

General Qualifying Exam Solutions

2020

Starkman, Nathaniel Lokken, Martine Ludwig, Bethany Winch, Harrison
Sheere, Connor

January 14, 2020

Preface

General Qualifying Exam Solutions: Physics and Fundamentals

Starkman, Nathaniel

Lokken, Martine

Ludwig, Bethany

Winch, Harrison

Sheere, Connor

This work is a collaborative effort.

It is based upon the works of:

- Miranda Herman
- Jeffery Emberson
- Charles Zhu
- Ned Wright

Contents

1 Cosmology	5
1.1 Introduction	6
1.1.1 Campbell Cosmo Intro	7
1.1.2 Zhu Intro	11
1.2 Q1) Recombination	14
1.2.1 Short Answer	14
1.2.2 Cosmology Class Fall 2019: Martine's Notes	14
1.2.3 Q1) Ludwig Cosmo Q1	15
1.2.4 Q1) Herman Cosmo Q1	16
1.2.5 Q1) Campbell Cosmo Q1	17
1.2.6 Q1) Zhu Cosmo Q1	19
1.3 Q2) Flat Universe Properties	21
1.3.1 Short Answer	21
1.3.2 Q2) Ludwig Cosmo Q2	22
1.3.3 Q2) Herman Cosmo Q2	23
1.3.4 Q2) Campbell Cosmo Q2	24
1.3.5 Q2) Zhu Cosmo Q2	27
1.4 Q3) Term Definitions	29
1.4.1 Short Answer	29
1.5 Q4) Big Bang Evidence	30
1.5.1 Short Answer	30
1.5.2 Relevant Equations	30
1.5.3 Q4) Ludwig Cosmo Q4	31
1.5.4 Q4) Herman Cosmo Q4	32
1.5.5 Q4) Campbell Cosmo Q4	33
1.5.6 Q4) Zhu Cosmo Q4	40
1.5.7 Cosmology Class Fall 2019: Martine's Notes	41
1.6 Q5) Big Bang Nucleosynthesis (BBN)	42
1.6.1 Q5) Ludwig Cosmo Q5	43
1.6.2 Q5) Herman Cosmo Q6	44
1.6.3 Q5) Campbell Cosmo Q6	45
1.6.4 Q5) Zhu Cosmo Q5	48
1.7 Q6) Type 1a Supernovae for Cosmology	50
1.7.1 Q6) Ludwig Cosmo Q6	51
1.7.2 Q6) Herman Cosmo Q7	52
1.7.3 Q6) Campbell Cosmo Q7	53
1.7.4 Q6) Zhu Cosmo Q8	60
1.8 Q7) The distance Ladder	64
1.8.1 Q7) Herman ExtraGal Q3	65
1.8.2 Q7) Campbell ExtraGal Q3	66
1.8.3 Q7) Zhu ExtraGal Q4	74
1.9 Q8) Other Methods for Parameter Constraints	77
1.9.1 Q8) Ludwig Cosmo Q7	78

1.9.2	Q8) Herman Cosmo Q8	79
1.9.3	Q8) Campbell Cosmo Q8	81
1.10	Q9) CMB Polarization	84
1.10.1	Ludwig Cosmo Q8	85
1.10.2	Herman Cosmo Q9	86
1.10.3	Q9) Campbell Cosmo Q9	87
1.11	Q10) CMB Secondary Anisotropies	89
1.11.1	Q10) Ludwig Cosmo Q9	90
1.11.2	Q10) Herman Cosmo Q10	91
1.11.3	Q10) Campbell Cosmo Q10	92
1.12	Q11) Cosmological Inflation	97
1.12.1	Ludwig Cosmo Q10	98
1.12.2	Herman Cosmo Q11	99
1.13	Q12) Two-Point Correlation Function (2PCF)	101
1.13.1	Q12) Herman Cosmo Q13	102
1.13.2	Q12) Campbell Cosmo Q13	103
1.14	Q13) Angular Power Spectrum	106
1.14.1	Short Summary	106
1.14.2	Campbell	106
1.15	Q14) Positive Cosmological Constant	107
1.15.1	Short Summary	107
1.15.2	Q14) Ludwig Cosmo Q13	108
1.15.3	Q14) Herman Cosmo Q14	109
1.15.4	Q14) Campbell Cosmo Q14	110
1.16	Q15) Epoch of Reionization	113
1.16.1	Short Summary	113
1.16.2	Q15) Ludwig Cosmo Q14	114
1.16.3	Q15) Herman Cosmo Q15	115
1.16.4	Q15) Campbell Cosmo Q15	117
1.16.5	Cosmology Class Fall 2019: Martine's Notes	126
1.17	Q16) 21 CM Hydrogen	126
1.17.1	Short Summary	127
1.17.2	Q16) Ludwig Cosmo Q15	128
1.17.3	Q16) Herman Cosmo Q16	129
1.17.4	Q16) Campbell Cosmo Q16	130
1.18	Q17) Cosmic Neutrino Background	134
1.18.1	Q17) Ludwig Cosmo Q17	135
1.18.2	Q17) Herman Cosmo Q18	136
1.18.3	Q17) Campbell Cosmo Q18	137
1.18.4	Cosmology Class Fall 2019: Martine's Notes	139
1.19	Q18) Dark Matter Candidates	140
1.19.1	Herman Cosmo Q20	141
1.19.2	Q18) Campbell Cosmo Q20	143
1.20	Q19) Galaxy Clusters	144
1.20.1	Herman ExtraGal Q13	145
1.21	Star Formation Quenching	147
1.21.1	Herman ExtraGal Q14	148
1.22	OLD COSMOLOGY QUESTIONS, NOT IN THIS QUAL	150
1.22.1	Herman Cosmo Q5	151
1.22.2	Herman Cosmo Q12	152
1.22.3	Herman Cosmo Q17	153
1.22.4	Herman Cosmo Q19	154
1.22.5	OLD Q3) CDM Spectrum evolution	155

2 Galaxies	174
2.1 Introduction	175
2.2 Q1) Mass of the Milky Way	176
2.2.1 Q1) Herman ExtraGal Q2	177
2.2.2 Q1) Emberson ExtraGal Q2	178
2.2.3 Q1) Campbell ExtraGal Q2	180
2.3 Q2) Nuclear Black Holes	182
2.4 Q3) AGN	182
2.5 Q4) Quasars	183
2.6 Q5) Methods of Galaxy Mass Determination	184
2.7 Q6) SEDs of Single-Burst Galaxies	185
2.8 Q7) Galactic Spiral Structure	186
2.9 Q8) Stellar Initial Mass Function (IMF)	187
2.10 Q9) Stellar Populations in the Galaxy	188
2.11 Q10) G-Dwarf Problem in the Solar Neighbourhood	189
2.12 Q11) Stellar Orbits in Potential	190
2.13 Q12) Dynamical Relaxation	191
2.14 Q13) Dynamical Friction	192
3 Stars and Planets	194
3.1 Introduction	195
4 Physics and Fundamentals	196
4.1 Introduction	197
4.2 Q1) Interferometry	197
4.2.1 Campbell Physics Q2	199
4.3 Q2) Detector Cooling	205
4.3.1 Campbell Physics Q11	206
4.4 Q3) Camera Field of View	215
4.4.1 Campbell Physics Q14	216
4.5 A4) Sidelobes in Diffraction-Limited Detectors	217
4.5.1 Herman Physics Q21	218

1. Cosmology

General Qualifying Exam Solutions: Cosmology

Starkman, Nathaniel

Lokken, Martine

Ludwig, Bethany

Winch, Harrison

Sheere, Connor

Sheere, Connor

INTRODUCTION

Campbell Cosmo Intro

1 Cosmology

1.1 Preface

Cosmology is the study of the Universe on the largest of scales – the origin, evolution, and possible fate of the Universe as a whole. Cosmology concerns itself with asking questions such as: “Where do we come from? What is the Universe made of? How did the elements form? Is it finite or infinite in spatial extent? Why is the Universe so smooth? How did galaxies form from such a smooth origin? Did the Universe have a beginning sometime in the past? Will it come to an end sometime in the future?” As a branch of science that studies the entirety of the Universe, it often deals with distances that are very big, objects that are very large, and timescales that are very long. For this reason, standard units of distance (meters), mass (kilogram), and time (seconds or years) are far too small and therefore conventionally work in much larger standard units.

One distance measure often employed by astronomers is the **astronomical unit** (AU) which is the average distance between the Earth and the Sun over a period of one year: $1\text{AU} = 1.5 \times 10^{11} \text{ m}$. Such a distance scale is useful on the scale of the Solar System, but is small compared to the distance between stars. For interstellar distances, another unit of measure called the **parsec** (pc) is most useful which is defined to be the distance at which one AU subtends an angle of one arcsecond: $1\text{pc} = 3.1 \times 10^{16} \text{ m}$. For example, we are located approximately 1.3 pc away from the nearest star, Proxima Centauri and 85 pc from the center of the Galaxy. This distance measure is therefore most useful for within the Galaxy but is small compared to the distance between neighbouring galaxies. For such intergalactic scales, units of megaparsecs (Mpc) are often used: $1\text{Mpc} = 10^6 \text{ pc} = 3.1 \times 10^{22} \text{ m}$. For example, we are located roughly 0.7 Mpc from the nearest galaxy Andromeda (M31), and 15 Mpc from the nearest cluster of galaxies, the Virgo cluster.

The standard unit of mass often used by astronomers is the Solar mass: $1\text{M}_\odot = 2.0 \times 10^{30} \text{ kg}$. While the mass of the Galaxy isn't as well known as the Solar mass, it is approximately $\text{M}_{\text{gal}} \approx 10^{12} \text{ M}_\odot$. Incidentally, the Sun also provides the standard unit of power (units of energy per second; Watts) in astronomy. The Sun's luminosity (i.e., the rate at which it radiates energy in the form of light) is $1\text{L}_\odot = 3.8 \times 10^{26} \text{ W}$. The total luminosity of our Galaxy is $\text{L}_{\text{gal}} = 3.6 \times 10^{10} \text{ L}_\odot$.

Astronomers often measure timescales in years, the time it takes for the Earth to orbit the Sun: $1\text{yr} = 3.2 \times 10^7 \text{ s}$. However since this is very short in terms of cosmological timescales, gigayears are more often employed: $1\text{Gyr} = 10^9 \text{ yr} = 3.2 \times 10^{16} \text{ s}$.

While cosmology deals with very large measurements, it also deals with very small ones particularly in the early Universe when it was still hot and dense. This has introduced some terminology and units of particle physics to enter the realm of cosmology. For example, measurements of energy are sometimes given in units of **electron volts** (eV) instead of Joules: $1\text{eV} = 1.6 \times 10^{-19} \text{ J}$. For example, the rest energy of the electron is $m_e c^2 = 511,000 \text{ eV} = 0.51 \text{ MeV}$ while the proton has a rest energy of $m_p c^2 = 938.3 \text{ MeV}$.

A more universal, less culturally-biased system of units is the *Planck system* based on the universal constants G , c , and \hbar . Combining the Newtonian gravitational constant, $G = 6.7 \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}$, the speed of light, $c = 2.998 \times 10^8 \text{ m s}^{-1}$, and the reduced Planck constant, $\hbar \equiv (h/2\pi) = 1.1 \times 10^{-34} \text{ J s} = 6.6 \times 10^{-16} \text{ eV s}$, yields a unique length scale known as the **Planck length**:

$$\ell_P = \left(\frac{G}{\hbar c^3} \right)^{1/2} = 1.6 \times 10^{-35} \text{ [m].}$$

The same constants can be combined to form the **Planck mass**,

$$M_P = \left(\frac{\hbar c}{G} \right)^{1/2} = 2.2 \times 10^{-8} \text{ [kg],}$$

and the **Planck time**,

$$t_P = \left(\frac{G \hbar}{c^5} \right)^{1/2} = 5.4 \times 10^{-44} \text{ [s].}$$

Using Einstein's relation between mass and energy, we can also define the **Planck energy**,

$$E_P = M_P c^2 = 2.0 \times 10^9 \text{ J} = 1.2 \times 10^{28} \text{ [eV].}$$

By bringing the Boltzmann constant, $k_B = 8.6 \times 10^{-5} \text{ eV K}^{-1}$, into the act, we can also define the **Planck temperature**,

$$T_P = E_P / k_B = 1.4 \times 10^{32} \text{ [K].}$$

The current Standard Model for the Universe is the “Hot Big Bang” model, which states that the

Campbell Cosmo Intro

Universe has expanded from an initially hot and dense state to its current relatively cool and tenuous state, and that the expansion is still going on today. This Standard Model is based upon three observational pillars: (1) the Hubble diagram exhibiting expansion; (2) light element abundances which are in accord with Big Bang nucleosynthesis; and (3) the blackbody radiation left over from the first few hundred thousand years, the cosmic microwave background. Developments in the last two decades of the 20th century – both theoretical and observational – point to several aspects that require an understanding beyond the Standard model: the existence of dark matter and perhaps even dark energy, the evolution of perturbations around the zero order smooth Universe, and inflation, the generator of these perturbations. The theory encompassing all these Beyond the Standard Model ingredients – dark matter plus evolution of structure plus inflation – is called Cold Dark Matter, or CDM. The “Cold” part of this moniker comes from requiring the dark matter particles to be able to clump efficiently in the early Universe. If they are hot instead, i.e., have large pressure, structure will not form at the appropriate levels.

Since temperature scales with redshift as $T = (1+z)T_0$, the very early Universe was hot and dense. As a result, interactions among particles occurred much more frequently than they do today. As an example, a photon today can travel across the observable Universe without deflection or capture, so it has a mean free path greater than 10^{28} cm. When the age of the Universe was equal to 1 second, though, the mean free path of a photon was about the size of an atom. Thus in the time it took the Universe to expand by a factor of 2, a given photon interacted many, many times. These multiple interactions kept the constituents in the Universe in equilibrium in most cases. Nonetheless, there were times when reactions could not proceed rapidly enough to maintain equilibrium conditions. These times are – perhaps not coincidentally – of the utmost interest to cosmologists today. This out-of-equilibrium phenomena played a role in (i) the formation of the light elements during Big Bang nucleosynthesis; (ii) recombination of electrons and protons into neutral hydrogen when the temperature was of order 1/4 eV; and quite possibly in (iii) production of dark matter in the early Universe

Inflation was introduced partly to explain how regions which could not have been in causal contact with each other have the same temperature. It was soon realized that the very mechanism that explains the uniformity of the temperature in the Universe can also account for the origin of perturbations. Inflation predicts that quantum-mechanical perturbations in the very early Universe are first produced when the relevant scales are causally connected. Then these scales are whisked outside the horizon by inflation, only to reenter much later to serve as initial conditions for the growth of structure and anisotropy in the Universe. We are not actually sure that inflation is the mechanism that generated the initial perturbations as it is very difficult to test a theory based on energy scales well beyond the reach of particle accelerators. Nonetheless, it is by far the most plausible explanation and the next generation of CMB and large-scale structure observations will put inflation to some stringent tests. There is also no known scalar field which can drive inflation. Therefore, it may well be true that the idea of inflation is correct but it is driven by something other than a scalar field.

