, such that the particles are non-relativistic, we find the equilibrium ratio (assuming $|\mu_{n,p}| \ll T$)

$$\left(\frac{n_n}{n_p}\right)_{\text{eq}} = \left(\frac{m_n}{m_p}\right)^{3/2} e^{-Q/T} \simeq e^{-Q/T}, \quad (115)$$

where $Q \equiv m_n - m_p \simeq 1.293$ MeV is the mass-difference between neutrons and protons. For $T \gg Q$, this gives $n_n \approx n_p$ (equal numbers of neutrons and protons), whilst for $T \ll Q$, the equilibrium abundance of neutrons is exponentially suppressed. Without freeze-out, this exponential suppression would drive $n_n \rightarrow 0$, leaving no neutrons to form helium (and thus life).

Phase 2: Neutron Freeze-Out An important quantity is the **neutron fraction**, defined via

$$X_n(T) \equiv \frac{n_n(T)}{n_n(T) + n_p(T)}. \quad (116)$$

Above $T \approx 1$ MeV the weak interactions controlling (inverse) β decay are in equilibrium, thus X_n attains the equilibrium value $X_n^{\text{eq}}(T) = e^{-Q/T}/(1 + e^{-Q/T})$ (from above). At lower temperatures, the weak interactions shut-off, and the neutrons fall out of equilibrium.³¹

To describe the non-equilibrium system, we use the integrated Boltzmann equation. Considering the general reaction $n + \ell \rightleftharpoons p + \ell'$ where ℓ, ℓ' are lepton species (which remain in equilibrium, since they are light with $n_\ell = n_\ell^{\text{eq}}$), we find

$$\frac{1}{a^3} \frac{d}{dt}(n_n a^3) = -\Gamma_n \left[n_n - \left(\frac{n_n}{n_p}\right)_{\text{eq}} n_p \right], \quad (117)$$

where $\Gamma_n = n_\ell \langle \sigma v \rangle$ is the total weak interaction rate for neutrons. Since the total baryon number $(n_n + n_p)a^3$ is conserved (implying that $d \log(n_n a^3) = d \log X_n$), we can rewrite this as an equation for X_n :

$$\frac{dX_n}{dt} = -\Gamma_n [X_n - (1 - X_n)e^{-Q/T}]. \quad (118)$$

From particle physics, the total rate Γ_n receives contributions from both ν_e and e^+ scattering channels. Explicitly, defining $x \equiv Q/T$:

$$\Gamma_n(x) = \frac{255}{\tau_n} \left(\frac{12 + 6x + x^2}{x^5} \right), \quad (119)$$

³¹Note that the neutrino decoupling temperature $T_{\text{dec}} \approx 1$ MeV and the proton-neutron mass-difference $Q \approx 1.3$ MeV are unrelated, but similar in magnitude. This is a happy coincidence with important consequences for humanity.

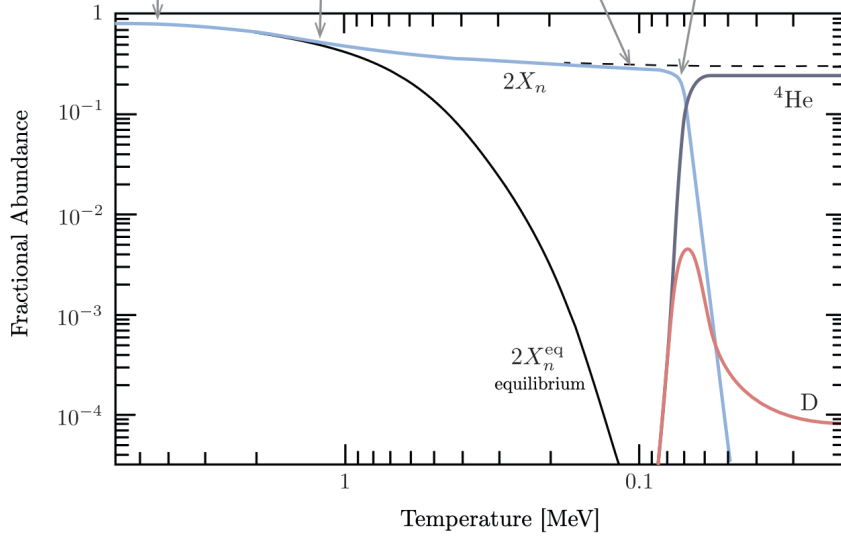


Figure 4: The abundance of free neutrons (blue), deuterium (red) and helium (black) as a function of temperature during the BBN epoch. Neutrons freeze-out at around 0.8 MeV, with their abundance falling afterwards due to decay processes. Deuterium forms around 0.1 MeV which is quickly converted into ${}^4\text{He}$ (and a trace amount of other elements), which uses up almost all the free neutrons. Note that the neutron abundance is far larger than the equilibrium result (black) due to freeze-out. Reproduced from D. Baumann, *Cosmology* (Cambridge University Press, 2022).

where $\tau_n \simeq 886.7$ s is the lifetime of free neutrons. During radiation domination, the expansion rate scales as T^2 , such that $H(x) = H(x=1)/x^2 \approx 1.13 \text{ s}^{-1}/x^2$. In combination, we find that $\Gamma_n(x)/H(x) \propto x^{-1}$ at large x : the interaction rate drops faster than the expansion rate, which implies that freeze-out occurs (with $\Gamma_n(x) \approx H(x)$ at $T \approx Q/2$).

Switching to x as the time variable (using $d \ln x = -d \ln T = d \ln a$, such that $dx/dt = xH$), the Boltzmann equation becomes

$$\frac{dX_n}{dx} = \frac{\Gamma_n(x)}{xH(x)} [e^{-x} - X_n(1 + e^{-x})]. \quad (120)$$

This can be solved numerically, yielding $X_n^\infty \equiv X_n(T \ll Q) \approx 0.15$. The net result is that the neutron fraction freezes out at roughly 15%. A plot of X_n is shown in Fig. 4.

This result can be obtained in a simpler (though approximate) manner by noting that nucleon freeze-out occurs when $\Gamma_n \lesssim H$, which gives $T_{\text{dec},n} \simeq 0.8$ MeV (slightly below the neutrino decoupling temperature, since the n - ν cross-section is somewhat larger than the ν - ν cross-section). Evaluating the equilibrium neutron fraction at this temperature, we find

$$X_n \simeq \frac{e^{-Q/T_{\text{dec}}}}{1 + e^{-Q/T_{\text{dec}}}} = \frac{e^{-1.293/0.8}}{1 + e^{-1.293/0.8}} \simeq 0.16 \approx \frac{1}{6} \quad (121)$$

in rough agreement with the numerical result.

Phase 3: Neutron Decay After freeze-out, free neutrons are unstable and decay via $n \rightarrow p + e^- + \bar{\nu}_e$ with lifetime $\tau_n \simeq 886.7$ s. This becomes important for $t \gtrsim 100$ s (when $T \lesssim 0.2$ MeV),

and reduces the neutron fraction as

$$X_n(t) \simeq X_n^\infty e^{-t/\tau_n} \simeq \frac{1}{6} e^{-t/887\text{s}}. \quad (122)$$

This process destroys the neutrons that would otherwise be available to form helium, so the final helium abundance depends sensitively on how quickly nucleosynthesis proceeds and the value of τ_n .

Phase 4: Deuterium Production and the Deuterium Bottleneck At $t \sim 100$ s, the Universe is composed mostly of free neutrons and protons. The density is far too low for direct three-body reactions to produce helium (which is the route used in stellar fusion); instead, heavier nuclei must be built up sequentially, starting with deuterium:



At high temperatures, any deuterium forming via the first reaction is immediately photodissociated. This process is fast since (a) the binding energy of deuterium is small ($B_D \approx 2.22$ MeV) and (b) there are $\sim 10^{10}$ photons per baryon, meaning that ionization can occur even when T falls below B_D .³² As a result, very little deuterium is present for $T \gtrsim 0.1$ MeV. This is known as the **deuterium bottleneck**: heavy element production cannot proceed until a sufficient abundance of deuterium accumulates.

The abundance of deuterium can be computed using chemical equilibrium (which holds as long as the first reaction dominates, which is true if there are free neutrons and protons available and the abundance of deuterium is not too large). Since all species are non-relativistic, their densities are given by the Maxwell-Boltzmann distribution (*i.e.* $n \propto g(mT)^{3/2} e^{-(m-\mu)/T}$). In chemical equilibrium, $\mu_p + \mu_n = \mu_D$, which implies that

$$\left(\frac{n_D}{n_p n_n} \right)_{\text{eq}} = \frac{g_D}{g_p g_n} \left(\frac{m_D}{m_n m_p} \right)^{3/2} \left(\frac{T}{2\pi} \right)^{-3/2} e^{B_D/T}, \quad (124)$$

where $B_D = m_n + m_p - m_D \approx 2.22$ MeV is the deuterium binding energy, and the degeneracy factors are $g_D = 3$ (spin-1), $g_n = g_p = 2$ (spin-1/2). To simplify, we note that $m_D \simeq 2m_p$, and $n_n \sim n_b$, where the density of baryons is given by $\eta n_\gamma = \eta \times 2\zeta(3)T^3/\pi^2$, where $\eta \approx 6 \times 10^{-18}$ is the baryon-to-photon ratio. We find

$$\left(\frac{n_D}{n_p} \right)_{\text{eq}} \sim \eta \left(\frac{T}{m_p} \right)^{3/2} e^{B_D/T}. \quad (125)$$

Despite the large Boltzmann factor $e^{B_D/T}$, the tiny baryon-to-photon ratio suppresses deuterium production until $T \ll B_D$ (matching the above conclusions). Setting $n_D^{\text{eq}} \sim n_p^{\text{eq}}$ gives the nucleosynthesis temperature $T_{\text{nuc}} \simeq 0.06\text{--}0.08$ MeV, as seen in Fig. 4. This corresponds to $t_{\text{nuc}} \simeq 200\text{--}300$ s (using $H(T)$ with $g_*(T) \approx 3.36$, since electrons and positrons have decoupled).

By the time of deuterium formation, a fraction of neutrons have already decayed. Inserting $t_{\text{nuc}} \approx 220$ s, we find the final neutron fraction:

$$X_n(t_{\text{nuc}}) \simeq \frac{1}{6} e^{-t_{\text{nuc}}/\tau_n} \simeq \frac{1}{6} e^{-220/887} \simeq 0.12, \quad (126)$$

which is fixed, since neutron decay requires *free* neutrons.

³²Due to the exponential tail of the distribution function, one in 10^{10} photons has energy above $30T$.

Phase 5: Production of Helium (and Beyond) Once deuterium accumulates, the subsequent reactions proceed rapidly. ${}^4\text{He}$ has a binding energy of $B_{{}^4\text{He}} = 28.3 \text{ MeV} \gg B_D$, so the Boltzmann factor heavily favors production of helium over deuterium. This implies that nearly all neutrons present in the universe at t_{nuc} are incorporated into ${}^4\text{He}$. Since each helium nucleus contains 2 neutrons, $n_{\text{He}} = n_n(t_{\text{nuc}})/2$, and the primordial helium mass fraction is

$$Y_p \equiv \frac{4n_{\text{He}}}{n_b} = 2X_n(t_{\text{nuc}}) \simeq 0.24. \quad (127)$$

Observationally, $Y_p \simeq 0.2458 \pm 0.0013$, in excellent agreement with the theoretical prediction. Since the prediction depends on both nuclear physics and the cosmological expansion history, it provides a powerful test of the consistency of the Hot Big Bang model.

In addition to ${}^4\text{He}$, BBN produces trace amounts of other light elements. A small residual of deuterium ($D/H \sim 10^{-5}$) remains when the Helium fusion reactions become inefficient (as the density drops), along with a comparable amount of ${}^3\text{He}$. Furthermore, tiny quantities of ${}^7\text{Li}$ and ${}^7\text{Be}$ ($\sim 10^{-10}$) are produced via reactions such as ${}^4\text{He} + {}^3\text{H} \rightarrow {}^7\text{Li} + \gamma$ and ${}^4\text{He} + {}^3\text{He} \rightarrow {}^7\text{Be} + \gamma$. Computing these abundances precisely requires solving the full set of coupled Boltzmann equations for all relevant nuclear species (which is done by a range of BBN codes).

Observations BBN predicts the primordial abundances of light elements, which can be compared to observational data. The various abundances can be probed as follows:

1. **${}^4\text{He}$** : Measured from emission lines in metal-poor ionized gas clouds (to minimize contamination from stellar helium production).
2. **D**: Measured from Lyman- α absorption in high-redshift quasar spectra. Deuterium is only produced in BBN, and can be easily destroyed in stars, thus the observed fraction is a lower-bound on the primordial abundance.
3. **${}^3\text{He}$** : This is difficult to constrain, since it is both created and destroyed in stars.
4. **${}^7\text{Li}$** : Measured in metal-poor stars in the Galactic halo. Since this is destroyed in stellar interiors, the observed abundance is a lower bound on the primordial value.

In general, we find excellent agreement between the predicted and observed abundances (possibly with a slight difference for ${}^7\text{Li}$, though this is most likely due to modeling of stellar chemistry).

Dependence on Early-Universe Physics The predicted abundances depend on several key input parameters:

1. **Neutron lifetime (τ_n)**: Increasing τ_n reduces the rate of β -decay (which turns $n \rightarrow p$), so neutrons decouple earlier (at higher T_{dec}), leading to a larger X_n^∞ and thus a larger Y_p .
2. **Effective degrees of freedom (g_*)**: Increasing g_* increases the expansion rate $H \propto g_*^{1/2} T^2$, causing earlier freeze-out (with $T_{\text{dec}} \sim g_*^{1/6}$) and thus a larger Y_p . This provides a constraint on the number of extra light relativistic species at the time of BBN, parameterized by ΔN_{eff} (with $g_*(T_{\text{nuc}}) = 10.75 + \Delta N_{\text{eff}}$).
3. **Baryon density (η)**: Y_p depends only weakly on $\eta = n_b/n_\gamma$, but the deuterium abundance is very sensitive: increasing η means the deuterium bottleneck ends earlier (since there are

more baryons to fuse with), leaving less residual deuterium. The observed D/H ratio therefore provides a precise measurement of $\Omega_b h^2$. This measurement is routinely used in cosmological analyses.

4. **Weak Interaction Rate:** Increasing G_F , and thus the rate of β -decay processes, reduces the neutron freeze-out temperature ($T_{\text{dec}} \sim G_F^{-2/3}$). This reduces the freeze-out density of neutrons, and thus Y_p .

4 Inflation

In the previous sections, we have outlined the ‘Hot Big Bang’ model, which accurately describes the evolution of the Universe from temperatures around 100 GeV down to today, encompassing the electroweak and QCD phase transitions, neutrino decoupling, nucleosynthesis, recombination, and the subsequent growth of cosmological structure (which we will describe in detail below). When pushed further backwards in time (towards the primordial singularity at $t \lesssim t_{\text{Pl}} \approx 10^{-43}$ s), however, this model runs into a number of important problems that cannot be addressed within the standard framework. In particular:

1. Why is the Universe so **homogeneous**?
2. Why is the Universe so **flat**?
3. Why are there no **topological defects** such as magnetic monopoles?
4. Why do we see **correlated fluctuations** on large scales?

In this section, we will discuss these puzzles in some detail, and introduce the theory of **inflation**: a primordial epoch of accelerated expansion that simultaneously resolves all four problems and explains the origin of structure in the Universe.

4.1 Problems with the Hot Big Bang

4.1.1 The Horizon Problem

At time t , the fraction of the Universe in **causal contact** with a given observer is set by the **particle horizon**. From §2, the comoving horizon is given by

$$\chi_{\text{ph}}(t) = \eta(t) - \eta_i = \int_{t_i}^t \frac{dt'}{a(t')}, \quad (128)$$

where η indicates conformal time and t_i is some initial time, which we’ll set to zero (defining the Big Bang). This is the comoving distance that a photon could have traveled since t_i , and thus the maximum range over which any signal could have propagated. Two points separated by $\chi > \chi_{\text{ph}}$ are said to be **acausal**: no information could have passed between them (since we don’t allow for superluminal information transfer!) and any thermal differences could not have been washed away (see Fig. 1 for a depiction of this). It is useful to rewrite χ_{ph} in terms of the scale-factor:

$$\chi_{\text{ph}}(a) = \eta(a) - \eta_i = \int_{\log a_i}^{\log a} \frac{d \log a'}{a' H(a')}, \quad (129)$$

where $[aH(a)]^{-1}$ is the **comoving Hubble radius**, which gives the current ‘size’ of the Universe.³³

Let’s assume that the early Universe was dominated by a fluid with equation of state parameter $w = P/\rho$. Following §2, $H^2 \propto \rho \propto a^{-3(1+w)}$, thus $(aH)^{-1} \propto a^{(1+3w)/2}$. If w is fixed, we can simply compute the comoving horizon:

$$\chi_{\text{ph}}(a) = \eta(a) - \eta_i \propto \frac{2}{1+3w} \left[a^{(1+3w)/2} - a_i^{(1+3w)/2} \right]. \quad (130)$$

In General Relativity, the **strong energy condition** states that $\rho + 3P \geq 0$, which is satisfied for all physical types of matter.³⁴ This implies that $w \geq -1/3$, which leads to a number of important conclusions:

1. $\eta \rightarrow 0$ as we approach the initial singularity ($a \rightarrow 0$), thus $\eta_i \approx 0$, and the particle horizon is equal to the conformal time, $\eta(a)$.
2. The comoving Hubble radius is an increasing function of time, *i.e.* $\frac{d}{dt}([aH]^{-1}) \geq 0$.
3. $\chi_{\text{ph}}(a)$ increases monotonically with a , *i.e.* the causally connected region grows over time.

A natural consequence is that the particle horizon at recombination ($a_{\text{rec}} \approx 10^{-3}$) is much smaller than the particle horizon today ($a_0 = 1$). Indeed, inserting the explicit form for $H(a)$ in terms of Λ CDM cosmological parameters, and assuming that the Universe was radiation dominated up to matter-radiation-equality ($a_{\text{eq}} \approx 3 \times 10^{-4}$), we find $\chi_{\text{ph}}(\eta_{\text{rec}}) \approx 280 \text{ Mpc} \ll \chi_{\text{ph}}(\eta_0) \approx 14000 \text{ Mpc}$.

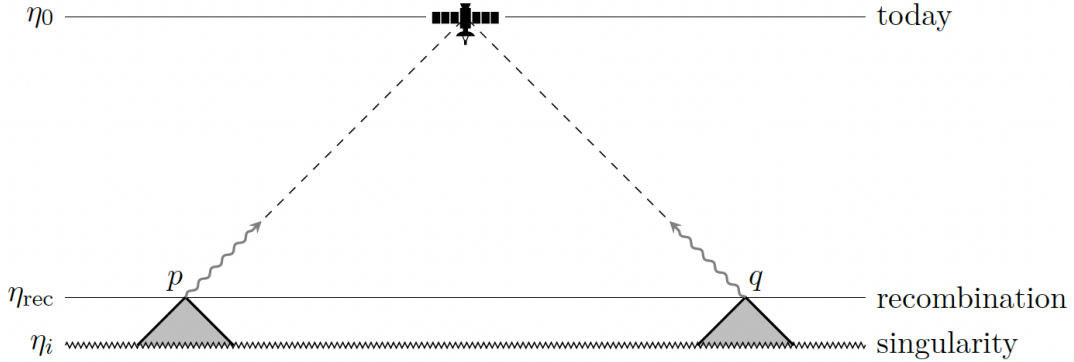


Figure 5: Sketch of the horizon problem in the Hot Big Bang model. We consider photons emitted at two points, p and q , whose particle horizons at the time of emission are shown in grey. Since $\eta_i \approx 0$, the comoving distance between the p and q is much larger than the particle horizon, thus the two points are causally disconnected at η_{rec} , and should exhibit different CMB temperatures. Reproduced from D. Baumann, *Cosmology* (Cambridge University Press, 2022).

To see why this is a problem, consider two points on opposite sides of the surface of last scattering, labelled p and q . As shown in Fig. 5, these points emit photons at η_{rec} that can be observed today,

³³Note the difference between the Hubble radius and the particle horizon: two points with $\chi > [aH(a)]^{-1}$ are not currently in causal contact, whilst two points with $\chi > \chi_{\text{ph}}(a)$ have *never* been in causal contact.

³⁴Practically, this is just a statement that gravity is an attractive force, which formally requires $(T_{\mu\nu} - \frac{1}{2}Tg_{\mu\nu})t^\mu t^\nu \geq 0$ for any timelike t^μ . It is famously violated by dark energy and certain types of scalar fields (which will be important below).

at η_0 . At the time of emission, p and q have a particle horizon of size $\chi_{\text{ph}}(\eta_{\text{rec}}) = \eta_{\text{rec}} - \eta_i \approx \eta_{\text{rec}}$ that is much smaller than their comoving separation $2\eta_0 - 2\eta_{\text{rec}}$ (under the standard Hot Big Bang assumptions).³⁵ This implies that p and q were not in causal contact for the entire history of the Universe up to recombination. Observationally, we find that the temperature of photons emitted from p and q is *identical* to $\mathcal{O}(0.001\%)$ – this is a paradox, since acausality implies that no physical processes could transfer energy from p to q and thus wash out primordial temperature imbalances (arising from quantum effects in the high-energy Universe).³⁶

Using the horizon size at recombination, we can estimate the angular extent of a causally-connected patch in the CMB:

$$\theta_{\text{hor}} = \frac{2\chi_{\text{ph}}(\eta_{\text{rec}})}{d_A^{\text{com}}(\eta_{\text{rec}})} \approx 2.3^\circ, \quad (131)$$

where $d_A^{\text{com}}(\eta_{\text{rec}}) = \eta_0 - \eta_{\text{rec}}$ is the comoving angular-diameter distance to recombination (assuming a flat Universe). Causality thus implies that any pair of points separated by $\theta > \theta_{\text{hor}}$ should be independent, yet we find $\sim 10^4$ pairs with identical temperatures! This is known as the **horizon problem**, and was first pointed out by Robert Dicke in the late 1970s.

4.1.2 The Flatness Problem

The second puzzle concerns the observed flatness of the Universe. From the Friedmann equation, we can define the time-dependent curvature parameter via

$$\Omega_k(t) = \frac{\rho_{\text{crit}}(t) - \rho(t)}{\rho_{\text{crit}}(t)} = -\frac{k}{R_0^2} [aH(a)]^{-2} \quad (132)$$

given the critical density $\rho_{\text{crit}}(t) = 3H^2(t)/(8\pi G)$ (which is the density required for a flat Universe). Notably, this depends on the comoving Hubble radius at time t ; scaling to the value today, we find

$$\Omega_k(a) = \left(\frac{a_0 H_0}{aH(a)} \right)^2 \Omega_{k,0}. \quad (133)$$

Observations of the CMB and galaxy clustering constrain $|\Omega_{k,0}| \lesssim 10^{-3}$; since $[aH(a)]^{-1}$ is an increasing function of a in the Hot Big Bang model, the upper limit on $|\Omega_k(a)|$ must strengthen as a decreases. In other words, the observed flatness of the Universe today implies that the primordial Universe was *extremely* flat.

To quantify this, let us consider the Universe at recombination. Assuming matter domination for all $a > a_{\text{eq}}$ (which is inaccurate at both low and high redshifts, but a fair approximation for our purposes), we have $H^2(a) \propto \rho_m(a) \propto a^{-3}$ and $[aH(a)]^{-1} \propto a^{1/2}$, thus

$$|\Omega_k(a_{\text{rec}})| \simeq |\Omega_{k,0}| \cdot 10^{-3} \lesssim 10^{-6}, \quad (134)$$

thus the energy density in curvature at recombination was less than one part in 10^6 . Noting that $[aH(a)]^{-1} \propto a$ in radiation domination, we can extend this constraint to the BBN epoch:

$$|\Omega_k(a_{\text{BBN}})| \simeq 10^{-12} |\Omega_k(t_{\text{rec}})| \lesssim 10^{-18}, \quad (135)$$

³⁵Note that this assumes that times close to the initial singularity contribute negligibly to $\chi_{\text{ph}}(\eta)$, *i.e.* assumes that $\int_0^{\delta t} \frac{dt}{a(t)}$ is small. While this could be avoided by invoking suitably bizarre quantum processes in the early Universe, this does not solve the curvature and monopole problems discussed below.

³⁶It is also important to note that the temperature of the CMB today is primarily set by the time at which recombination occurred. Regardless of the initial conditions, the temperature of recombination is fixed by atomic processes; however, a Universe with higher initial energy will recombine at later time, implying less time for the photons to redshift before being observed today.

and, at least heuristically, to the Planck time, finding $|\Omega_k(a_{\text{Pl}})| \lesssim 10^{-60}$. The radius of curvature of the early Universe (which is constrained to $R_k \gtrsim 21/H_0$ today) must therefore have been *many* orders of magnitude larger than the Hubble radius. This extreme fine-tuning of the initial curvature (plus the inherent dynamical instability of curvature) is known as the **flatness problem**.

4.1.3 The Monopole Problem

Grand Unified Theories (GUTs), which aim to describe the unification of the electroweak and strong forces at $\mathcal{O}(10^{16})$ GeV, predict that topological defects, such as magnetic monopoles, will be produced as the early Universe goes through a phase transition. At the time of production, these are expected to have a density of $\mathcal{O}(1)$ per Hubble patch, which would source a humungous density today (with $\Omega_{\text{mon},0} \sim 10^{10}$). Moreover, their energy density would quickly dominate the expansion history of the Universe, and overclose the Universe. Despite extensive searches, no magnetic monopoles have been detected, which is in clear tension with the theoretical predictions, and leads to the **monopole problem**.

4.1.4 Superhorizon Correlations

The final problem concerns the $\mathcal{O}(10^{-5})$ fluctuations observed in the CMB temperature (and the much larger fluctuations seen in the galaxy density today). Working in comoving coordinates, the wavelength of a given perturbation, λ , is fixed as we trace the physics back from recombination to the primordial singularity, whilst the comoving Hubble radius $[aH(a)]^{-1}$ shrinks. Any mode that lies inside the Hubble sphere today ($\lambda < [a_0 H(a_0)]^{-1}$; subhorizon) must have been outside at earlier times ($\lambda > [aH(a)]^{-1}$; superhorizon).³⁷ Moreover, a CMB fluctuation with $\lambda > \eta_{\text{rec}} = 280$ Mpc was outside the horizon for all $\eta < \eta_{\text{rec}}$, meaning that it could not be generated by causal physics (which is restricted to the horizon, since one cannot travel faster than the speed of light). Empirically, we detect fluctuations up to $\lambda_{\text{max}} \approx 14000$ Mpc in the form of large-scale correlations in the CMB, providing a further challenge for the conventional cosmological model.

4.2 The Inflationary Solution

The four problems described arise because we require that the comoving Hubble radius, $[aH(a)]^{-1}$, is an increasing function of time. The solution is clear: add a new epoch to the early Universe in which the Hubble radius *shrinks*, *i.e.*

$$\boxed{\frac{d}{dt}([a(t)H(t)]^{-1}) < 0.} \quad (136)$$

Noting that $H = \dot{a}/a$, this implies that $-\ddot{a}/\dot{a}^2 < 0$, thus $\ddot{a} > 0$ and the Universe undergoes accelerated expansion. This scenario is known as **inflation**, and replaces the Big Bang singularity by a reheating event at $\eta = 0$ (which we'll discuss below), preceded by a phase of accelerated expansion extending back to $\eta = -\infty$.

Inflationary Horizons To understand the implications of this, let's return to the definition of the comoving particle horizon:

$$\chi_{\text{ph}}(\eta) = \eta - \eta_i = \int_{\log a_i}^{\log a} \frac{d \log a'}{a' H(a')} \propto \frac{2}{1+3w} \left[a^{(1+3w)/2} - a_i^{(1+3w)/2} \right], \quad (137)$$

³⁷Note that this assumes $[aH(a)]^{-1} \sim \chi_{\text{ph}}(a)$, which is true in the Hot Big Bang model.

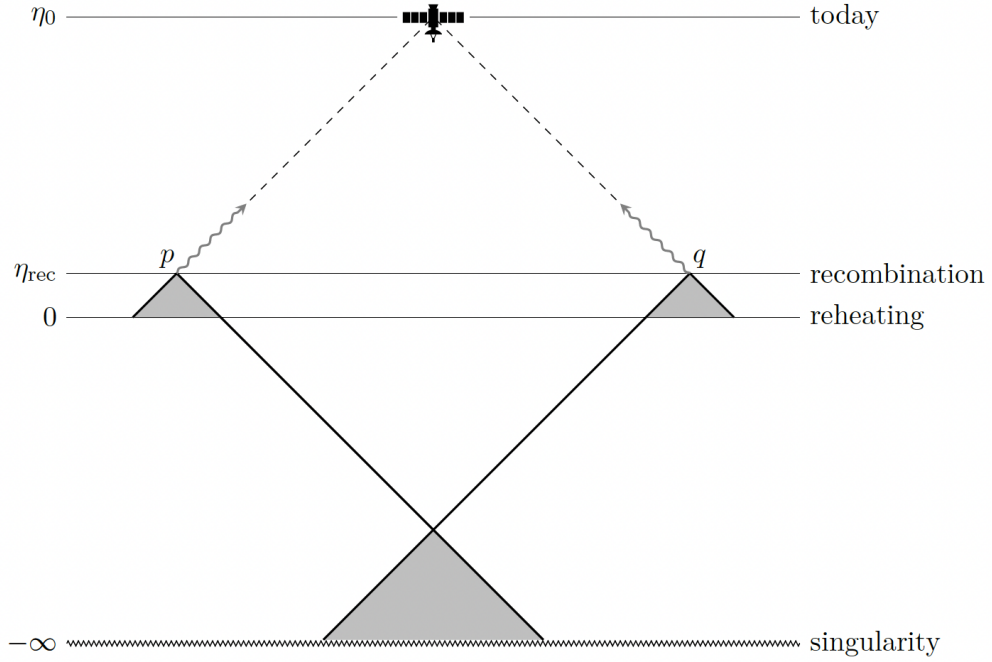


Figure 6: Sketch of the horizon solution in the inflationary model. Inflation extends the conformal diagram shown in Fig. 5 to negative conformal times, introducing a primordial period of shrinking Hubble radius (and thus accelerated expansion). Though points p and q are not in causal contact at reheating, they were in contact deep in the past (shown by the central grey region). This allowed their temperatures to equilibrate, solving the horizon problem. Reproduced from D. Baumann, *Cosmology* (Cambridge University Press, 2022).

introducing an equation-of-state parameter on the right-hand side. If $[aH(a)]^{-1}$ is large at early times (which is forbidden in the Hot Big Bang model), the integral can be dominated by the lower limit, such that $\chi_{\text{ph}}(\eta) \gg \eta$. In terms of the equation of state, this requires $w < -1/3$, which implies a violation of the strong energy condition.

In the inflationary scenario, the limit $a_i \rightarrow 0$ therefore corresponds to $\eta_i \rightarrow -\infty$ rather than $\eta_i \rightarrow 0$, and the spacetime diagram shown in Fig. 5 must be extended to negative conformal time, as shown in Fig. 6. This implies that points on the last scattering surface separated by comoving distances $\chi > \eta_{\text{rec}}$ are causally connected, provided that $\chi < \eta_{\text{rec}} - \eta_i \gg \eta_{\text{rec}}$. This provides a natural explanation for the horizon problem: if the entire observed Universe was in causal contact at $\eta \ll 0$, the system could equilibrate and exchange energy, sourcing the uniform CMB temperature we see today.