As the temperature of the Universe cools to 1 MeV, the cosmic plasma consists of (i) relativistic particles in equilibrium (i.e., photons, electrons and positrons), (ii) decoupled relativistic particles (i.e., neutrinos), and (iii) non-relativistic particles (i.e., baryons). The first simplification is that essentially no elements heavier than helium are produced at appreciable levels. So the only nuclei that need to be traced are hydrogen and helium, and their isotopes: deuterium, tritium, and ^3He . The second simplification is that, even in the context of this reduced set of elements, the physics splits up neatly into two parts since above $T \sim \text{MeV}$, no light nuclei form: only free protons and neutrons exist. The light elements in the Universe formed when the temperature of the cosmic plasma was of order 0.1 MeV. During **Big Bang nucleosynthesis** (BBN) where the primordial chemical elements are formed in the early Universe, roughly a quarter of the mass of the baryons is in the form of ^4He , the remaining in the form of free protons with only trace amounts of deuterium, ^3He , and lithium.

As the temperature drops to $T \sim 1 \text{ eV}$, photons remain tightly coupled to electrons via Compton scattering and electrons to protons via Coulomb scattering. It will come as no surprise that at these temperatures, there is very little neutral hydrogen. Energetics of course favors the production of neutral hydrogen with a binding energy of 13.6 eV, but the high photon/baryon ratio ensures that any hydrogen atom produced will be instantaneously ionized. These elements remain ionized until the temperature of the Universe drops well below the ionization energy of hydrogen. The **Epoch of Recombination** – at which time electrons and protons combine to form neutral hydrogen – is at redshift $z \sim 1,100$ corresponding to a temperature $T \sim 1,000 \text{ K}$ (or 0.25 eV). Before recombination, photons, electrons and protons are tightly coupled with one another because of Compton (the scattering of a photon by a charged particle, usually an electron) and Coulomb (elastic scattering of charged particles by the Coulomb interaction) scattering. After this time, photons travel freely through the Universe without interacting, so the photons in the CMB we observe today offer an excellent snapshot of the Universe at $z \sim 1,100$. The importance of this snapshot cannot be overstated. Our understanding of structure is based upon the observation of small perturbations in the temperature maps of the CMB. These indicate that the early

Campbell Cosmo Intro

Universe was inhomogeneous at the level of 1 part in 100,000. Over time the action of gravity causes the growth of these small perturbations into larger non-linear structures, which collapse to form sheets, filaments, and halos. These non-linear structures provide the framework within which galaxies form via the collapse and cooling of gas until the density required for star formation is reached.

The cosmic ‘**Dark Ages**’ is a period characterized by the absence of discrete sources of light via the first stars. The Λ CDM model predicts that nonlinear baryonic structure first emerges during this period, with virialized halos of dark and baryonic matter that span a range of masses from less than $10^4 M_\odot$ to about $10^8 M_\odot$ that are filled with neutral hydrogen atoms. The atomic density n_H and kinetic temperature of this gas T_K are high enough that T_K collisions populate the hyperfine levels of the ground state of these atoms in a ratio close to that of their statistical weights (3:1), with a spin temperature T_S that greatly exceeds the excitation temperature $T = 0.0681$ K. Since, as we shall show, $T_S > T_{\text{CMB}}$, the temperature of the cosmic microwave background (CMB), as well, for the majority of the halos, these “minihalos” can be a detectable source of redshifted 21 cm line emission. In addition to learning about galaxies and reionization, 21 cm observations have the potential to inform us about fundamental physics too. Part of the signal traces the density field giving information about neutrino masses and the initial conditions from the early epoch of cosmic inflation in the form of the power spectrum.

The emergence of the first sources of light in the Universe and the subsequent **Epoch of Reionization** of hydrogen mark the end of the Dark Ages. Despite its remote timeline, this epoch is currently under intense theoretical investigation and is beginning to be probed observationally. There are various reasons why studying this epoch is important. The first reason is that the reionization of hydrogen is a global phase transition affecting the range of viable masses for galaxies. Before reionization, small galaxies will be shielded by neutral hydrogen from ionizing UV radiation and therefore will be able to form more easily. After reionization and the establishment of a UV background, the formation of very small galaxies is hampered. The second reason to study this epoch is that it makes it possible to probe the power spectrum of density fluctuations emerging from recombination at scales smaller than are accessible by current cosmic microwave background experiments. Finally, in a Universe where structures grow hierarchically, the first sources of light act as seeds for the subsequent formation of larger objects. Thus, the third reason to study this period is that by doing so we may learn about processes relevant to the formation of the nuclei of present-day giant galaxies and perhaps on the connection between the growth of black holes and evolution of their host galaxies. Direct detection of Population III objects and of the first galaxies will be very challenging and it will be attempted by future deep imaging survey using techniques now in use at lower redshift, like the Lyman-break technique. Individual Population III stars could be detected most easily as supernovae. Early objects may leave a signature in the backgrounds that could either be detected directly or through a fluctuation analysis; detecting this signature may be simpler than detecting individual objects. Polarization measurements with a microwave background experiment like WMAP enable us to constrain the Thomson optical depth which is essentially a density-weighted number of free electrons along the line of sight. We can also probe directly the presence of neutral hydrogen by using the Gunn-Peterson trough and the properties of Lyman- α emitters. The Gunn-Peterson trough is essentially resonant Lyman- α absorption of the UV continuum of distant objects for wavelengths below that of Lyman- α . While diffuse neutral hydrogen present within some redshift interval will scatter the continuum, local hydrogen can scatter line emission and provide a somewhat complementary test to the Gunn-Peterson test. Gunn-Peterson trough constraints from distant quasars indicate that hydrogen is reionized at $z < 6$. Finally, a new promising area is that of 21 cm studies aiming at probing the distribution of neutral hydrogen at high redshift through detection of the 21 cm line emission or, in the most ambitious cases, of 21 cm line absorption over the cosmic microwave background.

The Hubble constant of the Benchmark Model is assumed to be $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$. The radiation in the Benchmark Model consists of photons and neutrinos. The photons are assumed to be provided solely by a CMB with current temperature $T_0 = 2.725$ K and density parameter $\Omega_{\gamma,0} = 5.0 \times 10^{-5}$. The energy density of the CNB is theoretically calculated to be 68% of that of the CMB, as long as neutrinos are relativistic. The matter content of the Benchmark Model consists partly of baryonic matter (that is, matter composed of protons and neutrons, with associated electrons), and partly of nonbaryonic dark matter. The evidence indicates that most of the matter in the Universe is nonbaryonic dark matter. The baryonic material that we are familiar with from our everyday existence has a density parameter of $\Omega_{b,0} \approx 0.04$ today. The density parameter of the nonbaryonic dark matter is roughly six times greater: $\Omega_{c,0} \approx 0.26$. The bulk of the energy density in the Benchmark Model, however, is not provided by radiation or matter, but by a cosmological constant, with $\Omega_{\Lambda,0} = 1 - \Omega_{m,0} - \Omega_{r,0} \approx 0.70$.

The Benchmark Model was first radiation-dominated, then matter-dominated, and is now entering into its lambda-dominated phase. Radiation gave way to matter at a scale factor $a_{rm} = \Omega_r, 0 / \Omega_m, 0 = 2.8 \times 10^{-4}$, corresponding to a time $t_{rm} = 4.7 \times 10^4$ yr. Matter, in turn, gave way to the cosmological constant at $a_{m\Lambda} = (\Omega_m, 0 / \Omega_{\Lambda,0})^{1/3} = 0.75$, corresponding to $t_{m\Lambda} = 9.8$ Gyr. The current age of the Universe, in the Benchmark Model, is $t_0 = 13.5$ Gyr.

Campbell Cosmo Intro

With $\Omega_{r,0}$, $\Omega_{m,0}$ and $\Omega_{\Lambda,0}$ known, the scale factor $a(t)$ can be computed numerically using the Friedmann equation. The transition from the $a \propto t^{1/2}$ radiation-dominated phase to the $a \propto t^{2/3}$ matter-dominated phase is not an abrupt one; neither is the later transition from the matter-dominated phase to the exponentially growing lambda-dominated phase. One curious feature of the Benchmark Model is that we are living very close to the time of matter-lambda equality.

1. COSMOLOGY (EARLY UNIVERSE, CMB, LARGE-SCALE STRUCTURE)

1.1. A Very Brief Primer on Cosmology

Just like in stars far too many question depend on the equations of stellar structure, in cosmology too many questions depend on the basic underpinnings of the FLRW universe. We will summarize the results below. This information comes from Ch. 4 and 5 of Ryden (2003).

1.1.1. The FLRW Universe

In accordance with the cosmological principle (that there be a set of observers that see the universe as homogeneous and isotropic), the spatial extent of the universe must have uniform curvature (unless we move to truly non-trivial geometries). This restricts our metric to be of a form known as the Robertson-Walker metric

$$ds^2 = cdt^2 - a(t)^2 \left(\frac{dx^2}{1 - \kappa x^2/R^2} + x^2 d\Omega^2 \right) \quad (1)$$

where $\kappa = -1, 0, 1$ and R scales κ . Another way of writing this metric (and making it perhaps more palatable) is

$$ds^2 = cdt^2 - a(t)^2 (dr^2 + S_\kappa(r)d\Omega^2) \quad (2)$$

where

$$S_\kappa = \begin{cases} R \sin(r/R) & \text{if } \kappa = 1 \\ r & \text{if } \kappa = 0 \\ R \sinh(r/R) & \text{if } \kappa = -1 \end{cases}, \quad (3)$$

Writing the metric in this form shows that if κ is non-zero, angular lengths are either decreased (for positive curvature) or increased (for negative). Just as importantly, this metric indicates that, like in Minkowski space, time is orthogonal to position (meaning we can foliate the spacetime such that each hypersurface slice can be associated with a single time t) and radial distances are independent of curvature S_κ .

The solution to the RW metric is known as the Friedmann-Lemaître equation, and describes how the scale factor $a(t)$ changes with time:

$$\left(\frac{\dot{a}}{a} \right)^2 = H^2 = \frac{8\pi G}{3}\rho - \frac{\kappa c^2}{R^2} \frac{1}{a^2} + \frac{\Lambda c^2}{3} \quad (4)$$

where ρ is the matter-energy density of the universe, $\kappa c^2/R^2$ the curvature term, λ the cosmological constant and a the scale factor in the RW metric. This expression can also be derived (but becomes unscrutable because the terms make little sense) by representing the universe by a homologously expanding sphere (i.e. an explosion at $t = 0$), and considering the dynamics of a test mass within this universe. If $r = r_0 a$ (noting this automatically produces $v = r_0 \dot{a}$, so $v/r = \frac{\dot{a}}{a} = H$, reproducing Hubble's law), we can integrate $\frac{d^2r}{dt^2} = \frac{4\pi G \rho}{3} - \frac{\kappa c^2}{R^2 a^2}$ ¹ to get Eqn. 4 (with Λ subsumed into a constant of integration). Even more easily, we can do the same with energy balance, $K + U = E$.

In a Λ -free universe, if a value of H^2 is given, ρ and κ/R^2 are linked, and there is a critical density

$$\rho_c = \frac{3H^2}{8\pi G}, \quad (5)$$

for which the universe is flat. We define $\Omega \equiv \rho/\rho_c$. We can then rewrite the FL solution as $1 - \Omega = \frac{-\kappa c^2}{R^2 a^2 H^2}$. Note that the right side cannot change sign! This means that if at any time $\rho > \rho_c$, the universe will forever be closed; if $\rho < \rho_c$, the universe will forever be open, and if $\rho = \rho_c$, the equality will hold for all time and the universe will be flat.

1.1.2. The Fluid and Acceleration Equations

From the first law of thermodynamics and an assumption of adiabaticity, $dE + PdV = 0$, $\dot{V}/V = 3\dot{a}/a$, and $dE/dt = \rho c^2 dV/dt + c^2 V d\rho/dt$. This gives us

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + P/c^2). \quad (6)$$

This can be combined with the FL equation to get

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3} \left(\rho + 3\frac{P}{c^2} \right) + \frac{\Lambda c^2}{3} \quad (7)$$

¹ Recall that $2\ddot{a}\dot{a} = \frac{d}{dt}\dot{a}^2$.

1.1.3. Equations of State

Cosmologists generally use equations of state that look like

$$P = \omega c^2 \rho. \quad (8)$$

The ideal gas law, for example, has an ω that is dependent on temperature, and therefore time. “Dust”, which is pressure-free, has $\omega = 0$ - stars exert little enough pressure to be considered a dust. A relativistic equation of state always has $P = \frac{1}{3}\rho c^2$, including photons. Dark energy has $\omega = -1$. In substances with positive ω , $\sqrt{\frac{P}{\rho}} = c_s \leq c$, which restricts $\omega \leq 1$.

Combining Eqn. 6 with $P = \omega c^2 \rho$ gives us $\rho = \rho_0 a^{-3(1+\omega)}$. From this we determine that matter density $\rho = \rho_0 a^{-3}$ and radiation density $\rho_r = \rho_{r,0} a^{-4}$. We can compare the densities of any component to the critical density to obtain Ω . For example, $\rho/\rho_c = \rho/(3H^2/8\pi G) = \frac{8\pi G}{3}\rho/H^2$. We then note that $\frac{8\pi G}{3}\rho = \frac{8\pi G}{3}\rho \frac{\rho_0}{\rho_0} = H_0^2 \Omega_{m,0} a^{-3}$ - conversions such as this will be useful in the following section. Note that $\rho_\Lambda = \frac{\Lambda c^2}{8\pi G}$, giving $\Omega_\Lambda = \frac{\Lambda c^2}{3H^2}$.

Taking ratios of Ω s gives us the energy component that dominates (ex. radiation to matter is $\Omega_r/\Omega_m = \rho_{r,0}/\rho_{m,0} \frac{1}{a} \approx \frac{1}{3600a}$ if $a_0 = 1$, indicating there was a time when radiation dominated the energy of the universe).