How much inflation do we need? To qualitatively solve the horizon problem, we require that the entire observable Universe today fits inside a single Hubble patch at the start of inflation:

$$[a_i H_i]^{-1} > [a_0 H_0]^{-1} \quad (138)$$

(writing $H_i \equiv H(a_i)$ *et cetera*). By relating the Hubble rate today to that at the end of inflation (noting that the interim physics is fairly well understood), we can estimate the required duration of the inflationary period, and thus the amount of accelerated expansion. Let us suppose that inflation ends via a period of ‘reheating’ at temperature $T_R \lesssim 10^{15}$ GeV.³⁸ From reheating until matter-radiation equality at $a_{\text{eq}} \sim 10^{-3}$, the Universe is dominated by radiation, thus $H^2 \propto \rho \propto a^{-4}$ and the Hubble radius scales as a^{-1} . After equality, the Hubble radius scales as $a^{-1/2}$ (ignoring dark energy domination): in combination, we find

$$\frac{a_0 H_0}{a_R H_R} \sim \frac{a_0 H_0}{a_{\text{eq}} H_{\text{eq}}} \frac{a_{\text{eq}} H_{\text{eq}}}{a_R H_R} \sim \frac{a_R}{a_{\text{eq}}} \left(\frac{a_{\text{eq}}}{a_0} \right)^{1/2} \sim \frac{T_{\text{eq}}}{T_R} \left(\frac{T_0}{T_{\text{eq}}} \right)^{1/2} \sim 10^{-28} \left(\frac{10^{15} \text{ GeV}}{T_R} \right), \quad (139)$$

setting $a \sim T^{-1}$, ignoring the small changes in $g_{*S}^{1/3}(T)$. Identifying the end of inflation with reheating, $a_e H_e \approx a_R H_R$, we therefore require

$$[a_i H_i]^{-1} \gtrsim 10^{28} \left(\frac{T_R}{10^{15} \text{ GeV}} \right) [a_e H_e]^{-1}. \quad (140)$$

Over the course of inflation, the comoving Hubble radius must shrink by a factor $\gtrsim e^{60 + \log(T_R/10^{15} \text{ GeV})}$. Since $H \simeq \text{const}$ during inflation (see below), this implies that the scale factor must grow by $\gtrsim 60$ e-folds, *i.e.* that the Universe expands by a factor greater than e^{60} in each dimension.³⁹

Other Problems Inflation also explains the large-scale correlations seen in the CMB. Due to the large Hubble radius in the early Universe, all modes visible today start subhorizon (with $\lambda < [aH(a)]^{-1}$), then become superhorizon as the Universe inflates (with $\lambda > [aH(a)]^{-1}$) before eventually re-entering the horizon in the post-inflationary universe. This implies that the correlations can be generated by standard causal physics (in particular, the quantum vacuum fluctuations discussed below). Moreover, inflation explains why the Universe is flat: since $\Omega_k(a) \propto [aH(a)]^{-2}$, a

³⁸An upper bound on T_R is given by the non-detection of primordial gravitational waves. BBN provides a lower bound of $T_R \gtrsim 1$ MeV, though it is likely much earlier, to allow for other processes such as electroweak symmetry breaking and baryogenesis.

³⁹This all happens in around 10^{-35} seconds of coordinate time, though the conformal time interval is much larger (from $\eta_i \lesssim -10^{60}$ to $\eta_e \approx 0$).

primordial curvature of $\mathcal{O}(1)$ is diluted by $\gtrsim e^{2 \times 60}$ by the end of inflation, causing no tension with the observational bounds from the late Universe. The density of magnetic monopoles in the early Universe will be similarly diluted by the expansion (reducing the comoving density by $\gtrsim e^{3 \times 60}$), explaining why none have been detected today.

4.3 Conditions for Inflation

Inflation solves the horizon, flatness, monopole, and super-horizon correlation problems by adding an early epoch with a decreasing comoving Hubble radius. This dilutes any primordial curvature, stretches the particle horizon at recombination far beyond its Hot Big Bang value, and extends η_i to $-\infty$, providing the extra conformal time needed to bring distant regions of the observable Universe into causal contact. Below, we'll discuss how to obtain such an epoch in practice.

To quantify inflation, it is useful to introduce the **slow-roll parameters**, $\{\epsilon, \eta\}$,⁴⁰ where

$$\epsilon \equiv -\frac{d \log H}{dN} = -\frac{\dot{H}}{H^2}, \quad \eta \equiv \frac{d \log \epsilon}{dN} = \frac{\dot{\epsilon}}{H\epsilon}, \quad (141)$$

where $dN = d \log a$ parametrizes the number of *e-folds* in inflation (with $a \sim e^N$). The condition of a shrinking Hubble radius can be written as

$$\frac{d}{dt}([aH(a)]^{-1}) = -\frac{\dot{a}H + a\dot{H}}{(aH)^2} = -\frac{1}{a}(1 - \epsilon) < 0 \quad \Rightarrow \quad \epsilon < 1; \quad (142)$$

this implies that the Hubble rate varies slowly as the Universe expands (*i.e.* the change in $\log H$ per e-fold of expansion is small). To solve the horizon, curvature, monopole and correlation problems, we require that inflation persists for $N \gtrsim 60$ e-folds: this implies that $|\eta| < 1$, such that ϵ is also slowly varying. The conditions

$$\epsilon < 1, |\eta| < 1 \quad (143)$$

are known as the **slow-roll** conditions and can be used to define valid inflationary models.⁴¹

In the limit of small ϵ (which is suggested by observations), we have $d \log H / dN \approx 0$, so $H \equiv d \log a / dt$ is approximately constant. This implies that the scale factor grows exponentially, with $a(t) \propto e^{Ht}$, and that the geometry of the Universe approaches **de Sitter space**, with line element

$$ds^2 = -dt^2 + e^{2Ht} d\vec{x}^2. \quad (144)$$

Realistic inflationary solutions correspond to **quasi-de Sitter** geometries, with ϵ small but non-zero (such that inflation eventually ends).

4.4 Scalar Field Inflation

To drive inflation, we require a fluid with equation-of-state parameter $w < -\frac{1}{3}$. From the continuity equation, $H^2 \propto \rho \propto a^{-3(1+w)}$, thus

$$\epsilon = -\frac{d \log H}{dN} = \frac{3}{2}(1 + w), \quad (145)$$

⁴⁰This is not to be confused with the conformal time, which is confusingly also labelled η .

⁴¹Strictly, $|\eta| < 1$ is not required at all times (see, e.g., ultra-slow-roll models). Observations suggest that $\epsilon \ll 1$ across the range of e-folds probed by CMB anisotropies.

which recovers $\epsilon < 1$, as expected. The required fluid has two notable properties: (1) it has negative pressure: $P < -\rho/3$, (2) it has almost constant energy density (since $|d \log \rho / d \log a| = |d \log H^2 / d \log a| = 2|\epsilon|$). As discussed above, these conditions violate the strong energy condition, thus inflation cannot be driven by conventional gravitating fields.

The simplest realization of inflation involves a scalar field, $\phi(\vec{x}, t)$, imaginatively labeled the **inflaton**.⁴² The dynamics of ϕ are specified by the action, which reads

$$S_\phi = \int dt d^3x a^3(t) \left[\frac{1}{2} \dot{\phi}^2 - V(\phi) \right], \quad (146)$$

where we have assumed homogeneity and isotropy (such that ϕ depends only on t), and inserted the FLRW metric.⁴³ This is specified by the potential $V(\phi)$ and the (canonical) kinetic energy $\frac{1}{2} \dot{\phi}^2$.

To compute the equation of motion for ϕ , we use the principle of least action, solving for $\frac{\delta S}{\delta \phi} = 0$. Setting $\phi \rightarrow \phi + \delta\phi$, we find

$$\delta S = \int dt d^3x a^3 \left[\dot{\phi} \delta \dot{\phi} - V'(\phi) \delta \phi \right] = \int dt d^3x \left[-\frac{d}{dt} (a^3 \dot{\phi}) - a^3 V'(\phi) \right] \delta \phi, \quad (147)$$

using integration by parts and denoting $V' \equiv dV/d\phi$. Setting $\delta S = 0$, we obtain the **Klein-Gordon equation** in an expanding Universe:

$$\boxed{\ddot{\phi} + 3H\dot{\phi} = -\frac{dV}{d\phi}}. \quad (148)$$

This describes a damped harmonic oscillator, with the $3H\dot{\phi}$ term describing **Hubble friction**, which causes the inflaton to lose energy as the Universe expands.⁴⁴

To compute the effect of ϕ on the expansion rate, we use the Friedmann equations. These require the scalar-field density and pressure, which can be derived from the stress-energy tensor, itself obtained from the scalar field action.⁴⁵ Skipping some calculation, we find

$$\rho_\phi = \frac{1}{2} \dot{\phi}^2 + V(\phi), \quad P_\phi = \frac{1}{2} \dot{\phi}^2 - V(\phi). \quad (149)$$

Using the Klein-Gordon equation, it is straightforward to show that these satisfy the continuity equation, $\dot{\rho}_\phi + 3H(\rho + P) = 0$. Inserting the energy density and pressure into the first Friedmann equation gives

$$H^2 = \frac{1}{3M_{\text{Pl}}^2} \left[\frac{1}{2} \dot{\phi}^2 + V(\phi) \right], \quad (150)$$

⁴²The theory of inflation has its origins in the study of scalar fields in an expanding universe – it was quickly realized that these could solve the myriad of problems plaguing the Hot Big Bang, though the precise mechanism took a few years to develop.

⁴³Formally, this follows from the covariant action for a minimally-coupled scalar field $S = \frac{M_{\text{Pl}}^2}{2} \int d^4x \sqrt{-g} R + \int d^4x \sqrt{-g} \mathcal{L}(\phi, \nabla_\mu \phi)$, where $\mathcal{L}$ is the scalar-field Lagrangian and the first piece is the Einstein-Hilbert term. Assuming a canonical kinetic term, the Lagrangian is given by $\mathcal{L}(\phi, \nabla_\mu \phi) = -\frac{1}{2} g^{\mu\nu} \nabla_\mu \phi \nabla_\nu \phi - V(\phi)$, where V is the potential. Neglecting fluctuations in ϕ and inserting the FLRW metric (with $\sqrt{-g} = a^3$), this reduces to (146), dropping the Einstein-Hilbert term.

⁴⁴This can also be obtained from the covariant action given in the previous footnote. Varying with respect to ϕ via the Euler-Lagrange equation, we obtain $\square \phi = dV/d\phi$ (for d'Alembertian $\square = g^{\mu\nu} \nabla_\mu \nabla_\nu$), which reduces to the Klein-Gordon equation on the FLRW background. Varying with respect to the metric leads to the Einstein's equations, which simplify to the Friedmann equations on the FLRW background.

⁴⁵In full, the stress-energy tensor is defined as $T_{\mu\nu}^{(\phi)} \equiv -(2/\sqrt{-g}) \delta S / \delta g^{\mu\nu}$. For the minimally coupled action given in the preceding footnotes, this gives $T_{\mu\nu}^{(\phi)} = \partial_\mu \phi \partial_\nu \phi - g_{\mu\nu} [\frac{1}{2} \partial^\alpha \phi \partial_\alpha \phi + V(\phi)]$. The density and pressure can be read-off by comparing to the standard covariant fluid action.

(where $M_{\text{Pl}} = (8\pi G)^{-1/2}$ is the reduced Planck mass), which relates the inflaton dynamics to the expansion rate. By differentiating this expression with respect to time, we can compute the first slow-roll parameter:

$$2H\dot{H} = \frac{\dot{\phi}}{3M_{\text{Pl}}^2}(\ddot{\phi} + V') = -\frac{H\dot{\phi}^2}{M_{\text{Pl}}^2} \Rightarrow \epsilon \equiv -\frac{\dot{H}}{H^2} = \frac{3}{2} \frac{\dot{\phi}^2}{\frac{1}{2}\dot{\phi}^2 + V(\phi)}, \quad (151)$$

using the Klein-Gordon equation. The inflationary condition $\epsilon < 1$ implies that the kinetic energy of the inflaton must be small compared to the potential energy, *i.e.* the field should be **slowly rolling**. Moreover, the second slow-roll parameter can be written

$$\eta \equiv -\frac{\dot{\epsilon}}{H\epsilon} = 2 \left[\frac{\ddot{\phi}}{H\dot{\phi}} - \frac{\dot{H}}{H^2} \right] \equiv 2(\epsilon - \delta), \quad (152)$$

for $\delta \equiv -\ddot{\phi}/(H\dot{\phi})$, which satisfies $|\eta| \ll 1$ if the acceleration term ($\propto \ddot{\phi}$) is small.

4.5 Slow-Roll Solution

Above, we obtained the ϵ, η parameters for an arbitrary inflationary potential and obtained the conditions under which inflation occurs. Motivated by this, we define the **slow-roll** regime by the two conditions

$$\frac{1}{2}\dot{\phi}^2 \ll V(\phi), \quad |\ddot{\phi}| \ll |3H\dot{\phi}|, \quad (153)$$

i.e. negligible kinetic energy and acceleration.⁴⁶ In these limits, the Friedmann and Klein-Gordon equations simplify to

$$H^2 \approx \frac{V(\phi)}{3M_{\text{Pl}}^2}, \quad 3H\dot{\phi} \approx -V'(\phi), \quad (154)$$

and the slow-roll parameter becomes

$$\epsilon \simeq \frac{\frac{1}{2}\dot{\phi}^2}{M_{\text{Pl}}^2 H^2} \simeq \frac{M_{\text{Pl}}^2}{2} \left(\frac{V'}{V} \right)^2. \quad (155)$$

We find that $\epsilon \ll 1$ requires $|V'/V| \ll M_{\text{Pl}}^{-1}$, *i.e.* the inflationary potential must be sufficiently **flat** (in Planck units). Differentiating the Klein-Gordon equation with respect to time, we find:

$$\frac{d}{dt} \left(3H\dot{\phi} + V'(\phi) \right) = -3H^2\dot{\phi} \left(\epsilon + \delta - \frac{V''}{3H^2} \right) = 0 \Rightarrow \epsilon + \delta = M_{\text{Pl}}^2 \frac{V''}{V}, \quad (156)$$

which constrains the second derivative of the potential. Based on this, we define the **potential slow-roll parameters**:

$$\boxed{\epsilon_V \equiv \frac{M_{\text{Pl}}^2}{2} \left(\frac{V'}{V} \right)^2, \quad \eta_V \equiv M_{\text{Pl}}^2 \frac{V''}{V};} \quad (157)$$

if these are both small, we obtain slow-roll inflation (with $\epsilon \approx \epsilon_V$ and $\eta_V \approx 2\epsilon - \eta/2$). These parameters are useful since they can be computed directly from the microphysical model without solving the full equations of motion.

⁴⁶These correspond to $\epsilon \ll 1$ and $|\delta| \ll 1$.

Finally, we can measure the total number of e-folds of inflation (*i.e.* the total expansion). Integrating from a_i to a_e , with $d \log a = H dt = (H/\dot{\phi}) d\phi$, we find

$$N_{\text{tot}} = \int_{a_i}^{a_e} d \log a = \int_{\phi_i}^{\phi_e} \frac{H}{\dot{\phi}} d\phi, \quad (158)$$

which simplifies to

$$N_{\text{tot}} \approx \int_{\phi_e}^{\phi_i} \frac{1}{\sqrt{2\epsilon_V}} \frac{|d\phi|}{M_{\text{Pl}}} \quad (159)$$

under the slow-roll approximation (since $H/\dot{\phi} = -(V/V')/M_{\text{Pl}}^2 = -2\sqrt{\epsilon_V}$). Recalling that $N_{\text{tot}} \gtrsim 60$ is needed to solve the horizon and flatness problems, we thus constrain the field excursion $\Delta\phi = |\phi_i - \phi_e|$ during inflation (with super-Planckian excursions if ϵ_V is large).

4.6 Summary of Inflation

In the past few sections, we have presented a variety of inflationary conditions. Here, we briefly pause to summarize these points, before proceeding with the post-inflationary dynamics. In its standard form, inflation has the following properties:

- **Exponential expansion:** Inflation requires accelerated expansion ($\ddot{a} > 0$), typically assumed to be exponential with $a(t) \propto e^{Ht}$. This corresponds to almost constant H ($\epsilon, |\eta| \ll 1$), and a (quasi-)de Sitter geometry.
- **The inflaton:** Inflation can be generated by a scalar field, ϕ , evolving along a potential $V(\phi)$. The evolution of ϕ is set by the Klein-Gordon equation, and informs the expansion history via the Friedmann equations.
- **Slow-Roll:** In the most common models for inflation, the inflaton ‘slowly-rolls’ down its potential. This corresponds to subdominant kinetic energy ($\epsilon \ll 1$) and small acceleration ($|\eta|, |\delta| \ll 1$), with an effective equation-of-state around $w_\phi \approx -1$.
- **Potential:** To obtain a slow-roll solution, the inflaton potential must be flat ($\epsilon_V \ll 1$) over a fairly wide range of ϕ ($|\eta_V| \ll 1$). This must hold at least for the ≈ 7 e-folds visible in the CMB, but is usually assumed for the $N_{\text{tot}} \gtrsim 60$ e-folds required to solve the flatness and horizon problems.

4.7 Reheating

Inflation ends when the slow-roll parameters, ϵ, η , become $\mathcal{O}(1)$. The transition phase that follows is known as **reheating**, and is responsible for converting the inflaton energy density into Standard Model particles. This phase is amongst the most mysterious in the Universe and is difficult to probe directly, though primordial gravitational waves can provide some insight. Here, we will present only a brief overview of the reheating scenario.

At the end of inflation, the kinetic and potential energies of the inflaton become comparable (since $\epsilon \sim 1$), and the inflaton begins to oscillate about the minimum of its potential (see Fig. 7). Assuming a quadratic minimum, $V(\phi) \approx V_0 + \frac{1}{2}m^2(\phi - \phi_0)^2$ for some mass-scale m , the Klein-Gordon equation becomes

$$\ddot{\phi} + 3H\dot{\phi} = -m^2\phi, \quad (160)$$

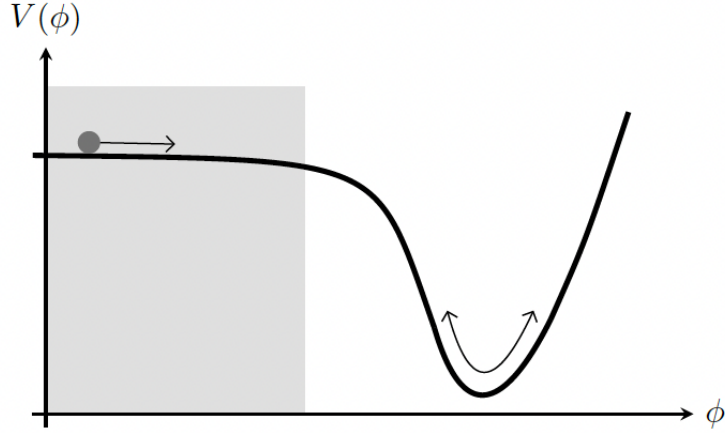


Figure 7: Cartoon of the inflaton potential. During the inflationary period (indicated by the gray region), ϕ slowly rolls along the potential surface. The almost-flatness of $V(\phi)$ translates into almost-exponential expansion (and an almost-scale-invariant spectrum of CMB fluctuations). At the end of inflation, the field oscillates around the potential minimum (the gulch on the right-hand-side) sourcing reheating. This transfers energy to the standard model sector, through decay and thermalization processes. Reproduced from D. Baumann, *Cosmology* (Cambridge University Press, 2022).

which defines a damped simple harmonic oscillator, with subdominant Hubble friction if $H \ll m$.

From the continuity equation, $\dot{\rho}_\phi + 3H(\rho_\phi + P_\phi) = 0$, which can be rearranged as

$$\frac{1}{a^3} \frac{d}{dt} (a^3 \rho_\phi) = -3HP_\phi = -\frac{3H}{2} (\dot{\phi}^2 - m^2 \phi^2), \quad (161)$$

(dropping the V_0 term). If the oscillations are rapid, the right-hand-side averages to zero (from the virial theorem), implying that $\rho_\phi \propto a^{-3}$ and the inflaton behaves as pressureless matter.

As the energy density of the inflaton drops, interactions between the inflaton and other fields become important, transferring energy into the Standard Model sector. The simplest models involve a slow decay of the inflaton into light fermionic species, which can be modeled via the Boltzmann equation

$$\frac{1}{a^3} \frac{d}{dt} (a^3 \rho_\phi) = -\Gamma_\phi \rho_\phi, \quad (162)$$

for decay rate Γ_ϕ . Other models of inflation involve fast decay into bosons via parametric resonance. This out-of-equilibrium process is usually known as ‘preheating’. Whichever mechanism operates, the decay products quickly interact and thermalize, forming a radiation bath with energy density

$$\rho_R = \frac{\pi^2}{30} g_\star(T_R) T_R^4 < \rho_\phi, \quad (163)$$

where the reheating temperature T_R depends on the duration of the reheating phase (set by Γ_ϕ and the precise interaction vertices). A lower bound of $T_R \gtrsim 1$ MeV is sourced by BBN (which requires neutrino equilibrium to be established); the reheating temperature is likely substantially higher to accommodate additional phenomena such as baryogenesis. Once the thermalized radiation bath dominates, the thermal history described in §3 begins.

4.8 Inflationary Fluctuations

In addition to resolving the horizon, flatness, and monopole problems, inflation provides a natural origin for the structure we observe in the Universe today, through the quantum fluctuations of the inflaton field ϕ . The full theory of inflationary perturbations is mathematically rich and grounded in relativistic quantum field theory. Rather than diving into the details (which really requires its own course), we here simply present a heuristic outline of the main ingredients and the key results, which we'll use in later sections.⁴⁷

The main result of inflationary perturbation theory can be understood from the Heisenberg uncertainty principle, $\Delta E \Delta t \simeq \hbar/2$. Here, $\Delta E \sim \delta\phi$ (since the inflaton field has units of energy), and $\Delta t \sim 1/H$ (since the only relevant timescale during (quasi-de Sitter) inflation is the Hubble scale), we conclude that the inflaton fluctuations must have characteristic size

$$\delta\phi \sim \frac{H}{2\pi} \quad (164)$$

(adding a factor of π to match the results of a more careful calculation). Due to the Einstein equations, the quantum fluctuations in ϕ source fluctuations in the spacetime metric (in particular, the local curvature), leading to an inhomogeneous Universe at the end of inflation. These curvature perturbations later grow under gravity and accrete material, forming the overdensities we observe today as galaxies, clusters, and CMB anisotropies. Below, we'll explore this picture in more detail, starting from a classical description then adding quantization.

4.8.1 Classical Perturbations

Up to now, we have assumed that the inflaton field (in addition to all other quantities) is independent of space, *i.e.* $\phi(\vec{x}, t) = \bar{\phi}(t)$. To describe inflationary perturbations, we must go beyond this assumption, writing

$$\phi(\vec{x}, t) \equiv \bar{\phi}(t) + \delta\phi(\vec{x}, t) \quad (165)$$

where $\bar{\phi}$ is the background solution discussed previously, and $\int d\vec{x} \delta\phi(\vec{x}, t) = 0$. Note that this explicitly breaks homogeneity.

In the previous sections, we derived the equation of motion for the background solution:

$$\ddot{\bar{\phi}}(t) + 3H(t)\dot{\bar{\phi}}(t) = -\frac{dV}{d\phi}, \quad (166)$$

starting from the background action $S_\phi = \int dt d^3x a^3(t) \left[\frac{1}{2} \dot{\bar{\phi}}^2 - V(\bar{\phi}) \right]$. Switching to conformal time (with $ad\eta = dt$), this can be written

$$\bar{\phi}''(\eta) + 2\mathcal{H}\bar{\phi}'(\eta) = -a^2 \frac{dV}{d\phi}, \quad (167)$$

where $X'(\eta) \equiv dX/d\eta$ and $\mathcal{H}(\eta) \equiv a'/a = aH$ is the conformal Hubble parameter. Working at first-order in fluctuations (*i.e.* neglecting terms of order $(\delta\phi)^2$), one can derive a similar expression

⁴⁷Excellent references for this section include Daniel Baumann's TASI lectures and Enrico Pajer's Part III notes, both of which feature an in-depth treatment of inflationary perturbation theory.

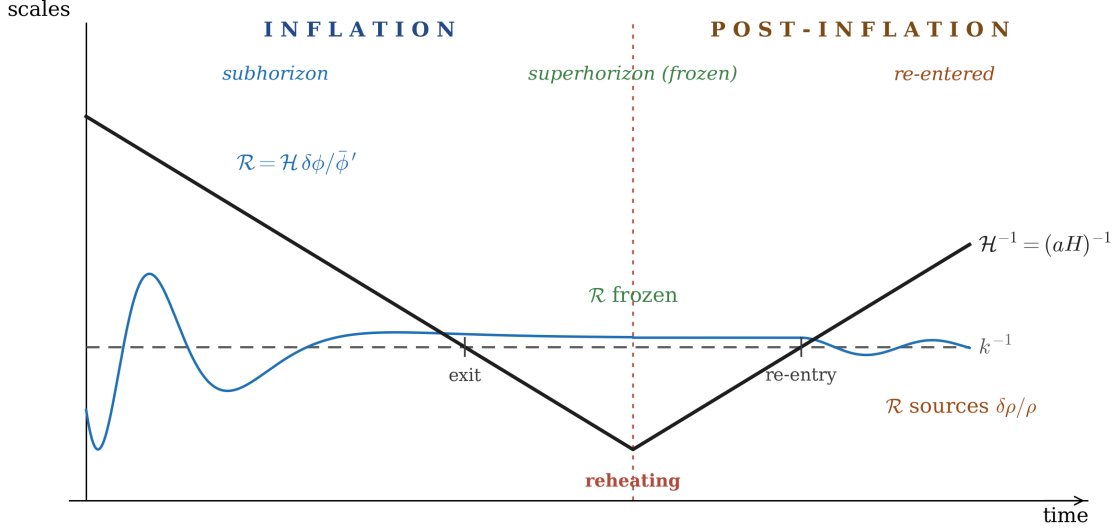


Figure 8: Schematic of the perturbation evolution during inflation. The blue line shows the curvature fluctuation, $\mathcal{R}$, sourced by the inflationary perturbation $\delta\phi$ for a given Fourier-mode k . At early times, the mode is subhorizon ($k \gg \mathcal{H}$) and oscillates as a free field. After horizon-crossing ($k \gtrsim \mathcal{H}$), the mode freezes and remains fixed until horizon re-entry ($k \lesssim \mathcal{H}$) in the late-Universe, where it resumes oscillating and sources physical density fluctuations.

for the fluctuations:⁴⁸

$$\delta\phi''(\vec{x}, \eta) + 2\mathcal{H}\delta\phi'(\vec{x}, \eta) - \nabla^2\delta\phi(\vec{x}, \eta) = \left[2\tilde{\epsilon}(3 - \tilde{\epsilon} + \tilde{\eta}) - \frac{a^2}{\mathcal{H}^2} \frac{d^2V}{d\phi^2}\right] \mathcal{H}^2 \delta\phi(\vec{x}, \eta). \quad (168)$$

This takes a similar form to the background solution with two key differences: (1) it includes spatial derivatives due to the broken homogeneity; (2) it explicitly depends on the slow-roll parameters $\tilde{\epsilon}, \tilde{\eta}$ (adding tildes to distinguish from conformal time) which encode metric fluctuations in the spatially flat gauge (see §5).

The above equation can be greatly simplified by introducing the rescaled field $f(\vec{x}, \eta) \equiv a(\eta)\delta\phi(\vec{x}, \eta)$ and Fourier-transforming, with $f(\vec{k}, \eta) \equiv \int d^3x e^{-i\vec{k}\cdot\vec{x}} f(\vec{x}, \eta)$. Following some tedious manipulations, we find the **Mukhanov–Sasaki equation**:

$$f''(\vec{k}, \eta) + \left(k^2 - \frac{z''(\eta)}{z(\eta)}\right) f(\vec{k}, \eta) = 0 \quad (169)$$

defining $z(\eta) \equiv \frac{a\bar{\phi}'}{\mathcal{H}}$. This is a highly general expression, describing the evolution of inflaton perturbations during inflation. Notably, the Mukhanov-Sasaki equation takes the form of a wave equation with an effective mass term $-z''/z$, whose solution is shown in Fig. 8. In the slow-roll limit (which is *not* assumed by the above expression), the mass reduces to $z''/z \simeq 2\mathcal{H}^2 = 2a^2H^2$, which demarcates two regimes of interest:

⁴⁸To derive this, one extremises the relativistic action $S_\phi = \int d^4x \sqrt{-g} [-\frac{1}{2}g^{\mu\nu}\nabla_\mu\phi\nabla_\nu\phi - V(\phi)]$, with respect to ϕ , which yields the Klein-Gordon equation $\frac{1}{\sqrt{-g}}\partial_\mu(\sqrt{-g}g^{\mu\nu}\partial_\nu\phi) = -dV/d\phi$. This can be simplified by inserting the metric (at first-order in fluctuations, working in the spatially-flat gauge) and rewriting metric fluctuations in terms of slow-roll parameters and $\delta\phi$.

- **Subhorizon** ($k \gg \mathcal{H} \equiv aH$): The effective mass is negligible and the solutions are plane waves with, $f(k, \eta) \sim e^{\pm ik\eta}$. In this case, which corresponds to early in inflation, when the characteristic wavelength $\lambda \sim 1/k$ is much smaller than the comoving Hubble radius $1/(aH)$, the modes oscillate freely, akin to free fields in Minkowski spacetime.
- **Superhorizon** ($k \ll \mathcal{H} \equiv aH$): In the opposite limit, the k^2 term is negligible and the equation becomes $f'' \approx (z''/z) f$, which admits the growing and decaying solutions $f \propto z$, $f \propto z^{-2}$. This regime is reached late in inflation, when the (rapidly shrinking) Hubble radius falls below the characteristic wavelength, and leads to the perturbations *freezing*, with f/z becoming constant in time, as shown in Fig. 8.

The qualitative behavior of $\delta\phi(\vec{k}, \eta)$ matches that discussed above: modes are generated on subhorizon scales, then exit the horizon mid-way through inflation. These modes subsequently re-enter the horizon after inflation ends, at which point they can be observed.

4.8.2 Quantum Perturbations

In the above section, we have presented the equation of motion for $\delta\phi$ using classical field theory techniques. Analogously to the particles in the standard model, ϕ is a quantum field, thus it is properly described by quantum (and not classical) field theory. Practically, this corresponds to replacing $\delta\phi$ with a quantized equivalent

$$\delta\phi(\vec{k}, \eta) \rightarrow \hat{\delta\phi}(\vec{k}, \eta) = \delta\phi(k, \eta) \hat{a}_{\vec{k}} + \delta\phi^*(k, \eta) \hat{a}_{-\vec{k}}^\dagger, \quad (170)$$

where $\hat{a}, \hat{a}^\dagger$ are creation and annihilation operators, and $\delta\phi$ is the mode-function (enforcing reality of ϕ).⁴⁹ Here, the creation and annihilation operators satisfy the commutation relation

$$[\hat{a}_{\vec{k}}, \hat{a}_{\vec{k}'}^\dagger] = (2\pi)^3 \delta_D(\vec{k} - \vec{k}') \quad (171)$$

and $\delta\phi(k, \eta)$ satisfy the Mukhanov–Sasaki equation given above.