A related question is how the radiation field temperature scales with time. Assuming adiabaticity, $dQ = dE + PdV$ (the work is being done by the radiation field on the universe). Since $P = \frac{1}{3}U = \frac{1}{3}aT^4$ we obtain $\frac{1}{T} \frac{dT}{dt} = -\frac{1}{3V} dVdT$, which implies (through integration and the fact that V scales like a^{-3}) that $T \propto a^{-1}$.

1.1.4. History of Expansion

Let us consider several possibilities:

- An empty flat universe is static. An empty, open universe goes like $a = ct/R$. An empty, closed universe is impossible.
- A flat universe with a single component would have $\dot{a}^2 = \frac{8\pi G}{3}\rho_0 a^{-(1+3\omega)}$. This gives $a \propto t^{2/(3+3\omega)}$.
- A Λ -dominated universe would have $\dot{a}^2 = \frac{\Lambda c^2}{3}a^2$, which gives $a \propto \exp(\sqrt{\Lambda c^2/3}t)$. We note we could have snuck Λ into the energy density of the universe if we set $\omega = -1$ and $\rho_0 = c^2/8\pi G$.

We may now consider a universe with radiation, stars, and a cosmological constant. Since $\kappa/R^2 = \frac{H_0^2}{a^2}(\Omega_0 - 1)$ (so that R may be written as $\frac{c}{H_0} \sqrt{|\Omega_\kappa|}$), we can actually write the FL equation as $H^2 = \frac{8\pi G}{3}\rho - \frac{H_0^2}{a^2}(\Omega_0 - 1)$, where ρ includes radiation, matter and dark energy, and if we divided by H_0^2 , we get

$$\frac{H^2}{H_0^2} = \frac{\Omega_{r,0}}{a^4} + \frac{\Omega_{m,0}}{a^3} + \Omega_{\Lambda,0} + \frac{1 - \Omega_0}{a^2} \quad (9)$$

where $\Omega_0 = \Omega_{r,0} + \Omega_{m,0} + \Omega_{\Lambda,0}$. Note how the curvature still responds to the total matter-energy density in the universe, but the expansion history may now be altered by Λ . Assuming that radiation is negligible, Fig. 16 describes the possible fates and curvatures of the universe.

Using measured values of the Ω s, we find the universe to have the expansion history given in Fig. 2.

1.1.5. Distance and Size Measurements

The redshift z is given by

$$1 + z = \frac{\lambda_0}{\lambda_e} = \frac{a_0}{a_e} \quad (10)$$

where subscript e stands for “emitted”.

Taylor expanding the current a , we obtain $a(t) \approx a(t_0) + \dot{a}|_{t=t_0}(t - t_0) + \frac{1}{2}\ddot{a}|_{t=t_0}(t - t_0)^2$. Dividing both sides by a_0 (which is equal to 1) we get $1 + H_0(t - t_0) - \frac{1}{2}q_0 H_0^2(t - t_0)^2$. $q_0 = -\ddot{a}/aH^2|_{t=t_0}$ is known as the deceleration parameter, and helps constrain the makeup of the universe, since $q_0 = \frac{1}{2} \sum_\omega \Omega_\omega (1 + 3\omega)$.

The comoving distance (interpretable as how distant the object would be today) to an object whose light we are seeing is given by

$$d_c(t_0) = c \int_{t_e}^{t_0} \frac{dt}{a}, \quad (11)$$

which can be converted into $d_c = c \int_0^z \frac{dz}{H}$. Since radial distances are not curvature-dependent, the proper (physical) distance is simply given by (Davis & Lineweaver 2004)

$$d_p(t) = a(t)d_c \quad (12)$$

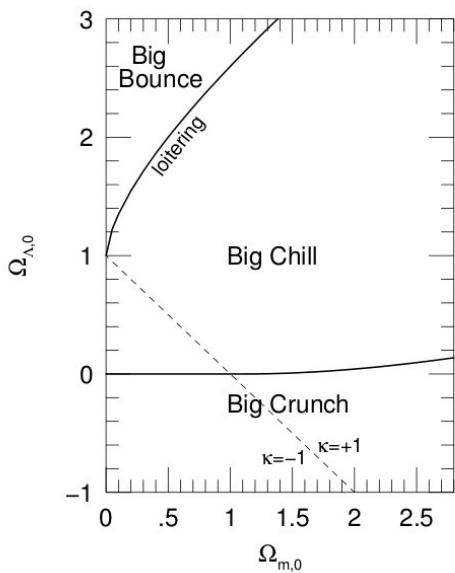


FIG. 1.— Fate of the universe as a function of Ω_m and Ω_Λ . From Ryden (2003), her Fig. 6.3.

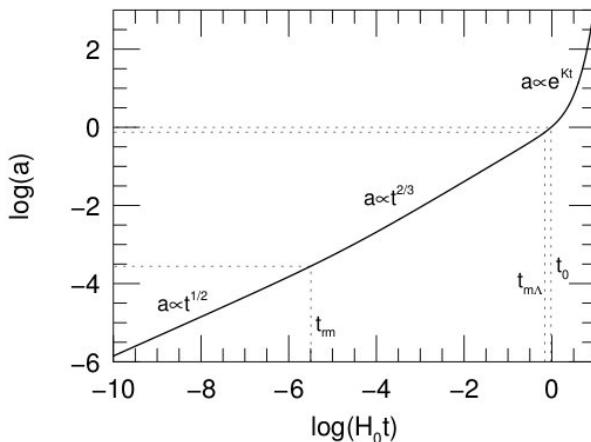


FIG. 2.— Fate of the universe, using measured values of Ω_i . From Ryden (2003), her Fig. 6.5.

where $a_0 = 1$ is assumed, and t represents time since the beginning of the universe. The luminosity distance is defined as

$$d_L = \sqrt{\frac{L}{4\pi F}} = S_\kappa(r)(1+z). \quad (13)$$

The second expression is due to two factors - first, the expansion of space drops energy with $1+z$, and increases the thickness of any photon shell dr by $1+z$ as well. The area covered by the wave of radiation is $4\pi S_\kappa^2(r)$ ($S_\kappa = r$ for a flat universe), where r should be interpreted as the comoving distance d_c . The angular diameter distance $d_A = \frac{l}{d\theta}$ (l is the physical length of an object at the time the light being observed was emitted) is given by the fact that $ds = a(t_e)S_\kappa(r)d\theta$. If the length l is known, then $ds = l$ and we obtain

$$d_A = \frac{S_\kappa}{1+z}. \quad (14)$$

Note that the angular diameter distance is related to the luminosity distance by $d_A = d_L/(1+z)^2$. For $z \rightarrow 0$, all these distances are equal to cz/H_0 , but at large distances they begin to differ significantly.

1.2. Question 1

Q1) RECOMBINATION

What is recombination? At what temperature did it occur? Explain why this does not match the ionization potential of hydrogen.

Short Answer

(From Campbell Answers)

Recombination refers to the time at which the temperature of the early Universe became cool enough such that it was thermodynamically favorable for the ionized plasma of free electrons and ions to couple and form neutral atoms. Numerically, this might be defined as the moment when the number density of ions is equal to the number density of neutral atoms.

Temperature: $T \sim 1,000\text{ K}$ (corresponding to an energy of $\sim 0.1\text{ eV}$) at a redshift of $z \sim 1,100$. This does not match the ionization potential of hydrogen because the early Universe (as it was hot and dense) can be described by a blackbody with a characteristic distribution of photon energies including an exponential tail of high energy photons (Wein's tail). While the peak of the blackbody spectrum describing the temperature of the early Universe is below the ionization energy of hydrogen, the photons in the high-energy exponential tail of the blackbody spectrum have sufficient energies for photoionization.

Cosmology Class Fall 2019: Martine's Notes

Steps: Integrate blackbody to find what percentage of photons have energies higher than 13.6 eV. Extrapolate measured Ω_b back in time to get the photon to baryon ratio. The number of photons above 13.6 eV equals number of baryons at 5600 K.

$$X_p = \frac{N_p}{N_p + N_H} \approx 0.1 \quad (1.1)$$

at T_{rec} .

Saha equation:

$$\frac{X^2}{1 - X} = \frac{(2\pi m_e k T)^{3/2}}{(n_e + n_H)(2\pi h)^3} e^{-13.6eV/kT} \quad (1.2)$$

At what temperature is X 10%? Depends on $\Omega_b h^2$ because of the $(n_e + n_H)$ in denominator. Higher density of baryons makes recombination happen earlier. Solving for when X_p is 0.1 gives $T \sim 3600\text{K}$ or 0.3 eV. This is recombination.

Then ask, when will the universe be 'transparent'? Can define that by when optical depth $\tau = 1$. This comes out to be $T \sim 3200\text{K}$, $z \sim 1100$. This is decoupling.

Barth's Follow Up Questions Approximately how long did the process of recombination take (define however is convenient)?

Q1) Ludwig Cosmo Q1

Question 1 - Recombination

What is recombination? At what temperature did it occur? Explain why this does not match the ionization potential of hydrogen.

Relevant Equations

- $E = k_b T$ Thermal Energy
- $\frac{n_p n_e}{n_e} \propto T^{3/2} e^{-1/T}$ Saha Equation
- $x_e = \frac{n_e}{n_p + n_H}$ Free Electron Fraction

Solution

What is recombination?

- The epoch where a sea of electrons, atomic nuclei, and photons became neutral atoms for the first time.
- Temperature dropped due to expansion. Photoionization no longer occurred often enough to maintain the ionization fraction of Hydrogen.

At what temperature did it occur?

- Assuming that the reaction happens fast enough to keep things in thermal equilibrium we can use the Saha equation to relate the ionization fraction to the temperature.

$$\frac{n_p n_e}{n_e} \propto T^{3/2} e^{-1/T}$$

- This can be further simplified if we require that $n_e = n_p$ and we use the free electron fraction. The temperature ends up being about $\sim 1/4$ eV or

$$13.6 \text{ eV} \frac{1}{8.6e - 5 \text{ eV}} K \approx 150000 K$$

Explain why this does not match the ionization potential of hydrogen

- The ionization potential of hydrogen is 13.6 eV or

$$13.6 \text{ eV} \frac{1}{8.6e - 5 \text{ eV}} K \approx 150000 \text{ kelvin}$$

- The temperature of recombination is significantly lower than this because CMB photons aren't uniform. The cmb is very close to a blackbody spectrum which has an exponential tail known as the Wien's Tail. This tail contains a non negligible number of photons which are still able to ionize hydrogen. This is made worse by the fact that photons also outnumbered baryons $10^9 : 1$ because of baryogenesis and the resultant particle/antiparticle annihilation.

Follow Up

- The process was not instantaneous. As recombination progressed, the number of free electrons available for Thomson scattering which kept photons and baryons in thermal equilibrium, decreased. This caused a drop in the opacity for the photons until the optical depth was low enough ($\tau = 1$) that photons could stream freely through the universe: The CMB. This corresponds to the surface of last scattering where radiation and matter decoupled.
- If the photon:baryon ratio were higher there would be more ionization and recombination would happen later.
- If the ionization energy of Hydrogen was lower, it would take longer to cool off significantly past this temperature due to wien's tail so recombination would happen later.

Q1) Herman Cosmo Q1

C1

- expansion $\rightarrow \downarrow T \rightarrow \downarrow$ photoionization $\rightarrow \uparrow \text{HI}$
- Wien's tail of CMB blackbody spectrum
- $\uparrow \text{HI} \rightarrow \downarrow \text{free } e^- \rightarrow \downarrow \text{Thomson scattering} \rightarrow \downarrow \text{opacity}$
 $\rightarrow \text{photon-matter decouple} \rightarrow \text{CMB}$

- Recombination

- Recombination was the period ($t \sim 380 \text{ Kyr}$, $z \sim 1/100$) during which the universe transitioned from a sea of e^- 's and atomic nuclei (and γ 's) to neutral atoms as the temperature ($T \sim 3700 \text{ K}, 0.3 \text{ eV}$) decreased due to expansion such that photoionization no longer occurred often enough to maintain the ionization fraction of H. The Saha eqn relates ionization fraction to T:

$$\frac{n_{\text{H}^+}}{n_i} \propto T^{3/2} \exp(-1/T)$$

- The temperature had to fall far below the ionization T of H ($\sim 10^5 \text{ K}$, 13.6 eV) because the Wien's tail of the CMB blackbody distribution still contained a non-negligible # of γ 's able to ionize H, made even worse by the fact that γ 's outnumbered baryons $\sim 10^9 : 1$ (due to baryogenesis and resultant particle/antiparticle annihilation).

- The process was not instantaneous. As recombination progressed, the # of free e^- 's available for Thomson scattering, which kept γ 's and baryons in thermal equilibrium, decreased. This caused a drop in the opacity for the γ 's until the optical depth was low enough ($\tau \approx 1$) that photons could stream freely through the universe: The CMB. This corresponds to the surface of last scattering, when radiation and matter decoupled.

- If the γ :baryon ratio was higher, there would be more ionization and thus recombination would happen later

- If the ionization energy of H was lower, it would take longer to cool off significantly past this temperature (due to Wien's tail), so recombination would happen later.

Q1) Campbell Cosmo Q1

1.2 Question 1

What is recombination? At what temperature did it occur? Explain why this does not match the ionization potential of hydrogen.

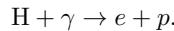
1.2.1 Short answer

Recombination¹ refers to the time at which the temperature of the early Universe became cool enough such that it was thermodynamically favourable for the ionized plasma of free electrons and ions to couple and form neutral atoms. Numerically, this might be defined as the moment when the number density of ions is equal to the number density of neutral atoms. This occurred at a temperature of $T \sim 1,000\text{ K}$ (corresponding to an energy of $\sim 0.1\text{ eV}$) at a redshift of $z \sim 1,100$. This does not match the ionization potential of hydrogen because the early Universe (as it was hot and dense) can be described by a blackbody with a characteristic distribution of photon energies including an exponential tail of high energy photons (Wein's tail). While the peak of the blackbody spectrum describing the temperature of the early Universe is below the ionization energy of hydrogen, the photons in the high-energy exponential tail of the blackbody spectrum have sufficient energies for photoionization.

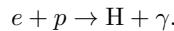
1.2.2 Additional context

The term *recombination* is a common misnomer because this was in fact the first time that free electrons combined with ions to produce a neutral medium – some cosmologists hold that “combination” would have been a more appropriate term. Prior to this time, the mean free path of the photons against scattering off the free electrons is much less than the Hubble distance. This means that gravitational forces attempting to compress the plasma must also increase the photon density. This produces an increase in temperature and hence in radiation pressure. Any perturbation in the baryon-photon plasma thus behaves as an acoustic wave.