To solve the system, we must specify the Initial conditions. A natural choice is the analogue of the ground state of a simple harmonic oscillator, in which every mode approaches the positive-frequency Minkowski vacuum on deep sub-horizon scales ($\delta\phi \sim e^{-ik\eta}$). This choice is known as the **Bunch–Davies vacuum**, and is the standard initial condition for inflationary perturbations. Under the slow-roll assumptions, we can solve the Mukhanov-Sasaki equation to find

$$f(k, \eta) = a(\eta) \delta\phi(k, \eta) = \frac{1}{\sqrt{2k}} \left(1 - \frac{i}{k\eta}\right) e^{-ik\eta} \quad (172)$$

We now have all the ingredients we need to compute the statistics of the inflaton fluctuations. As usual in quantum mechanics, we compute averages by sandwiching the operators between vacuum states, $|0\rangle$. First, we can compute the expectation value of the perturbation itself:

$$\langle 0 | \hat{\delta\phi}(\vec{k}, \eta) | 0 \rangle = 0, \quad (173)$$

noting that $\hat{a}_{\vec{k}}|0\rangle = \langle 0|\hat{a}_{\vec{k}}^\dagger = 0$. As expected, the fluctuations in ϕ vanish on average. Next, we can compute the **two-point function / power spectrum**, which encodes the variance of the field as a function of scale:

$$\langle 0 | \hat{\delta\phi}(\vec{k}, \eta) \hat{\delta\phi}^*(\vec{k}', \eta) | 0 \rangle = |\delta\phi(k, \eta)|^2 (2\pi)^3 \delta_D(\vec{k} - \vec{k}'). \quad (174)$$

⁴⁹These redefinitions are common in quantum field theory. If you haven't seen this before, you can think of this as the field-theory analog of promoting the position and momentum variables to operators in quantum mechanics (*i.e.* $q, p \rightarrow \hat{q}, \hat{p}$ and introducing creation and annihilation operators, $\hat{q} \sim a + a^\dagger, \hat{p} \sim a - a^\dagger$).

This is easily derived from the commutation relation, and satisfies two important symmetries: (1) it is homogeneous (since $\vec{k} = \vec{k}'$), (2) it is isotropic (since the expectation depends only on $|\vec{k}| = k$).⁵⁰ Extending to higher-order, we quickly discover that the inflaton field is **Gaussian**, such that any odd correlators (such as the three-point function) vanish, and any even correlators are fully determined by the power spectrum.

Inserting the mode functions derived above, we can explicitly compute the power spectrum of inflation. Usually, this is expressed in terms of the *dimensionless* spectrum, $\Delta_{\delta\phi}^2(k, \eta)$, which satisfies

$$\Delta_{\delta\phi}^2(k, \eta) \equiv \frac{k^3}{2\pi^2} |\delta\phi(k, \eta)|^2 = \left(\frac{H}{2\pi}\right)^2 [1 + (k\eta)^2]. \quad (175)$$

In the superhorizon limit, $k\eta \rightarrow 0$, leading to the asymptotic power spectrum

$$\Delta_{\delta\phi}^2(k) \approx \left(\frac{H}{2\pi}\right)_{k=aH}^2, \quad (176)$$

where the right-hand-side is evaluated as the mode crosses the horizon and thus freezes (recalling that H must have some dependence on time, else inflation will never end). Notably, this (a) matches the intuition derived at the beginning of this section from the Heisenberg uncertainty principle, and (b) is approximately scale-invariant.

4.8.3 Curvature Perturbations

Inflaton fluctuations induce perturbations in the metric, $\delta g_{\mu\nu}(\vec{x}, t)$, and thus the local expansion properties. To understand this, recall that inflation ends when the slow-roll parameter reaches unity ($\tilde{\epsilon} = 1$), which is set by the Hubble parameter $H^2 \sim V(\phi)$ and thus the value of ϕ . A local fluctuation in ϕ towards the potential minimum will lead to inflation ending earlier in that region of space; the net effect is a pattern of local curvature imprinted by the inflaton fluctuations.

To make this intuition precise, we introduce the **comoving curvature perturbation**,⁵¹

$$\mathcal{R}(\vec{k}, \eta) = \mathcal{H}(\eta) \frac{\delta\phi(\vec{k}, \eta)}{\phi'(\eta)} = \frac{f(\vec{k}, \eta)}{z(\eta)} \quad (177)$$

(working in the spatially flat gauge, see §5), which parametrizes the local curvature, with $\delta g_{ij}(\vec{x}, t) = a^2(t) e^{2\mathcal{R}(\vec{x}, t)} \delta_{ij}$ (in the comoving gauge). This has a number of useful properties including conservation on superhorizon scales (meaning that $\mathcal{R}$ is fixed once $k \gtrsim \mathcal{H}(\eta)$). We can compute the dimensionless power spectrum of $\mathcal{R}$, denoted by

$$\langle \mathcal{R}(\vec{k}) \mathcal{R}^*(\vec{k}') \rangle \equiv (2\pi)^3 \delta_{\text{D}}(\vec{k} - \vec{k}') \frac{2\pi^2}{k^3} \Delta_{\mathcal{R}}^2(k), \quad (178)$$

which depends only on $|\vec{k}|$ due to homogeneity and isotropy. Using the results from the previous section, we can compute the curvature power spectrum explicitly:

$$\Delta_{\mathcal{R}}^2(k) \equiv \frac{k^3}{2\pi^2} P_{\mathcal{R}}(k) = \frac{1}{8\pi^2 \tilde{\epsilon}} \frac{H^2}{M_{\text{Pl}}^2} \Big|_{k=aH}. \quad (179)$$

⁵⁰To understand this, it is useful to consider the real-space correlator $\langle 0 | \delta\hat{\phi}(\vec{x}) \delta\hat{\phi}(\vec{y}) | 0 \rangle$. By homogeneity, this can depend only on the separation between points $\vec{x} - \vec{y}$ (and not their absolute position); by isotropy, the correlator can depend only the distance $|\vec{x} - \vec{y}|$. Rewriting the functions in Fourier-space, these two symmetries imply that the power spectrum can depend only on $k = |\vec{k}| = |\vec{k}'|$.

⁵¹In the comoving gauge, the Friedmann equation can be written as $H^2 = \frac{8\pi G}{3} \rho + \frac{2}{3} \nabla^2 \mathcal{R}$, which shows that $\mathcal{R}$ encodes the local radius of curvature.

which holds for all times satisfying $k \gtrsim \mathcal{H}(\eta)$. In the slow-roll regime, both H and $\tilde{\epsilon}$ are (approximately) constant (since $\tilde{\epsilon} \approx \tilde{\eta} \approx 0$), thus the power spectrum is approximately *scale-invariant*.

Motivated by the above relation, one can define an empirical model for the curvature power spectrum

$$\Delta_{\mathcal{R}}^2(k) = A_s \left(\frac{k}{k_*} \right)^{n_s-1}, \quad (180)$$

where A_s, n_s parametrize the power spectrum amplitude and spectral index, and we evaluate at the pivot scale $k_* = 0.05 \text{ Mpc}^{-1}$. Within slow-roll inflation,

$$A_s = \frac{1}{8\pi^2 \tilde{\epsilon}_*} \frac{H_*^2}{M_{\text{Pl}}^2}, \quad n_s = 1 - 2\tilde{\epsilon}_* - \tilde{\eta}_* \quad (181)$$

where the asterisk indicates quantities evaluated at k_* . For a pure de Sitter Universe, $n_s = 1$ and the power spectrum is exactly scale-invariant ($\Delta_{\mathcal{R}}^2 \sim k^0$); in all realistic models of inflation, n_s is *slightly* below 1 (known as a ‘red tilt’). This is a generic prediction of inflation and holds under minimal assumptions. Observationally, the amplitude and spectral index are constrained by cosmic microwave background observations, finding

$$A_s \approx 2.1 \times 10^{-9}, \quad n_s \approx 0.96, \quad (182)$$

in excellent agreement with the slow-roll expectations. This, coupled with the observed adiabaticity of CMB fluctuations (discussed below), represents the strongest current evidence for the inflationary model.

4.8.4 Gravitational Waves, Non-Gaussianity, and Beyond

In addition to the scalar curvature perturbations discussed above, inflaton fluctuations induce tensor fluctuations of the metric, h_{ij} , *i.e.* primordial gravitational waves. These can be computed analogously to curvature perturbations: the net result is a Gaussian distribution of fluctuations with dimensionless power spectrum

$$\Delta_h^2(k) = \frac{2\pi^2}{k^3} |h_{ij}(\vec{k}) h^{ij}(\vec{k})| = \frac{2}{\pi^2} \left(\frac{H(t)}{M_{\text{Pl}}} \right)^2_{k=aH}, \quad (183)$$

evaluating the right-hand-side at horizon exit. As for scalars, this can be parametrized in terms of an amplitude, A_t , and a spectral index, n_t :⁵²

$$\Delta_h^2(k) = A_t \left(\frac{k}{k_*} \right)^{n_t}, \quad (184)$$

where $A_t \propto H^2 \sim V$ directly measures the energy scale of inflation. Within slow-roll inflation, $n_t \approx -2\tilde{\epsilon}$, implying that the tensor spectrum should be close to scale-invariant. Combining A_t and A_s , we can define the **tensor-to-scalar ratio**

$$r \equiv \frac{A_t}{A_s} = 16 \epsilon_*, \quad (185)$$

⁵²Due to an historical quirk of notation, scale-invariance requires $n_t = 0$ in the tensor sector, whilst $n_s = 1$ for scalars.

which depends only on the first slow-roll parameter. Measuring r is a key target of current and future cosmic microwave background observatories (through the measurement of B -mode polarization). This is often heralded as a *smoking gun* of inflation, but has so far remained elusive, with modern experiments constraining $r \lesssim 0.04$.

In the above discussion, we have focussed on the simplest inflationary model: a single field slowly-rolling down its potential. Whilst useful, and in full agreement with observational constraints, this scenario is maximally boring: it sources only an almost-scale-invariant scalar and tensor power spectrum. The inflationary literature abounds with examples of models beyond single-field slow-roll, including additional fields, interactions, non-flat potentials, and beyond. A key feature of many such scenarios is that they induce novel phenomenology, such as:

- **Primordial Features:** If the potential, $V(\phi)$, contains bumps, steps, or wiggles, we can induce transient departures from the slow-roll regime, *i.e.* brief regions with large $\tilde{\eta}$. These manifest as additional scale-dependence in the curvature power spectrum, such as oscillations in $\Delta_{\mathcal{R}}^2(k)$ or enhancements in power on very small scales (sourcing primordial black holes if the effect is sufficiently large).
- **Non-Gaussianity:** In the single-field slow-roll scenario, the curvature perturbations inherits a Gaussian distribution from the vacuum fluctuations of ϕ . If ϕ evolves non-linearly during inflation (due to interactions with itself or with other fields), the distribution function becomes non-Gaussian, and higher-order statistics are generated, such as the bispectrum $\propto \langle \mathcal{R}^3 \rangle$ or trispectrum $\propto \langle \mathcal{R}^4 \rangle$.⁵³ A detection of primordial non-Gaussianity would provide evidence for non-linear effects in inflation, with the particular amplitude and shape encoding the type of processes at work.
- **Isocurvature:** Additional fields present during inflation can generate entropy (isocurvature) perturbations, sourcing fluctuations in the relative number densities of different species (e.g., dark matter, baryons, and radiation) at the end of inflation. These can be correlated or uncorrelated with the adiabatic curvature fluctuation, and lead to distinct modifications to the cosmic microwave background anisotropies.

Each of these phenomena can be probed using observational data from the cosmic microwave background and galaxy surveys. Despite extensive study, there have been no compelling detections: the mysteries of inflation remain a mystery!

5 Relativistic Perturbation Theory

Inflation imprints a nearly scale-invariant spectrum of curvature fluctuations onto the early Universe. What happens to these perturbations after inflation? In particular, how do they grow and interact with matter and radiation? Schematically, metric perturbations (including curvature) mediate gravitational interactions between photons, neutrinos, baryons, dark matter, and dark energy, which in turn modify the evolution and local expansion. To describe this system, we must treat the perturbations in all species entering the Einstein equations consistently: here, we will attack this problem using the tools of relativistic perturbation theory.

In these notes, we will restrict to **linear perturbation theory**, restricting to first-order in perturbations. This approximation is accurate on scales $\gtrsim 10$ Mpc, which covers essentially all of the observational window relevant for the CMB, and much of the regime relevant for large-scale

⁵³Even in the simplest slow-roll models, the inherent non-linearity of Einsteinian gravity imparts a small level of non-Gaussianity to the distribution. This is slow-roll suppressed, however, and likely unobservably small.

structure observations.⁵⁴ Our eventual aim is to describe the growth of fluctuations in dark matter (which source the distribution of galaxies) and in baryons and photons around recombination (which set the statistics of the CMB). We will adopt a five-step process: (1) consider the fluctuations in the metric, $g_{\mu\nu}$; (2) consider fluctuations in matter, via the stress-energy tensor, $T_{\mu\nu}$; (3) use the Einstein equations to relate $g_{\mu\nu}$ and $T_{\mu\nu}$; (4) add the initial conditions from inflation; (5) solve for the evolution of dark matter, photons, and beyond.

5.1 Introduction to Perturbation Theory

We begin by splitting every quantity into a background piece (which we solved for in §2) and a small perturbation, e.g.,

$$\begin{aligned} g_{\mu\nu}(\vec{x}, \eta) &= \bar{g}_{\mu\nu}(\eta) + \delta g_{\mu\nu}(\vec{x}, \eta), \\ \rho(\vec{x}, \eta) &= \bar{\rho}(\eta) + \delta\rho(\vec{x}, \eta), \\ P(\vec{x}, \eta) &= \bar{P}(\eta) + \delta P(\vec{x}, \eta), \\ T_{\mu\nu}(\vec{x}, \eta) &= \bar{T}_{\mu\nu}(\eta) + \delta T_{\mu\nu}(\vec{x}, \eta); \end{aligned} \tag{186}$$

this is identical to our treatment of the inflaton field in the last section. Here, and throughout this section, we'll work in conformal time (with $dt = a d\eta$), which simplifies the equations somewhat.

The background quantities, hereafter denoted with bars, are functions only of time, and are set by the continuity and Friedmann equations. For later reference, let's collect the key equations satisfied by the background quantities (written in conformal time):

$$\begin{aligned} \bar{g}_{\mu\nu} dx^\mu dx^\nu &= a^2(\eta) [-d\eta^2 + d\vec{x}^2] && \text{(FLRW metric),} \\ \bar{T}^\mu{}_\nu &= (\bar{\rho} + \bar{P}) \bar{u}^\mu \bar{u}_\nu + \bar{P} \delta^\mu{}_\nu && \text{(stress-energy),} \\ \mathcal{H}^2 &= \frac{8\pi G}{3} \bar{\rho} a^2 && \text{(Friedmann I),} \\ \mathcal{H}' &= -\frac{4\pi G}{3} (\bar{\rho} + 3\bar{P}) a^2 && \text{(Friedmann II),} \\ \bar{\rho}' &= -3\mathcal{H}(\bar{\rho} + \bar{P}) && \text{(continuity),} \end{aligned} \tag{187}$$

where $\mathcal{H} \equiv a'/a$ is the conformal Hubble parameter and u^μ is the four-velocity of a comoving observer. For simplicity, we have dropped spatial curvature, which is a fair approximation given the tight empirical bounds.

The above equations were obtained by solving the Einstein equations ($G_{\mu\nu} = 8\pi G \bar{T}_{\mu\nu}$) at the background level, together with the conservation law $\nabla_\mu T^{\mu\nu} = 0$; to obtain analogous equations for the perturbations, we must repeat this calculation, working at first-order in δ , *i.e.* we must solve

$$\delta G_{\mu\nu} = 8\pi G \delta T_{\mu\nu}, \quad \nabla^\mu T_{\mu\nu} - \bar{\nabla}^\mu \bar{T}_{\mu\nu} = 0 \tag{188}$$

(noting that ∇ depends on the metric). This is conceptually straightforward, though algebraically somewhat painful.

5.2 Metric Perturbations

In §2, we derived the FLRW metric by assuming that the Universe was homogeneous and isotropic. At first-order in perturbations, both of these assumptions break down and the metric gains explicit

⁵⁴In practice, many large-scale structure observables depend on *quasi-linear* scales, which can be accurately described by including higher-order perturbations in a self-consistent manner (and restricting to the Newtonian limit). In these notes, we will not discuss such complications: a detailed treatment can be found in the Effective Field Theory of Large Scale Structure lecture notes available on my website.

dependence on the spatial coordinates $\vec{x}$.⁵⁵ The most general perturbation around the FLRW metric can be written as

$$ds^2 = a^2(\eta) \left[-(1 + 2A) d\eta^2 + 2B_i dx^i d\eta + (\delta_{ij} + 2E_{ij}) dx^i dx^j \right], \quad (189)$$

where A , B_i and E_{ij} are arbitrary functions of $(\vec{x}, \eta)$, which describe (small) perturbations to conformal time, the time-space slicing and spatial coordinates respectively. This form reduces to the FLRW metric in the limit $A, B_i, E_{ij} \rightarrow 0$. Notably, A contributes a single scalar degree of freedom, B_i contributes three (one per spatial direction), and E_{ij} contributes six (since it is a symmetric 3×3 matrix), giving a total of $1 + 3 + 6 = 10$ independent metric perturbations, as expected.

5.2.1 SVT Decomposition

To simplify this expression, we can write the metric perturbations in terms of their scalar, vector, and tensor components (hereafter ‘SVT’). To motivate this, consider the Helmholtz decomposition of a vector field into a gradient and a divergence-free piece:

$$\vec{v} = \vec{\nabla}\phi + \vec{w}, \quad \Rightarrow \quad v_i = \partial_i\phi + w_i. \quad (190)$$

Here, ϕ carries the scalar component, whilst $\vec{w}$ represents the *bona fide* vectorial piece (subject to the constraint $\partial_i w^i = 0$). We can apply a similar decomposition to the B_i and E_{ij} metric perturbations:

$$\begin{aligned} B_i &= \partial_i B + \hat{B}_i, \\ E_{ij} &= C \delta_{ij} + \left(\partial_i \partial_j - \frac{1}{3} \delta_{ij} \nabla^2 \right) E + \frac{1}{2} \left(\partial_i \hat{E}_j + \partial_j \hat{E}_i \right) + \hat{E}_{ij}, \end{aligned} \quad (191)$$

where $C \equiv \frac{1}{3} E^i_i$ encodes the trace of E_{ij} , $\hat{B}_i, \hat{E}_i$ are divergence-free vectors (satisfying $\partial_i \hat{B}^i = \partial_i \hat{E}^i = 0$), and $\hat{E}_{ij}$ is a transverse, traceless tensor (with $\partial_i \hat{E}^{ij} = 0$, $\hat{E}^i_i = 0$). The full metric is described by:

- 4 **scalars**: A, B, C, E ;
- 2 **vectors**: $\hat{B}_i, \hat{E}_i$, each with two independent components;
- 1 **tensor**: $\hat{E}_{ij}$, with two independent components (describing the two gravitational wave polarizations);

giving a total of ten degrees of freedom, as before. In practice, it’s useful to write each perturbation in Fourier-space, defining, for example, $A(\vec{x}, \eta) = \int_{\vec{k}} A(\vec{k}, \eta) e^{i\vec{k} \cdot \vec{x}}$, with the SVT decomposition

$$\begin{aligned} B_i(\vec{k}, \eta) &= i\hat{k}_i B + \hat{B}_i, \\ E_{ij}(\vec{k}, \eta) &= C \delta_{ij} - \left(\hat{k}_i \hat{k}_j - \frac{1}{3} \delta_{ij} \right) E + \frac{i}{2} (\hat{k}_i \hat{E}_j + \hat{k}_j \hat{E}_i) + \hat{E}_{ij}, \end{aligned} \quad (192)$$

for $\hat{k}^i \equiv k^i/|\vec{k}|$, absorbing factors of $|\vec{k}|$ into the metric perturbations. At linear order, the equations of motion for each metric perturbation, $A(\vec{k}, \eta)$, reduce to ordinary differential equations, with each Fourier-mode evolving independently.

The utility of the SVT decomposition is that, at linear order, scalar, vector, and tensor modes decouple: the equations of motion for scalar perturbations depend only on scalar perturbations (in both matter and the metric) *et cetera*. Moreover, the sectors differ in their relevance:

⁵⁵Homogeneity and isotropy remain valid in a spatially-averaged sense, e.g., the statistical correlations between pairs of points can depend only on the distance between the points, not on the absolute position or the orientation.

- **Scalars** are important: they describe curvature perturbations that seed large-scale structure.
- **Tensors** are important but small: these describe as-yet-undetected gravitational waves.
- **Vectors** are unimportant: these decay rapidly during both matter- and radiation-domination, such that any primordial vector perturbations will be unobservable today..

In the remainder of this section, we will focus primarily on scalar perturbations: this reduces the number of degrees of freedom to just four.

5.2.2 Gauge Freedom

The above treatment contains an important subtlety: the metric perturbations, A, B_i, E_{ij} , depend explicitly on our choice of coordinate system, or equivalently, our choice of *gauge*. By changing coordinates, we can reshuffle the perturbations (for example mixing the temporal and spatial sectors) and can even create perturbations from a seemingly homogeneous system. To see this more clearly, let's consider two examples:

(1) Creating a perturbation from nothing. Let's consider the FLRW line-element, $ds^2 = a^2(\eta) [-d\eta^2 + \delta_{ij} dx^i dx^j]$ in (η, x^i) coordinates. Under the spatial coordinate transformation $x^i \rightarrow \tilde{x}^i \equiv x^i + \xi^i(\vec{x}, \eta)$, for some small distortion vector ξ^i , the line-element transforms as

$$ds^2 \rightarrow a^2(\eta) [-d\eta^2 - 2(\partial_\eta \xi_i) d\tilde{x}^i d\eta + d\tilde{x}^i d\tilde{x}^j (\delta_{ij} - \partial_i \xi_j - \partial_j \xi_i)], \quad (193)$$

since $dx^i = d\tilde{x}^i - \partial_k \xi^i d\tilde{x}^k - \partial_\eta \xi^i d\eta$. This corresponds to the perturbed metric with $B_i = -\xi'_i$ and $\hat{E}_{ij} = -\partial_{(i} \xi_{j)}$. The two line-elements represent the same Universe – however, one contains perturbations whilst the other does not! Notably, these perturbations are not physical: they are pure **gauge modes**, generated by the coordinate change.

(2) Trading physical perturbations. Take a uniform density field $\rho(\vec{x}, \eta) = \bar{\rho}(\eta)$ in (η, x^i) coordinates. Under the coordinate shift $\eta \rightarrow \tilde{\eta} \equiv \eta + \xi^0(\vec{x}, \eta)$, the density becomes $\bar{\rho}(\eta[\tilde{\eta}, \vec{x}]) = \bar{\rho}(\tilde{\eta}) - \bar{\rho}'(\tilde{\eta}) \xi^0$, which includes an apparent density perturbation $\delta\rho = -\bar{\rho}' \xi^0$. Reversing the logic, we can ‘hide’ a physical density perturbation $\delta\rho$ via a judicious choice of coordinates (practically absorbing it into the metric).

We can understand the gauge-dependence of relativistic quantities by applying a general coordinate shift, $x^\mu \rightarrow \tilde{x}^\mu = x^\mu + \xi^\mu(x)$. By definition, the value of a scalar quantity ϕ (including both background and perturbations) cannot depend on the gauge, *i.e.* $\phi(x) = \tilde{\phi}(\tilde{x})$. Expanding to first order, we find

$$\bar{\phi}(\eta) + \delta\phi(x) = \bar{\phi}(\tilde{\eta}) + \delta\tilde{\phi}(\tilde{x}) \Rightarrow \delta\tilde{\phi}(x) = \delta\phi(x) - \xi^0(x) \bar{\phi}'(x). \quad (194)$$

noting that there are no gauge ambiguities for background fields (thus $\bar{\phi} = \tilde{\bar{\phi}}$, if evaluated at the same location). The right-hand-side tells us how scalars transform under an infinitesimal coordinate shift. Since the line-element is coordinate invariant, *i.e.* $ds^2 = g_{\mu\nu}(x) dx^\mu dx^\nu = \tilde{g}_{\alpha\beta}(\tilde{x}) d\tilde{x}^\alpha d\tilde{x}^\beta$, the metric must transform according to⁵⁶

$$g_{\mu\nu}(x) = \tilde{g}_{\alpha\beta}(\tilde{x}) \frac{\partial \tilde{x}^\alpha}{\partial x^\mu} \frac{\partial \tilde{x}^\beta}{\partial x^\nu}. \quad (195)$$

⁵⁶As an example, consider the time-time piece of the metric. From the transformation rule, $a^2(\eta) [1 + 2A(x)] = (1 + T')^2 a^2(\eta + T) (1 + 2\tilde{A}(\tilde{x}))$, with $a(\eta + T) \approx a(\eta) [1 + T\mathcal{H}]$ at first order.

By writing both $g_{\mu\nu}$ and $\tilde{g}_{\mu\nu}$ in terms of the metric perturbations A, B_i, E_{ij} and decomposing the shift vector as $\xi^\mu = (T, \partial^i L + \hat{L}^i)$, the scalar, vector, and tensor perturbations can be shown to transform as

$$\begin{aligned} A &\rightarrow A - T' - \mathcal{H}T, & C &\rightarrow C - \mathcal{H}T - \frac{1}{3}\nabla^2 L, \\ B &\rightarrow B + T - L', & E &\rightarrow E - L, \\ \hat{B}_i &\rightarrow \hat{B}_i - \hat{L}'_i, & \hat{E}_i &\rightarrow \hat{E}_i - \hat{L}_i, & \hat{E}_{ij} &\rightarrow \hat{E}_{ij}. \end{aligned} \quad (196)$$

Note that the tensor perturbation $\hat{E}_{ij}$ is manifestly gauge-invariant, while the scalar and vector modes mix non-trivially with the coordinate shifts (though there is not mixing *between* the two sectors).

5.2.3 Choosing the Gauge

Keeping track of the coordinate dependence of metric perturbations is tedious in practice, and it is often cleaner to sidestep the gauge dependence altogether, using one of two commonly applied strategies.

First, we can construct combinations of the metric perturbations that are invariant under (small) coordinate transformations. One such set of variables are the **Bardeen variables**:

$$\begin{aligned} \Psi &= A + \mathcal{H}(B - E') + (B - E')', \\ \Phi &= -C + \frac{1}{3}\nabla^2 E - \mathcal{H}(B - E'), \\ \hat{\Phi}_i &= \hat{B}_i - \hat{E}'_i, \\ \hat{E}_{ij} &= \hat{E}_{ij}. \end{aligned} \quad (197)$$

The scalar Bardeen potentials, Φ, Ψ , the vector perturbation $\hat{\Psi}$, and the tensor perturbation $\hat{E}_{ij}$, represent genuine physical degrees of freedom that cannot be changed by a coordinate transformation.⁵⁷ Notably, there are only two scalar degrees of freedom, with the other two simply describing gauge modes.

Secondly, we can fix the gauge by selecting an appropriate coordinate system (*i.e.* choice of L, T), and keeping track of all relevant perturbations (akin to performing electromagnetic calculations in, say, the Coulomb gauge). At the end of the calculation, we can express the final results in terms of Bardeen variables, ensuring that the outcomes do not depend on the coordinate system we selected. Three common scalar gauge choices are:

- (a) **Newtonian gauge**: This sets $B = E = 0$, such that the Bardeen potentials are given by $\Psi = A$ and $\Phi = -C$. The line element becomes

$$ds^2 = a^2(\eta) [-(1 + 2\Psi) d\eta^2 + (1 - 2\Phi) d\vec{x}^2], \quad (198)$$

(dropping the vector and tensor modes). This gauge is conceptually simple: since there are no $d\eta dx^i$ contributions, constant-time hypersurfaces are orthogonal to observers at rest ($B = 0$), and the geometry on spatial hypersurfaces is isotropic ($E = 0$). In this gauge, the equations of motion also match those of Newtonian gravity, which makes the calculations particularly simple.

⁵⁷This is straightforward to verify: substitute the above transformation rules into the Bardeen potential definitions.

(b) **Spatially flat gauge:** This sets $C = E = 0$, giving the line-element

$$ds^2 = a^2(\eta) [-(1 + 2A) d\eta^2 + 2B_i dx^i d\eta + d\vec{x}^2]. \quad (199)$$

In this gauge, there are no fluctuations in the spatial metric, which is a particularly convenient choice for describing inflationary fluctuations (indeed, the Mukhanov-Sasaki equation discussed in §4.8 is derived in this gauge).

(c) **Synchronous gauge:** This sets $A = B = 0$, giving the line-element

$$ds^2 = a^2(\eta) [-d\eta^2 + (\delta_{ij} + 2E_{ij}) dx^i dx^j]. \quad (200)$$

Here, there are no fluctuations in the time component of the metric, *i.e.* the time coordinate coincides with the proper time of comoving observers. This is the gauge used by most Boltzmann codes including CLASS.

5.3 Matter Perturbations

Having defined the metric at first order in perturbations, we now turn to the stress-energy tensor, $T^\mu{}_\nu$, which encodes the particle content of the Universe. As discussed in §2, the stress-energy for a general fluid can be written as

$$T^\mu{}_\nu = (\rho + P) u^\mu u_\nu + P \delta^\mu{}_\nu + \Pi^\mu{}_\nu, \quad (201)$$

where ρ is the energy density, P is the isotropic pressure, u^μ is the fluid velocity, and $\Pi^\mu{}_\nu$ is the (traceless) anisotropic stress, all of which depend on x^μ . As before, we can split each quantity into background and perturbed contributions, which define the zeroth and first-order components of $T^\mu{}_\nu$ via

$$\bar{T}^0{}_0(\eta) = -\bar{\rho}(\eta), \quad \bar{T}^i{}_0(\eta) = 0, \quad \bar{T}^i{}_j(\eta) = P(\eta) \delta^i{}_j \quad (202)$$

and

$$\delta T^0{}_0(x) = -\delta\rho(x), \quad \delta T^i{}_0(x) = [\bar{\rho}(\eta) + \bar{P}(\eta)] v^i(x), \quad \delta T^i{}_j(x) = \delta P(x) \delta^i{}_j + \Pi^i{}_j(x), \quad (203)$$

introducing the bulk velocity, v^i .⁵⁸ In practice, the above equations involve the total stress-energy perturbation, summed over all species of matter and radiation present in the Universe. Explicitly:

$$\delta\rho = \sum_a \delta\rho_{(a)}, \quad q^i \equiv (\bar{\rho} + \bar{P})v^i = \sum_a q^i_{(a)}, \quad \delta P = \sum_a \delta P_{(a)}, \quad \Pi^{ij} = \sum_a \Pi^{ij}_{(a)}, \quad (204)$$

where the sum runs over baryons, photons, dark matter, dark energy, *et cetera*.

The stress-energy tensor is characterized by four perturbations: the density $\delta\rho$, the pressure δP , the fluid velocity v^i , and the anisotropic stress $\Pi^i{}_j$, totalling ten degrees-of-freedom. The latter two can be decomposed into scalar, vector, and tensor components as before:

$$\begin{aligned} v^i &= \partial^i v + \hat{v}^i, \\ \Pi^{ij} &= (\partial^i \partial^j - \tfrac{1}{3} \delta^{ij} \nabla^2) \Pi + \tfrac{1}{2} (\partial^i \hat{\Pi}^j + \partial^j \hat{\Pi}^i) + \hat{\Pi}^{ij}, \end{aligned} \quad (205)$$

where $\hat{v}^i, \hat{\Pi}^i$ are divergence-free (*i.e.* $\partial_i \hat{v}^i = \partial_i \hat{\Pi}^i = 0$) and $\hat{\Pi}^{ij}$ is transverse-traceless (*i.e.* $\partial_i \Pi^{ij} = \Pi^i{}_i = 0$). Once again, the scalar, vector, and tensor components evolve independently, implying that scalar physics can be described by just four degrees-of-freedom ($\delta\rho, \delta P, v, \Pi$).

⁵⁸Here, we work in the ‘upstairs-downstairs’ convention, using $\delta T^\mu{}_\nu$, rather than $\delta T_{\mu\nu}$. This isolates the matter perturbations from the metric perturbations (which enter through $g_{\mu\nu}$, $u^\mu = a^{-1}(1 - A, v^i)$ and $u_\mu = a^{-1}(-(1 + A), v_i + B_i)$).