The ionization energy of hydrogen is 13.6 eV . A photon γ with an energy exceeding this ionization energy is capable of ionizing a hydrogen atom to produce a free electron e and proton p through the following process:



This process can of course run in reverse through the process of *radiative recombination* where a free electron and proton can combine to produce a bound neutral hydrogen atom:



A crude approximation for the temperature at which recombination occurred could be made by assuming that the average photon energy must have been at least the ionization energy of hydrogen. Since the current temperature of the CMB is 2.7 K , this should yield a lower limit on the recombination temperature:

$$T_{\text{rec}} \sim \frac{E_{\text{ion}}}{E_{\text{CMB}}} = \frac{13.6\text{ eV}}{k_B \cdot 2.7\text{ K}} \sim 60,000\text{ [K].}$$

Clearly this is a very crude estimate since this is off by an order of magnitude, the recombination temperature being closer to $T \sim 1,000\text{ K}$. This is because this doesn't take into consideration the fact that the CMB photon energies are not single-valued as assumed above. When a dense, opaque object is in thermal equilibrium (such as the early Universe), the distribution of photon energies only depends on temperature following the blackbody function (or Planck function):

$$B_\nu(T) = \frac{2h\nu^3}{c^2} \frac{1}{\exp(h\nu/k_B T) - 1} \text{ [erg s}^{-1} \text{ cm}^{-2} \text{ Hz}^{-1} \text{ sr}^{-1}\text{].} \quad (1)$$

Figure 1 shows blackbody spectra for various temperatures in log-log space, where the peak of each blackbody spectrum is related to its associated temperature via $h\nu_{\text{peak}} \approx 2.82 k_B T$. While the mean photon energy is $2.82 k_B T$, approximately one in every 500 photons will have energies exceeding $10 k_B T$, one in 3 million will exceed $20 k_B T$, and one in 30 billion will exceed $30 k_B T$. While only a small fraction of CMB photons are found in the high-energy tail-end of the Planck distribution, the overall number of CMB photons is enormous – outnumbering baryons by nearly 2 billion to one. Therefore, the vast number of CMB photons surrounding neutral hydrogen atoms greatly increases the probability of photoionization, even with a mean photon energy less than the ionization energy of hydrogen.

¹A complete misnomer as the plasma has always been completely ionized up to this point

Q1) Campbell Cosmo Q1

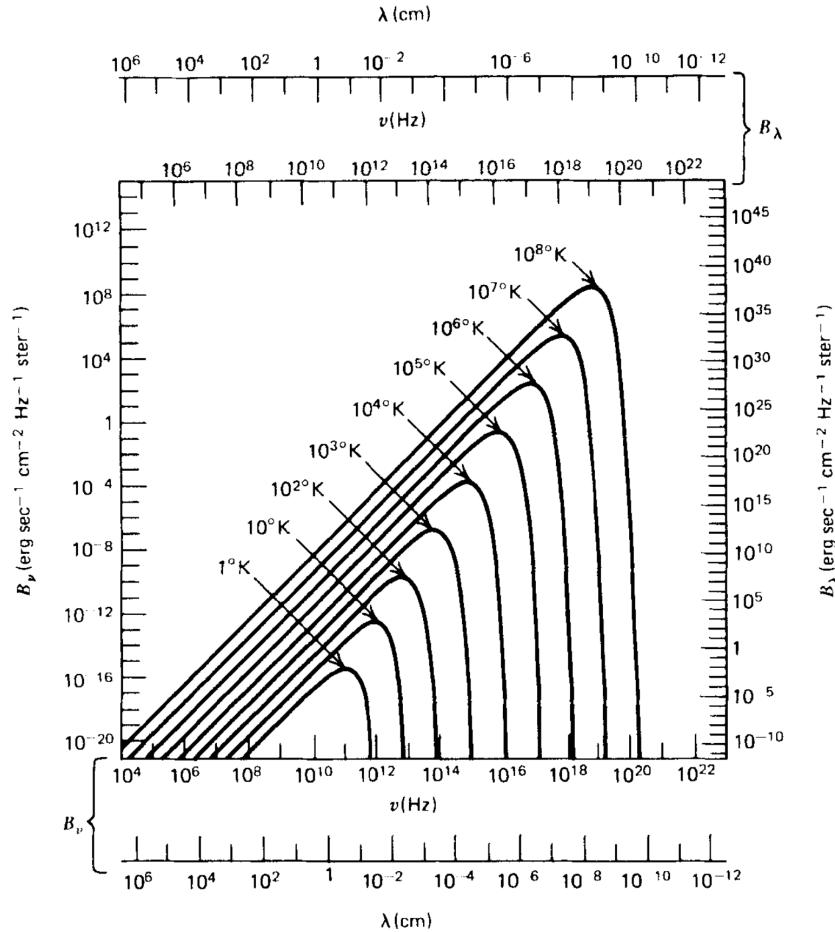


Figure 1: Blackbody spectra at various temperatures. Figure taken from Ryden (2017).

1.2.3 Follow-up Questions

- How is the CMB related to recombination?
- Around how long ago or at what redshift did this occur?
- What is the last scattering surface?
- Is the surface of last scattering a well-defined surface (i.e., did recombination happen suddenly)?

Q1) Zhu Cosmo Q1

10

QUESTION: What is recombination? At what temperature did it occur? How does this relate to the ionization potential of Hydrogen?

Most of this information comes from Ryden (2003), filtered through Emberson (2012). Subscript 0 will represent present-day values.

Recombination is when the universe cooled to the point at which protons combined with electrons to form hydrogen atoms. The resulting massive decrease in opacity caused the universe to become optically thin, and the photon field of the universe decoupled from its matter counterpart². Recombination does not refer to a single event or process: the epoch of recombination is the time at which the baryonic component of the universe went from being ionized to being neutral (numerically, one might define it as the instant in time at which the number density of ions is equal to the number density of neutral atoms). The epoch of photon decoupling is the time at which the rate at which photons scatter from electrons becomes smaller than the Hubble parameter (at the time). When photons decouple, they cease to interact with the electrons, and the universe becomes transparent. Third, the epoch of last scattering is the time at which a typical CMB photon underwent its last scattering from an electron. The three processes are related because hydrogen opacity is the driver of all three.

For simplification, let us assume the universe is made completely of hydrogen atoms. ${}^4\text{He}$ has a higher first ionization energy significantly higher than that of H, and therefore helium recombination would have occurred at an earlier time.

The degree of H ionization is determined by the Saha equation, which can be derived (especially for H, where it is easy) from the grand partition function Ξ (Carroll & Ostlie 2006),

$$\frac{N_{i+1}}{N_i} = \frac{2Z_{i+1}}{n_e Z_i} \left(\frac{m_e k_B T}{2\pi\hbar^2} \right)^{3/2} e^{-\chi_i/k_B T} \quad (15)$$

where i indicates the degree of ionization, χ_i is the ionization energy from degree i to degree $i + 1$. Suppose we ignore excited internal energy states (note that Ryden does this, but does not make it explicit); then $Z_{H^+} = Z_p = g_p = 2$ and $Z_H = g_H = 4$, for all possible spin states of the nucleus and electron. This gives us (multiplying the lefthand side of Eqn. 15 by V/V)

$$\frac{n_H}{n_p} = n_e \left(\frac{m_e k_B T}{2\pi\hbar^2} \right)^{-3/2} e^{\chi/k_B T} \quad (16)$$

where $\chi = 13.6$ eV and $n_e = n_p$. Using the fact that the number of photons is $0.243 \left(\frac{k_B T}{\hbar c} \right)^3$ and $n_p = n_\gamma \eta$, with $\eta \approx 5.5 \times 10^{-10}$ (this does not change by much throughout the era of recombination), we may eliminate n_e and solve for when the left side is equal to one, i.e. when X , the ionization fraction, is equal to 1/2. This gives us $T = 3740$ K, $z = 1370$ (0.24 Myr for a matter dominated and flat universe). Past this temperature, photons became too cold to ionize H. The evolution of X with redshift is given in Fig. 3.

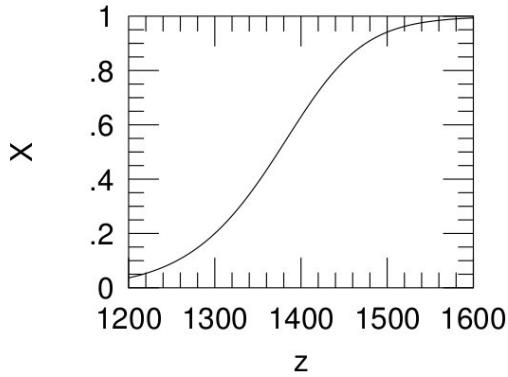


FIG. 3.— Change in ionized fraction X as a function of redshift. From Ryden (2003), her Fig. 9.4.

1.2.1. Couldn't photons have decoupled from baryons before recombination?

Photons are coupled to baryons through photoionization and recombination, though in the very hot universe the dominant interaction would have been Thomson scattering off free electrons, with a rate given by $\Gamma = n_e \sigma_e c$. Values can be calculated using $n_e = \frac{0.22m^3}{a^3}$ (if $a = 1$ today) and $\sigma_e = 6.65 \times 10^{-29}$. Photons decouple (gradually) from baryons when Γ exceeds H , equivalent to saying the mean free path λ exceeds c/H . The critical point $\Gamma = H$ occurred at $z \approx 42$, $T \approx 120$ K, long past recombination.

² Since atoms were always ionized before this, “recombination” is almost a misnomer!

If we perform the same calculation during the era of recombination, setting n_e using the analysis above and obtaining H from $H(t) = H_0 \Omega_m \frac{a_0^3}{a^3} = H_0 \Omega_m \frac{1}{(1+z)^3}$, we obtain $z \approx 1100$ and $T \approx 3000$ (exact answers are difficult without modelling, since during the final stages of recombination the system was no longer in LTE).

1.2.2. What is the last scattering surface?

The last scattering surface is the $\tau = 1$ surface for photons originally trapped in the optically thick early universe. The age $t_0 - t$ of this surface can be found using

$$\tau = 1 = \int_t^{t_0} \Gamma(t) dt \quad (17)$$

In practice this is difficult, and so we again estimate that $z \approx 1100$ for last scattering.

Q2) FLAT UNIVERSE PROPERTIES

The universe is said to be “flat”, or, close to flat. What are the properties of a flat universe and what evidence do we have for it?

Short Answer

The three properties of a flat universe are:

1. parallel lines never converge or diverge
2. angles in a triangle sum to 180 degrees
3. the universe is at the critical density

The first two are geometric properties of a flat space. The third is a statement of the universe’s energy content.

The Friedmann Equation is $H(t)^2 = \frac{8\pi G}{3} \left[\rho(t) + \frac{\rho_{crit} - \rho_0}{a(t)^2} \right]$, $\rho_k \equiv \rho_{crit} - \rho_0$.

Therefore, $\rho_k = 0 \rightarrow \rho_0 = \rho_{crit}$

There is a lot of evidence for a flat universe

1. direct measurements of the universe’s density: CMB power spectrum, galaxy cluster distribution.
2. Angular scale of CMB peak: positive curvature moves peak to larger angular scales (lower l). Vice versa for negative curvature.
3. map the expansion rate with Type Ia SNR.

Q2) Ludwig Cosmo Q2

Question 2 - Flat Universe

The universe is said to be “flat”, or, close to flat. What are the properties of a flat universe and what evidence do we have for it?

Relevant Equations:

- $H(t)^2 = \frac{8\pi G}{3} \left[\rho(t) + \frac{\rho_{cr} - \rho_0}{a(t)^2} \right]$ Friedmann Equation
- $H(a)^2 = H_0^2 \left[\Omega_r a^{-4} + \Omega_m a^{-3} + \Omega_k a^{-2} + \Omega_\Lambda \right]$ Friedmann Equation
- $\rho_c = \frac{3H_0^2}{8\pi G}$ Critical Density

Solution

What are the properties of a flat universe?

- Parallel Lines do not Converge or Diverge
- Sum of Angles in a triangle = 180 deg
- Density of the universe is the critical density

What evidence do we have for it?

- By measuring the density of the universe we can determine the curvature. We can probe the different components many ways (CL Spectrum of CMB, galaxy cluster distribution, etc.). We can also look at the angular scale of CMB fluctuations which depend on the universe's geometry.
- If the universe were positively curved, the peak of the CMB Cl power spectrum would occur at larger angular scales (lower multipoles)
- If it was negatively curved, the peaks would be at smaller angular scales (higher multipoles).
- Since the expansion rate evolves differently for different densities we can also use Type Ia SNR to map distance to z and infer Ω_0 .
- From the Friedman Eq the density of the universe determines its geometry and tells us how the scale factor evolves over time.

$$\frac{H^2}{H_0^2} = \Omega_r a^{-4} + \Omega_m a^{-3} + \Omega_k a^{-2} + \Omega_\Lambda$$

- $\Omega_0 = \Omega_r + \Omega_m + \Omega_\Lambda, \quad 1 - \Omega_0 = \Omega_k \rightarrow \text{ if } \Omega_0 = 1, \Omega_k = 0$

Follow Up

- The universe was likely much flatter in the early universe as even slight deviations in the density would've changed our expansion significantly. Inflation solves this flatness problem by flattening everything out through exponential expansion, so it doesn't matter how curved the universe was pre inflation.

Q2) Herman Cosmo Q2

C2

Flat Universe

- density of universe determines its geometry
- for flat universe, $\Omega_0 = 1$ and angles $\approx 1^\circ$
- Measure density components of universe (SNe, TSZ, BAOs, grav. lensing), and angular size of CMB fluctuations.

Properties

- parallel lines don't converge or diverge $\parallel \cap \cap$
- sum of angles in triangle = 180° \triangle
- density of the universe = critical density ($\rho_c = \frac{3H^2}{8\pi G}$; $\Omega_0 = \frac{\rho_0}{\rho_c} = 1$)

* By the Friedmann eqn, the density of the universe determines its geometry and tells us how the scale factor evolves over time.

$$\frac{H^2}{H_0^2} = \Omega_r a^{-4} + \Omega_m a^{-3} + \Omega_k a^{-2} + \Omega_\Lambda$$

$$\Omega_0 = \Omega_r + \Omega_m + \Omega_\Lambda, \quad 1 - \Omega_0 = \Omega_k \rightarrow \text{if } \Omega_0 = 1, \Omega_k = 0 \rightarrow \text{flat!}$$

* By measuring the density of the universe we can determine the curvature. We can probe the different components many ways (CMB spectrum of CMB, galaxy cluster distribution, etc). We can also look at the angular scale of CMB fluctuations, which depend on the universe's geometry.

- if the universe was \oplus ly curved, the peak of the CMB CMB power spectrum would occur at larger angular scales (lower multipoles) \odot
and if it was \ominus ly curved, the peaks would be at smaller angular scales (higher multipoles) Δ .

- since the expansion rate evolves differently for different densities, we can also use Type Ia SNe to map distance to z and infer Ω_0 .

* The universe must have been even flatter in the early universe, as even slight deviations in the density would've changed our expansion significantly by now. Inflation solves this flatness problem by flattening everything out through exponential expansion, so it didn't matter how curved the universe was pre-inflation.

$$\Omega = 1 + \left(\frac{K c_{te}}{a(t)} \right)^2 \quad \text{if } a \propto e^t, \Omega \approx 1, \text{ so } K \text{ isn't important.}$$

Hilary

Q2) Campbell Cosmo Q2

1.3 Question 2

The Universe is said to be “flat,” or, close to flat. What are the properties of a flat Universe and what evidence do we have for it?

1.3.1 Short answer

There are three simple properties of a flat Universe: (1) parallel lines never converge nor diverge; (2) the sum of all energy densities is equal to its critical value; and (3) the sum of angles within a triangle is always 180° . Two common techniques for measuring the curvature (i.e., topology) of the Universe include measuring the total energy density of the Universe (Ω_0), and using the main peak in the CMB angular power spectrum (C_ℓ) as a standard ruler for the size of the sound horizon at the surface of last scattering.

1.3.2 Additional context

In developing a mathematical theory of general relativity, in which spacetime geometry/curvature is related to the mass-energy density, Einstein needed a way of mathematically describing curvature. Since picturing the curvature of a four-dimensional spacetime is, to say the least, difficult, let’s start by considering ways of describing the curvature of two-dimensional spaces, then extend what we have learned to higher dimensions.

The simplest of two-dimensional spaces is a plane, on which Euclidean geometry holds (see (a) of Figure 2). On a plane, a geodesic is a straight line. If a triangle is constructed on a plane by connecting three points with geodesics, the sum of the angles made with its vertices obeys:

$$\Delta_{\text{flat}} = \alpha + \beta + \gamma = \pi \quad [\text{rad}].$$

Note that a plane has infinite area, and has no upper limits on the possible distance between points.

Now consider another simple two-dimensional space, the surface of a sphere (see (b) of Figure 2). On the surface of a sphere, a geodesic is a portion of a great circle – that is, a circle whose center corresponds to the center of the sphere. If a triangle is constructed on the surface of the sphere by connecting three points with geodesics, the sum of the angles made with its vertices obeys:

$$\Delta_{\text{closed}} = \alpha + \beta + \gamma = \pi + A/R^2 \quad [\text{rad}],$$

where A is the area of the triangle, and R is the radius of the sphere. All spaces in which $\alpha + \beta + \gamma > \pi$ are called “positively curved” spaces. The surface of a sphere is a positively curved two-dimensional space. Moreover, it is a space where the curvature is homogeneous and isotropic; no matter where you draw a triangle on the surface of a sphere, or how you orient it, it must always satisfy the above equation for Δ_{closed} . Note that a sphere has a maximum possible distance between points; the distance between antipodal points, at the maximum possible separation, is πR .

In addition to flat spaces and positively curved spaces, there exist negatively curved spaces. An example of a negatively curved two-dimensional space is the hyperboloid, or saddle-shape (see (c) of Figure 2). If a triangle is constructed on this surface by connecting three points with geodesics, the sum of the angles made with its vertices obeys:

$$\Delta_{\text{open}} = \alpha + \beta + \gamma = \pi - A/R^2 \quad [\text{rad}],$$

where A is the area of the triangle, and R is the radius of curvature. All spaces in which $\alpha + \beta + \gamma < \pi$ are called “negatively curved” spaces. The surface of a hyperboloid is a negatively curved two-dimensional space. A surface of constant negative curvature has infinite area, and has no upper limit on the possible distance between points.

If you want a two-dimensional space to be homogeneous and isotropic, there are only three possibilities that fit the bill: the space can be uniformly flat, it can have uniform positive curvature, or it can have uniform negative curvature. Thus, if a two-dimensional space has curvature which is homogeneous and isotropic, its geometry can be specified by two quantities, κ , and R . The number κ , called the curvature constant, is $\kappa = 0$ for a flat space, $\kappa = +1$ for a positively curved space, and $\kappa = -1$ for a negatively curved space. If the space is curved, then the quantity R , which has dimensions of length, is the radius of curvature.

The results for two-dimensional space can be extended straightforwardly to three dimensions. A three-dimensional space, if its curvature is homogeneous and isotropic, must be flat, have uniform positive curvature, or have uniform negative curvature. If a three-dimensional space is flat ($\kappa = 0$), it has the following metric:

$$ds_{\text{flat}}^2 = dx^2 + dy^2 + dz^2.$$

Campbell Cosmo Q2

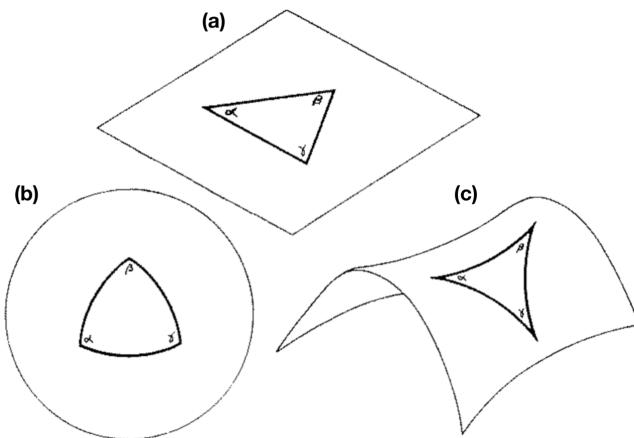


Figure 2:
 (a) Flat geometry for which $\Delta_{\text{flat}} = \alpha + \beta + \gamma = \pi$.
 (b) Closed geometry for which $\Delta_{\text{closed}} = \alpha + \beta + \gamma = \pi + A/R^2$, where A is the area of the triangle, and R is the radius of the sphere. (c) Open geometry for which $\Delta_{\text{open}} = \alpha + \beta + \gamma = \pi - A/R^2$, where A is the area of the triangle, and R is the radius of curvature.

By making the simple coordinate substitution $x = r \cos \theta$, $y = r \sin \theta$, this can be written in spherical coordinates as:

$$ds_{\text{flat}}^2 = dr^2 + r^2[d\theta^2 + \sin^2 \theta d\phi^2].$$

If a three-dimensional space has uniform positive curvature ($\kappa = +1$), its metric is

$$ds_{\text{closed}}^2 = dr^2 + R^2 \sin^2(r/R)[d\theta^2 + \sin^2 \theta d\phi^2].$$

A positively curved three-dimensional space has finite volume, just as a positively curved two-dimensional space has finite area. The point at $r = \pi R$ is the antipodal point to the origin, just as the south pole, at $r = \pi R$, is the antipodal point to the north pole, at $r = 0$, on the surface of a sphere. By traveling a distance $C = 2\pi R$, it is possible to “circumnavigate” a space of uniform positive curvature.

If a three-dimensional space has uniform negative curvature ($\kappa = -1$), its metric is

$$ds_{\text{open}}^2 = dr^2 + R^2 \sinh^2(r/R)[d\theta^2 + \sin^2 \theta d\phi^2].$$

Like flat space, negatively curved space has infinite volume.

The three possible metrics for a homogeneous, isotropic, three-dimensional space can be written more compactly in the form:

$$ds^2 = dr^2 + S_\kappa(r)^2 d\Omega^2$$

where

$$d\Omega \equiv d\theta^2 + \sin^2 \theta d\phi^2$$

and

$$S_\kappa(r) = \begin{cases} R \sin(r/R), & (\kappa = +1) \\ r, & (\kappa = 0) \\ R \sinh(r/R), & (\kappa = -1) \end{cases}.$$

Here, $\kappa = 0, -1, +1$ is the sign of curvature for a flat, open, and closed Universe, respectively. The coordinate system (r, θ, ϕ) is not the only possible system. For instance, if we switch the radial coordinate from r to $x \equiv S_\kappa(r)$, the metric for a homogeneous, isotropic, three-dimensional space can be written in the form:

$$ds^2 = \frac{dx^2}{1 - \kappa x^2/R^2} + x^2 d\Omega^2.$$

Energy density of the Universe: The **critical density** of the Universe is defined to be

$$\rho_c \equiv \left(\frac{3H_0^2}{8\pi G} \right) [\text{g cm}^{-3}]$$

which allows the density parameter to be defined as

$$\Omega_0 \equiv \frac{\rho}{\rho_c} [\text{dimensionless}].$$

Campbell Cosmo Q2

If the total energy density is greater than the critical density (i.e., $\Omega_0 > 1$), then the Universe is said to be **closed**: initially parallel lines eventually converge, just as lines of constant longitude meet at the North and South poles. A closed Universe, much like the surface of a sphere, has positive curvature. In a low-density Universe whose total energy density is less than critical value (i.e., $\Omega_0 < 1$), the Universe is said to be **open**: initially parallel lines eventually diverge, as would marbles rolling off a saddle. While a closed Universe has positive curvature, an open Universe has negative curvature.

CMB angular power spectrum: The comparison between the predicted acoustic peak scale and its angular extent provides a measurement of the angular diameter distance to recombination. The angular diameter distance in turn depends on the spatial curvature and expansion history of the Universe. Assuming the size of the Universe's horizon at the time of recombination and the distance to the last scattering surface, the geometry (or curvature) of the Universe can be measured using the angular size of the first peak of the angular power spectrum. If the first peak is at $\ell \sim 220$, the Universe is flat, whereas if the first peak is at $\ell < 220$ or $\ell > 220$, the Universe is open or closed, respectively.

Q2) Zhu Cosmo Q2

11

If we perform the same calculation during the era of recombination, setting n_e using the analysis above and obtaining H from $H(t) = H_0 \Omega_m \frac{a_0^3}{a^3} = H_0 \Omega_m \frac{1}{(1+z)^3}$, we obtain $z \approx 1100$ and $T \approx 3000$ (exact answers are difficult without modelling, since during the final stages of recombination the system was no longer in LTE).

1.2.2. What is the last scattering surface?

The last scattering surface is the $\tau = 1$ surface for photons originally trapped in the optically thick early universe. The age $t_0 - t$ of this surface can be found using

$$\tau = 1 = \int_t^{t_0} \Gamma(t) dt \quad (17)$$

In practice this is difficult, and so we again estimate that $z \approx 1100$ for last scattering.

1.3. Question 2

QUESTION: The universe is said to be "flat", or, close to flat. What are the properties of a flat universe and what evidence do we have for it?

This information comes mainly from Emberson (2012), with supplement from Carroll & Ostlie (2006).

As noted in Sec. ?, the FLRW universe may only have three types of curvature. When $\kappa = 1$, the universe is positively curved, since $R \sin(r/R) < r$ (i.e. the actual size of the object would be smaller than its physical size, consistent with the fact that two straight lines intersecting on a circle will eventually meet again) and when $\kappa = -1$ the universe is negatively curved, since $R \sinh(r/R) > r$.

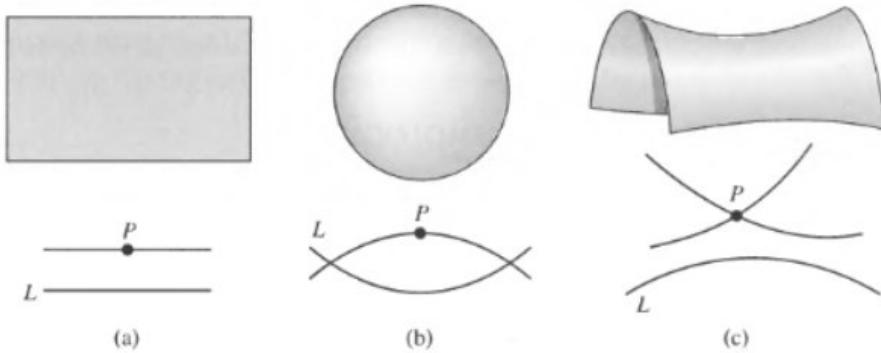


FIG. 4.— Schematic of two dimensional analogues to possible curvature in an FLRW universe. The behaviour of lines parallel at a point P within each space is also drawn. A Euclidian plane has no curvature, a sphere has positive curvature and a saddle has negative curvature. From Carroll & Ostlie (2006), their Fig. 29.15.

Fig. 4 shows the two primary geometric features of a flat, closed and open universe. In open and closed universes, parallel lines tend to diverge (open) or converge (closed), while for a flat universe two parallel lines remain parallel indefinitely. Open and flat universes are infinite, while a closed universe may have a finite extent, since it "curves back" on itself. In $\Lambda = 0$ universes, the geometry of the universe is intimately related to the matter-energy density of the universe.

Measurement of the curvature of the universe is difficult. In a $\Lambda = 0$ universe it actually is greatly simplified, since curvature and expansion history are linked, and the age of the universe, combined with H_0 , can be used to determine the curvature (or H_0 and q_0). In a universe with a non-zero cosmological constant, however, the age of the universe is decoupled from the curvature. Instead, we use a standard ruler: the first peak of the CMB power spectrum. This peak, due to the length of the sound horizon at decoupling, is

$$r_s(z_{rec}) = c \int_{z_{rec}}^{\infty} \frac{c_s}{H(z)} dz \quad (18)$$

where $c_s = (3(1 + 3\rho_{bary}/\rho_{ph}))^{-1/2}$ (Vardanyan et al. 2009). Detailed measurements of higher order peaks and their spacings in the CMB allow us to constrain both $H(z)$ and c_s , and obtain a preferred length scale (Eisenstein 2005). This is our standard ruler, and if we measure its current angular size θ , we can use Eqn. 14 alongside Eqn. 3 to determine

$$\frac{\theta}{1+z} = \frac{l}{S_\kappa} \quad (19)$$

Note that l is known, but S_κ depends on the co-moving distance between us and the CMB. This requires some knowledge of the subsequent expansion history of the universe, or else there is a degeneracy between Ω_m , Ω_Λ and Ω_κ (Komatsu et al. 2009). An additional constraint, such as a measurement of H_0 , or the series of luminosity distance measurements using high- z SNe, allows us to constrain Ω_κ (Komatsu et al. 2009). See Fig. 5.