5.3.1 Gauge Transformations

As with the metric, the matter perturbations depend on our choice of coordinate system. Under a coordinate shift $x^\mu \rightarrow \tilde{x}^\mu = x^\mu + \xi^\mu$, the stress-energy tensor transforms according to the usual tensor transformation rule:

$$T^\mu{}_\nu(x) = \frac{\partial x^\mu}{\partial \tilde{x}^\alpha} \frac{\partial \tilde{x}^\beta}{\partial x^\nu} \tilde{T}^\alpha{}_\beta(\tilde{x}) \quad (206)$$

Setting $\xi^\mu = (T, L^i)$ as before (for small T, L^i), we can thus relate the density, pressure, velocity, and anisotropic stress perturbations in the original and shifted frames, finding

$$\begin{aligned} \delta\rho &\rightarrow \delta\rho - \bar{\rho}' T, & \delta P &\rightarrow \delta P - \bar{P}' T, \\ q_i &\rightarrow q_i + (\bar{\rho} + \bar{P}) L'_i, & v_i &\rightarrow v_i + L'_i, & \Pi_{ij} &\rightarrow \Pi_{ij}. \end{aligned} \quad (207)$$

The anisotropic stress is manifestly gauge-invariant, whilst the density, pressure, and velocity perturbations all mix non-trivially with the coordinate shift T, L^i (though the scalar and vector components remain decoupled).

Analogously to the metric perturbations, we can construct a set of **gauge-invariant** variables which are unaffected by the above coordinate transformations. There are three commonly used gauge-invariant scalars:

- (a) The **comoving density contrast**:

$$\Delta(x) \equiv \frac{1}{\bar{\rho}} [\delta\rho + \bar{\rho}' (v + B)]. \quad (208)$$

If $v + B = 0$ (known as the comoving gauge), this corresponds to the fractional density perturbation $\delta\rho/\bar{\rho}$. As we will see in §5.5, Δ satisfies the usual Poisson equation, $\nabla^2\Phi = 4\pi G a^2 \bar{\rho} \Delta$ on all scales (not just in the subhorizon regime, where relativistic corrections are small).

- (b) The **uniform-density curvature perturbation**:

$$\zeta(x) \equiv -C + \frac{1}{3} \nabla^2 E + \mathcal{H} \frac{\delta\rho}{\bar{\rho}'}. \quad (209)$$

- (c) The **comoving curvature perturbation**:

$$\mathcal{R}(x) \equiv -C + \frac{1}{3} \nabla^2 E - \mathcal{H} (v + B). \quad (210)$$

The latter two variables, ζ and $\mathcal{R}$, characterize the intrinsic curvature of spatial hypersurfaces, via the spatial Ricci scalar ${}^{(3)}R = (4/a^2) \nabla^2 (-C + \frac{1}{3} \nabla^2 E)$ (for spatial metric perturbations C and E_{ij}). On superhorizon scales ($k \ll \mathcal{H}$), we will later show that the two variables are approximately conserved and asymptotically equal, making them natural variables for tracking primordial perturbations across multiple cosmic epochs.

Relatedly, we can define two new gauges, complementing those listed in §5:

- (a) **Uniform-density gauge**: This sets $\delta\rho = 0$, which corresponds to a particular choice of the time and space coordinates. In this gauge, the (scalar) spatial metric can be simply expressed as $\delta g_{ij} = a^2(1 - 2\zeta)\delta_{ij}$, involving the uniform-density curvature parameter ζ .
- (b) **Comoving gauge**: This fixes $v + B = 0$, such that the fluid with respect to the observer. This leads to the (scalar) spatial metric $\delta g_{ij} = a^2(1 - 2\mathcal{R})\delta_{ij}$, which involves the comoving curvature perturbation $\mathcal{R}$. This gauge is commonly used in inflationary studies, since it cleanly separates the temporal and spatial perturbations.

5.4 Conservation Equations

Equipped with the metric and matter perturbations outlined in the previous sections, we can now derive the dynamical equations governing the evolution of the matter sector, invoking conservation of stress-energy: $\nabla_\mu T^\mu{}_\nu = 0$. As noted above, the full stress-energy tensor is a sum of the individual stress-energy tensors for each components of the Universe; in the absence of direct interactions between the species (except through gravity), this relation holds for each species independently.

Up to first order in perturbations, the conservation law can be written

$$\nabla_\mu T^\mu{}_\nu = \partial_\mu T^\mu{}_\nu + \Gamma^\mu_{\mu\alpha} T^\alpha{}_\nu - \Gamma^\alpha_{\mu\nu} T^\mu{}_\alpha = 0. \quad (211)$$

Due to the appearance of Christoffel symbols, this depends on both matter and metric perturbations. To obtain the perturbed conservation equations, we adopt a three-step procedure: (1) define the Christoffel symbols; (2) substitute $\Gamma^\mu_{\nu\rho}$ and $T^\mu{}_\nu$ into the above expression and collect the first-order terms; (3) simplify the $\nu = 0$ and $\nu = i$ expressions.⁵⁹ In order to streamline the algebra, we work in the Newtonian gauge (defined by $B = E = 0$), which has the line-element

$$ds^2 = a^2(\eta) \left[-(1 + 2\Psi) d\eta^2 + (1 - 2\Phi) \delta_{ij} dx^i dx^j \right], \quad (212)$$

where Ψ, Φ are the Bardeen potentials. Up to first order, the associated Christoffel symbols are given by

$$\begin{aligned} \Gamma_{00}^0 &= \mathcal{H} + \Psi', & \Gamma_{i0}^0 &= \partial_i \Psi, & \Gamma_{00}^i &= \partial^i \Psi, \\ \Gamma_{ij}^0 &= \mathcal{H} \delta_{ij} - [\Phi' + 2\mathcal{H}(\Phi + \Psi)] \delta_{ij}, \\ \Gamma_{j0}^i &= \mathcal{H} \delta^i_j - \Phi' \delta^i_j, \\ \Gamma_{jk}^i &= -[\delta_j^i \partial_k + \delta_k^i \partial_j] \Phi + \delta_{jk} \partial^i \Phi. \end{aligned} \quad (213)$$

using the definition $\Gamma_{\nu\lambda}^\mu \equiv \frac{1}{2} g^{\mu\sigma} (\partial_\nu g_{\sigma\lambda} + \partial_\lambda g_{\sigma\nu} - \partial_\sigma g_{\nu\lambda})$ and indicating the background contributions in **red**.

5.4.1 Continuity Equation

When the dust settles, we find a fairly straightforward form for the $\nu = 0$ conservation equation:

$$\nabla_\mu T^\mu{}_0 = 0 \quad \Rightarrow \quad \boxed{\delta\rho' = -3\mathcal{H}(\delta\rho + \delta P) - \partial_i q^i + 3(\bar{\rho} + \bar{P})\Phi'}. \quad (214)$$

This is the **relativistic continuity equation**, and describes the evolution of density fluctuations in the perturbed Universe. Each describes a different physical effect:

- $-3\mathcal{H}(\delta\rho + \delta P)$ encodes the dilution induced by the Universe's expansion, which appears also in the background continuity equation $\bar{\rho}' = -3\mathcal{H}(\bar{\rho} + \bar{P})$.
- $-\partial_i q^i$ encodes the flow of energy into a comoving volume element due to fluid motion (with momentum density $q^i = (\bar{\rho} + \bar{P})v^i$).
- $3(\bar{\rho} + \bar{P})\Phi'$ encodes the relativistic correction to the local expansion rate. From the metric, Φ acts as a local rescaling of the scale-factor $a \rightarrow (1 - \Phi)a$, which modifies the dilution of ρ by a factor $3\Phi'(\bar{\rho} + \bar{P})$.⁶⁰

⁵⁹Whilst conceptually straightforward, this is algebraically arduous. To streamline these notes, we present only the final results in this section.

⁶⁰This can be seen explicitly by replacing $a \rightarrow (1 - \Phi)a$ in the background continuity equation and expanding to first-order.

The continuity equation can be rewritten in terms of the density contrast, $\delta \equiv \delta\rho/\bar{\rho}$. Using also the background continuity equation, we find

$$\delta' = -\left(1 + \frac{\bar{P}}{\bar{\rho}}\right)(\nabla \cdot \vec{v} - 3\Phi') - 3\mathcal{H}\left(\frac{\delta P}{\delta\rho} - \frac{\bar{P}}{\bar{\rho}}\right)\delta. \quad (215)$$

For non-relativistic species ($\bar{P} \ll \bar{\rho}$), this reduces to $\delta' \approx -\nabla \cdot \vec{v} + 3\Phi'$; further restricting to subhorizon scales (with $k \gg \mathcal{H}$), we recover the usual Newtonian continuity equation in conformal time, $\delta' \approx -\nabla \cdot \vec{v}$.

5.4.2 Euler Equation

The spatial components ($\nu = i$) of the conservation equation give the **relativistic Euler equation**:

$$\nabla_\mu T^\mu_i = 0 \quad \Rightarrow \quad q'_i = -4\mathcal{H} q_i - (\bar{\rho} + \bar{P}) \partial_i \Psi - \partial_i \delta P - \partial_j \Pi^j_i, \quad (216)$$

which describes the cosmological evolution of momentum. Here, the leading term, $-4\mathcal{H} q_i$, reflects the dilution of momentum density, which scales as a^{-4} for all species. The remaining terms encode gravitational forces (via $\partial_i \Psi$, cf. Newton's second law), pressure gradients ($\partial_i \delta P$), and anisotropic stresses ($\partial_j \Pi^j_i$). Rewriting in terms of velocity, with $q_i \equiv (\bar{\rho} + \bar{P})v_i$, we find

$$v'_i = -\left(\mathcal{H} + \frac{\bar{P}'}{\bar{\rho} + \bar{P}}\right)v_i - \frac{1}{\bar{\rho} + \bar{P}}(\partial_i \delta P + \partial^j \Pi_{ij}) - \partial_i \Psi. \quad (217)$$

For non-relativistic species without anisotropic stress, this simplifies to $v'_i \approx -\mathcal{H}v_i - (1/\bar{\rho}) \partial_i \delta P - \partial_i \Psi$, which matches the expected Newtonian limit, and implies that velocities decay as a^{-1} in the absence of pressure gradients or external forces.

5.4.3 Application to Matter and Radiation

The continuity and Euler equations presented above apply to any gravitationally interacting species and can be used to understand the evolution of baryons, photons, and dark matter. In many cases, we can simplify the equations considerably by assuming that the fluids of interest are **adiabatic**: this holds if the pressure perturbation can be expressed in terms of the density perturbation via $\delta P = c_s^2 \delta\rho$, with sound-speed $c_s^2 = \bar{P}'/\bar{\rho}'$ (which may itself be a function of ρ). This is an excellent approximation for pressureless matter (cold dark matter and baryons) and the tightly-coupled photon-baryon fluid prior to recombination, but fails for free-streaming species such as neutrinos, where δP depends non-trivially on the full distribution function. Below, we discuss the first two cases in more detail.

Matter On cosmological scales, cold dark matter and baryons are usually assumed to be pressureless, with an equation of state $\bar{w}_m = 0$, and vanishing anisotropic stress, $\Pi^\mu_\nu = 0$. In this case, the conservation equations reduce to

$$\begin{aligned} \delta'_m &= -\nabla \cdot \vec{v}_m + 3\Phi', \\ v'_m &= -\mathcal{H} v_m - \nabla \Psi, \end{aligned} \quad (218)$$

which can be combined into the second-order equation:

$$\delta''_m + \mathcal{H} \delta'_m = \nabla^2 \Psi + 3(\Phi'' + \mathcal{H}\Phi'). \quad (219)$$

Here, the left-hand-side describes the friction induced by cosmological expansion, whilst the right-hand-side gives the gravitational source terms. This equation governs the gravitational growth of dark matter perturbations, and we will later use it to describe the clustering properties of dark matter on linear scales.

Radiation For a perfect radiation fluid with no anisotropic stress, $\bar{P}_r = \frac{1}{3}\bar{\rho}_r$, thus by adiabaticity, $\delta P_r = c_s^2 \delta \rho_r$ with $c_s^2 = 1/3$. The continuity and Euler equations become

$$\begin{aligned}\delta'_r &= -\frac{4}{3}\nabla \cdot \vec{v}_r + 4\Phi', \\ v'_r &= -\frac{1}{4}\nabla \delta_r - \nabla \Psi,\end{aligned}\tag{220}$$

where the factor of $1/4$ in the Euler equation arises from $\delta P_r/(\bar{\rho}_r + \bar{P}_r) = [c_s^2 \delta \rho_r]/[\bar{\rho}_r(1 + 1/3)] = \delta_r/4$, and the Hubble-friction term has dropped out exactly (since $\bar{P}'_r/(\bar{\rho}_r + \bar{P}_r) = -\mathcal{H}$, using the background continuity equation). Combining the two relations, we find

$$\boxed{\delta''_r - \frac{1}{3}\nabla^2 \delta_r = \frac{4}{3}\nabla^2 \Psi + 4\Phi''}.\tag{221}$$

This takes the form of a wave equation with sound speed $c_s = 1/\sqrt{3}$, sourced by gravitational forces ($\nabla^2 \Psi$) and relativistic effects (Φ''). In contrast to matter, there is no friction term; instead, the $\nabla^2 \delta_r$ term leads to propagating sound waves, which will later find to source acoustic peaks in the cosmic microwave background, as well as the BAO feature in galaxy distributions.

5.5 Einstein Equations

Using conservation of stress-energy, we have derived equations for the evolution of the cosmological density and velocity perturbations, given the scalar Bardeen potentials Φ and Ψ . To close the system, we need to know how the metric perturbations themselves evolve: this can be achieved using the perturbed Einstein equations, $\delta G_{\mu\nu} = 8\pi G \delta T_{\mu\nu}$, as outlined below.

To derive the relevant equations, we start from the perturbed metric defined above, and successively derive the Christoffel symbols, $\Gamma^\mu_{\nu\rho}$, the Ricci tensor, $R_{\mu\nu}$, the Ricci scalar, $R = g^{\mu\nu} R_{\mu\nu}$, and finally the Einstein tensor $G^\mu{}_\nu$, which is then related to the perturbed stress-energy tensor, $T^\mu{}_\nu$. As for the continuity equation, we will work in the Newtonian gauge, and neglect both vector and tensor metric perturbations.

Starting from the Christoffel symbols listed in §5.4, we find the Ricci tensor up to first-order in Φ, Ψ :

$$\begin{aligned}R_{00} &= -3\mathcal{H}' + \nabla^2 \Psi + 3\Phi'' + 3\mathcal{H}(\Phi' + \Psi'), \\ R_{0i} &= 2\partial_i(\Phi' + \mathcal{H}\Psi), \\ R_{ij} &= \delta_{ij}[(\mathcal{H}' + 2\mathcal{H}^2)] - \delta_{ij}[\Phi'' + \nabla^2 \Phi - 2(\mathcal{H}' + 2\mathcal{H}^2)(\Phi + \Psi) - \mathcal{H}\Psi' - 5\mathcal{H}\Phi'] \\ &\quad + \partial_i \partial_j (\Phi - \Psi),\end{aligned}\tag{222}$$

(marking the background contributions in red), along with the Ricci scalar:

$$a^2 R = 6(\mathcal{H}' + \mathcal{H}^2) - 12(\mathcal{H}' + \mathcal{H}^2)\Psi - 6\mathcal{H}(\Psi' + 3\Phi') + 4\nabla^2 \Phi - 2\nabla^2 \Psi - 6\Phi''.\tag{223}$$

Finally, we can form the Einstein tensor via $G^\mu{}_\nu = R^\mu{}_\nu - \frac{1}{2}\delta^\mu{}_\nu R$, whose components can be written

$$\begin{aligned}G^0{}_0 &= -a^{-2}(1 - 2\Psi)R_{00} - \frac{1}{2}R, \\ G^i{}_0 &= a^{-2}(1 + 2\Phi)R^i{}_j - \frac{1}{2}\delta^i{}_j R, \\ G^0{}_i &= -2a^{-2}\partial_i(\Phi' + \mathcal{H}\Psi).\end{aligned}\tag{224}$$

Equating to the stress-energy tensor at first-order leads to four relations (complementing the two Friedmann equations at zeroth order):

1. $\delta G^0_0 = 8\pi G \delta T^0_0$: We obtain the **relativistic Poisson equation**:

$$\boxed{\nabla^2 \Phi - 3\mathcal{H}(\Phi' + \mathcal{H}\Psi) = 4\pi G a^2 \delta\rho,} \quad (225)$$

where $\delta\rho = \sum_a \delta\rho_{(a)}$ is the total density perturbation, summed over species. In the deep subhorizon regime (with $k \gg \mathcal{H}$), the first term on the left-hand-side dominates,⁶¹ and we recover the familiar Newtonian Poisson equation, $\nabla^2 \Phi = 4\pi G a^2 \delta\rho$. This indicates that relativistic corrections are important only for modes with wavelengths close-to (and above) the horizon scale.

2. $\delta G^0_i = 8\pi G \delta T^0_i$: We obtain the **momentum constraint**:

$$\boxed{\Phi' + \mathcal{H}\Psi = -4\pi G a^2 q,} \quad (226)$$

which relates the evolution of Φ to the total momentum density q (with $q^i = \partial^i q$). Combining this with the Poisson equation derived above, one obtains the simple relation

$$\boxed{\nabla^2 \Phi = 4\pi G a^2 \bar{\rho} \Delta,} \quad (227)$$

noting that $q = (\bar{\rho} + \bar{P})v = -3\mathcal{H}\bar{\rho}'v$, from the background continuity equation. Here, we have introduced the gauge-invariant comoving density contrast Δ (see §5.3), which is equal to $(\delta\rho + \bar{\rho}'v)/\bar{\rho}$ in this gauge. This matches the usual Newtonian Poisson equation: however, it holds on both subhorizon *and* superhorizon scales, with the relativistic corrections absorbed into the redefined density variable.⁶²

3. $\delta G^i_j = 8\pi G \delta T^i_j$ (trace-free). Subtracting the trace component via $\delta G^i_j \rightarrow \delta G^i_j - \frac{1}{3}\delta^i_j \delta G^k_k$, we obtain the **anisotropic stress equation**:

$$\boxed{(\partial_i \partial_j - \frac{1}{3}\delta_{ij} \nabla^2)(\Phi - \Psi) = 8\pi G a^2 \Pi_{ij},} \quad (228)$$

which relates the anisotropic stress to the Bardeen potentials. Rewriting in Fourier-space, with $\Pi_{ij}(\vec{k}, \eta) = -(\bar{\rho} + \bar{P})[\hat{k}_i \hat{k}_j - \frac{1}{3}\delta_{ij}] \Pi(\vec{k}, \eta)$, this becomes $k^2(\Phi - \Psi) = 8\pi G a^2 (\bar{\rho} + \bar{P}) \Pi$, where $\Pi = \Pi_b + \Pi_c + \Pi_\gamma + \Pi_\nu$ is the combined scalar anisotropic stress. On large-scales, dark matter and baryons act as a perfect fluid with $\Pi_{b,c} \rightarrow 0$; moreover, photons are tightly coupled to baryons at early times (with $\Pi_\gamma \rightarrow 0$) and subdominant at late times. Further neglecting the neutrino contribution (which has $\Pi_\nu \neq 0$ at all times), we find $\Pi \approx 0$, which implies that the two metric potentials are equal: $\Phi \approx \Psi$.⁶³

4. $\delta G^i_j = 8\pi G \delta T^i_j$ (trace). This leads to the **pressure equation**:

$$\boxed{\Phi'' + \mathcal{H}(\Psi' + 2\Phi') + (2\mathcal{H}' + \mathcal{H}^2)\Psi + \frac{1}{3}\nabla^2(\Psi - \Phi) = 4\pi G a^2 \delta P.} \quad (229)$$

⁶¹To see this, note that $\nabla^2 \Phi \sim k^2 \Phi$, and $\Phi' \sim \mathcal{H}\Phi$, with $\Phi \sim \Psi$.

⁶²On superhorizon scales, $\Delta \sim (k/\mathcal{H})^2 \Phi \ll \Phi$, thus the gauge-invariant density is heavily suppressed compared with δ . Observationally, we only have access to subhorizon modes, whereupon $(\Delta - \delta)/\Delta \sim \mathcal{H}^2/k^2 \rightarrow 0$, thus we can use δ to define the density contrast in practice.

⁶³See the appendix of Daniel Baumann's book for more details.

Setting $\Psi = \Phi$ (*i.e.* neglecting anisotropic stress) and assuming an adiabatic fluid with $\delta P = c_s^2 \delta \rho$, this can be simplified to

$$\Phi'' + 3[1 + c_s^2]\mathcal{H}\Phi' + (2\mathcal{H}' + \mathcal{H}^2[1 + 3c_s^2] + k^2 c_s^2)\Phi = 0 \quad (230)$$

in Fourier-space, where we have combined with the relativistic Poisson equation to eliminate $\delta \rho$ (225). This is a second order ODE for Φ alone, that can be used to obtain the evolution of Φ in both radiation-dominated ($c_s^2 \rightarrow 1/3$) and matter-dominated ($c_s^2 \rightarrow 0$) regimes.

In summary, the four Einstein equations listed above combined with the conservation equations for each species lead to a closed (and redundant) system of equations: the four scalar variables (δ, v, Φ, Ψ) are uniquely determined once the equation of state w and the sound speed c_s^2 of each species are specified, along with suitable initial conditions.

5.6 Initial Conditions

The final ingredient in our theory of perturbations is the set of initial conditions for matter and metric fluctuations. As discussed in §4.8, these are provided by inflation, which sources a (nearly) scale-invariant spectrum of curvature perturbations, $\mathcal{R}$, which are generated by vacuum fluctuations early in inflation and freeze as they exit the horizon. Below, we relate these to the fluid initial conditions by considering the limiting form of the above equations in the superhorizon regime.

5.6.1 Superhorizon Limits

Setting $k \ll \mathcal{H}$, the conservation equations simplify considerably. For matter and radiation (using 218 & 220), we find

$$\delta'_m \approx 3\Phi', \quad \delta'_r \approx 4\Phi', \quad (231)$$

since the velocity terms become small.⁶⁴ Integrate these equations directly, we find the solutions for photons, neutrinos, cold dark matter, and baryons:

$$\begin{aligned} \delta_\gamma(\vec{k}, \eta) &= 4\Phi(\vec{k}, \eta) + C_\gamma(\vec{k}), & \delta_\nu(\vec{k}, \eta) &= 4\Phi(\vec{k}, \eta) + C_\nu(\vec{k}) \\ \delta_b(\vec{k}, \eta) &= 3\Phi(\vec{k}, \eta) + C_b(\vec{k}), & \delta_c(\vec{k}, \eta) &= 3\Phi(\vec{k}, \eta) + C_c(\vec{k}) \end{aligned} \quad (k \ll \mathcal{H}), \quad (232)$$

where $C_a(\vec{k})$ are time-independent integration constants, which cannot be determined from the conservation equations alone. Eliminating Φ gives

$$\begin{aligned} \delta_\nu(\vec{k}, \eta) &= \delta_\gamma(\vec{k}, \eta) + S_\nu(\vec{k}), & \delta_c(\vec{k}, \eta) &= \frac{3}{4}\delta_\gamma(\vec{k}, \eta) + S_c(\vec{k}) \\ \delta_b(\vec{k}, \eta) &= \frac{3}{4}\delta_\gamma(\vec{k}, \eta) + S_b(\vec{k}), \end{aligned} \quad (233)$$

introducing a set of **isocurvature perturbations** $\{S_\nu, S_c, S_b\}$. These encode relative density fluctuations between species that leave the total energy density (and thus the curvature) unchanged. In the simplest inflationary models, all perturbations are derived from the same quantum field: this is known as the **adiabatic** limit, and corresponds to $S_a = 0$ for all species.

This can be understood as follows. As discussed in §4, quantum fluctuations in single-field inflation (also known as ‘single-clock’ inflation) lead to inflation ending at different times in different locations. This induces perturbations in the (initially homogeneous) pressure and density fields,

⁶⁴This follows from the Euler equation: $v' + \mathcal{H}v \approx -\partial_i \Psi$ (dropping pressure and anisotropic stress contributions), thus $v \sim (k/\mathcal{H}) \Psi$, implying that $\nabla \cdot \vec{v} \sim (k^2/\mathcal{H}^2) \Psi'$, which is small for superhorizon modes.

with, for example, $\bar{\rho}_a(\eta) \rightarrow \bar{\rho}_a(\eta + \pi) \approx \bar{\rho}_a(\eta) + \bar{\rho}'_a \pi$, where π is the time perturbation (or more formally, the Goldstone mode of broken time translations). Since all perturbations are sourced by the same shift, we find $\delta\rho_a/\delta P_a = \bar{\rho}'_a/\bar{P}'_a$, which corresponds to the adiabatic scenario discussed previously. Moreover, $\delta_a = (\bar{\rho}'_a/\bar{\rho}_a)\pi = -3\mathcal{H}(1+w_a)\pi$, which implies that $\delta_a/(1+w_a)$ is the same for all species, and thus $\delta_\gamma = \frac{4}{3}\delta_c$, with $S_c = 0$ *et cetera*.. We shall assume this going forward, setting $\delta_\gamma = \delta_\nu = \frac{4}{3}\delta_b = \frac{4}{3}\delta_c$ on superhorizon scales.

Next, we use the Einstein equations to fix the overall amplitude of the density perturbations. On superhorizon scales, the Poisson equation reduces to

$$3\mathcal{H}(\Phi' + \mathcal{H}\Phi) \approx -4\pi G a^2 \delta\rho \quad (k \ll \mathcal{H}), \quad (234)$$

where we have set $\Psi = \Phi$, neglecting anisotropic stress from neutrinos. At early times, the Universe is radiation-dominated with $\delta\rho \approx \bar{\rho}\delta_\gamma$ (since $\delta\rho = \delta\rho_\gamma + \delta\rho_\nu \approx (\bar{\rho}_\gamma + \bar{\rho}_\nu)\delta_\gamma$, using adiabaticity); inserting the Friedmann equation $\frac{3}{2}\mathcal{H}^2 = 4\pi G a^2 \bar{\rho}$, one finds

$$\eta\Phi''(\vec{k}, \eta) + 4\Phi'(\vec{k}, \eta) = 0, \quad (235)$$

noting that $\mathcal{H} = 1/\eta$ in radiation domination. This has two solutions: a constant ‘growing’ mode and a rapidly-decaying mode $\Phi \propto \eta^{-3}$. Discarding the latter, we conclude that Φ is approximately constant on superhorizon scales during radiation domination. This fixes the above integration constants, giving the superhorizon limits

$$\boxed{\delta_\gamma(\vec{k}, \eta) = \delta_\nu(\vec{k}, \eta) = \frac{4}{3}\delta_b(\vec{k}, \eta) = \frac{4}{3}\delta_c(\vec{k}, \eta) = -2\Phi(\vec{k}, \eta) \quad (k \ll \mathcal{H}).} \quad (236)$$

This is an important result: the initial conditions for the post-inflationary fluid are set entirely by the primordial gravitational potential $\Phi(\vec{k}, \eta)$ on superhorizon scales.

5.6.2 The Curvature Perturbation

In the previous section, we demonstrated that Φ is approximately constant on superhorizon scales during radiation domination (up to a rapidly-decaying mode $\propto \eta^{-3}$). The same conclusion holds during matter domination;⁶⁵ however, the two constants differ, since the potential is sensitive to the changing equation of state (from $w_r = \frac{1}{3}$ to $w_m = 0$). A careful calculation shows that

$$\Phi_{\text{MD}}(\vec{k}) = \frac{9}{10} \Phi_{\text{RD}}(\vec{k}) \quad (k \ll \mathcal{H}) \quad (237)$$

for any mode that is superhorizon in both regimes. The variation of Φ across the matter-radiation transition (and similarly any transition where the equation of state varies, e.g., reheating), motivates using a different variable to track the perturbations on superhorizon scales, and thus relate early- and late-Universe physics. The uniform density and comoving curvature perturbations, ζ and $\mathcal{R}$, provide excellent candidates, since they are conserved on superhorizon scales (regardless of the background evolution) and asymptotically equal. Below we demonstrate the key properties of the curvature perturbations and relate them to the initial conditions of Φ and δ_m .

⁶⁵This can be easily verified by setting $\mathcal{H} = 2/\eta$ and $\delta\rho \approx \bar{\rho}\delta_m$, thus $\eta\Phi'' + 6\Phi' = 0$ and $\Phi = a + b\eta^{-5}$, since $\delta'_m = 3\Phi'$ on superhorizon scales.

Conservation of Curvature Working in the Newtonian gauge (with $B = E = 0$, as in §5.3), the two curvature perturbations introduced in §5.3 take the explicit form

$$\begin{aligned}\zeta &= -C + \frac{1}{3}\nabla^2 E + \mathcal{H} \frac{\delta\rho}{\bar{\rho}'} &\Rightarrow \quad \zeta &= \Phi - \frac{\delta\rho}{3(\bar{\rho} + \bar{P})} \\ \mathcal{R} &= -C + \frac{1}{3}\nabla^2 E - \mathcal{H}(v + B) &\Rightarrow \quad \mathcal{R} &= \Phi - \mathcal{H}v,\end{aligned}\tag{238}$$

where v is the scalar contribution to the velocity (with $v^i = \partial^i v + \hat{v}^i$). Differentiating the uniform density curvature perturbation with respect to η and substituting the background and perturbed continuity equations ($\bar{\rho}' = -3\mathcal{H}(\bar{\rho} + \bar{P})$ and $\delta\rho' = -3\mathcal{H}(\delta\rho + \delta P) - \partial_i q^i + 3(\bar{\rho} + \bar{P})\Phi'$), one finds

$$\frac{\bar{\rho} + \bar{P}}{\mathcal{H}}\zeta' = \underbrace{\left(\frac{\delta P}{\delta\rho} - \frac{\bar{P}'}{\bar{\rho}'}\right)\delta\rho}_{\text{non-adiabatic pressure}} + \underbrace{\frac{\partial_i q^i}{3\mathcal{H}}}_{\sim (k/\mathcal{H})^2 q},\tag{239}$$

where the first term vanishes for an adiabatic fluid and the second is suppressed by $(k/\mathcal{H})^2$ on superhorizon scales. We thus conclude that

$$\boxed{\zeta'(\vec{k}, \eta) \approx 0 \quad (k \ll \mathcal{H}, \text{ adiabatic}),}\tag{240}$$

i.e. ζ is conserved on superhorizon scales, regardless of the background expansion history (which has not entered the above derivation). Via a similar argument, we find $\mathcal{R}' \rightarrow 0$ for $k \ll \mathcal{H}$, thus the comoving curvature perturbation is also conserved.

Equivalence of ζ and $\mathcal{R}$ The difference between the two curvature perturbations is given by

$$\zeta - \mathcal{R} = \frac{\mathcal{H}}{\bar{\rho}'} [\delta\rho - \bar{\rho}'(v + B)] \equiv \frac{\mathcal{H}}{\bar{\rho}'} \bar{\rho} \Delta = -\frac{1}{3(1+w)}\Delta,\tag{241}$$

in terms of the comoving density contrast Δ , introducing the equation-of-state $w = \bar{P}/\bar{\rho}$. Combining with the Poisson equation $\nabla^2\Phi = 4\pi G a^2 \bar{\rho} \Delta = \frac{3}{2}\mathcal{H}^2\Delta$, we find

$$\zeta(\vec{k}, \eta) - \mathcal{R}(\vec{k}, \eta) = \frac{2}{9(1+w)} \left(\frac{k}{\mathcal{H}}\right)^2 \Phi(\vec{k}, \eta) \rightarrow 0\tag{242}$$

which quickly decays to zero on superhorizon scales. We can thus use either ζ or $\mathcal{R}$ to describe the superhorizon perturbations, and smoothly connect to the inflationary regime.