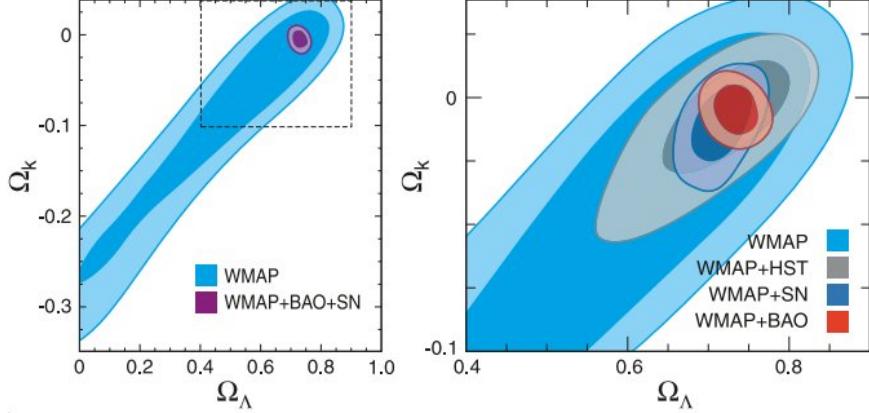


FIG. 5.— Joint two-dimensional marginalized constraint on the dark energy density Ω_Λ , and the spatial curvature parameter, Ω_κ . The contours show the 68% and 95% confidence levels. Additional data is needed to constrain Ω_κ : HST means H_0 from Hubble measurements, SN means luminosity distances from high- z SN, and BAO means baryon acoustic oscillation measurements from galaxy surveys. From Komatsu et al. (2009), their Fig. 6.

1.4. Question 3

QUESTION: Outline the development of the Cold Dark Matter spectrum of density fluctuations from the early universe to the current epoch.

Most of this information is from Schneider (2006), Ch. 7.3 - 7.5.

The growth of a single perturbation (described as one of the follow-up questions) in a matter-dominated universe can be described in the following way. We define the relative density contrast $\delta(\vec{r}, t) = (\rho(\vec{r}, t) - \bar{\rho})/\bar{\rho}$; from this $\delta(\vec{r}, t) \leq -1$. At $z \sim 1000$ $|\delta(\vec{r}, t)| \ll 1$. The mean density of the universe $\bar{\rho}(t) = (1 + z^3)\bar{\rho}_0 = \bar{\rho}_0/a(t)^3$ from Hubble flow. Like in the classic Newtonian stability argument of an infinite static volume of equally space stars, any overdense region will experience runaway collapse (and any underdense region will become more and more underdense). In the linear perturbative regime, the early stages of this collapse simply make it so that the the expansion of the universe is delayed, so $\delta(\vec{r}, t)$ increases. As it turns out, $\delta(\vec{r}, t)$ can be written as $D_+(t)\delta_0(\vec{x})$ in the linear growth regime. $D_+(t)$ is normalized to be unity today, and $\delta_0(\vec{x})$ is the linearly-extrapolated (i.e. no non-linear evolution taken into account) density field today.

The two-point correlation function $\xi(r)$ (Sec. 1.18) describes the over-probability of, given a galaxy at $r = 0$, there will be another galaxy at r (or x , here). It describes the clustering of galaxies, and is key to understanding the large-scale structure of the universe. We define the matter power spectrum (often shortened to just “the power spectrum”) as

$$P(k) = \int_{-\infty}^{-\infty} \xi(r) \exp(-ikr) r^2 dr \quad (20)$$

Instead of describing the spatial distribution of clustering, the power spectrum decomposes clustering into characteristic lengths $L \approx 2\pi/k$, and describes to what degree each characteristic contributes to the total overprobability.

Since the two-point correlation function depends on the square of density, if we switch to co-moving coordinates and stay in the linear regime,

$$\xi(x, t) = D_+^2(t) \xi_0(x, t_0). \quad (21)$$

Likewise,

$$P(k, t) = D_+^2 P(k, t_0) \equiv D_+^2 P_0(k), \quad (22)$$

i.e. everything simply scales with time. This the evolution of the power spectrum is reasonably easily described.

The initial power spectrum $P_0(k)$ was generated by the quantum fluctuations of inflation. It can be argued (pg. 285 of Schneider (2006)) that the primordial power spectrum should be $P(k) = Ak^{n_s}$, where A is a normalization factor

Q3) TERM DEFINITIONS

Define and describe the following terms: Comoving Distance, Proper Distance, Angular Diameter Distance, Luminosity Distance, Proper Time, Coordinate Time

Short Answer

$$E(Z) = \sqrt{\Omega_r a^{-4} + \Omega_m a^{-3} + \Omega_k a^{-2} + \Omega_\Lambda}$$

$$d_H = c/H_0$$

Comoving Distance: “the distance between two points measured along a path defined at the present cosmological time. For objects moving with the Hubble flow, it is deemed to remain constant in time.” (Wikipedia)

$$\chi = \int_{t_e}^t c \frac{dt'}{a(t')} = d_H \int_0^z c \frac{dz'}{E(z')}$$

$$\chi = \begin{cases} |\kappa|^{-1/2} \sinh^{-1} \sqrt{|\kappa|} r & \kappa < 0 (\text{negatively curved, hyperbolic universe}) \\ r & \kappa = 0 (\text{flat universe}) \\ |\kappa|^{-1/2} \sin^{-1} \sqrt{|\kappa|} r & \kappa > 0 (\text{positively curved, spherical universe}) \end{cases}$$

Solving for $r \equiv d_M$, the *transverse comoving distance*.

Transverse Comoving Distance:

$$d_M = \begin{cases} \frac{d_H}{\sqrt{\Omega_k}} \sinh \left(\sqrt{\Omega_k} d_C(z)/d_H \right) & \Omega_k > 0 (\text{negatively curved, hyperbolic universe}) \\ d_C(z) & \kappa = 0 (\text{flat universe}) \\ \frac{d_H}{\sqrt{|\Omega_k|}} \sin \left(\sqrt{|\Omega_k|} d_C(z)/d_H \right) & \Omega_k > 0 (\text{positively curved, spherical universe}) \end{cases}$$

Proper Distance: $d_p(t) = a(t)\chi$. The proper distance between two points at time t is just the distance that would be measured by rulers between them at that time.

Angular Diameter Distance: $d_A = a(t)d_M$

Luminosity Distance: $d_L = d_M/a(t)$

Light travel Distance: $d_T(z) = d_H \int_0^z z \frac{dz'}{(1+z')E(z')}$

Proper time: the observer’s time

conformal time: $d\tau = dt/a$

Q4) BIG BANG EVIDENCE

State and explain three key pieces of evidence for a Big Bang origin for the observable Universe.

Short Answer

from Campbell's Notes

The success of the Big Bang (BB) rests on three observational pillars:

1. Hubble's Law exhibiting expansion: If the Universe is expanding at the present time, then by 'turning back the clock' the Universe must have been much smaller in the past. Hence, the BB.
2. Light element abundances which are in accord with Big Bang nucleosynthesis: When the Universe was still a very hot plasma, the extreme radiation field ensured that any nucleus produced would be immediately photoionized by a high energy photon. As the Universe cooled (via expansion) well below the typical binding energies of nuclei, light elements began to form. Knowing the conditions of the early Universe and the relevant cross sections, one can calculate the expected primordial abundances of these light elements. Such predictions are consistent with measurements.
3. The blackbody radiation left over from the first few hundred thousand years, the cosmic microwave background: The fact that the early Universe was very hot and dense meant that the baryonic matter was well coupled with the radiation field implying thermal equilibrium (TE) of photons. In TE, photons should follow the blackbody (or Planck) function in which the energy density is only dependent on temperature. As it turns out, the CMB radiation is the most accurate BB curve yet to be measured!

Relevant Equations

For Hubble, $v = H_0 d$

for the CMB, $T = T_0/a = 2.7K/a \rightarrow \infty$ as $a \rightarrow 0$

Q4) Ludwig Cosmo Q4

Question 4 - Big Bang

State and explain three key pieces of evidence for a Big Bang origin for the observable Universe.

Solution

- Hubble's Law
 - At large scales all galaxies are receding away from us ($v = H_0 d$) implying that the universe is expanding. If we extrapolate backward, there must have been a time where everything was much closer together. Hence, the big bang.
- CMB \Rightarrow Cosmic Microwave Background
 - Thermal black body radiation from when baryons and photons were in thermal equilibrium. It's homogeneous and isotropic, implying that at some point the entirety must have been in causal contact and therefore the universe must have been much smaller and hotter in the past.
 - $T = T_0(1 + z) = 2.7K(1 + z)$, as Z goes to ∞ , really freaking hot.
- BBN \Rightarrow Big Bang Nucleosynthesis
 - Subatomic particles combined to form the lightest elements first as the universe expanded and cooled enough for the photon energy to decrease past the atomic binding energies
 - By knowing or assuming some initial conditions for the universe and knowing relevant interaction cross section, you can calculate the expected primordial abundances of elements and compare the current measurements.

Q4) Herman Cosmo Q4

C4

- recessional velocities of galaxies
- CMB homogeneity and isotropy
- primordial abundances of elements predicted by BB theory.

• Big Bang evidence

• Hubble's Law

- at large scales, all galaxies are receding from us ($v = H_0 d$), implying the universe is expanding. If we extrapolate backward, there must have been a time when everything was much closer together. Hence, the Big Bang.

• CMB

- thermal blackbody radiation from when baryons and photons were in thermal equilibrium. It's homogeneous and isotropic, implying that at some point the entirety must have been in causal contact and therefore the universe must have been much smaller and hotter in the past
- $$T = (1+z)T_0 = (1+z) \cdot 2.73 \text{ K} \quad \therefore z \rightarrow \infty \Rightarrow T \text{ really hot}$$

• BBN

- subatomic particles combined to form the lightest elements first, as the universe expanded and cooled enough for the photon energy to decrease past the atomic binding energies
- by knowing / assuming the universe's initial conditions, and knowing the relevant interaction cross sections, you can calculate the expected primordial abundances of elements and compare to current measurements.

Q4) Campbell Cosmo Q4

1.5 Question 4

State and explain three key pieces of evidence for a Big Bang origin for the observable Universe.

1.5.1 Short answer

The success of the Big Bang (BB) rests on three observational pillars:

1. Hubble's Law exhibiting expansion: The first key observation to the modern era of cosmology that the Universe is expanding. If the Universe is expanding at the present time, then by ‘turning back the clock’ the Universe must have been much smaller in the past. Hence, the BB.
2. Light element abundances which are in accord with Big Bang nucleosynthesis: When the Universe was still a very hot plasma, the extreme radiation field ensured that any nucleus produced would be immediately photoionized by a high energy photon. As the Universe cooled (via expansion) well below the typical binding energies of nuclei, light elements began to form. Knowing the conditions of the early Universe and the relevant cross sections, one can calculate the expected primordial abundances of these light elements. Such predictions are consistent with measurements.
3. The blackbody radiation left over from the first few hundred thousand years, the cosmic microwave background: The fact that the early Universe was very hot and dense meant that the baryonic matter was well coupled with the radiation field implying thermal equilibrium (TE) of photons. In TE, photons should follow the blackbody (or Planck) function in which the energy density is only dependent on temperature. As it turns out, the CMB radiation is the most accurate BB curve yet to be measured!

1.5.2 Additional context

1. Hubble's Law

We have good evidence that the Universe is expanding. This means that early in its history, the distances between galaxies was smaller than it is today. It's convenient to describe this expansion effect by introducing the **scale factor** a , whose present value is equal to one ($a(t_0) \equiv 1$). At earlier times, a was much smaller than it is today – hence, the Big Bang.

The first key observation to the modern era of cosmology was the discovery of an expanding Universe. This is popularly credited to Edwin Hubble in 1929, but in fact the honour lies with Vesto Slipher more than 10 years earlier. Slipher was measuring spectra of nebulae whose nature was still under hot debate at that time. Observations of Hubble settled this debate in 1924 when he discovered *Cepheid variables* in M31 (Andromeda) establishing a distance of roughly 1 Mpc. More than a decade earlier in 1913, Slipher had measured the spectrum of M31 and found that it was approaching Earth at a velocity of over 200 km s^{-1} . Over the next decade, he measured Doppler shifts for dozens of galaxies: with only a few exceptions, they were redshifted. By the time Hubble arrived, the basics of relativistic cosmology were already worked out and predictions existed that galaxy redshifts should increase with distance. It's hard to know how much these influenced Hubble, but by 1929 he had obtained Cepheid distances towards 24 galaxies along with their redshifts and claimed that they followed the empirical linear relationship:

$$v = H_0 d \text{ [km s}^{-1}\text]},$$

citing theoretical predictions as a possible explanation. At the time, Hubble estimated $H_0 \approx 500 \text{ km s}^{-1} \text{ Mpc}^{-1}$ because his calibration of Cepheid variables was in error. The best modern value is currently $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

Recall that the wavelength of light or sound emitted from a receding object is stretched out (i.e., Doppler shifted) so that the observed wavelength is larger than the emitted one. It is convenient to define this stretching factor as the redshift z :

$$1 + z \equiv \frac{\lambda}{\lambda_0} = \frac{1}{a} \text{ [dimensionless]},$$

or

$$(1 + z)^{-1} \equiv \frac{\lambda_0}{\lambda} = a \text{ [dimensionless]}.$$

For low redshifts, the standard Doppler formula applies and $z \sim v/c$. Therefore, a measurement of the amount by which absorption and/or emission lines are redshifted is a direct measure of how fast the structures in which they reside are receding from us. Hubble's diagram is shown in Figure 8, which shows not only that distant galaxies appear to be receding from us, but that the trend increases linearly with distance which is exactly what we would expect for an expanding Universe.

Q4) Campbell Cosmo Q4

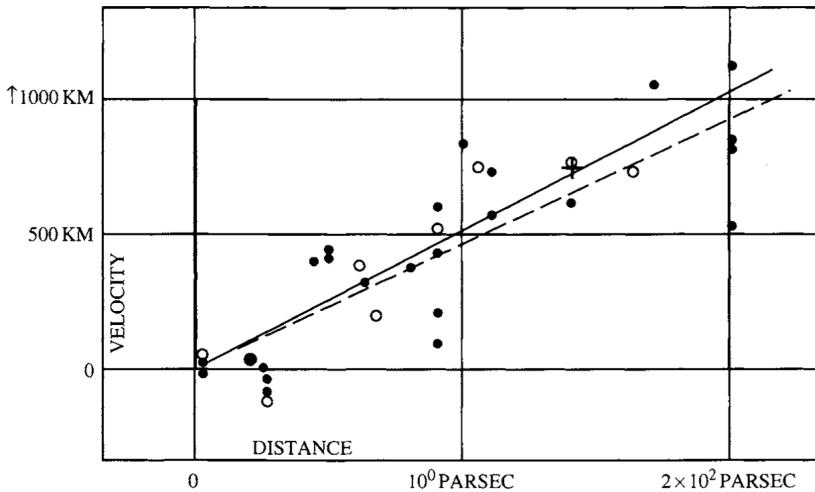


Figure 8: The original Hubble diagram (Hubble, 1929). Velocities of distant galaxies (units should be km s^{-1}) are plotted vs distance (units should be Mpc). Solid (dashed) line is the best fit to the filled (open) points which are corrected (uncorrected) for the Sun's motion. Image taken from Dodelson (2003).

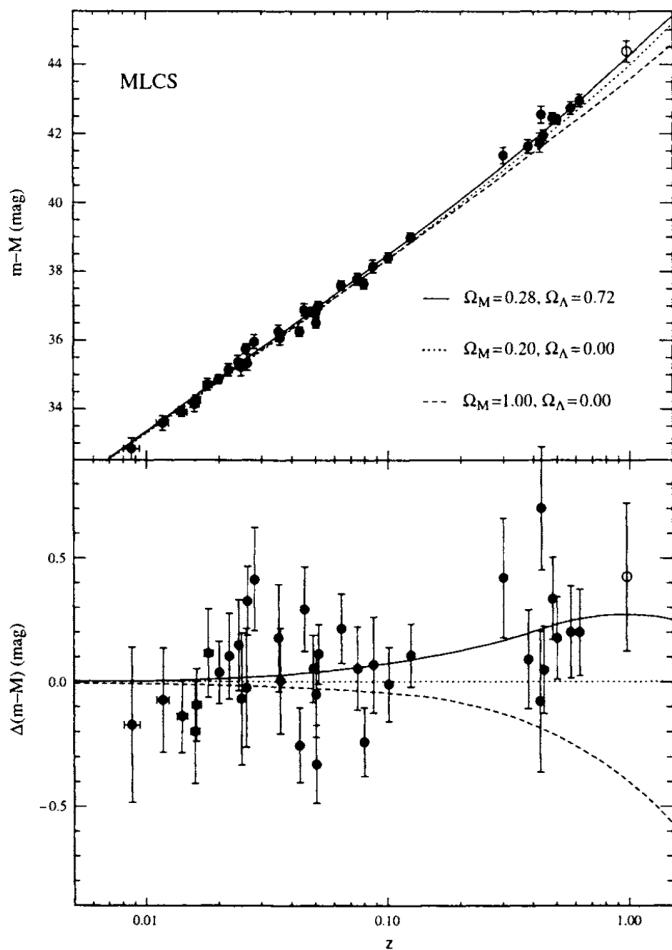


Figure 9:
Hubble diagram from distant Type Ia supernovae. Top panel shows apparent magnitude (an indicator of the distance) vs redshift. Lines show the predictions for different energy contents in the Universe, with Ω_M the ratio of energy density today in matter compared to the critical density and Ω_Λ the ratio of energy density in a cosmological constant to the critical density. Bottom panel plots the residuals, making it clear that the high-redshift supernovae favor a Λ -dominated Universe over a matter-dominated one. Figure taken from Dodelson (2003).

The Hubble diagram is still the most direct evidence we have that the Universe is expanding. Current incarnations use the same principle as the original: find the distance and the redshift of distant objects. Measuring redshifts is straightforward; the hard part is determining distances for objects of unknown intrinsic brightness. One of the most popular techniques is to try to find a standard candle, a class of objects which have the same intrinsic brightness. Any difference between the apparent brightness of two such objects then is a result of their different distances from us. This method is typically generalized to find a correlation between an observable and intrinsic brightness. For example, **Cepheid variables** are stars for which intrinsic brightness is tightly related to their pulsation period.

Q4) Campbell Cosmo Q4

As seen in Figure 9, the standard candle that can be seen at largest distances is a Type Ia supernova. Since they are so bright, supernovae can be used to extend the Hubble diagram out to very large redshifts (the current record is of order $z \sim 1.7$), a regime where the simple Doppler law ceases to work. Figure 9 shows a recent Hubble diagram using these very distant objects. The three curves in Figure 9 depict three different possibilities: flat matter dominated; open; and flat with a cosmological constant (Ω). The high-redshift data are now good enough to distinguish among these possibilities, strongly disfavoring the previously favored flat, matter-dominated Universe. The current best fit is a Universe with about 70% of the energy in the form of a cosmological constant, or some other form of dark energy.

2. Big Bang Nucleosynthesis

When the Universe was much hotter and denser, when the temperature was of order $h \text{ MeV}/k_B$, there were no neutral atoms or even bound nuclei. The vast amounts of radiation in such a hot environment ensured that any atom or nucleus produced would be immediately destroyed by a high energy photon. As the Universe cooled well below the binding energies of typical nuclei, light elements began to form. Knowing the conditions of the early Universe and the relevant nuclear cross-sections, we can calculate the expected primordial abundances of all the elements.

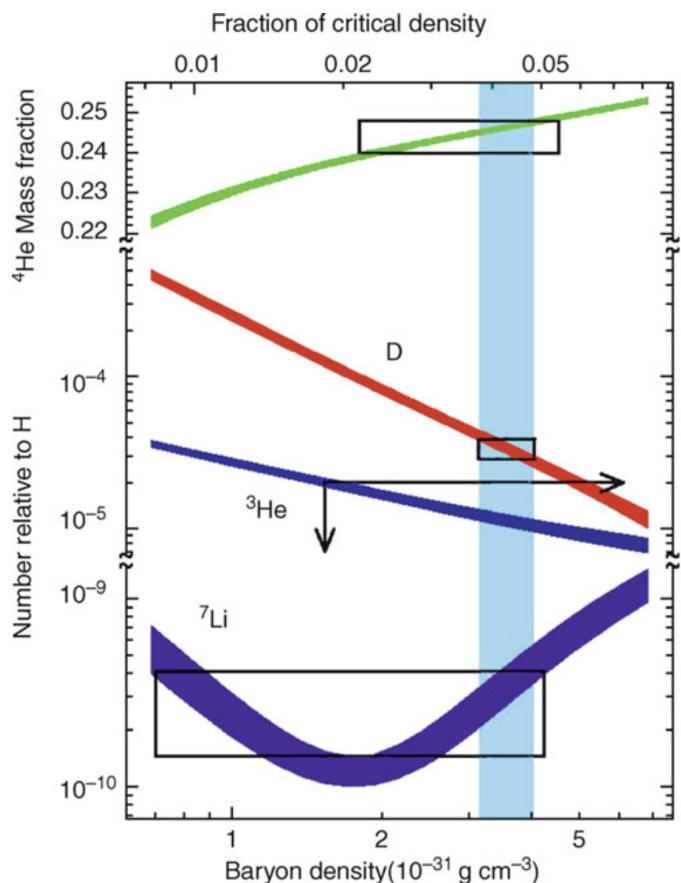


Figure 10: BBN predictions of the primordial abundances of light elements as a function of today's baryon density ($\rho_{b,0}$ lower axis) and the corresponding density parameter Ω_b where $h = 0.65$ was assumed. The vertical extent of the rectangles marks the measured values of the abundances (top: ${}^4\text{He}$, center: D, bottom: ${}^7\text{Li}$). The horizontal extent results from the overlap of these intervals with curves computed from theoretical models. The ranges in Ω_b that are allowed by these three species do overlap, as is indicated by the vertical strip. The deuterium measurements yield the most stringent constraints for Ω_b . Figure taken from Schneider (2006).

Figure 10 shows the predictions of Big Bang Nucleosynthesis (BBN) for the light element abundances³. The boxes and arrows show the current estimates for the light element abundances. These are consistent with the predictions, and this consistency test provides yet another ringing confirmation of the Big Bang. The theoretical predictions depend on the density of protons and neutrons at the time of nucleosynthesis. The combined proton plus neutron density is called the **baryon density** (ρ_b) since both protons and neutrons have baryon number one and these are the only baryons around at the time. Thus, BBN gives us a way of measuring the baryon density in the Universe. Since we know how those densities scale as the Universe evolves (they fall as a^{-3}), we can turn the measurements of light element abundances into measures of the baryon density today.

In particular, the measurement of primordial deuterium pins down the baryon density extremely accurately to only a few percent of the critical density. Ordinary matter (baryons) contributes at most 5%

³Recall nuclear notation: The 4 in ${}^4\text{He}$ refers to the total number of nucleons (protons and neutrons). So ${}^4\text{He}$ has two neutrons and two protons, while ${}^3\text{He}$ has two protons and one neutron.

Q4) Campbell Cosmo Q4

of the critical density (i.e., $\Omega_b = 0.005$). Since the total matter density today is almost certainly larger than this – direct estimates give values of order 20-30% nucleosynthesis provides a compelling argument for non-baryonic dark matter.

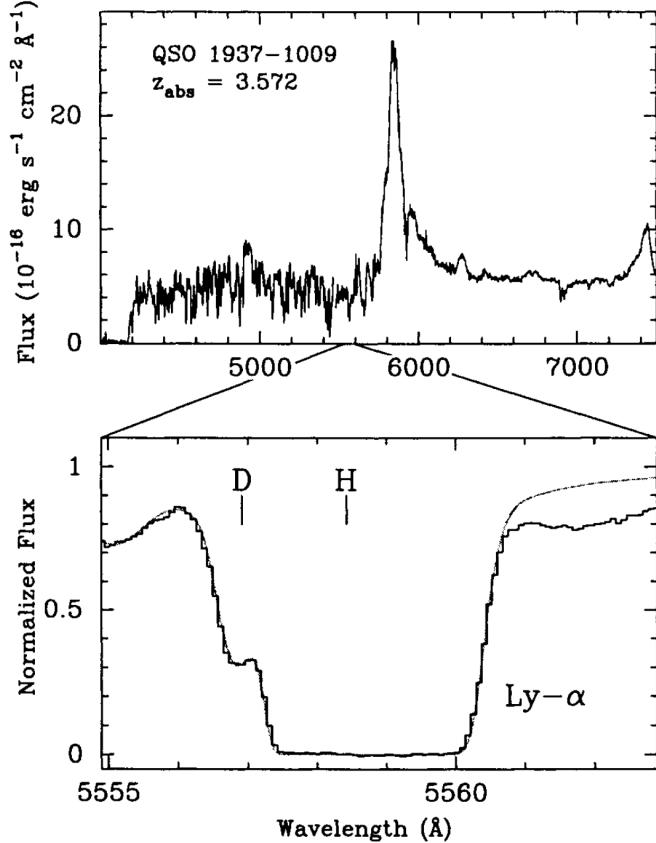


Figure 11: Spectrum from a distant QSO (Burles, Nollett, and Turner, 1999). Absorption of photons with rest wavelength 1216 Å corresponding to the ($n = 1$) to ($n = 2$) state of hydrogen is redshifted up to $1216(1 + 3.572)$ Å. Bottom panel provides details of the spectrum in this range, with the presence of deuterium clearly evident. Figure taken from Dodelson (2003).

The deuterium measurements (Burles & Tytler, 1998) are the new developments in the field. These measurements are so exciting because they explore the deuterium abundance at redshifts of order 3-4, well before much processing could have altered the primordial abundances. Figure 11 shows one such detection. The basic idea is that light from distant QSOs is absorbed by intervening neutral hydrogen systems. The key absorption feature arises from transition from the ground state ($n=1$) of hydrogen to the first excited state ($n = 2$), requiring a photon with wavelength $\lambda = 1215.7$ Å. Since photons are absorbed when exciting hydrogen in this fashion, there is a trough in the spectrum at Å, redshifted by a factor of $(1 + z)$. The corresponding line from deuterium should be (i) shifted over by $0.33(1 + z)$ Å and (ii) much less damped since there is much less deuterium. Figure 11 shows just such a system; there are now half a dozen with detections precisely in the neighborhood shown in Figure 10. Note that the steep decline in deuterium as a function of baryon density helps here: even relatively large errors in deuterium measurements translate into small errors on the baryon density.

3. Cosmic Microwave Background (CMB)

The CMB offers us a look at the Universe when it was only $\sim 380,000$ years old. The photons in the CMB last scattered off electrons at $z \sim 1100$; since then they have traveled freely through space. When we observe them today, they literally come from the earliest moments of time. They are therefore the most powerful probes of the early Universe. If an object is opaque then the protons, neutrons, electrons, and photons which it contains frequently interact and attain *thermal equilibrium*. A crucial fact about the CMB is that the collisions between electrons and photons before last scattering ensured that the photons were in equilibrium. That is, they should have a blackbody spectrum. When a system is in thermal equilibrium, the density of photons in the system as a function of photon energy depends only on the system temperature T . It doesn't matter whether the system is a tungsten filament or a sphere of ionized hydrogen and helium.

The specific intensity of a gas of photons following a blackbody spectrum is called the **Planck function**:

$$I_\nu = B_\nu(T) = \frac{2h\nu^3}{c^2} \frac{1}{\exp(h\nu/k_B T) - 1} [\text{erg s}^{-1} \text{ cm}^{-2} \text{ Hz}^{-1} \text{ sr}^{-1}].$$

Q4) Campbell Cosmo Q4

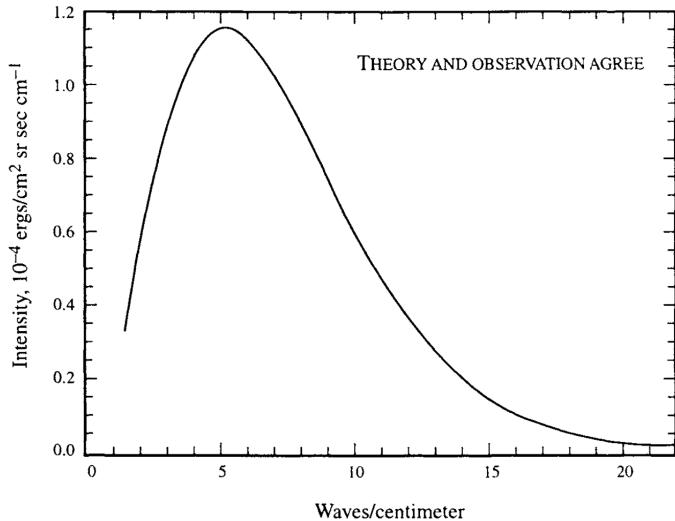


Figure 12: Intensity of cosmic microwave radiation as a function of wavenumber from Far InfraRed Absolute Spectrophotometer (FIRAS) (Mather et al., 1994), an instrument on the COBE satellite. Hidden in the theoretical blackbody curve are dozens of measured points, all of which have uncertainties smaller than the thickness of the curve! Figure taken from Dodelson (2003).