Relation to Φ Using either definition, one can derive a general relation between the curvature perturbation and the Newtonian-gauge potential Φ valid on superhorizon scales. By combining the Einstein and conservation equations, we find the superhorizon relation

$$\mathcal{R}(\vec{k}, \eta) \approx \frac{5+3w}{3(1+w)}\Phi(\vec{k}, \eta) \quad (k \ll \mathcal{H});\tag{243}$$

this gives $\mathcal{R} = \frac{3}{2}\Phi$ in radiation-domination ($w = 1/3$) and $\mathcal{R} = \frac{5}{3}\Phi$ in matter-domination, sourcing the previously-stated ratio $\Phi_{\text{MD}}/\Phi_{\text{RD}} = 9/10$. Finally, we can express the adiabatic initial conditions for matter in terms of $\mathcal{R}$ (working in radiation-domination):

$$\boxed{\Phi_i(\vec{k}) = \frac{2}{3}\mathcal{R}_i(\vec{k}), \quad \delta_{\gamma,i}(\vec{k}) = \delta_{\nu,i}(\vec{k}) = \frac{4}{3}\delta_{b,i}(\vec{k}) = \frac{4}{3}\delta_{c,i}(\vec{k}) = -\frac{4}{3}\mathcal{R}_i(\vec{k}).}\tag{244}$$

This provides an explicit link between primordial physics (encoded in the inflationary power spectrum of $\mathcal{R}$, $\Delta_{\mathcal{R}}^2 = H^2/(8\pi^2 M_{\text{Pl}}^2 \epsilon)$) and the late-Universe matter and radiation perturbations.

5.7 Evolution of the Gravitational Potential

We finally have all of the ingredients in place to solve for the evolution of species through cosmic time. We begin by solving for the (Fourier-space) potential, $\Phi(\vec{k}, \eta)$, on all scales, using the initial conditions derived above. An important quantity will be the time at which a given mode enters the horizon (and thus becomes observationally accessible); due to the changing expansion history, we will find different behavior for modes entering in radiation- and matter-dominated eras (*i.e.* if $k > k_{\text{eq}} = \mathcal{H}(\eta_{\text{eq}})$ or $k < k_{\text{eq}}$ for matter-radiation-equality at η_{eq} , cf. Fig. 8).

By combining the trace and trace-free components of the spatial Einstein equation (and neglecting anisotropic stress), we previously derived the following evolution equation for Φ :

$$\Phi'' + 3[1 + c_s^2]\mathcal{H}\Phi' + (2\mathcal{H}' + \mathcal{H}^2[1 + 3c_s^2] + k^2c_s^2)\Phi = 0 \quad (245)$$

(cf. 245) assuming an adiabatic fluid with sound-speed $c_s^2 = \delta P / \delta \rho$. Combining with the second Friedmann equation (which gives $2\mathcal{H}' = -(1 + 3w)\mathcal{H}^2$), and assuming a fixed equation-of-state $w = c_s^2$, we find a closed-form equation for the potential:

$$\Phi'' + 3(1 + w)\mathcal{H}\Phi' - w\nabla^2\Phi = 0, \quad (246)$$

which holds on both sub- and superhorizon scales, assuming constant w .

Matter domination In the matter-dominated regime ($z \lesssim 3500$), the equation of state is $w \approx 0$, and the equation reduces to $\Phi'' + 3\mathcal{H}\Phi' = 0$. Noting that $a \propto \eta^2$ and $\mathcal{H} = 2/\eta$ during matter domination, we find the solutions

$$\Phi(\vec{k}, \eta) = C_1(\vec{k}) + C_2(\vec{k})a(\eta)^{-5/2} \quad (\text{RD}), \quad (247)$$

for integration constants $C_i(\vec{k})$, set by the initial conditions. Note that the growing mode is constant on both sub- and superhorizon scales, while the decaying mode falls off rapidly (and is typically discarded).

Radiation domination At earlier times, the effective equation of state is $w \approx 1/3$, thus the equation becomes $\Phi'' + 4\mathcal{H}\Phi' + \frac{1}{3}k^2\Phi = 0$. Using $a \propto \eta$ and $\mathcal{H} = 1/\eta$, this can be written

$$\frac{d^2u}{d\varphi^2} + \frac{2}{\varphi}\frac{du}{d\varphi} + \left(1 - \frac{2}{\varphi^2}\right)u = 0, \quad (248)$$

defining the rescaled variables through $\Phi(\vec{k}, \eta) = u(\varphi)/\varphi$ and $\varphi = k\eta/\sqrt{3}$. This has solutions $u(\varphi) \propto j_1(\varphi), n_1(\varphi)$ for spherical Bessel and Neumann functions j_1 and n_1 . Discarding the singular solution and fixing the normalization by matching to the superhorizon limit $\Phi(k \rightarrow 0, \eta) \rightarrow \frac{2}{3}\mathcal{R}_i$, we obtain

$$\Phi(\vec{k}, \eta) = 2\mathcal{R}_i(\vec{k}) \left[\frac{\sin(k\eta/\sqrt{3}) - (k\eta/\sqrt{3})\cos(k\eta/\sqrt{3})}{(k\eta/\sqrt{3})^3} \right] \quad (\text{RD}). \quad (249)$$

On superhorizon scales ($k\eta \ll 1$), $\Phi \rightarrow \frac{2}{3}\mathcal{R}_i$, recovering the constant initial condition discussed above, whilst in the subhorizon regime ($k\eta \gg 1$), the cosine term dominates, giving

$$\Phi(\vec{k}, \eta) \rightarrow -6\mathcal{R}_i \frac{\cos(k\eta/\sqrt{3})}{(k\eta)^2}, \quad (250)$$

exhibiting oscillations with sound-speed $c_s = 1/\sqrt{3}$ decaying as $a(\eta)^2$.⁶⁶

⁶⁶Physically, the decay occurs since the relativistic fluid cannot collapse on scales smaller than the Jeans length, due to pressure support.

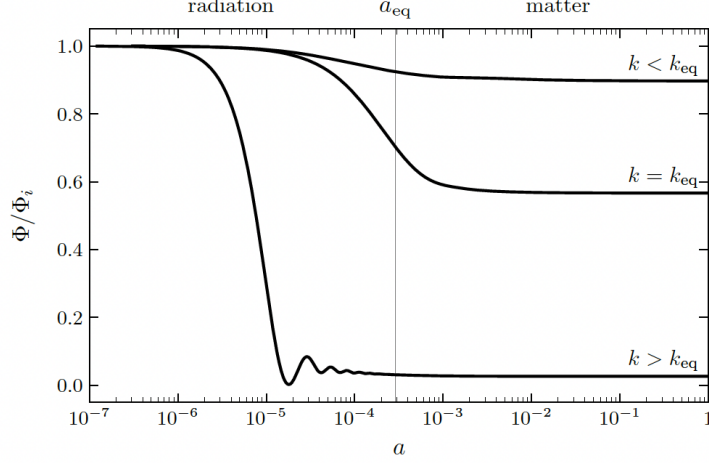


Figure 9: Evolution of the gravitational potential from radiation domination (left) to matter domination (right). Modes with $k < k_{\text{eq}} \approx 0.01 \text{ Mpc}^{-1}$ are superhorizon throughout radiation-domination, thus their amplitude is fixed (up to a 10% decay through matter-radiation equality), while modes with $k > k_{\text{eq}}$ enter the horizon during radiation-domination and feature decaying oscillations, before reaching a constant plateau in matter-domination. Reproduced from D. Baumann, *Cosmology* (Cambridge University Press, 2022).

Full Solution The full evolution of Φ interpolates between the radiation- and matter-domination solutions found above and must be solved numerically. At even later times, we approach dark-energy domination, in which $\Phi(\vec{k}, \eta) \sim a(\eta)^{-1}$ on all scales, thus potentials eventually decay (which gives rise to the late-time integrated Sachs-Wolfe effect).⁶⁷ The key regimes are summarized below and shown in Fig. 9:

	RD ($w = 1/3$)	MD ($w = 0$)	Λ D ($w = -1$)
Superhorizon	const.	const., $\frac{9}{10} \Phi_{\text{RD}}$	$\sim a^{-1}$
Subhorizon	$\sim a^{-2} \cos(k\eta/\sqrt{3})$	const.	$\sim a^{-1}$

Importantly, modes that enter the horizon after matter-radiation equality have large amplitudes (set by primordial physics), whilst those entering before equality are suppressed due to the decay of Φ in radiation-domination. As we will see below, this scale-dependent suppression generates the characteristic shape of the matter power spectrum, with a turnover at $k_{\text{eq}} \equiv \mathcal{H}_{\text{eq}}$.

5.8 Evolution of Dark Matter Perturbations

Next, we describe the growth of cold dark matter (CDM) perturbations, δ_m , through both radiation- and matter-domination. In §5.6, we demonstrated that δ_m is constant on superhorizon scales; here, we will solve for the subhorizon evolution, which itself sources the matter power spectrum.

Here, our starting point will be equation (219):

$$\delta_m'' + \mathcal{H}\delta_m' = \nabla^2\Phi + 3(\Phi'' + \mathcal{H}\Phi') \approx \nabla^2\Phi \quad (k \gg \mathcal{H}), \quad (251)$$

⁶⁷To derive this, note that $\delta\rho \approx \delta\rho_m$, since dark energy does not cluster.

which we derived in §5.4.3 by combining the conservation and Einstein equations. The gravitational potential can be decomposed into contributions sourced by matter and radiation perturbations, *i.e.* $\Phi = \Phi_m + \Phi_r$; from the Poisson equation, these are given by

$$\nabla^2 \Phi_m \approx 4\pi G a^2 \bar{\rho}_m \delta_m, \quad \nabla^2 \Phi_r \approx 4\pi G a^2 \bar{\rho}_r \delta_r \quad (252)$$

for $k \gg \mathcal{H}$ (whence $\Delta_m \approx \delta_m$). In matter-domination, $\bar{\rho}_m \gg \bar{\rho}_r$, thus the first contribution dominates, and we can straightforwardly solve for δ_m . In radiation-domination, however, $\bar{\rho}_r \gg \bar{\rho}_m$, thus we must also consider the effects of Φ_r .

As discussed in §5.4.3, the radiation perturbations satisfy $\delta_r'' - \frac{1}{3}\nabla^2 \delta_r \approx 0$ on subhorizon scales, which leads to oscillatory solutions with $\delta_r \sim \cos(k\eta/\sqrt{3})$ (as discussed below). Since, $\nabla^2 \Phi_r \sim a^2 \bar{\rho}_r \delta_r \sim a^{-2} \delta_r$, the radiation-sourced potential exhibits decaying oscillations with $\Phi_r \sim (ka)^{-2} \cos(k\eta/\sqrt{3})$.⁶⁸ Physically, we are interested only in the gravitational potential averaged over oscillation cycles (since matter cannot respond to a rapidly-oscillating forces), thus we conclude that Φ_r is highly suppressed relative to Φ_m , even in radiation domination. Setting $\Phi \approx \Phi_m$ leads to the closed-form equation

$$\delta_m'' + \mathcal{H} \delta_m' \approx 4\pi G a^2 \bar{\rho}_m \delta_m = \frac{3}{2} \Omega_m \mathcal{H}^2 \delta_m, \quad (253)$$

using the first Friedmann equation with $\bar{\rho}_m(\eta) = \Omega_m(\eta) \bar{\rho}_{\text{tot}}(\eta)$.

To solve this through the radiation- and matter-dominated eras, we change variables to $y \equiv a/a_{\text{eq}}$ (where $a_{\text{eq}} = \Omega_{r,0}/\Omega_{m,0}$ is the equality scale), with

$$\frac{\mathcal{H}}{\mathcal{H}_0} = \frac{aH}{a_0 H_0} = a \sqrt{\Omega_{m,0}/a^3 + \Omega_{r,0}/a^4} = \frac{\Omega_{m,0}}{\sqrt{\Omega_{r,0}}} \frac{\sqrt{1+y}}{y}, \quad \Omega_m(a) = \frac{y}{1+y}, \quad \frac{d}{d\eta} = y a_e \mathcal{H} \frac{d}{da} y \mathcal{H} \frac{d}{dy} \quad (254)$$

ignoring dark energy. This leads to the **Mészáros equation**:

$$\boxed{\frac{d^2 \delta_m}{dy^2} + \frac{2+3y}{2y(1+y)} \frac{d\delta_m}{dy} - \frac{3}{2y(1+y)} \delta_m = 0.} \quad (255)$$

which describes the subhorizon growth of dark matter perturbations.

The Mészáros equation has two exact solutions:

$$\delta_m(y) \propto \begin{cases} (1 + \frac{3}{2}y) \\ (1 + \frac{3}{2}y) \log \left[\frac{\sqrt{1+y}+1}{\sqrt{1+y}-1} \right] - 3\sqrt{1+y} \end{cases}, \quad (256)$$

which describe both growing and decaying modes. In the radiation-dominated regime ($y \ll 1$), the growing mode behaves as $\delta_m \propto \log y \propto \log a$, while the decaying mode tends to a constant: this implies that dark matter perturbations grow only *logarithmically* during radiation domination, since the rapidly-oscillating radiation pressure dominates the energy budget and prevents the matter from clumping efficiently.⁶⁹ In the matter-dominated regime ($y \gg 1$), the growing mode behaves as $\delta_m \propto y \propto a$, while the decaying mode falls off as $y^{-3/2} \propto a^{-3/2}$: matter perturbations now grow linearly with the scale factor, seeded by logarithmically-enhanced growth in radiation-domination.

Typically, one expresses the subhorizon growth of dark matter perturbations using a pair of scale-independent ‘growth factors’ $D_{\pm}(t)$ via

$$\delta_m(\vec{k}, \eta) = D_+(\eta) \delta_+(\vec{k}) + D_-(\eta) \delta_-(\vec{k}), \quad (257)$$

⁶⁸This matches the scaling found in the previous section for the total potential during radiation-domination.

⁶⁹This is known as the Mészáros effect.

(switching to coordinate time t), where $\delta_{\pm}(\vec{k})$ encode the spatial variation, set by the initial conditions. As demonstrated above, the growing mode scales as $\log a \propto \log t$ during radiation-domination and $a \propto t^{2/3}$ in matter-domination, whilst $D_{-}(t)$ is constant during radiation-domination and decays as $a^{-3/2} \propto t^{-1}$ in matter-domination.⁷⁰ Finally during dark energy domination, D_{+} is constant and $D_{-} \propto a^{-2}$ (with exponential decay in time), since $\delta\rho/\bar{\rho}_{\text{tot}} \rightarrow 0$ (as $\delta\rho_{\Lambda} = 0$), halting structure growth.

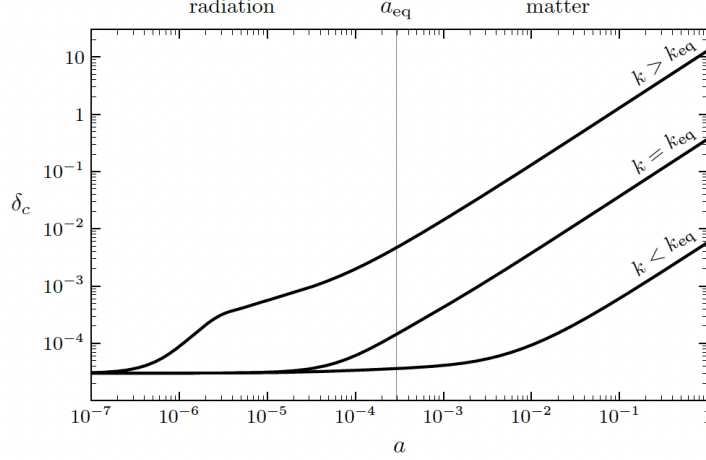


Figure 10: Evolution of cold dark matter perturbations over time. Modes start frozen on superhorizon scales, then grow as they enter the horizon. Modes with $k < k_{\text{eq}}$ enter in matter-domination and grow as $a \propto t^{2/3}$, while those with $k > k_{\text{eq}}$ have an initial phase of $\log a \propto \log t$ growth during radiation-domination (which is somewhat distorted by the logarithmic axis choice). Reproduced from D. Baumann, *Cosmology* (Cambridge University Press, 2022).

In summary, we find the following behavior for the growing mode of δ_m in radiation-, matter-, and dark-energy-dominated regimes:

	RD ($w = 1/3$)	MD ($w = 0$)	AD ($w = -1$)
Superhorizon	const.	const.	const.
Subhorizon	$\sim \log a$	$\sim a$	const.

which is shown graphically in Fig. 12.

5.9 The Matter Power Spectrum

By combining the time-dependence of δ_m derived in the previous section with the initial conditions from inflation, we can solve for the statistics of dark matter in the late Universe. Here, we will focus on the **matter power spectrum**, $P_m(k, \eta)$, defined via

$$\langle \delta_m(\vec{k}, \eta) \delta_m^*(\vec{k}', \eta) \rangle = (2\pi)^3 \delta_D(\vec{k} - \vec{k}') P_m(k, \eta), \quad (258)$$

⁷⁰This can also be derived from the original equation for δ_m : setting $\bar{\rho}_r \delta_r \rightarrow 0$ on subhorizon scales (where δ_r oscillates rapidly), the Mészáros equation reduces to $\ddot{\delta}_m + (1/t)\dot{\delta}_m \approx 0$ in radiation domination (using $H = 1/2t$), with solutions $\delta_m \propto \{1, \log t\} \propto \{1, \log a\}$.

where the Dirac delta appears due to translation-invariance, as before.⁷¹ Physically, $P_m(k, \eta)$ represents the variance of the density field smoothed on scale k^{-1} .

As discussed above, the evolution of δ_m depends on whether a given mode is (a) superhorizon, (b) subhorizon during radiation-domination or (c) subhorizon during matter-domination (neglecting dark-energy domination). Recalling that a mode of wavenumber k enters the horizon at when $\mathcal{H} = aH = k$, this defines three wavenumber regimes: (a) $k < \mathcal{H}_0$; (b) $\mathcal{H}_0 < k < \mathcal{H}_{\text{eq}}$; (c) $k > \mathcal{H}_{\text{eq}}$, where $\mathcal{H}_{\text{eq}} \approx 0.01 \text{ Mpc}^{-1}$ defines the wavenumber of modes entering at matter-radiation equality. Below, we will describe the power spectrum in each regime, building up the model over time.

Early Times At sufficiently early times, all Fourier modes of interest today are outside the horizon, thus the matter density contrast tracks the primordial curvature perturbation, $\delta_m(\vec{k}, \eta_i) \approx -\mathcal{R}_i(\vec{k})$ (cf. §5.6). In this regime, the density field inherits the primordial curvature spectrum generated during inflation (§4.8):

$$P_m(k, \eta_i) = P_{\mathcal{R}}(k) \equiv \frac{2\pi^2}{k^3} \Delta_{\mathcal{R}}^2(k) = \frac{2\pi^2}{k^3} A_s k^{n_s-1}, \quad (259)$$

which scales as k^{-3} in the scale-invariant limit (if $n_s \rightarrow 1$).

Radiation-Domination Next, consider the Universe at matter-radiation equality ($\eta = \eta_{\text{eq}}$). At this time, modes with $k < k_{\text{eq}}$ are still superhorizon and are thus frozen (with $\delta_m(\vec{k}, \eta_{\text{eq}}) = -\mathcal{R}_i(\vec{k})$ as before). Modes with $k > k_{\text{eq}}$, on the other hand, entered the horizon at some earlier time $a_k < a_{\text{eq}}$ (defined by $k = (aH)_k \propto a_k^{-1}$) and are growing logarithmically due to the Mészáros effect, with a total enhancement of $\log(a_{\text{eq}}/a_k) \sim \log(k/k_{\text{eq}})$ by η_{eq} . This leads to the power spectrum

$$P_m(k, \eta_{\text{eq}}) \sim P_{\mathcal{R}}(k) \begin{cases} 1, & k < k_{\text{eq}}, \\ \log^2(k/k_{\text{eq}}), & k > k_{\text{eq}}, \end{cases} \quad (260)$$

scaling as $k^{-3} \log^2(k/k_{\text{eq}})$ on small-scales.

Matter-Domination Finally, we consider the power spectrum at the current time, η_0 (assuming matter domination). As mentioned above there are three regimes: (a) modes with $k < k_0 \equiv \mathcal{H}_0$ are still outside the horizon, giving $\delta_m(\vec{k}, \eta_0) = -\mathcal{R}_i(\vec{k})$; (b) modes with $k_0 < k < k_{\text{eq}}$ entered the horizon during matter-domination (with $k = (aH)_k \propto a_k^{-1/2}$), and have thus grown by a factor $a_0/a_k \propto (k/k_0)^2$; (c) modes with $k > k_{\text{eq}}$ entered during radiation-domination and grew by $\log(a_{\text{eq}}/a_k) \sim \log(k/k_{\text{eq}})$, before growing by a further $a_0/a_{\text{eq}} \sim (k_{\text{eq}}/k_0)^2$ in matter-domination. In combination, we find the late-time matter power spectrum

$$P_m(k, \eta_0) \sim P_{\mathcal{R}}(k) \begin{cases} 1 & (k < k_0) \\ (k/k_0)^4 & (k_0 < k < k_{\text{eq}}) \\ (k_0/k_{\text{eq}})^4 \log^2(k/k_{\text{eq}}), & (k > k_{\text{eq}}). \end{cases} \quad (261)$$

which scales as $k^{-3}, k^1, k^{-3} \log^2(k/k_{\text{eq}})$ for a scale-invariant primordial power spectrum. As shown in Fig. 11, these scalings reproduce the characteristic ‘broken-power-law’ shape of the full dark

⁷¹The right-hand-side can also be written as $(2\pi)^3 \delta_{\text{D}}(\vec{k} - \vec{k}') \frac{2\pi^2}{k^3} \Delta_m^2(k, \eta)$, where Δ_m^2 is the *dimensionless* power spectrum (cf. §4). We use the large-scale-structure convention of working with $P_m(k, \eta)$ rather than $\Delta_m^2(k, \eta)$ here.

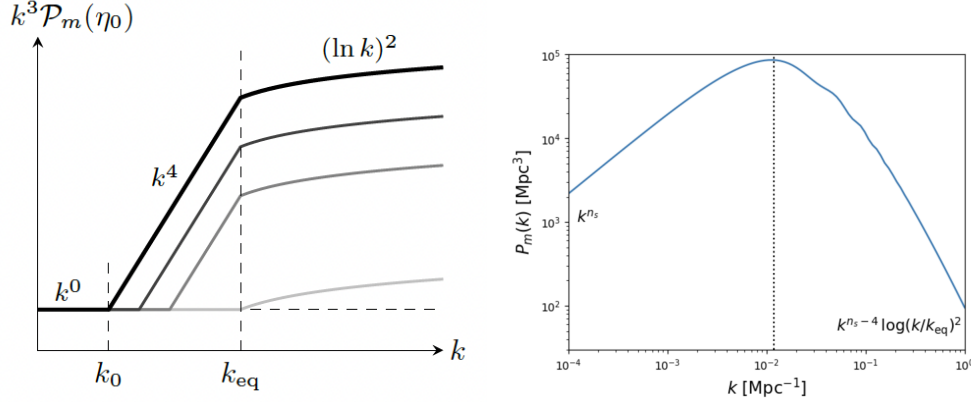


Figure 11: **Left:** Idealized view of the matter power spectrum, scaled by k^3 . Modes with $k < k_0$ are superhorizon, thus trace the initial conditions, whilst those with $k_0 < k_{\text{eq}}$ entered the horizon during matter domination, and grow as k^4 . Finally, modes with $k > k_{\text{eq}}$ entered during radiation domination, thus scale as $\log(k)^2$. **Right:** Full computation of the matter power spectrum on subhorizon scales, computed using CLASS with $n_s \approx 0.96$. Up to the smooth behavior around k_{eq} and the wiggles at $k \sim 10^{-1}$ (which we discuss in §5.10.4), this reproduces the phenomenology from the simplified picture. The transition from radiation- to matter-domination sources the ‘hump’ at $k = k_{\text{eq}}$. Left panel reproduced from D. Baumann, *Cosmology* (Cambridge University Press, 2022).

matter power spectrum, with a prominent turnover at k_{eq} sourced by the different growth history of modes in matter- and radiation-domination.⁷²

Often, one extracts the scale-dependent features of the power spectrum using a **transfer function**, $T(k)$, defined via

$$T(k) \equiv \frac{D_+(\eta_i)}{D_+(\eta)} \frac{\delta_m(\vec{k}, \eta)}{\delta_m(\vec{k}, \eta_i)}, \quad (262)$$

where η is some time after equality. This relates the late-time density contrast to its primordial value, isolating the effects of scale-dependent growth. By construction, $T(k) \rightarrow 1$ for $k \ll k_{\text{eq}}$ and $T(k) \propto k^{-2} \log k$ for $k \gg k_{\text{eq}}$. With this notation, one can rewrite the late-time density directly in terms of the primordial field:

$$\delta_m(\vec{k}, \eta > \eta_{\text{eq}}) = D_+(\eta) T(k) \mathcal{R}(\vec{k}), \quad (263)$$

which cleanly isolates the time-dependence ($D_+ \sim a$), scale-dependence ($T \sim 1 + k^{-2} \log k$) and dependence on the initial conditions ($\mathcal{R}$). This form is extensively used to explore the impacts of modified inflationary initial conditions on large-scale structure statistics (e.g., non-Gaussianity and primordial features).

5.10 Evolution of the Photons and Baryons

Lastly, we solve for the evolution of photons and baryon fluctuations through cosmic history. These source the CMB fluctuations and will be important for the discussion in §6. Whilst baryons behave

⁷²In our discussion, we have ignored the effects of dark-energy domination. This freezes δ_m , leading to a slight suppression of the present-day power spectrum amplitude, though not the shape.

analogously to dark matter today, their distribution in the early Universe was very different. As discussed in §3, rapid Compton scattering of photons and electrons kept the baryons tightly coupled to photons prior to decoupling at $z \approx 1100$; in this regime, the species behave as a single fluid, with a shared bulk velocity and correlated perturbations.⁷³ Due to radiation pressure, the photon-baryon fluid oscillates in the early Universe, with baryons providing inertia and dark matter providing a driving force through potential wells. The interplay between pressure and gravity leads to the rich structure of the CMB acoustic peaks as well as the baryon acoustic oscillation (BAO) feature in the late-time matter power spectrum (sourcing the ‘wiggles’ in the right panel of Fig. 11).

5.10.1 Photon Evolution

We begin by studying the distribution of photons, δ_γ , in the absence of baryon perturbations. This is a good approximation at sufficiently early times, whence $\bar{\rho}_b \ll \bar{\rho}_\gamma$.

Radiation-Domination Deep in the radiation-dominated regime, the Poisson equation gives $\nabla^2 \Phi \approx 4\pi G a^2 \bar{\rho}_r \Delta_r$, where Δ_r is the comoving density contrast.⁷⁴ Using the first Friedmann equation and $\mathcal{H} = 1/\eta$, this can be written

$$\Delta_r(\vec{k}, \eta) = -\frac{2}{3}(k\eta)^2 \Phi(\vec{k}, \eta); \quad (264)$$

inserting the radiation-domination solution for Φ derived in §5.7, we immediately obtain

$$\Delta_r(\vec{k}, \eta) = -4\mathcal{R}_i(\vec{k}) \frac{\sin \varphi - \varphi \cos \varphi}{\varphi}, \quad \varphi \equiv k\eta/\sqrt{3}, \quad (265)$$

which has the super- and subhorizon limits

$$\Delta_r(\vec{k}, \eta) \rightarrow \mathcal{R}_i(\vec{k}) \begin{cases} -\frac{4}{9}(k\eta)^2 & (k\eta \ll 1), \\ 4 \cos(k\eta/\sqrt{3}) & (k\eta \gg 1). \end{cases} \quad (266)$$

On large scales, Δ_r is heavily suppressed, while on small scales, Δ_r oscillates with sound speed $c_s = 1/\sqrt{3}$. Switching to gauge-dependent variables, the photon perturbation $\delta_\gamma(\vec{k}, \eta)$ takes a constant value in the superhorizon limit ($-\frac{4}{3}\mathcal{R}_i(\vec{k})$, as found previously), whilst $\delta_\gamma(\vec{k}, \eta)$ is equal to $\Delta_r(\vec{k}, \eta)$ on subhorizon-scales and thus oscillates as a sound wave.⁷⁵

Matter-Domination At late-times, the Poisson equation is dominated by the dark matter contributions, and cannot be used to solve for δ_γ . In §5.4.3, we showed that the continuity and Euler equations for δ_γ reduce to

$$\delta'_\gamma = -\frac{4}{3}\vec{\nabla} \cdot \vec{v}_\gamma + 4\Phi', \quad \vec{v}'_\gamma = -\frac{1}{4}\vec{\nabla} \delta_\gamma - \vec{\nabla} \Phi, \quad (267)$$

which can be combined to form a wave equation:

$$\delta''_\gamma - \frac{1}{3}\nabla^2 \delta_\gamma = \frac{4}{3}\nabla^2 \Phi + 4\Phi'', \quad (268)$$

⁷³This description is accurate if the optical depth between successive scatterings is much shorter than a Hubble time, which holds until decoupling (by definition).

⁷⁴These equations apply also to neutrinos, though we ignore them here for the sake of simplicity.

⁷⁵It is straightforward to show that $(\delta - \Delta)/\Delta \sim (\mathcal{H}/k)^2$, which vanishes on small-scales.

where the final term vanishes, since Φ is constant in matter domination. This is a simple harmonic oscillator equation with driving provided by the gravitational potential, and has the general solution is given by

$$\delta_\gamma(\vec{k}, \eta) = C(\vec{k}) \cos(k\eta/\sqrt{3}) + D(\vec{k}) \sin(k\eta/\sqrt{3}) - 4\Phi(\vec{k}), \quad (269)$$

for integration constants $C(\vec{k}), D(\vec{k})$ set by the initial conditions and matter-radiation equality. Note that the oscillations occur around $\delta_\gamma^{\text{eq}} = -4\Phi$ rather than around zero;⁷⁶ due the potential well induced by dark matter perturbations. On superhorizon scales, $\delta_\gamma(\vec{k}, \eta) \rightarrow \text{const.}$, as before. This solution is shown graphically in Fig. 12.

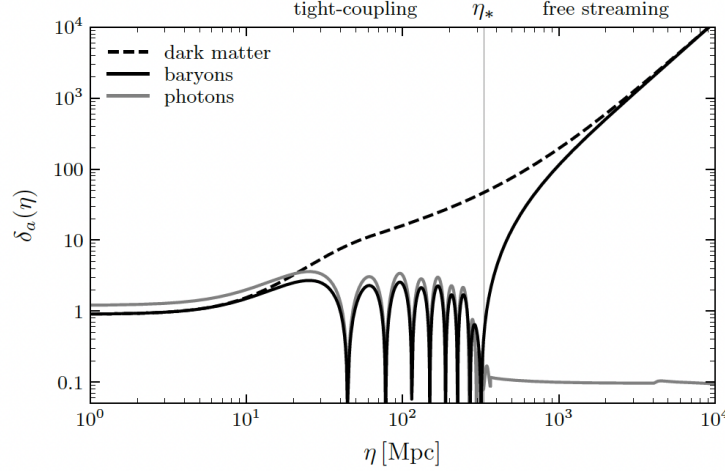


Figure 12: Evolution of photons, baryons, and dark matter perturbations over time. Dark matter perturbations grow as $\log a$ in radiation-domination then a in matter-domination, while photons oscillate in radiation-domination before gradually freezing in radiation-domination. Before the epoch of baryon drag, baryons track the fluctuations in photons (up to a factor of 3/4 from the adiabatic initial conditions); after decoupling, they grow as a and asymptote to the dark matter limit. Reproduced from D. Baumann, *Cosmology* (Cambridge University Press, 2022).

5.10.2 Tight Coupling

In the limit of highly-efficient Compton scattering (known as *tight coupling*), baryons and photons behave as a single fluid, whose evolution can be treated using similar techniques to those for photons. As we show below, the addition of baryons modifies the picture in two ways: (a) the momentum density is modified due to the additional inertia provided by baryons (b) the effective sound speed decreases relative to the pure-radiation case (which had $c_s^2 = 1/3$).

In the tight-coupling regime, baryons and photons share a common velocity $\vec{v}_\gamma = \vec{v}_b$, giving the combined momentum density

$$\vec{q} = (\bar{\rho}_\gamma + \bar{P}_\gamma)\vec{v}_\gamma + (\bar{\rho}_b + \bar{P}_b)\vec{v}_b \equiv \frac{4}{3}(1 + R)\bar{\rho}_\gamma\vec{v}_\gamma, \quad R \equiv \frac{\bar{\rho}_b}{\bar{\rho}_\gamma + \bar{P}_\gamma}, \quad (270)$$

where R measures the baryon-to-photon momentum ratio (which grows as a , reaching values of order unity at recombination). Inserting into the Euler equation $\vec{q}' + 4\mathcal{H}\vec{q} = -(\bar{\rho} + \bar{P})\vec{\nabla}\Phi - \vec{\nabla}\delta P$

⁷⁶More strictly, $\delta_\gamma^{\text{eq}} = -4\Phi_0$, where Φ_0 is the time-averaged potential during matter-domination.

(and noting that $\bar{\rho}_\gamma \propto a^{-4}$), we find

$$[(1+R)\vec{v}_\gamma]' = -\frac{1}{4}\vec{\nabla}\delta_\gamma - (1+R)\vec{\nabla}\Phi, \quad (271)$$

which describes the evolution of photon velocity due to pressure gradients and gravity, with baryons entering through the $1+R$ factors (corresponding to an increased inertia, *i.e.* mass). Since the scattering processes are elastic, there is no exchange of energy between photons and baryons: this implies that the continuity equations are unchanged, thus $\delta'_\gamma = -\frac{4}{3}\vec{\nabla} \cdot \vec{v}_\gamma + 4\Phi'$. Combining the above equations, we find

$$\boxed{\delta''_\gamma + \frac{R'}{1+R}\delta'_\gamma - \frac{1}{3(1+R)}\nabla^2\delta_\gamma = \frac{4}{3}\nabla^2\Phi + 4\Phi'' + \frac{4R'}{1+R}\Phi'.} \quad (272)$$

This describes a damped and driven acoustic waves with effective sound speed

$$c_s^2 = \frac{1}{3(1+R)}, \quad (273)$$

which is suppressed compared to the radiation prediction ($c_s^2 = 1/3$) due to the baryon loading.

5.10.3 Acoustic Waves

On sufficiently small scales, we can ignore metric perturbations, thus the above equation reduces to the homogeneous wave equation

$$\delta''_\gamma + \frac{R'}{1+R}\delta'_\gamma - \frac{1}{3(1+R)}\nabla^2\delta_\gamma = 0, \quad (274)$$

with frequency $\omega = kc_s(\eta)$. For $\omega \gg \mathcal{H}$, the oscillations are rapid compared to the expansion-induced damping, thus we can solve using a WKB approximation, finding⁷⁷

$$\delta_\gamma(\vec{k}, \eta) \approx \frac{A(\vec{k})}{(1+R)^{1/4}} \exp[\pm ikr_s(\eta)], \quad r_s(\eta) \equiv \int_0^\eta c_s(\eta') d\eta'. \quad (275)$$

Notably, this depends on the **sound horizon**, $r_s(\eta)$, which is equal to the comoving distance a sound wave can travel between reheating and time η . At decoupling, $r_s(\eta_*) \approx 150$ Mpc, which sets the fundamental scale of oscillations in the CMB and the matter power spectrum.

Assuming adiabatic initial conditions, δ_γ is constant on superhorizon scales; this implies that only the cosine branch is excited, giving the tight-coupling solution:

$$\boxed{\delta_\gamma(\vec{k}, \eta) \approx \frac{C(\vec{k})}{(1+R)^{1/4}} \cos(kr_s(\eta)),} \quad (276)$$

where $C(\vec{k})$ is a (real) integration constant set by the initial conditions, *i.e.* $\mathcal{R}(\vec{k})$. The conclusion is clear: the photon-baryon fluid exhibits acoustic waves with phase velocity c_s , propagating from density perturbations seeded by inflation.

There are a number of important implications of this result:

⁷⁷Explicitly, one inserts the ansatz $\delta_\gamma^{\text{WKB}}(\vec{k}, \eta) = B(\vec{k}, \eta) \exp[\pm i \int^\eta \omega(k, \eta') d\eta']$ into the equation and solves for the slowly-varying function B . Noting that (a) the B'' term is subdominant for $\omega \gg \mathcal{H}$, and (b) the frequency evolves according to $d \log \omega = -(1/2)d \log(1+R)$, this yields $B \propto (1+R)^{-1/4}$.

1. **Phase coherence:** Since only the cosine solution is excited, modes at all frequencies oscillate coherently, with a common phase $kr_s(\eta)$. This is a direct prediction of single-field inflation (and would be violated by isocurvature perturbations, see §5.6) and leads to the characteristic series of peaks and troughs in the CMB power spectrum.
2. **Velocity oscillations** From the continuity equation, the velocity perturbations scale as $\delta\gamma'$, thus $v_\gamma \sim \sin(kr_s)$. This implies that velocity and density perturbations are out-of-phase by $\pi/2$ and generates Doppler shifts that contribute to the CMB temperature anisotropies (see §6).
3. **Silk damping:** On scales smaller than the photon mean free path, photons diffuse out of high-density regions, which leads to an exponential suppression $\delta_\gamma \propto \exp[-(k/k_{\text{Silk}})^2]$ in the oscillation envelope. This damps the small-scale acoustic peaks in the CMB.

5.10.4 Baryon Acoustic Oscillations

Up to redshift $z_{\text{drag}} \approx 1060$, the tight coupling between baryons and photons leads to acoustic oscillations in both the baryon densities and velocities, with $\delta_b \approx \frac{3}{4}\delta_\gamma$ (assuming adiabaticity) and $v_b \approx v_\gamma$ (as shown in Fig. 12). After the epoch of baryon drag (which occurs slightly after recombination due to the large excess of photons over baryons, see §3), the baryons decouple from the fluid and evolve according to the conservation equations for pressureless matter, with perturbations growing as $\delta_b \propto a$.

Importantly, the distribution of baryons retains information on the acoustic oscillations, which provide initial conditions for the decoupled evolution. Furthermore, this information propagates to the total matter perturbation, $\delta_m \equiv f_b\delta_b + (1 - f_b)\delta_c$ (where $f_b = \Omega_{b,0}/\Omega_{m,0} \approx 0.16$ is the baryon fraction), both directly and through changes to δ_c induced by the baryonic contribution to the gravitational potential. The net effect is that the late-time density field contains a weak acoustic signature $\propto f_b$, known as **Baryon Acoustic Oscillation** (BAO).

To understand this in more detail, let's consider the baryon distribution just before η_{drag} . As derived above, $\delta_b(\vec{k}, \eta_{\text{drag}}) \propto \cos(kr_d)\Phi_i(\vec{k})$ and $v_b(\vec{k}, \eta_{\text{drag}}) \propto \sin(kr_d)\Phi_i(\vec{k})$, where $r_d \equiv r_s(\eta_{\text{drag}})$ is the sound-horizon at baryon drag. After η_{drag} , the oscillations cease, and the *combined* matter density evolves according to the Mészáros equation, which has the general solution

$$\delta_m(\vec{k}, \eta) = C_1(\vec{k})y + C_2(\vec{k})y^{-3/2} \quad (y > a_{\text{drag}}/a_{\text{eq}}) \quad (277)$$

where $y = a/a_{\text{eq}}$ (neglecting radiation contributions for simplicity). Matching both δ_m and δ'_m at η_{drag} (where the baryonic contributions have the forms discussed above), we find the baryonic contribution to the growing mode

$$\delta_m(\vec{k}, \eta) \supset f_b [\delta_b(\vec{k}, \eta_{\text{drag}}) + (k\eta_d) v_b(\vec{k}, \eta_{\text{drag}})] \frac{a}{a_{\text{drag}}}, \quad (\eta > \eta_{\text{drag}}) \quad (278)$$

which expands with a , as expected. On small scales, the velocity term dominates,⁷⁸ producing an oscillatory contribution of the form $\delta_m(\vec{k}, \eta) \supset f_b(a/a_{\text{drag}}) \sin(kr_d)\Phi_i(\vec{k})$.

The net effect is that the matter power spectrum contains a small oscillatory component in addition to the smooth features described above, known as the BAO feature:

$$P_m(k) \supset 2f_b \sin(kr_d) e^{-(k/k_{\text{Silk}})^2} P_m^{\text{smooth}}(k), \quad (279)$$

⁷⁸Physically, this arises due to the baryonic contribution to the gravitational potential (with $\Phi \supset (\mathcal{H}/k)^2 f_b \delta_b$).

including the exponential suppression at high- $k > k_{\text{Silk}}$. This feature can be seen from Fig. 11 and plays a central role in large-scale structure cosmology, since $r_d \approx 150 \text{ Mpc}$ acts as a **standard ruler**: its physical size is fixed by early-Universe physics (and is calibrated by the CMB to sub-percent precision), and its angular size can be measured from galaxy populations at various redshifts, giving a clean probe of the expansion history through the angular diameter distance $d_A(z)$ (and thus the relative expansion rate, $E(z)$).

5.11 Summary

In the previous sections, we have mapped out the evolution of dark matter, baryon, and photon distributions through radiation- and matter-domination, and connected them to their quantum initial conditions using the curvature perturbation, $\mathcal{R}$. For reference, we provide a brief summary of the various forms of evolution, focusing on a perturbation with $k < k_{\text{eq}}$ that enters the horizon during matter domination.

1. **Photons** (δ_γ): On superhorizon scales, $\delta_\gamma = -2\Phi$ is frozen, then acoustic oscillations begin when the mode enters the horizon at $k = \mathcal{H}$, with $\delta_\gamma \sim \cos(kr_s)\Phi$. After recombination, the photons free-stream out of the potential wells to form the CMB we observe today (see §6).
2. **Baryons** (δ_b): Initially, the distribution of baryons is frozen with $\delta_b = \frac{3}{4}\delta_\gamma$ on superhorizon scales due to adiabaticity. Upon horizon entry, the baryons oscillate along with the photons, before decoupling at $z_{\text{drag}} \approx 1060$ and growing as $\delta_b \propto a$, with their statistics encoding the BAO features with wavelength $r_d = r_s(\eta_{\text{drag}})$.
3. **Dark matter** (δ_c): Cold dark matter starts with $\delta_b = \frac{3}{4}\delta_\gamma$ on superhorizon scales, before growing logarithmically during the radiation era (due to the Mészáros effect) and finally linearly during the matter era, sourcing the matter power spectrum, $P_m(k, \eta)$.

For modes with $k > k_{\text{eq}}$, the evolution is similar, but features an additional Mészáros suppression in δ_m , sourcing a characteristic $\log^2 k$ scaling of the small-scale matter power spectrum. In practice, one usually solves for the evolution of species numerically using codes such as CAMB and CLASS; at heart, these implement the Einstein equations discussed above, in combination with the power-law initial conditions from inflation, and various asymptotic limits above (e.g., tight-coupling).

6 The Cosmic Microwave Background

Having developed the linear-evolution machinery for photons, baryons, and dark matter, we are now in a position to use it to make our first predictions for an observable signature of the primordial Universe: the temperature anisotropies of the **cosmic microwave background** (CMB). At a conceptual level, the CMB is the cleanest snapshot of the early Universe available to us: the photons we observe today were emitted at last scattering ($z_* \simeq 1090$, when Compton scattering ceased) and have travelled freely ever since, carrying with them information on both the inflationary initial conditions (encoded by the curvature perturbation $\mathcal{R}$) and the subsequent linear evolution of δ_γ , $\vec{v}_\gamma$, Φ , and Ψ up to the last-scattering surface (LSS). CMB physics is therefore an enormously rich subject; in this final section, we give a brief introduction to its essentials, focusing on the large-scale ‘‘Sachs-Wolfe’’ regime that is most readily connected to the perturbation theory developed in earlier sections.⁷⁹

⁷⁹We restrict ourselves to the temperature anisotropies, ignoring CMB polarization, which encodes additional information about the velocity field at LSS and the primordial gravitational-wave background.

6.1 CMB Statistics

The CMB is a **two-dimensional** observable: we measure the temperature perturbation $\delta T(\hat{n}) \equiv T(\hat{n}) - \bar{T}$ as a function of position on the sky, $\hat{n} \equiv (\theta, \phi)$, with mean temperature $\bar{T} \simeq 2.73$ K. Since the relevant statistical isotropy is on the sphere (rather than in three-dimensional space), it is natural to expand δT in spherical harmonics rather than Fourier modes. The two-dimensional analog of the 3D Fourier decomposition reads

$$\delta T(\hat{n}) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\hat{n}), \quad a_{\ell m} = \int d\hat{n} \delta T(\hat{n}) Y_{\ell m}^*(\hat{n}), \quad (280)$$

where $Y_{\ell m}(\theta, \phi) \propto P_{\ell}^m(\cos \theta) e^{im\phi}$ are the spherical harmonics, the multipole index $\ell = 0, 1, 2, \dots$ indicates angular scales $\theta \sim \pi/\ell$, and the azimuthal index $m \in \{-\ell, \dots, \ell\}$ encodes orientation.

Just as in three dimensions, statistical isotropy of the underlying physics implies that the two-point function depends only on the multipole index ℓ , and we can define the **angular power spectrum**

$$\langle a_{\ell m} a_{\ell' m'}^* \rangle = \delta_{\ell \ell'} \delta_{m m'} C_{\ell}, \quad (281)$$

the 2D analog of the matter power spectrum $P_m(k)$ introduced in §5.10. The quantity $\ell(\ell+1)C_{\ell}/(2\pi)$ measures the contribution of each angular scale to the temperature variance per logarithmic interval in ℓ , and is the canonical way to display CMB data. We will compute it for the large-angle Sachs-Wolfe regime below.

6.2 The Temperature Anisotropy: Sachs-Wolfe and Beyond

To predict C_{ℓ} , we need to relate $\delta T(\hat{n})$ to the linear perturbations $(\delta_{\gamma}, \vec{v}_b, \Phi, \Psi)$ derived in earlier sections. The starting point is the photon geodesic equation, which describes how photon energies evolve under the influence of metric perturbations. In an affine parameter λ ,

$$\frac{dp^{\mu}}{d\lambda} = -\Gamma_{\alpha\beta}^{\mu} p^{\alpha} p^{\beta}, \quad p^{\mu} \equiv \frac{dx^{\mu}}{d\lambda}, \quad (282)$$

where p^{μ} is the photon 4-momentum. To connect to observable quantities, we relate p^{μ} to the energy E measured by a comoving observer via $p^{\hat{\mu}} = (E, p^{\hat{i}})$ in a local inertial frame.⁸⁰ After some algebra (see Baumann), the relevant equation governing the redshift of aE along a photon trajectory takes the compact form

$$\frac{d \log(aE)}{d\eta} = \Psi' + \Phi' - \frac{d\Psi}{d\eta}, \quad (283)$$

where the total derivative $d/d\eta = \partial_{\eta} + \hat{p}^i \partial_i$ is taken along the photon path. Integrating from emission at η_e to today at η_0 , and using the boundary terms to capture the gravitational redshift,

$$\log(a_0 E_0) - \log(a_e E_e) = \Psi_e - \Psi_0 + \int_{\eta_e}^{\eta_0} (\Psi' + \Phi') d\eta, \quad (284)$$

where $\Psi_e \equiv \Psi(\vec{x}_*, \eta_*)$ is evaluated at the (last-scattering) emission event and Ψ_0 at the observer (today).

⁸⁰Explicitly, $p^0 = (E/a)(1 - \Psi)$ and $p^i = (E/a)(1 + \Phi)\hat{p}^i$ in the Newtonian gauge, ensuring $\eta_{\hat{\mu}\hat{\nu}} p^{\hat{\mu}} p^{\hat{\nu}} = g_{\mu\nu} p^{\mu} p^{\nu}$.

To extract a temperature anisotropy, we note that CMB photons retain their thermal spectrum after decoupling (since the Bose-Einstein distribution is preserved under free-streaming), so $T \propto E$. The temperature variation observed today comes from three sources: the intrinsic temperature at emission (the photon overdensity δ_γ contributes $\delta T/T = \delta_\gamma/4$, since $\bar{\rho}_\gamma \propto T^4$), the kinematic Doppler shift from the line-of-sight velocity of the baryon-photon fluid at last scattering, and the metric corrections in the equation above. Combining these, and dropping the unobservable constant Ψ_0 contribution, we arrive at the celebrated formula for the CMB temperature anisotropy:

$$\boxed{\frac{\delta T}{T}(\hat{n}) = \underbrace{\left[\frac{1}{4}\delta_\gamma + \Psi\right]_*}_{\text{Sachs-Wolfe}} - \underbrace{(\hat{n} \cdot \vec{v}_b)_*}_{\text{Doppler}} + \underbrace{\int_{\eta_*}^{\eta_0} (\Psi' + \Phi') d\eta}_{\text{Integrated Sachs-Wolfe}}}, \quad (285)$$

where the subscript $*$ denotes evaluation at the last-scattering surface and the line integral is taken along the photon trajectory $\vec{x}(\eta) = \hat{n}(\eta_0 - \eta)$. The three terms have distinct physical origins:

1. The **Sachs-Wolfe** term, $(\delta_\gamma/4 + \Psi)|_*$, combines the intrinsic temperature fluctuation at LSS with the gravitational redshift experienced by photons climbing out of the potential wells at last scattering. Since $\Psi < 0$ inside overdense regions, the SW term *cools* photons emerging from such regions, partially compensating the larger intrinsic δ_γ there. This term dominates on large angular scales.
2. The **Doppler** term, $-(\hat{n} \cdot \vec{v}_b)_*$, records the line-of-sight peculiar velocity of the baryon-photon fluid at LSS. It vanishes on superhorizon scales (where $\vec{v}_b \sim (k/\mathcal{H})\Psi$ is suppressed by $k/\mathcal{H} \rightarrow 0$) and provides a $\pi/2$ -out-of-phase contribution to the acoustic peaks (cf. §5.10).
3. The **integrated Sachs-Wolfe (ISW)** term records the cumulative effect of time-varying metric potentials along the photon's path from LSS to today. It vanishes in pure matter domination (where Φ and Ψ are constant) but is sensitive to any departure from MD, including the early radiation-to-matter transition (the “early ISW”) and the late-time onset of dark-energy domination (the “late ISW”), as well as nonlinear evolution.

6.3 The Sachs-Wolfe Plateau

On large angular scales (corresponding to modes with $k\chi_* \lesssim 1$, where $\chi_* \simeq 14,000$ Mpc is the comoving distance to last scattering), the SW term in (285) dominates the Doppler and ISW contributions, and we can approximate

$$\frac{\delta T}{T}(\hat{n}) \approx \left[\frac{1}{4}\delta_\gamma + \Psi\right]_* . \quad (286)$$

For superhorizon modes at last scattering — the regime relevant for the lowest multipoles $\ell \lesssim 100$ — the adiabatic initial conditions from inflation (cf. §5.6) imply $\delta_\gamma = \frac{4}{3}\delta_m$, and from the integrated continuity equation through the matter-radiation transition, $\delta_m = -2\Psi$ on superhorizon scales in matter domination (one can verify this using the gauge-invariant comoving density contrast together with the conservation of $\mathcal{R}$). Hence $\delta_\gamma = -\frac{8}{3}\Psi$, and the SW combination becomes

$$\frac{\delta T}{T}(\hat{n}) \approx \left[\frac{1}{4} \cdot \left(-\frac{8}{3}\Psi\right) + \Psi\right]_* = \frac{1}{3}\Psi_* . \quad (287)$$

Substituting $\Psi_* = \frac{3}{5}\mathcal{R}_i$ from the conservation of $\mathcal{R}$ on superhorizon scales (cf. §5.6), we arrive at the **Sachs-Wolfe limit**:

$$\boxed{\frac{\delta T}{T}(\hat{n}) \approx \frac{1}{5}\mathcal{R}_i(\vec{x}_*) \quad (k\chi_* \ll 1),} \quad (288)$$

where $\vec{x}_* = -\hat{n}\chi_*$ is the LSS position along the line of sight. This is a remarkable result: the large-angle CMB temperature anisotropy is one-fifth of the primordial curvature perturbation, directly connecting the inflationary initial conditions to an observable temperature pattern on the sky.

To predict the angular power spectrum, we Fourier-transform $\mathcal{R}_i$ and integrate against spherical harmonics. Using the plane-wave expansion $e^{i\vec{k}\cdot\vec{r}} = 4\pi \sum_{\ell m} i^\ell j_\ell(kr) Y_{\ell m}(\hat{r}) Y_{\ell m}^*(\hat{k})$ (where j_ℓ are spherical Bessel functions) and the orthonormality of $Y_{\ell m}$, one finds

$$a_{\ell m} = \frac{4\pi(-i)^\ell}{5} \int \frac{d^3k}{(2\pi)^3} j_\ell(k\chi_*) Y_{\ell m}^*(\hat{k}) \mathcal{R}_i(\vec{k}). \quad (289)$$

Taking expectation values, using the primordial curvature power spectrum $\langle \mathcal{R}_i(\vec{k}) \mathcal{R}_i^*(\vec{k}') \rangle = (2\pi)^3 \delta_D(\vec{k} - \vec{k}') (2\pi^2/k^3) \Delta_{\mathcal{R}}^2(k)$, and integrating over the angular variables, the cross-terms vanish and we obtain

$$C_\ell = \frac{4\pi}{25} \int_0^\infty \frac{dk}{k} \Delta_{\mathcal{R}}^2(k) j_\ell^2(k\chi_*). \quad (290)$$

For a scale-invariant primordial spectrum, $\Delta_{\mathcal{R}}^2(k) = A_s$ (constant), the remaining integral is a standard Bessel identity:

$$\int_0^\infty \frac{dk}{k} j_\ell^2(k\chi_*) = \frac{1}{2\ell(\ell+1)}, \quad (291)$$

giving the celebrated **Sachs-Wolfe plateau**:

$$\boxed{\frac{\ell(\ell+1)C_\ell}{2\pi} = \frac{A_s}{25} \quad (\ell \lesssim 100).} \quad (292)$$

The fact that $\ell(\ell+1)C_\ell$ is approximately *independent* of ℓ at low multipoles — a flat plateau in the canonical $\ell(\ell+1)C_\ell/(2\pi)$ vs ℓ plot (see Fig. 13) — is a direct prediction of (a) scale-invariant initial conditions from inflation and (b) the dominance of the SW combination on large scales. The measured amplitude of the plateau pins down the primordial amplitude to $A_s \simeq 2.1 \times 10^{-9}$.⁸¹

6.4 Beyond the Sachs-Wolfe Limit

At multipoles $\ell \gtrsim 100$, corresponding to modes that have entered the horizon by last scattering, the simple Sachs-Wolfe formula must be extended to include the full subhorizon dynamics: the acoustic oscillations of the photon-baryon fluid, the Doppler contribution from the fluid velocity, the integrated Sachs-Wolfe effect, and the diffusive damping of small-scale power. Solving for all of these from scratch for a general mode is a formidable task, but there is one feature that renders the problem tractable: *all* of the linear evolution we have studied is linear in the primordial curvature perturbation. As a result, every quantity entering the master formula (285) is proportional to $\mathcal{R}_i(\vec{k})$, and we may capture the entire intervening physics by a set of multiplicative **transfer functions**. For the Sachs-Wolfe combination we write

$$\left[\frac{1}{4}\delta_\gamma + \Psi\right](\vec{k}, \eta_*) \equiv T_{\text{SW}}(k, \eta_*) \mathcal{R}_i(\vec{k}), \quad (293)$$

⁸¹The (near-) scale-invariance is itself a non-trivial prediction: a primordial tilt $\Delta_{\mathcal{R}}^2 \propto k^{n_s-1}$ gives $\ell(\ell+1)C_\ell \propto \ell^{n_s-1}$, and the observed scaling $n_s \simeq 0.97$ (slightly red-tilted) is one of the central confirmations of slow-roll inflation.

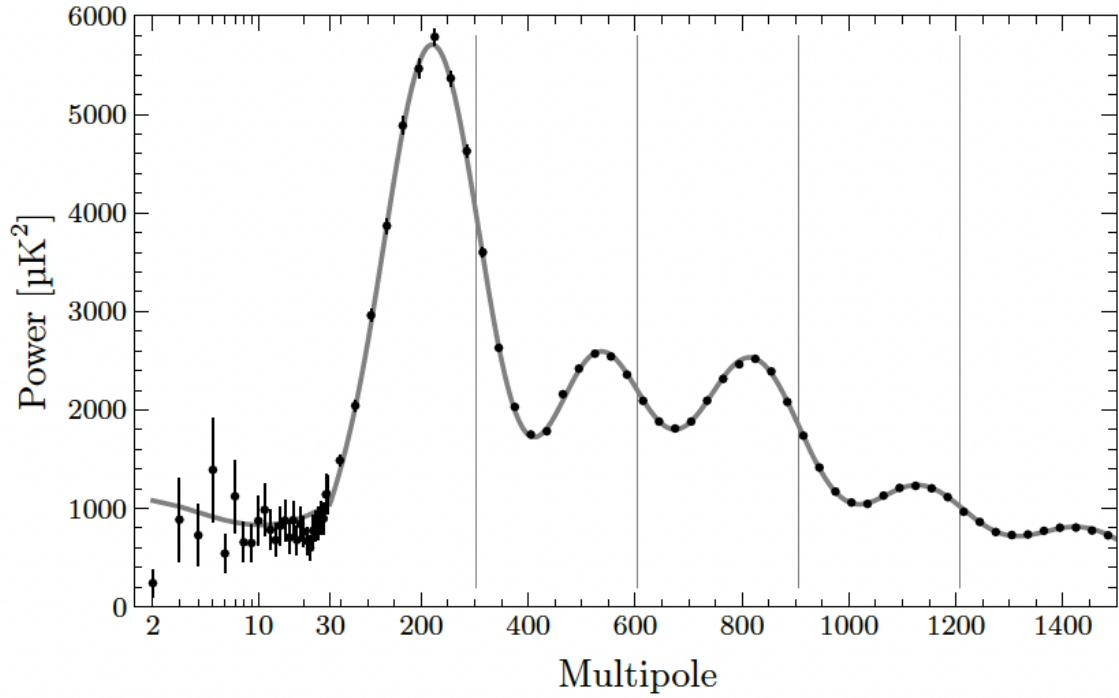


Figure 13: Plot of the CMB power spectrum, plus observational data from *Planck*. The flat region at low- ℓ is the Sachs-Wolfe plateau, which traces the primordial power spectrum amplitude. At higher- ℓ , we see the acoustic peaks, which are damped out by Silk damping. Note the asymmetry between odd and even peaks due to baryon loading. The Λ CDM model provides an exquisite fit to the data. Reproduced from D. Baumann, *Cosmology* (Cambridge University Press, 2022).

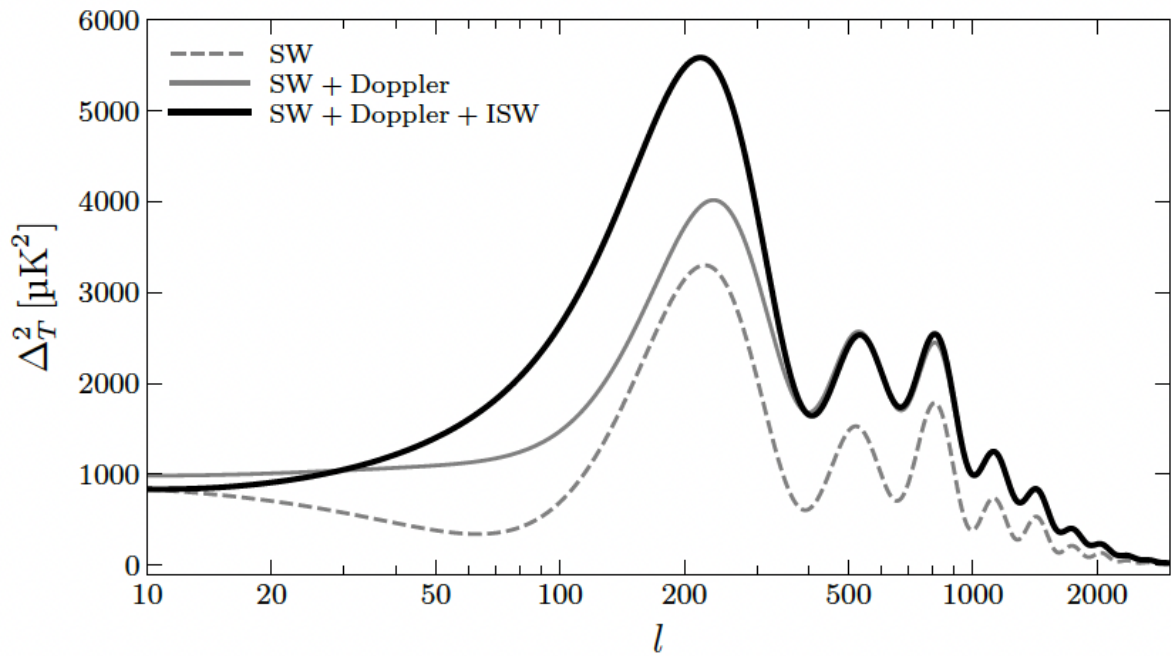


Figure 14: Contributions to the CMB power spectrum: Sachs Wolfe, Doppler, and integrated Sachs Wolfe. On large-scales, the Doppler effect is small, but becomes important at larger ℓ . The ISW effect adds power on large-scales around the first peak.

where $T_{\text{SW}}(k, \eta_*)$ encodes all of the linear physics — horizon entry, acoustic oscillations, baryon loading, potential decay — between the inflationary initial conditions and last scattering. The Doppler and ISW terms are treated identically. In Fourier space the line-of-sight velocity becomes $(\hat{n} \cdot \vec{v}_b)_* \rightarrow i \hat{k} \cdot \hat{n} T_{\text{Doppler}}(k, \eta_*) \mathcal{R}_i(\vec{k})$, while the ISW integral is written as $\int d\eta (\Psi' + \Phi') = \int d\eta T_{\text{ISW}}(k, \eta) \mathcal{R}_i(\vec{k}) e^{i\vec{k} \cdot \hat{n} \chi}$. In practice these transfer functions are obtained by solving the coupled Einstein-Boltzmann equations numerically for each mode, but their qualitative behaviour can be read off from the asymptotic solutions we have already derived.

6.4.1 Projection onto the Sky

To see how these transfer functions imprint themselves on the observed anisotropies, consider the temperature field with the ISW term temporarily dropped for clarity,

$$\frac{\delta T}{T}(\hat{n}) = \int \frac{d^3 k}{(2\pi)^3} e^{i\vec{k} \cdot \hat{n} \chi_*} \left\{ T_{\text{SW}}(k, \eta_*) - i \hat{k} \cdot \hat{n} T_{\text{Doppler}}(k, \eta_*) \right\} \mathcal{R}_i(\vec{k}). \quad (294)$$

The Doppler factor of $\hat{k} \cdot \hat{n}$ can be traded for a derivative of the plane wave using $i \hat{k} \cdot \hat{n} e^{i\vec{k} \cdot \hat{n} \chi_*} = k^{-1} \partial_{\chi_*} e^{i\vec{k} \cdot \hat{n} \chi_*}$, so that after projecting onto spherical harmonics with the plane-wave expansion (cf. §6.3) the multipole coefficients become

$$a_{\ell m} = 4\pi (-i)^\ell \int \frac{d^3 k}{(2\pi)^3} Y_{\ell m}^*(\hat{k}) \left\{ T_{\text{SW}}(k, \eta_*) j_\ell(k\chi_*) - T_{\text{Doppler}}(k, \eta_*) j'_\ell(k\chi_*) \right\} \mathcal{R}_i(\vec{k}). \quad (295)$$

Taking the variance and using the primordial spectrum, we obtain the general angular power spectrum

$$\boxed{C_\ell = 4\pi \int_0^\infty \frac{dk}{k} \Delta_{\mathcal{R}}^2(k) \left\{ T_{\text{SW}}(k, \eta_*) j_\ell(k\chi_*) - T_{\text{Doppler}}(k, \eta_*) j'_\ell(k\chi_*) \right\}^2 + (\text{ISW})}, \quad (296)$$

where the Doppler term enters through the derivative j'_ℓ of the spherical Bessel function (reflecting its $\pi/2$ phase offset), and the ISW term contributes an additional line-of-sight integral against j_ℓ . This expression makes the structure of CMB prediction transparent: the spectrum is built from the primordial input $\Delta_{\mathcal{R}}^2(k) \sim A_s k^{n_s-1}$, the **transfer functions** $T(k)$ that describe linear evolution, and the **projection functions** $j_\ell(k\chi_*)$ that map three-dimensional wavenumbers onto angular scales. Since $j_\ell(k\chi_*)$ is sharply peaked at $\ell \simeq k\chi_*$, a feature at wavenumber k appears at multipole $\ell \sim k\chi_*$, i.e. at angular scale $\theta \sim \pi/\ell \sim \pi/(k\chi_*)$. Schematically, the canonical plotting combination is therefore

$$\frac{\ell(\ell+1)}{2\pi} C_\ell \sim \Delta_{\mathcal{R}}^2(k) [T_{\text{SW}}^2(k) + T_{\text{Doppler}}^2(k)]_{k=\ell/\chi_*}, \quad (297)$$

so that the angular power spectrum is essentially a projected image of the squared transfer functions. We show the full solution in Fig. 14.

6.4.2 The Sachs-Wolfe Transfer Function

To estimate T_{SW} and T_{Doppler} , we recall the asymptotic evolution of the potential and photon perturbations derived in §5.7 and §5.10. For the gravitational potential, the radiation-era solution is

$$\frac{\Phi(\vec{k}, \eta)}{\mathcal{R}_i(\vec{k})} = 2 \frac{\sin \varphi - \varphi \cos \varphi}{\varphi^3}, \quad \varphi \equiv \frac{k\eta}{\sqrt{3}}, \quad (298)$$

which tends to $\frac{2}{3}$ on superhorizon scales and decays as $-6 \cos(k\eta/\sqrt{3})/(k\eta)^2$ once the mode enters the horizon; during matter domination it freezes to the constant $\Phi = \frac{2}{3} \cdot \frac{9}{10} \mathcal{R}_i = \frac{3}{5} \mathcal{R}_i$. For the photons, the radiation-era density contrast obeys $\delta_\gamma/\mathcal{R}_i = -\frac{4}{3}$ on superhorizon scales and $4 \cos(k\eta/\sqrt{3})$ once subhorizon, while in matter domination the tight-coupling sound-wave solution of §5.10 reads

$$\delta_\gamma(\vec{k}, \eta) \simeq \frac{C(\vec{k}) \mathcal{R}_i}{(1+R)^{1/4}} \cos(kr_s(\eta)) - \underbrace{4(1+R)\Phi}_{\text{slow mode}}, \quad (299)$$

where the second (slowly-varying) piece is the gravitational forcing term that was neglected in the leading WKB analysis. The Sachs-Wolfe source is the combination $S \equiv \frac{1}{4}\delta_\gamma + \Phi$.

On **large scales** ($k < k_{\text{eq}}$), the relevant modes enter the horizon during matter domination, where $\Phi = \frac{3}{5} \mathcal{R}_i$ is constant. Substituting the sound-wave solution and dropping the subdominant $(1+R)^{1/4}$ variation of the forcing term in the limit $R \ll 1$, the source becomes

$$S(\vec{k}, \eta_*) \simeq \frac{1}{5} \mathcal{R}_i(\vec{k}) \left[\frac{C}{(1+R)^{1/4}} \cos(kr_s(\eta_*)) - 3R \right], \quad (300)$$

so that

$$T_{\text{SW}}(k \ll k_{\text{eq}}) \simeq \frac{1}{5} \left[\frac{C}{(1+R)^{1/4}} \cos(kr_s(\eta_*)) - 3R \right]. \quad (301)$$

This transfer function captures two key features of the acoustic peaks. First, the $\cos(kr_s)$ dependence produces a series of **oscillations** — the imprint of the sound waves in the photon-baryon plasma — with extrema at

$$kr_s(\eta_*) = n\pi \quad (n \in \mathbb{Z}) \implies \ell \sim n\pi \frac{\chi_*}{r_s(\eta_*)} \sim 300, 600, 900, \dots \quad (302)$$

giving the characteristic peak spacing of the CMB. Second, the constant offset $-3R$ makes the oscillations **asymmetric**: because baryon loading ($R \neq 0$) enhances the gravitational compressions of the fluid but not the rarefactions (the photon pressure resists compression but does not aid expansion), odd “compression” peaks are boosted relative to even “rarefaction” peaks. The amplitude of this odd-even modulation (which is clearly seen in Fig. 14) is a sensitive probe of the baryon density ω_b .

On **small scales** ($k > k_{\text{eq}}$, $\ell \gtrsim 300$), the relevant modes enter the horizon during radiation domination, where the potentials decay away (as above). In this regime the Sachs-Wolfe source reduces to the bare oscillation

$$S(\vec{k}, \eta_*) \equiv \frac{1}{4}\delta_\gamma + \Phi \simeq \mathcal{R}_i(\vec{k}) \cos(kr_s(\eta_*)), \quad T_{\text{SW}}(k \gg k_{\text{eq}}) \simeq \cos(kr_s), \quad (303)$$

where the latter is only a very rough estimate. Notably, the decay of Φ during radiation domination drives the oscillations resonantly, boosting the overall amplitude of the small-scale peaks relative to the large-scale plateau — the so-called “radiation driving” effect.

6.4.3 The Doppler and ISW Contributions

The picture above describes the dominant Sachs-Wolfe contribution, but two corrections complete the story. The first is the **Doppler term**. From the continuity equation $\delta'_\gamma = \frac{k}{3} v_\gamma$ (so $T_{\text{Doppler}} \sim \frac{3}{4} T'_{\text{SW}}/k$), the velocity perturbation is $\pi/2$ out of phase with the density, giving

$$T_{\text{Doppler}}(k) \sim \frac{\sin(kr_s)}{(1+R)^{3/4}}. \quad (304)$$

This produces oscillations in quadrature with the SW peaks, partially filling in the troughs; it is negligible on large scales ($k < r_s^{-1} \approx (145 \text{ Mpc})^{-1}$) but becomes relevant near the acoustic peaks.

The second correction is the **integrated Sachs-Wolfe term**, sourced by time-varying potentials along the line of sight. As discussed above, Φ and Ψ are constant in pure matter domination, so the ISW effect probes departures from MD. It receives an *early* contribution from the residual potential decay around the radiation-to-matter transition (which boosts power near the first peak, since $\Phi' \rightarrow 0$ only after equality), a *late* contribution from the onset of dark-energy domination (which adds power on the very largest scales, near the $\sim 1^\circ$ peak, projected through $j_\ell(k(\eta_0 - \eta))$ with smaller $\eta_0 - \eta$), and a nonlinear contribution from the evolution of Φ in collapsed structures (the Rees-Sciama effect). Together, the SW, Doppler, and ISW transfer functions — supplemented by the damping physics discussed next — reproduce the full acoustic peak structure of C_ℓ in the range $\ell \sim 100$ –1500, one of the most precisely-measured and theoretically-understood signals in all of physics (see Fig. 13).

6.5 Diffusion and the Damping Tail

The transfer functions above predict acoustic peaks of undiminished height out to arbitrarily small scales. In reality, the peaks are exponentially suppressed at high ℓ by two related effects that smear out small-scale anisotropies: *Silk damping* and *Landau damping* (as seen in Fig. 14).

6.5.1 Silk Damping

The tight coupling between photons and baryons is never perfect: photons **diffuse** out of overdense regions by random-walking between scattering events. With mean free path $\ell_\gamma = 1/(a n_e \sigma_T)$, a photon undergoes $N = \Delta\eta/\ell_\gamma$ scatterings in a conformal time $\Delta\eta$, so the mean-squared distance travelled is the random-walk result

$$\langle \Delta r^2 \rangle = N \ell_\gamma^2 = \ell_\gamma \Delta\eta. \quad (305)$$

Integrating over the history from $\eta = 0$ to last scattering, the comoving **diffusion length** is

$$r_D^2(\eta_*) \sim \int_0^{\eta_*} \frac{d\eta}{a n_e \sigma_T}, \quad (306)$$

which evaluates to $r_D \simeq 7 \text{ Mpc}$. Density perturbations on scales smaller than r_D are washed out as photons diffuse from hot to cold regions, exponentially damping the transfer function,

$$T_{\text{SW}}(k) \longrightarrow e^{-(k/k_{\text{Silk}})^2} T_{\text{SW}}(k), \quad k_{\text{Silk}} \sim 0.2 \text{ Mpc}^{-1}. \quad (307)$$

This **Silk damping** suppresses the acoustic peaks at $\ell \gtrsim 1400$. Physically, the effect is more than a simple random walk: the imperfect coupling allows a *slip* between the electron and photon fluids, sourced by thermal conduction and photon viscosity (anisotropic stress). A careful treatment adds these viscosity and slip terms to the Boltzmann hierarchy, but the qualitative Gaussian damping envelope is captured by the diffusion estimate above.

6.5.2 Landau Damping

A second, closely-related effect arises because photon decoupling (and recombination) is *not* instantaneous. Rather than being emitted from an infinitely thin surface, the CMB photons we observe were last scattered over a finite *shell* in conformal time, described by the **visibility function**

$$g(\eta) = \frac{d}{d\eta} e^{-\tau(\eta)}, \quad (308)$$

which is sharply peaked around $z \simeq 1090$ but has nonzero width. Including this finite thickness replaces the sharp projection $j_\ell(k\chi_*)$ in the anisotropy integral by a weighted average over the shell,

$$\left. \frac{\Delta T}{T} \right|_{\text{SW}} \sim \int \frac{d^3 k}{(2\pi)^3} \cdots \left[\frac{1}{4} \delta_\gamma + \Psi \right] j_\ell(k\chi) \int d\eta g(\eta), \quad (309)$$

which averages over modes of different phase and thereby **damps** fluctuations on scales below the shell thickness ($\lesssim 5$ Mpc). This is known as **Landau damping**. Combining the diffusion and finite-thickness effects yields a total damping scale $k_D \simeq (8.8 \text{ Mpc})^{-1}$, which sets the exponential cutoff of the observed damping tail.

6.6 Summary So Far

Assembling all of these ingredients reproduces the familiar shape of the CMB temperature power spectrum: a flat Sachs-Wolfe plateau at low ℓ , the harmonic series of acoustic peaks at $\ell \sim 300, 600, 900, \dots$ derived above, and an exponentially-falling damping tail at high ℓ . Superimposed on this are the secondary modulations: the ISW effect dominates at the lowest multipoles and boosts the first peak, the Doppler term contributes peaks $\pi/2$ out of phase that partially fill the troughs, and baryon loading sets the relative heights of odd and even peaks.

6.7 Reionization

There is one further important late-time modification to the CMB. Around $z \sim 6$ –10, the first stars and galaxies **reionized** the neutral hydrogen of the intergalactic medium ($\text{H} \rightarrow \text{H}^+$), repopulating the Universe with free electrons off which CMB photons can again Compton scatter, albeit with small probability. The integrated probability is set by the **reionization optical depth**,

$$\tau_{\text{reio}} \sim \int_0^{\eta_*} n_e \sigma_T a d\chi, \quad (310)$$

which depends on the free-electron fraction along the line of sight. A fraction $e^{-\tau}$ of the primary photons pass through unscattered, while the rest are scattered out of (and an isotropic background scattered into) the line of sight, so that the observed temperature is

$$T(\hat{n}) = \underbrace{(\bar{T} + \Delta T)(\hat{n}) e^{-\tau}}_{\text{scattered out}} + \underbrace{\bar{T}(1 - e^{-\tau})}_{\text{scattered in}}. \quad (311)$$

Because the rescattered photons arrive with an averaged temperature drawn from many directions, the net effect on the anisotropies is a uniform suppression $\Delta T/T \rightarrow e^{-\tau} \Delta T/T$, and hence

$$\boxed{C_\ell \longrightarrow e^{-2\tau} C_\ell \quad (\ell \gtrsim 10).} \quad (312)$$

The restriction $\ell \gtrsim 10$ applies because reionization occurs late, so its effect is confined to scales inside the horizon at that epoch. Two important consequences follow. First, since the SW plateau amplitude scales as $A_s e^{-2\tau}$, there is a **degeneracy** between the primordial amplitude A_s and the optical depth τ ; this is broken by large-scale polarization, which is *generated* by reionization rather than suppressed by it (see below). Second, reionization need not be spatially uniform — it proceeds through ionized bubbles — so τ can be *patchy*, a subdominant effect that is difficult to detect but carries information about the epoch of reionization.

6.8 The Impact of Λ CDM Parameters

The power of the CMB lies in the fact that the rich structure described above depends sensitively on the parameters of the cosmological model, allowing them to be measured with exquisite precision. The standard Λ CDM model is specified by six parameters, which map onto distinct physical sectors:

$$\underbrace{A_s, n_s}_{\text{inflation}}, \quad \underbrace{\omega_b = \Omega_b h^2, \omega_c = \Omega_c h^2}_{\text{densities}}, \quad \underbrace{\Omega_\Lambda \text{ (or } h = H_0/100)}_{\text{dark energy}}, \quad \underbrace{\tau}_{\text{reionization}}.$$

These imprint themselves on the spectrum through two main observables: the *positions* and the *heights* of the acoustic peaks.

Peak positions. The peaks sit at the angular scale of the sound horizon,

$$\Theta_s = \frac{r_s(\eta_*)}{d_A(\eta_*)}, \quad (313)$$

where the numerator (the sound horizon at last scattering) depends on the densities ω_b, ω_c that set the expansion rate and sound speed before recombination, and the denominator (the angular-diameter distance to the LSS) depends on the late-time expansion through $\Omega_\Lambda, H_0, \Omega_m$. Measuring Θ_s thus ties together early- and late-Universe physics.

Peak heights. The relative heights of the peaks are governed by several effects: diffusion damping (set by ω_b, ω_m , which control the mean free path and the age of the Universe at decoupling); the early ISW boost of the first peak (set by ω_m, ω_b through the proximity of equality to recombination); baryon loading (the odd-even modulation $\propto (1+R)^{-1/4} \cos -3R$, set by ω_b); and radiation driving (which raises the overall small-scale amplitude through the decay of Φ in radiation domination).

The individual parameter dependences can then be summarized as follows:

1. A_s rescales the entire spectrum (with amplitude $\propto A_s e^{-2\tau}$), while n_s tilts it, raising or lowering large scales relative to small.
2. Increasing ω_b changes the sound horizon r_s (and hence Θ_s), enhances the odd “compression” peaks relative to the even ones, and reduces diffusion damping (stronger coupling pushes the damping tail to higher ℓ).
3. Increasing ω_m (and hence, at fixed $\omega_b + \omega_c$, the expansion rate and h) shifts r_s and χ_* (and thus Θ_s), moves matter-radiation equality earlier so that radiation driving is weaker (lowering the high- ℓ peaks), reduces the early ISW (lowering the first peak), and alters the diffusion scale.
4. Increasing Ω_Λ changes the distance to the CMB (degenerate with spatial curvature for a flat $k=0$ universe) and adds late-time ISW power on the largest scales.
5. Additional relativistic species, parameterized by N_{eff} via $\rho_{\text{rad}} = [1 + \frac{7}{8}(\frac{4}{11})^{4/3} N_{\text{eff}}] \rho_\gamma$, increase the expansion rate at last scattering (changing $H(\eta_*)$ and hence Θ_s) and enhance diffusion damping; N_{eff} is measured precisely by tracking the damping-tail amplitude at fixed Θ_s .

6.9 Additional Topics

The discussion so far has focused on the scalar temperature anisotropies, which are the richest and best-measured CMB observable. We close with a brief survey of several further effects — polarization, tensor modes, secondary anisotropies, and spectral distortions — each of which opens an additional window onto the early and late Universe.

6.9.1 Polarization

The Compton scattering that couples photons and electrons also **polarizes** the CMB. Because an electromagnetic wave has its polarization transverse to its propagation, radiation incident along $\hat{x}$ and scattered into $\hat{z}$ retains only the $\hat{y}$ polarization, generating *linear* polarization. If the radiation field incident on a given electron were isotropic, the contributions from all incoming directions would cancel and no net polarization would result. Polarization is therefore sourced only by an **anisotropic** — specifically *quadrupolar* — incident intensity, weighted by the differential cross section $d\sigma/d\Omega = (3\sigma_T/8\pi)|\hat{\epsilon} \cdot \hat{\epsilon}'|^2$ for $\hat{\epsilon} \rightarrow \hat{\epsilon}'$. The resulting polarization pattern is thus proportional to the temperature quadrupole at last scattering.

Formally, the polarization is described by the Stokes parameters $Q \propto |E_x|^2 - |E_y|^2$ and $U \propto 2|E_x||E_y|\cos\phi$, which specify the $+$ and $\times$ orientations. It is more convenient to decompose these into the coordinate-invariant ***E*-modes** (curl-free, $\sim \partial_a \partial_b P^{ab}$) and ***B*-modes** (divergence-free, $\sim \epsilon_{abc} P^{ab} \partial_c$). The polarization sourced by scalar perturbations is purely *E*-mode, with orientation parallel or perpendicular to $\hat{k}$. Estimating the quadrupole generated by the velocity gradient across a mean free path,

$$\left. \frac{\Delta T}{T} \right|_{\text{scatter}} = \hat{n} \cdot (\vec{v}_b(\vec{x}) - \vec{v}_b(\vec{x}_0)) \sim \hat{n}_i \hat{n}_j \ell_p \partial_i v_j^b \sim -\ell_p v_b (\hat{k} \cdot \hat{n})^2, \quad (314)$$

for scattering at $\vec{x}_0$ from $\vec{x} = \vec{x}_0 + \ell_p \hat{n}$, we see that the polarization is generated by diffusion (requiring $k \gtrsim \ell_p^{-1}$) and is sourced by velocities, so it is $\pi/2$ out of phase with the temperature peaks. Crucially, the quadrupole is *erased* in perfect tight coupling, so polarization probes precisely the imperfections of the photon-baryon fluid at last scattering. This produces both *EE* and *TE* power spectra as shown in Fig. 15.

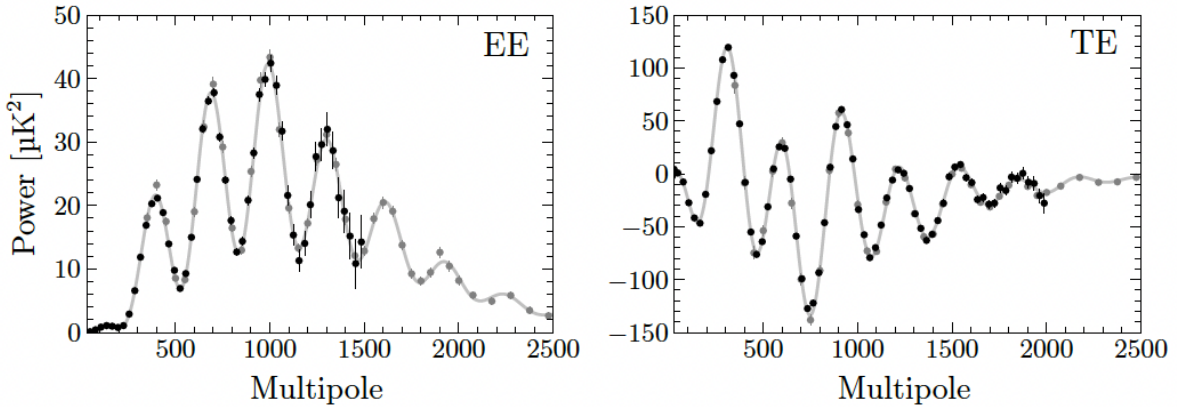


Figure 15: Power spectrum of CMB *E*-mode polarization (left) and the cross-spectrum with temperature anisotropies (right). We plot $\ell(\ell+1)C_\ell/(2\pi)$ in both cases. The data points are from Planck (dark) and ACT (light). The model is predicted from the temperature spectrum (and not a fit to the *E*-mode data). Reproduced from D. Baumann, *Cosmology* (Cambridge University Press, 2022).

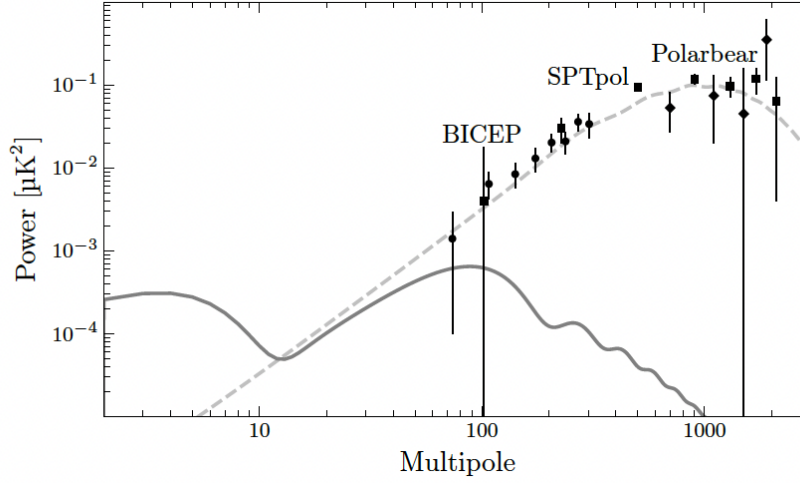


Figure 16: Power spectrum of CMB B-mode polarization, plus data from recent surveys. The solid curve shows the prediction from primordial tensors (*i.e.* inflationary gravitational waves) with $r = 0.01$, while the dotted line shows contributions from gravitational lensing. Reproduced from D. Baumann, *Cosmology* (Cambridge University Press, 2022).

6.9.2 Tensor Modes

So far we have considered only scalar fluctuations, but inflation also generates a stochastic background of **gravitational waves**. These distort the metric via $\delta g_{ij} \supset a^2 h_{ij}$ (with two polarizations h_+ , $h_\times$) and modify photon propagation through an extra term in the redshifting equation,

$$\frac{d \log(aT)}{d\eta} = -\frac{d\Psi}{d\eta} + \Phi' + \Psi' - \frac{1}{2} h'_{ij} \hat{n}^i \hat{n}^j, \quad (315)$$

so that passing gravitational waves induce a temperature *quadrupole* on the CMB. Unlike scalar perturbations, tensors generate *both* E - and B -mode polarization (the two tensor degrees of freedom), with oscillations at $\ell \simeq n\pi\chi_*/\eta_*$ that are *not* acoustic in origin (see Fig. 16). The B -mode signature is the cleanest probe of the primordial tensor-to-scalar ratio r and hence of the energy scale of inflation.

6.9.3 Secondary Anisotropies

Beyond the primary anisotropies imprinted at last scattering, many processes modulate the CMB photons as they travel from the LSS to us. These **secondaries** carry information about the late-time Universe, as sketched in Fig. 17.

CMB lensing. The *positions* of CMB photons on the sky are deflected by gravitational lensing from intervening structure, remapping the observed direction $\hat{n}_{\text{obs}}$ relative to the emitted $\hat{n}_{\text{emit}}$. From the spatial geodesic equation, $d\hat{n}^i/d\eta = -\nabla_\perp^i(\Phi + \Psi)$, the net deflection is the line-of-sight integral

$$\vec{\alpha} = \int_0^{\chi_*} \frac{(\chi_* - \chi)}{\chi \chi_*} \nabla_{\hat{n}}(\Phi + \Psi) d\chi \equiv \nabla \phi_{\text{lens}}, \quad (316)$$

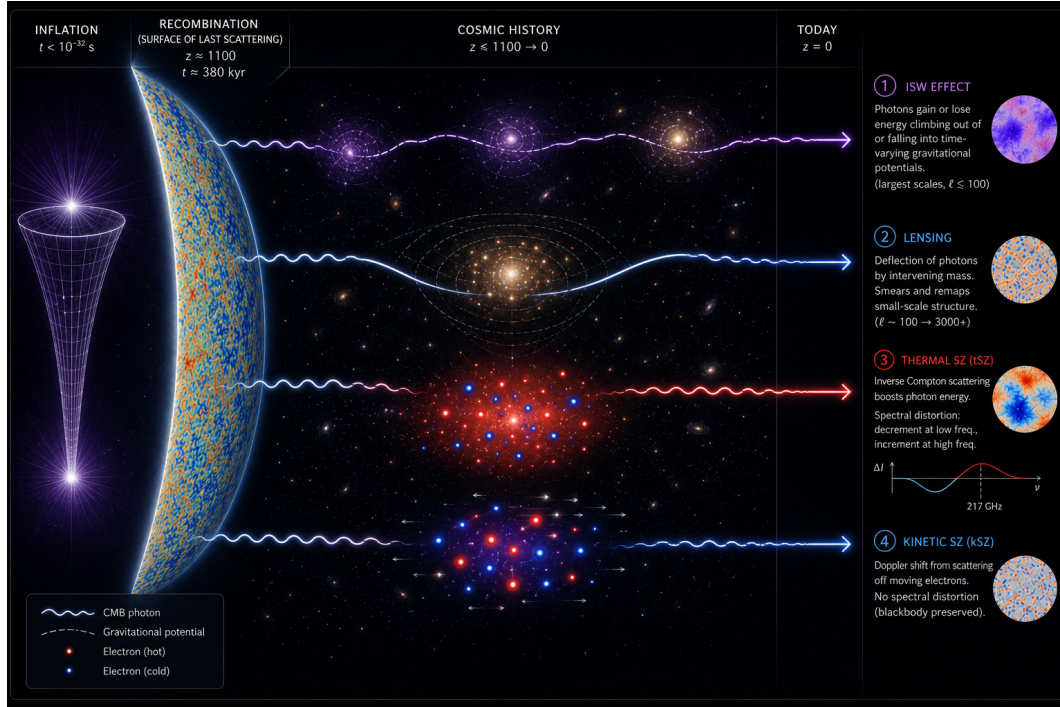


Figure 17: Sketch of various CMB secondaries. The main secondaries are the ISW effect (from time-varying potentials), CMB lensing (from the geodesic distortion induced by large-scale structure), and the thermal and kinematic Sunyaev-Zel'dovich signals. Respectively, these carry information about dark energy (and matter-radiation equality), the distribution of dark matter, electron pressure, and electron momenta, and have all been detected at high signal-to-noise. AI-generated image (ChatGPT).

so that $\delta T(\hat{n}) \rightarrow \delta T(\hat{n} + \vec{\alpha}) = \delta T(\hat{n}) + \nabla \phi_{\text{lens}} \cdot \nabla \delta T$. Lensing smooths the acoustic peaks, transfers E -modes into B -modes, and sources additional small-scale power. Most importantly, it renders the CMB slightly non-Gaussian and statistically anisotropic: the off-diagonal two-point function $\langle a_{\ell m} a_{\ell' m'}^* \rangle$ no longer vanishes but picks up a term $\propto \phi_{(\ell-\ell')(m-m')}$, equivalent to a four-point function $\langle TTTT \rangle \sim \langle TT \rangle^2 \langle \phi \phi \rangle$. This makes lensing a powerful probe of the integrated potential $\Phi + \Psi$ and hence of late-Universe physics. This additionally contributes to polarization, and is a major contaminant to searches for primordial gravitational waves using B -modes (see Fig. 16).

Moving lens. A lens moving transversely to the line of sight produces a time-varying potential, $\dot{\Phi} = -\vec{V}_\perp \cdot \vec{\nabla}_\perp \Phi$, which via the ISW mechanism imprints a *dipole*-like temperature pattern,

$$\frac{\Delta T}{T} = -2 \int d\chi \vec{V}_\perp \cdot \vec{\nabla}_\perp (\Phi + \Psi). \quad (317)$$

This moving-lens effect has recently been detected by combining CMB maps with galaxy surveys to predict the transverse velocities $\vec{V}_\perp$.

Shapiro time delay. Just as lensing distorts the *direction* of photons, the same potentials delay their arrival *time*, $\delta\eta = -\int d\chi (\Phi + \Psi)$, shifting the effective distance to last scattering, $\chi \rightarrow \chi + \delta\chi$. The induced temperature change $\delta\chi \partial_\chi T$ is tiny, since the CMB is emitted from a thin shell, making this effect very hard to observe.

Patchy screening. Recall that reionization suppresses the anisotropies by $e^{-\tau}$. Since τ is in reality spatially varying (reionization proceeds in bubbles), the screening is *patchy*: $\delta T_{\text{obs}} \simeq T_{\text{obs}} - \delta\tau(\hat{n}) T_{\text{obs}}(\hat{n})$, which — like lensing — modulates the CMB and induces non-Gaussianity. This effect has very recently been close to detection.

Sunyaev-Zeldovich effects. CMB photons can scatter off free electrons in the gas of the late Universe, producing the **Sunyaev-Zeldovich (SZ)** effects. In the *thermal* SZ effect, photons scatter off *hot* cluster gas, conserving photon number but transferring energy, which distorts the spectrum away from a blackbody:

$$\frac{\Delta T}{T} \approx \left(\underbrace{\int dl n_e \sigma_T}_{\text{scattering prob.}} \underbrace{\frac{k_B T_e}{m_e c^2}}_{\text{energy transfer}} \right) \times (\text{frequency factor}), \quad (318)$$

which depends on the integrated electron pressure and produces frequency-dependent hot-spots around galaxy clusters — a powerful probe of cluster gas profiles. In the *kinetic* SZ effect, photons scatter off the *bulk motion* of electrons via a Doppler shift, with no frequency change (so the spectrum remains a blackbody):

$$\frac{\Delta T}{T} = - \underbrace{\int dl n_e \sigma_T}_{\text{scattering prob.}} \underbrace{(\vec{v}_e \cdot \hat{n})}_{\text{LOS velocity}}, \quad (319)$$

which traces the large-scale velocity field and the electron distribution. Both SZ effects, along with lensing, the ISW effect, and patchy screening, can be organized by their characteristic line-of-sight kernels — $\int d\chi \dot{\Phi}$ (ISW), $\nabla \Phi \cdot \nabla T$ (lensing), $-\delta\tau T$ (screening), $\int n_e v_\parallel$ (kSZ), and $\int n_e T_e$ (tSZ). Numerous further secondaries exist, including polarized, rotational and relativistic SZ, Rayleigh scattering, the Rees-Sciama effect, cosmic birefringence, and post-Born lensing.

6.9.4 Spectral Distortions

Finally, we have throughout assumed the CMB to be a perfect blackbody, with photon occupation number $n(\nu, T) = (e^{h\nu/k_B T} - 1)^{-1}$. Departures from this — **spectral distortions** — arise whenever energy is injected into the photons or redistributed among them; note that adding energy distorts the spectrum unless photons are also created. The phenomenology depends on redshift, since it is set by the efficiency of the photon-number-changing and energy-redistributing interactions:

- At very high redshift ($z > 10^8$), all relevant processes — double Compton scattering and bremsstrahlung (which create photons) together with Compton scattering (which redistributes energy) — are efficient, so any injected energy is rapidly thermalized into a new blackbody at a slightly higher temperature, leaving *no* distortion.
- For $z \sim 10^5$ – 10^6 , photon-creating processes freeze out but Compton scattering still efficiently redistributes energy at fixed photon number. Energy injection then drives the occupation number towards a Bose-Einstein form with a nonzero chemical potential, $n(x) = (e^{x+\mu} - 1)^{-1}$, yielding a μ -**distortion**. Sources at these redshifts include the dissipation of acoustic waves (Silk damping), decaying particles, and primordial magnetic fields; the standard Λ CDM prediction is $\mu \sim 10^{-8}$.
- For $z \lesssim 10^4$, even Compton scattering becomes inefficient, and energy injection produces a y -**distortion** that probes lower-energy processes such as scattering off hot electrons; the Λ CDM prediction is $y \sim 10^{-6}$.

Spectral distortions offer a genuinely new view of the thermal history, but they remain undetected: the best constraints to date come from the COBE/FIRAS measurement of the blackbody spectrum.

7 Large-Scale Structure

The dark-matter and baryon perturbations developed in the previous chapter provide an excellent description of the early Universe, where $|\delta| \sim 10^{-5}$ and squared perturbations are completely negligible. As discussed there (see in particular §5.8 and Fig. 11), this linear evolution maps inflationary curvature perturbations onto a characteristic matter power spectrum

$$P_m^{\text{lin}}(k) \sim \begin{cases} k & k < k_{\text{eq}}, \\ k^{-3} \log^2(k/k_{\text{eq}}) & k > k_{\text{eq}}, \end{cases} \quad (320)$$

with the time dependence factored out through the **linear growth factor** $D(\eta)$: $\delta_m(\vec{k}, \eta) = D(\eta) \delta_L(\vec{k})$, with $D \propto \log a$ during radiation domination (the Mészáros effect) and $D \propto a$ during matter domination. As we saw in §6, this linear description is sufficient to predict the temperature anisotropies of the CMB, since the relevant photon-baryon perturbations at last scattering are still safely $|\delta| \ll 1$.

The situation at low redshift is more complicated. Although linear theory remains exact at $z \gg 1$ (when $\delta^2 \ll |\delta|$), gravitational collapse continually transfers power from large to small scales, and by today the modes with $k \gtrsim 0.1 h \text{ Mpc}^{-1}$ have grown to $|\delta| \sim \mathcal{O}(1)$. On these *nonlinear* scales the linear treatment fails, and we need a genuinely nonperturbative or higher-order description.⁸² In

⁸²Just as for the matter power spectrum, the late-time density field also retains an oscillatory BAO feature inherited from baryon-photon coupling at recombination (cf. §5.10.4); these oscillations are the standard ruler used by modern galaxy surveys to measure the expansion history.

this final section, we lay out the standard **perturbation theory** expansion for the matter density on subhorizon scales, develop it to one-loop order, extend it to galaxies via the bias expansion, and conclude with a brief discussion of the **effective field theory** reformulation that underpins all modern galaxy-survey analyses.

7.1 Newtonian Fluid Equations

The starting point is the system of equations governing the dark-matter density and velocity on subhorizon scales. We work throughout in the limit $k \gg \mathcal{H}$, where the typical wavelength of interest is much smaller than the Hubble radius and the relativistic effects of structure formation can be neglected — this is the **Newtonian limit** of cosmological perturbation theory. For simplicity we further restrict to a universe containing only dark matter and a cosmological constant, ignoring radiation and treating the baryons as part of the cold pressureless fluid. The dark matter then behaves as a non-relativistic perfect fluid governed by

$$\delta'_m + \vec{\nabla} \cdot [(1 + \delta_m) \vec{v}] = 0, \quad (\text{continuity}) \quad (321)$$

$$\vec{v}' + (\vec{v} \cdot \vec{\nabla}) \vec{v} = -\mathcal{H} \vec{v} - \vec{\nabla} \Phi, \quad (\text{Euler}) \quad (322)$$

$$\nabla^2 \Phi = 4\pi G a^2 \bar{\rho}_m \delta_m = \frac{3}{2} \mathcal{H}^2 \Omega_m \delta_m, \quad (\text{Poisson}) \quad (323)$$

where primes denote conformal-time derivatives, $\Omega_m \equiv \bar{\rho}_m / \bar{\rho}_{\text{crit}}$ is the matter density parameter, and we have dropped the Φ' source term in the Euler equation since on subhorizon scales $\Phi' \ll |\vec{\nabla} \Phi|$. The Hubble drag $-\mathcal{H} \vec{v}$ arises from the expansion of the background, the gravitational source on the right-hand side of Euler couples velocity to density via the potential, and the Poisson equation closes the system by relating Φ to δ_m . As we shall see, throughout this section it is convenient to work with the **velocity divergence**

$$\theta \equiv \vec{\nabla} \cdot \vec{v}, \quad (324)$$

since the irrotational component of $\vec{v}$ is fully captured by θ via $\vec{v} = \vec{\nabla}(\nabla^{-2}\theta)$, and any vorticity decays away with the expansion in the absence of shell crossing.

7.2 Linear Solution

At first order, dropping all quadratic and higher products of δ_m and $\vec{v}$, the fluid equations linearize to

$$\delta'_m + \theta = 0, \quad \theta' + \mathcal{H}\theta = -\nabla^2 \Phi = -\frac{3}{2} \mathcal{H}^2 \Omega_m \delta_m. \quad (325)$$

Differentiating the continuity equation and substituting Euler eliminates θ , yielding the second-order growth equation for the matter density contrast,

$$\boxed{\delta''_m + \mathcal{H} \delta'_m - \frac{3}{2} \mathcal{H}^2 \Omega_m \delta_m = 0}, \quad (326)$$

which is precisely the subhorizon limit of the relativistic growth equation derived in §5.8. Because this equation is scale-independent, its solution factorizes into a k -independent time evolution and a k -dependent initial profile,

$$\delta_m(\vec{k}, \eta) = D(\eta) \delta_L(\vec{k}), \quad \theta(\vec{k}, \eta) = -\mathcal{H} f D(\eta) \delta_L(\vec{k}), \quad (327)$$

where $\delta_L(\vec{k})$ is the linear density field set by the primordial initial conditions, and $D(\eta)$ is the **linear growth factor** satisfying

$$D'' + \mathcal{H}D' - \frac{3}{2}\mathcal{H}^2\Omega_m D = 0. \quad (328)$$

The velocity normalization follows from continuity, $\theta = -\delta'_m$, which expresses the velocity divergence in terms of the time-derivative of the density and motivates the definition of the **linear growth rate**

$$f \equiv \frac{d \log D}{d \log a}, \quad (329)$$

i.e. the logarithmic derivative of the growth factor. In matter domination ($\Omega_m = 1$), the growing-mode solution is $D \propto a$ and $f \rightarrow 1$, recovering the familiar $\delta_m \propto a$ result. As Λ takes over at $z \lesssim 0.5$, growth slows and D asymptotes to a constant, with $f \rightarrow 0$. In a generic flat Λ CDM background, the growing-mode solution can be written in closed form as

$$D(\eta) = D_0 \mathcal{H}(\eta) \int_0^{a(\eta)} \frac{da'}{[\mathcal{H}(a')]^3}, \quad (330)$$

where D_0 is a normalization constant (typically fixed by $D(\eta_0) = 1$, i.e. relative to today). This integral form makes manifest that D depends only on the background expansion history, and it reproduces the matter-era result $D \propto a$ as well as the gradual slowdown of growth in the dark-energy era.

7.3 Standard Perturbation Theory

The linear solution holds only while $|\delta_m| \ll 1$. To go beyond linear order, we expand the full nonlinear fluid equations perturbatively in powers of the linear field, treating $\langle \delta_L^2 \rangle_R^{1/2} < 1$ on some smoothing scale R as the formal small parameter. The strategy is the standard one: insert the linear solution into the quadratic source terms to generate a second-order correction, feed both into the cubic sources for a third-order correction, and so on, generating an entire tower of perturbative pieces $\delta^{(n)}, \theta^{(n)}$ each of order δ_L^n .

For the algebra, it is cleanest to move to Fourier space, where the nonlinear continuity equation takes the form

$$\delta'(\vec{k}, \eta) + \theta(\vec{k}, \eta) = - \int \frac{d^3 p}{(2\pi)^3} \alpha(\vec{p}, \vec{k} - \vec{p}) \theta(\vec{p}, \eta) \delta(\vec{k} - \vec{p}, \eta), \quad (331)$$

with the **advection kernel**

$$\alpha(\vec{p}, \vec{k} - \vec{p}) \equiv \frac{\vec{k} \cdot \vec{p}}{p^2}, \quad (332)$$

and the nonlinear Euler equation becomes

$$\theta'(\vec{k}, \eta) + \mathcal{H} \theta(\vec{k}, \eta) + \frac{3}{2} \mathcal{H}^2 \Omega_m \delta(\vec{k}, \eta) = - \int \frac{d^3 p}{(2\pi)^3} \beta(\vec{p}, \vec{k} - \vec{p}) \theta(\vec{p}, \eta) \theta(\vec{k} - \vec{p}, \eta), \quad (333)$$

with the **velocity-coupling kernel**

$$\beta(\vec{p}, \vec{k} - \vec{p}) \equiv \frac{k^2 \vec{p} \cdot (\vec{k} - \vec{p})}{2 p^2 |\vec{k} - \vec{p}|^2}. \quad (334)$$

The mode-coupling integrals on the right-hand sides describe how pairs of modes generate a third one: α controls the transport of density by velocity (the convective $\vec{\nabla} \cdot (\delta \vec{v})$ term in continuity), while β controls the nonlinear advection of momentum (the $(\vec{v} \cdot \vec{\nabla}) \vec{v}$ term in Euler). Schematically, the second-order solution then takes the form $\delta^{(2)}, \theta^{(2)} \sim \int_{\vec{p}} \delta_L \theta_L \propto D^2 \delta_L^2$, sourced by products of linear quantities.

7.3.1 The Perturbative Ansatz and the SPT Kernels

A clean way to organize this expansion is to write all higher-order pieces in terms of convolutions of δ_L alone, with the time and momentum dependence neatly factored apart:

$$\delta_m(\vec{k}, \eta) = \sum_{n=1}^{\infty} D^n(\eta) \int \left[\prod_{i=1}^n \frac{d^3 p_i}{(2\pi)^3} \delta_L(\vec{p}_i) \right] (2\pi)^3 \delta_D\left(\sum_i \vec{p}_i - \vec{k}\right) F_n(\vec{p}_1, \dots, \vec{p}_n), \quad (335)$$

$$\theta(\vec{k}, \eta) = - \sum_{n=1}^{\infty} [\mathcal{H} f D(\eta)]^n \int \left[\prod_{i=1}^n \frac{d^3 p_i}{(2\pi)^3} \delta_L(\vec{p}_i) \right] (2\pi)^3 \delta_D\left(\sum_i \vec{p}_i - \vec{k}\right) G_n(\vec{p}_1, \dots, \vec{p}_n). \quad (336)$$

The dimensionless functions F_n and G_n are the **SPT kernels**: F_n controls the n th-order density and G_n the corresponding velocity divergence, with the entire scale dependence encoded in the kernel arguments and the entire time dependence carried by D^n and $(\mathcal{H} f D)^n$ respectively. By construction, the n th-order correction grows as $\delta^{(n)} \propto [D \delta_L]^n$, so successive orders are suppressed by additional powers of the linear amplitude.

The kernels are fixed by substituting the ansatz back into the nonlinear fluid equations and matching order by order, which yields a set of algebraic recursion relations. At lowest order, $F_1 = G_1 = 1$ trivially recovers the linear solution. At second order, the explicit symmetric forms read

$$F_2(\vec{k}_1, \vec{k}_2) = \frac{5}{7} + \frac{1}{2}(\hat{k}_1 \cdot \hat{k}_2) \left(\frac{k_1}{k_2} + \frac{k_2}{k_1} \right) + \frac{2}{7}(\hat{k}_1 \cdot \hat{k}_2)^2, \quad (337)$$

$$G_2(\vec{k}_1, \vec{k}_2) = \frac{3}{7} + \frac{1}{2}(\hat{k}_1 \cdot \hat{k}_2) \left(\frac{k_1}{k_2} + \frac{k_2}{k_1} \right) + \frac{4}{7}(\hat{k}_1 \cdot \hat{k}_2)^2, \quad (338)$$

or equivalently, in terms of the symmetrized advection and velocity-coupling kernels,

$$F_2 = \frac{5}{7} \alpha + \frac{2}{7} \beta, \quad G_2 = \frac{3}{7} \alpha + \frac{4}{7} \beta. \quad (339)$$

The same procedure generates F_n, G_n to arbitrary order, with each successive kernel built recursively from products of lower-order ones — in principle, all of the structure of weakly-nonlinear gravitational clustering is encoded as a tower of algebraic kernel formulae.

7.4 The One-Loop Power Spectrum

The most important application of perturbation theory is the computation of the next-to-leading correction to the matter power spectrum — the so-called **1-loop** contribution. At leading order, the power spectrum is simply the variance of the first-order field,

$$P_{11}(k, \eta) \equiv \langle \delta^{(1)}(\vec{k}, \eta) \delta^{(1)*}(\vec{k}, \eta) \rangle = D^2(\eta) P_L(k), \quad (340)$$

where $P_L(k)$ is the linear power spectrum set by the initial conditions, $\langle \delta_L(\vec{k}) \delta_L^*(\vec{k}') \rangle = (2\pi)^3 \delta_D(\vec{k} - \vec{k}') P_L(k)$; this is the result that we have used throughout. To go beyond, we expand the full power spectrum in powers of the perturbative pieces,

$$P_m(k, \eta) = \langle (\delta^{(1)} + \delta^{(2)} + \delta^{(3)} + \dots) (\delta^{(1)} + \delta^{(2)} + \delta^{(3)} + \dots)^* \rangle, \quad (341)$$

and identify the lowest-order nontrivial corrections. The cross-term $\langle \delta^{(1)} \delta^{(2)*} \rangle$ vanishes by **Gaussianity**: since $\delta^{(2)} \sim \delta_L \delta_L$, this correlator reduces to the three-point function $\langle \delta_L^3 \rangle$, which vanishes identically for a Gaussian field. The Gaussianity of δ_L is in turn a direct prediction of single-field inflation. The leading nontrivial corrections are therefore the two pieces of order $D^4 P_L^2$:

$$P_{22}(k, \eta) \equiv \langle \delta^{(2)} \delta^{(2)*} \rangle, \quad P_{13}(k, \eta) \equiv 2 \text{Re} \langle \delta^{(1)} \delta^{(3)*} \rangle, \quad (342)$$

which together constitute the one-loop correction.

7.4.1 The P_{22} Contribution

Inserting the second-order ansatz into the variance gives

$$P_{22}(k, \eta) = D^4(\eta) \int \frac{d^3 p}{(2\pi)^3} \frac{d^3 p'}{(2\pi)^3} F_2(\vec{p}, \vec{k} - \vec{p}) F_2^*(\vec{p}', \vec{k} - \vec{p}') \langle \delta_L(\vec{p}) \delta_L(\vec{k} - \vec{p}) \delta_L^*(\vec{p}') \delta_L^*(\vec{k} - \vec{p}') \rangle. \quad (343)$$

For a Gaussian field, the four-point function factorizes into a sum of three two-point pairings via Wick's theorem,

$$\langle \delta_1 \delta_2 \delta_3 \delta_4 \rangle = \langle \delta_1 \delta_2 \rangle \langle \delta_3 \delta_4 \rangle + \langle \delta_1 \delta_3 \rangle \langle \delta_2 \delta_4 \rangle + \langle \delta_1 \delta_4 \rangle \langle \delta_2 \delta_3 \rangle. \quad (344)$$

Of the three contractions, the first — pairing the two “incoming” legs $\delta_L(\vec{p}) \delta_L(\vec{k} - \vec{p})$ together — enforces $\vec{k} = 0$ via the resulting Dirac delta and is dropped, since it contributes only at zero wavenumber where the mean has been subtracted. The remaining two contractions both pair an incoming leg with an outgoing leg, generating $\delta_D(\vec{p} - \vec{p}')$ and $\delta_D(\vec{p} - (\vec{k} - \vec{p}'))$ respectively, and they are equal by the symmetry $F_2(\vec{a}, \vec{b}) = F_2(\vec{b}, \vec{a})$. Carrying out the resulting momentum integrals yields a single loop integral with an overall combinatoric factor of two:

$$P_{22}(k, \eta) = 2 D^4(\eta) \int \frac{d^3 p}{(2\pi)^3} |F_2(\vec{p}, \vec{k} - \vec{p})|^2 P_L(p) P_L(|\vec{k} - \vec{p}|). \quad (345)$$

A useful check is that $F_2(\vec{x}, -\vec{x}) = 0$, which guarantees that $P_{22} \rightarrow 0$ as $\vec{k} \rightarrow 0$, as required by mass and momentum conservation: large-scale modes cannot be sourced by mode coupling alone.

7.4.2 The P_{13} Contribution

For P_{13} , the third-order field $\delta^{(3)} \sim \int F_3 \delta_L \delta_L \delta_L$ correlates with $\delta^{(1)} \sim \delta_L$ to give a single contraction of the internal δ_L pairs against the external δ_L . With three equivalent ways to pair the external leg with one of the three internal legs of $\delta^{(3)}$, the result is

$$2 \text{Re} \langle \delta^{(1)}(\vec{k}) \delta^{(3)*}(\vec{k}) \rangle = 6 D^4(\eta) P_L(k) \int \frac{d^3 p}{(2\pi)^3} F_3(\vec{p}, -\vec{p}, \vec{k}) P_L(p), \quad (346)$$

where F_3 is the symmetrized third-order SPT kernel. The structure is particularly clean: an external factor of $P_L(k)$ multiplies a single loop integral over $P_L(p)$, weighted by the kernel evaluated at the configuration $(\vec{p}, -\vec{p}, \vec{k})$.

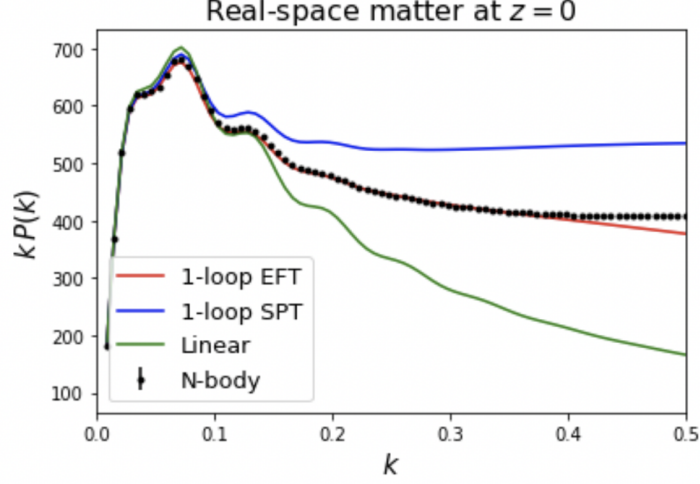


Figure 18: Comparison of the tree-level (green) and one-loop (blue and red) dark matter power spectra with the measurements from numerical simulations. On large-scales, the theory is dominated by the linear terms, with one-loop corrections important for $k \gtrsim 0.1 h \text{ Mpc}^{-1}$. Despite the addition of higher-order terms, standard perturbation theory (SPT) does not provide a better fit to the data than linear theory; this is resolved by the Effective Field Theory (EFT), which provides an excellent fit for $k \lesssim 0.3 h \text{ Mpc}^{-1}$.

7.4.3 The 1-Loop Result

Combining the two contributions, the matter power spectrum at next-to-leading order — the **1-loop** power spectrum — reads

$$P_m(k, \eta) = D^2(\eta) P_L(k) + D^4(\eta) \left\{ 2 \int \frac{d^3 p}{(2\pi)^3} |F_2|^2 P_L(p) P_L(|\vec{k} - \vec{p}|) + 6 P_L(k) \int \frac{d^3 p}{(2\pi)^3} F_3(\vec{p}, -\vec{p}, \vec{k}) P_L(p) \right\}. \quad (347)$$

The terminology *loop* is borrowed from quantum field theory: P_{22} and P_{13} correspond to diagrams with a single internal momentum integral, exactly analogous to one-loop Feynman graphs. Physically, the 1-loop correction enhances power on small scales — where mode coupling is important — while leaving the largest scales close to the linear prediction. This is shown in Figure 18.

Applications and limitations. The 1-loop power spectrum, together with its 2-loop extension (recently applied to lensing data from the Dark Energy Survey), provides a quantitatively useful description of the matter clustering on quasi-linear scales: it is the relevant theoretical input for **weak gravitational lensing** of background galaxies, for the broadband shape of $P_m(k)$ in the BAO regime ($k \lesssim 0.2 h \text{ Mpc}^{-1}$), and for a host of related observables. Its limitations are built into the logic of the expansion itself: once $|\delta| \gtrsim 1$, no finite number of terms can converge, and the perturbative series breaks down. SPT is therefore a controlled *large-scale* model, and must be supplemented by simulations, halo models, or (as we discuss below) the EFT extension on the deeply nonlinear scales of galaxy clusters.

7.5 Galaxy Bias

In practice, the dark-matter distribution is very hard to measure directly — the principal exception being gravitational lensing, which is sensitive to the projected matter density along the line of sight. Our best observational handle on large-scale structure is instead the spatial pattern of luminous tracers, especially **galaxies**. Their number density defines a separate overdensity field,

$$\delta_g(\vec{x}, \eta) \equiv \frac{n_g(\vec{x}, \eta) - \bar{n}_g(\eta)}{\bar{n}_g(\eta)}, \quad (348)$$

which can be analyzed by the same statistical machinery as δ_m . The key physical observation is that on sufficiently large scales galaxies *trace* the underlying matter distribution: regions of high dark-matter density are precisely the regions in which dark-matter halos form and galaxies subsequently emerge by the cooling of baryons within them. At leading order, this physics is captured by a single multiplicative coefficient, the **linear bias** b_1 ,

$$\boxed{\delta_g(\vec{k}, \eta) = b_1(\eta) \delta_m(\vec{k}, \eta) + (\text{noise})}, \quad (349)$$

where the noise is a stochastic contribution uncorrelated with δ_m , arising from the discrete and finite sampling of the underlying continuous field. Taking the two-point function, the galaxy power spectrum then takes the simple form

$$P_{gg}(k, \eta) = b_1^2(\eta) P_{mm}(k, \eta) + P_{\text{shot}}, \quad (350)$$

where the **shot-noise** contribution $P_{\text{shot}} \sim 1/\bar{n}_g$ is a white-noise term from the finite sample size. Crucially, b_1 is not predicted from first principles — it depends on all the messy details of galaxy formation — but it can be *measured* directly from the data and marginalized over in the cosmological analysis.

7.5.1 The Higher-Order Bias Expansion

The linear-bias model is only the leading term in a systematic expansion of δ_g in powers of the matter density and its derivatives. At second order in δ_m , the most general local expression consistent with the symmetries of the problem reads

$$\boxed{\delta_g(\vec{x}, \eta) = \underbrace{b_1(\eta) \delta_m(\vec{x}, \eta)}_{\mathcal{O}(\delta)} + \underbrace{\frac{1}{2} b_2(\eta) \delta_m^2(\vec{x}, \eta) + b_{s^2}(\eta) \left[\left(\frac{\partial_i \partial_j}{\partial^2} - \frac{1}{3} \delta_{ij} \right) \delta_m \right]^2}_{\mathcal{O}(\delta^2)} + \dots}, \quad (351)$$

plus many more terms at $\mathcal{O}(\delta^3)$ and beyond. The first new operator, δ_m^2 , captures the nonlinear response of galaxy formation to the local density; the second, the squared tidal tensor $s_{ij} \equiv (\partial_i \partial_j / \partial^2 - \frac{1}{3} \delta_{ij}) \delta_m$, encodes the response to anisotropic stretching of the matter field. These two operators exhaust the second-order possibilities consistent with three guiding principles:

1. **Isotropy and homogeneity:** δ_g must be a scalar built from δ_m and its derivatives, with no preferred direction — so single-derivative operators such as $\partial_i \delta_m$ are forbidden.
2. **Locality:** δ_g at a point depends only on δ_m in a small neighbourhood, not on far-away values $\delta_m(\vec{x} + \vec{R})$, since galaxy formation is a local process.⁸³

⁸³Higher-derivative terms such as $\nabla^2 \delta_m$ are nevertheless allowed because they arise locally; they enter at higher order in k and are suppressed by an additional factor of $k^2 R_{\text{gal}}^2$ for some characteristic galaxy scale R_{gal} .

3. **The equivalence principle:** δ_g cannot depend on Φ or $\partial_i \Phi$ (the gravitational acceleration), since these are coordinate artifacts of the freely-falling frame; only second and higher derivatives of Φ , which carry genuine tidal information, may appear.

7.5.2 The Galaxy Power Spectrum Pipeline

Putting all the pieces together, a model for the observed galaxy power spectrum is built up through a clean four-step pipeline:

$$\mathcal{R}(\vec{k}) \xrightarrow{\text{inflation}} \delta_L(\vec{k}), D(\eta) \xrightarrow{\text{linear theory}} \delta^{(1)}, \delta^{(2)}, \dots = \delta_m \xrightarrow{\text{SPT}} \delta_g = b_1 \delta_m + \dots \xrightarrow{\text{bias}} P_{gg}(k, \eta).$$

The inflationary curvature perturbation sets the initial conditions; the linear growth factor evolves them into a (still small) late-time linear field; SPT generates the nonlinear matter overdensity to the required loop order; and the bias expansion converts this into a galaxy overdensity, whose two-point function gives the observed power spectrum. In any real analysis, one must additionally include *redshift-space distortions*, noise contributions, and instrumental systematics.⁸⁴ At next-to-leading order, this 1-loop bias-expanded model is the standard analysis pipeline used in the current generation of galaxy surveys, including **DESI** and **Euclid**; the 2-loop extension is an active frontier.

7.6 The Effective Field Theory of Large-Scale Structure

The standard perturbation theory developed in the previous subsections suffers from a fundamental tension: it is constructed as an expansion in δ_m , but δ_m is large precisely on the small scales where nonlinear evolution is most important. Quantitatively, the variance of the matter density field receives a contribution from arbitrarily small scales,

$$\sigma^2(\vec{x}) \equiv \langle \delta_m^2(\vec{x}) \rangle = \int \frac{d^3k}{(2\pi)^3} P_m(k) \sim \log k_{\text{max}} \rightarrow \infty, \quad (352)$$

since $P_m \sim k^{-3}$ on small scales gives a logarithmic UV divergence. Although mild, this divergence signals that there is no scale on which $|\delta|$ is genuinely small, and so naively no scale on which the perturbative expansion is under control. Worse, the SPT loop integrals couple modes of all wavenumbers,

$$P_{22}(k) \sim 2 \int \frac{d^3p}{(2\pi)^3} |F_2|^2 P_L(p) P_L(|\vec{k} - \vec{p}|), \quad (353)$$

so even a large-scale prediction for $P_{22}(k)$ at $k \ll 1 h \text{ Mpc}^{-1}$ depends on integrals over loop momenta p extending into the deeply nonlinear regime, where the perturbative description has no business applying. Empirically, this manifests in the unhappy observation that vanilla 1-loop SPT does not in fact provide a better fit to data than the linear theory it was meant to improve upon (see Figure 18).

⁸⁴Redshift-space distortions arise because the radial coordinate of a galaxy is inferred from its redshift, which is the sum of a cosmological expansion piece and a peculiar-velocity piece. The resulting position is therefore shifted by $\hat{n} \cdot \vec{v}/H$ along the line of sight, breaking statistical isotropy and introducing an angular dependence in the galaxy power spectrum, $P_{gg}(k, \mu)$, with $\mu \equiv \hat{k} \cdot \hat{n}$.

7.6.1 The EFT Solution: Smoothing and Counterterms

The modern resolution is the **effective field theory of large-scale structure** (EFT of LSS), which does the perturbative expansion more carefully by acknowledging that small-scale physics is not under our control and explicitly *integrating it out* of the long-wavelength dynamics. The key technical step is to introduce a comoving wavenumber cutoff Λ and replace the bare fields by their smoothed counterparts,

$$\delta_\Lambda(\vec{k}) \equiv \begin{cases} \delta_m(\vec{k}), & k < \Lambda, \\ 0, & k > \Lambda, \end{cases} \quad (354)$$

and similarly for $\vec{v}_\Lambda$, with Λ chosen large enough that $\langle \delta_\Lambda^2 \rangle^{1/2} \lesssim 1$. By construction, the smoothed variance is finite and the perturbative expansion in δ_Λ is genuinely controlled.

Applying the smoothing to the fluid equations themselves — not merely to the final answer — yields modified equations of motion for $\delta_\Lambda, \vec{v}_\Lambda$, with new **counterterms** that account for the back-reaction of the integrated-out short modes on the long-wavelength dynamics. Schematically, the smoothing does not commute with the nonlinear products of the fluid equations: $[\delta \vec{v}]_\Lambda \neq \delta_\Lambda \vec{v}_\Lambda$, and the difference is precisely the (finite) effect of subgrid physics on the long-wavelength fluid. The recipe is then:

1. Smooth the fluid equations at scale Λ , generating new source terms for $\delta_\Lambda, \vec{v}_\Lambda$.
2. Solve the smoothed system perturbatively, exactly as in SPT.
3. Compute the power spectrum of δ_Λ , with all loop integrals now cut off at Λ .

At one loop, the resulting EFT power spectrum takes the form

$$P_m(k) = D^2 P_L(k) + D^4 \left\{ 2 \int_{p < \Lambda} |F_2|^2 P_L P_L + 6 \int_{p < \Lambda} F_3 P_L(k) P_L \right\} - c_s^2(\Lambda) k^2 P_L(k), \quad (355)$$

with two key modifications relative to bare SPT: (a) the loop integrals are now truncated at $p < \Lambda$, and (b) there is a new **counterterm** $-c_s^2(\Lambda) k^2 P_L(k)$, whose structure looks like that of an effective sound speed for the dark-matter fluid (hence the suggestive notation).

7.6.2 Renormalization and Physical Predictions

The deep insight at the heart of the EFT formalism is that the smoothing scale Λ was introduced as a purely **computational** trick and cannot appear in any physical observable. Demanding that the long-wavelength $P_m(k)$ at $k < \Lambda$ be independent of Λ then fixes the running of the counterterm coefficient $c_s^2(\Lambda)$ in such a way that the Λ -dependence cancels between the truncated loop integrals and the counterterm. This is precisely the mechanism of **renormalization** familiar from quantum field theory: short-distance physics is absorbed into a renormalized coefficient (here c_s^2), which encodes the response of the long-wavelength dynamics to subgrid effects and is treated as a free parameter to be fit to data or to simulations. Once fit, the EFT 1-loop power spectrum provides a reliable description of large-scale clustering well into the mildly nonlinear regime, with controlled theoretical errors at any chosen k . This is clearly demonstrated in Figure 18.

This formalism has reshaped modern galaxy-survey cosmology: all current cosmological analyses of galaxy clustering data, from BOSS through DESI to Euclid, are built on the 1-loop EFT framework, suitably extended by the bias expansion of §7.5 and by redshift-space distortions. The

full technical theory is substantially more elaborate than the sketch presented here, with deep ties to renormalization-group running, operator product expansions, and infrared resummation — but the basic philosophy is the simple one outlined above: *smooth, expand, and renormalize*.

The End

Thanks for joining me to explore Cosmology! For further reading, see the excellent books by Daniel Baumann (which follows these notes), Steven Weinberg (which goes a bit deeper into the mathematics), and Scott Dodelson & Fabian Schmidt (which has a particularly excellent CMB section). Also see Anthony Challinor’s online notes on CMB anisotropies and my own notes on the EFT of Large-Scale Structure. Please, [get in touch](#) with any questions!