Figure 12 shows the remarkable agreement between this prediction of Big Bang cosmology and the observations by the FIRAS instrument aboard the COBE spacecraft. We have been told that detection of the 3 K background by Penzias and Wilson in the mid-1960s was sufficient evidence to decide the controversy in favor of the Big Bang over the Steady State Universe. Penzias and Wilson, though, measured the radiation at just one wavelength. If even their one-wavelength result was enough to tip the scales, the current data depicted in Figure 12 should send skeptics from the pages of physics journals to the far reaches of radical Internet chat groups.

The photons which make up the cosmic microwave background (CMB) today have a well-measured temperature $T_{\text{CMB}} = 2.725 \pm 0.002 \text{ K}$. A photon with an energy $k_B T_{\text{CMB}}$ today has a wavelength $hc/k_B T_{\text{CMB}}$. Early on, when the scale factor was smaller than it is today, this wavelength would have been correspondingly smaller. Since the energy of a photon is inversely proportional to its wavelength, the photon energy would have been larger than today by a factor of $1/a$. This argument applied to the thermal bath of photons implies that the temperature of the plasma as a function of time is

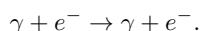
$$T(t) = \frac{T_0}{a(t)} = T_0(1+z) [\text{K}].$$

At early times, then, the temperature was higher than it is today.

The energy density of each component of the Universe is equal to the number density of each corresponding component times its average energy. The primary energy density components of the Universe includes radiation, matter (baryons and dark matter), and dark energy. Since photon energy scales as a^{-1} in addition to the number density which scales as a^{-3} , the energy density of radiation should scale as a^{-4} . The energy density of non-relativistic particles, such as baryons, have a constant rest mass energy. Since their number density scales inversely proportional to volume, the energy density of matter scales as a^{-3} . Dark energy (introduced as a cosmological constant) is believed to have a constant energy density. While matter, and possibly a cosmological constant, dominate the current cosmological landscape, radiation must have been the dominant constituent of the Universe due to its energy density scaling of a^{-4} . Figure 13 illustrates how radiation, matter, and the cosmological constant energy density components evolve with time.

The CMB was detected by Arno Penzias & Robert Wilson in 1965 and were awarded the 1978 Nobel prize in physics; they accomplished this by locating an unaccounted-for source of noise in a radio telescope at Bell Laboratories being used to study our own Galaxy. This excess “noise” was isotropic and constant with time, so it couldn’t be associated with an isolated celestial source. Wilson and Penzias were puzzled until they were put into touch with Robert Dicke and his research group at Princeton University. Dicke had deduced that the Universe, if it started in a hot dense state, should now be filled with microwave radiation. The timing of the discovery was especially ironic, given that instruments were underway at that time to test the theoretical prediction that such a background should exist.

When the Universe was fully ionized, photons interacted primarily with electrons via Thomson scattering (the low energy limit of Compton scattering):



Q4) Campbell Cosmo Q4

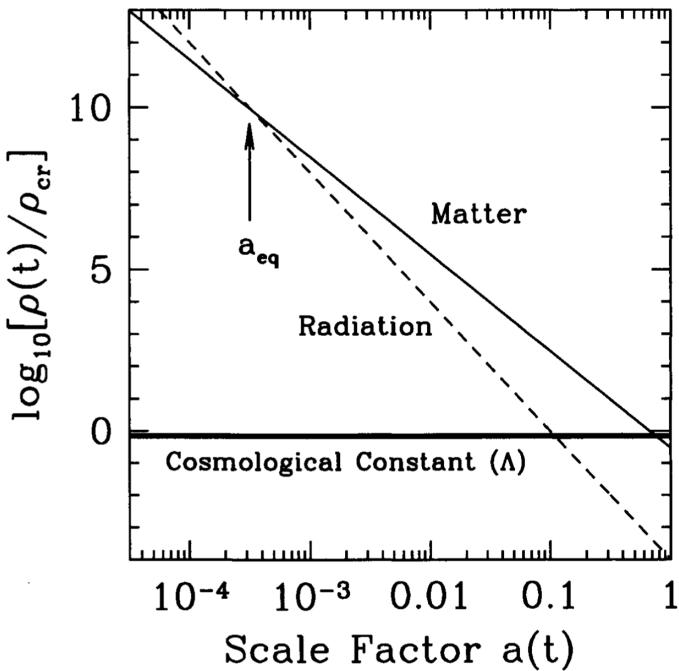


Figure 13: Energy density vs scale factor for different constituents of a flat Universe. Shown are non-relativistic matter, radiation, and a cosmological constant. All are in units of the critical density today. Even though matter and cosmological constant dominate today, at early times, the radiation density was largest. The epoch at which matter and radiation are equal is a_{eq} . Image taken from Dodelson (2003).

Thomson scattering is an elastic process whereby energy and momentum are transferred between the photon and electron. The cross-section for Thomson scattering is the Thomson cross-section of

$$\sigma_e = 6.65 \times 10^{-29} [\text{m}^2].$$

The mean free path of a photon – that is, the mean distance it travels before scattering from an electron – is therefore

$$\ell_{\text{mfp}} = \frac{1}{n_e \sigma_e} [\text{m}].$$

When the baryonic component of the Universe is fully ionized, $n_e = n_p = n_b$. Currently, the number density of baryons is $n_{b,0} = 0.22 \text{ m}^{-3}$. The number density of conserved particles, such as baryons, goes with the scale factor as $1/a^3$, so when the early Universe was fully ionized, the free electron density was

$$n_e = n_b = \frac{n_{b,0}}{a^3} [\text{m}^{-3}].$$

Since photons travel with speed c , the rate at which a photon undergoes scattering interactions is

$$\Gamma = \frac{c}{\ell_{\text{mfp}}} = n_e \sigma_e c [\text{s}^{-1}].$$

Using the fact that $n_e = n_b = n_{b,0}/a^3$ for a fully-ionized Universe,

$$\Gamma = \frac{n_{b,0} \sigma_e c}{a^3} = \frac{4.4 \times 10^{-21}}{a^3} [\text{s}^{-1}].$$

The photons remain coupled to the electrons as long as their scattering rate, Γ , is larger than H , the rate at which the Universe expands; this is equivalent to saying that their mean free path ℓ_{mfp} is shorter than the Hubble distance c/H .

Measuring the spectrum of the CMB, and confirming that it is indeed a blackbody, is not a simple task, even with modern technology. The current energy per CMB photon, $\sim 6 \times 10^4 \text{ eV}$, is tiny compared to the energy required to break up an atomic nucleus ($\sim 1 \text{ MeV}$) or even the energy required to ionize an atom ($\sim 10 \text{ eV}$). However, the mean photon energy is comparable to the energy of vibration or rotation for a small molecule such as H_2O . Thus, CMB photons can zip along for more than 13 billion years through the tenuous intergalactic medium, then be absorbed a microsecond away from the Earth's surface by a water molecule in the atmosphere. Microwaves with wavelengths shorter than $\lambda \sim 3 \text{ cm}$ are strongly absorbed by water molecules.

The CMB can be measured at wavelengths shorter than 3 cm by observing from high-altitude balloons or from the South Pole, where the combination of cold temperatures and high altitude keeps the atmospheric humidity low. The best way to measure the spectrum of the CMB, however, is to go completely above

Q4) Campbell Cosmo Q4

the damp atmosphere of the Earth. The CMB spectrum was first measured accurately over a wide range of wavelengths by the COsmic Background Explorer (COBE) satellite, launched in 1989, into an orbit 900 km above the Earth's surface. COBE actually contained three different instruments. The Diffuse InfraRed Background Experiment (DIRBE) was designed to measure radiation at the wavelengths $0.001 \text{ mm} < \lambda < 0.24 \text{ mm}$; at these wavelengths, it was primarily detecting stars and dust within our own Galaxy. The second instrument, called the Far InfraRed Absolute Spectrophotometer (FIRAS), was used to measure the spectrum of the CMB in the range $0.1 \text{ mm} < \lambda < 10 \text{ mm}$, a wavelength band which includes the peak in the CMB spectrum. The third instrument, called the Differential Microwave Radiometer (DMR), was designed to make full-sky maps of the CMB at three different wavelengths: $\lambda = 3.3 \text{ mm}, 5.7 \text{ mm}, \text{ and } 9.6 \text{ mm}$.

The most important fact we learned from our first 25 years of surveying the CMB was that the early Universe was very smooth. No anisotropies were detected in the early CMB. This period, while undoubtedly frustrating for observers searching for anisotropies, solidified the view of a smooth Big Bang. We are now moving on. We have discovered anisotropies in the CMB, indicating that the early Universe was not completely smooth. There were small perturbations in the cosmic plasma. To understand these, we must go beyond the Standard Model.

1.5.3 Follow-up Questions

- Is inflation needed to explain any of these observations?
- How is BBN used to constrain models of the Universe?
- Are there cosmological models in which there was no Big Bang?

Q4) Zhu Cosmo Q4

16

Since there is only one normalization, a survey of galaxy cluster total masses at different redshifts (using hot gas and a mass-to-light conversion, gravitational microlensing, etc) can be used to determine cosmological parameters. This is because the minimum overdensity for collapse δ_{\min} is dependent on both the growth rate of overdensities and the expansion of the universe. Increasing Ω_m , for example, decreases $n(M, z)/n(M, 0)$, since massive halo growth is more extreme the higher Ω_m is. A large Ω_Λ dampens massive halo growth.

1.5. Question 4

QUESTION: State and explain three key pieces of evidence for the Big Bang theory of the origin of the Universe.

This information is cribbed from Emberson (2012).

The Big Bang theory is the theory that the universe started off in an extremely hot, dense state, which then rapidly expanded, cooled, and became more tenuous over time. The Big Bang theory requires that at some point in the past a). the universe was born, b). the universe was extremely hot and c). objects were much closer together. The three key pieces of evidence are:

1. **Hubble's Law:** galaxies isotropically recede from our position with the relationship

$$\vec{v} = H_0 \vec{r} \quad (26)$$

known as Hubble's Law. As it turns out, moving into the frame of another galaxy ($\vec{r}' = \vec{r} - \vec{k}$, $\vec{v}' = \vec{v} - H_0 \vec{k} = H_0(\vec{r} - \vec{k}) = H_0 \vec{r}'$) does not change any observations. At larger distances, Hubble's Law breaks down (see Sec. 1.7), but the rate of expansion only increases with distance. Because of this isotropic radial motion outward, we can back-calculate a time when all the galaxies ought to be together at one point. This time is $t_0 = r/v = 1/H_0 \approx 14$ Gyr, the Hubble Time. This gives an age to the universe, and indicates that in the distant past everything was closer together.

2. **The Cosmic Microwave Background:** the cosmic microwave background (CMB) is a near perfect isotropic blackbody with a (current) $T_0 \approx 2.73$ K. For a blackbody, $\lambda_{\text{peak}} = 0.0029 \text{ mK}/T$, $U = aT^4$ and $n = \beta T^3$, which gives us $n \approx 400 \text{ cm}^{-3}$, $\epsilon \approx 0.25 \text{ eV cm}^{-3}$, and $\lambda \approx 2 \text{ mm}$. In Big Bang cosmology, this microwave background is the redshifted ($T \propto a^{-1}$) vestige of the surface of last scattering, when $T \approx 3000$ K and the universe became neutral enough for photons to travel unimpeded. This is evidence that the universe used to be hot.
3. **Big Bang Nucleosynthesis:** in the Big Bang theory, the lightest elements were created out of subatomic particles when the temperature dropped enough that the average photon was significantly below the binding energy of light elements. A detailed calculation of nucleosynthetic rates of H, D, He and Li during the first few minutes of the universe is consistent with the current abundances of light elements in the universe. See Sec. 1.8.

Additionally, no object has been found to be older than the currently accepted age of the universe, 13.7 Gyr. As we look back in time, we notice that the average galaxy in the universe looked considerably different - this evolution is consistent with Λ CDM cosmology, which has small, dense cores of dark matter forming due to gravitational instability, and then merging to form larger cores.

1.5.1. What is Olbers's Paradox?

Olbers's paradox is the apparent contradiction one has when an infinitely old, infinitely large universe with a fixed stellar density is assumed. In such a universe every single line of sight would eventually reach a star's photosphere. Since a typical photospheric temperature is ~ 5000 K and surface brightness is independent of distance, we would expect the entire sky to be at ~ 5000 K, roasting the Earth. Setting a finite age to the universe is one solution to the paradox; another would be that stars only formed in the last several billion years, and light from more distant stars have yet to reach us.

1.5.2. Are there Big Bang-less cosmologies?

It is impossible to generate a matter dominated universe for which there is no Big Bang. It is possible, however, for a Λ -dominated universe to be infinitely old, since an exponential (see Sec 1.1.4) never goes to zero. This is consistent with the steady state theory (Sec. 1.6).

1.6. Question 5

QUESTION: Define and describe the "tired light hypothesis" and the "steady state universe" as alternatives to the Big Bang. How have they been disproved observationally?

Cosmology Class Fall 2019: Martine's Notes

1. Expansion of the universe. However, this is still possible with the steady-state model where as the universe expands, matter keeps getting created and the same density is maintained. In this model, older galaxies redshift away into nothingness and newer galaxies form into existence from the created matter. The new galaxies look exactly like the old ones so you should see the same distribution of galaxies at any redshift. Observations of quasars at high-redshift (which we don't see at low-redshift) proved that the universe is changing over time so the steady-state model couldn't be true.

The CMB is the best counter to any steady-state model. The universe was clearly very different in the past. It was a hot dense gas. This is only possible if the universe was smaller in the past.

Big bang nucleosynthesis predicts the correct ratio of elements as is measured.

Q5) BIG BANG NUCLEOSYNTHESIS (BBN)

Describe Big Bang nucleosynthesis. Why are only very light elements (H, D, He, and traces of Li) produced?

Q5) Ludwig Cosmo Q5

Question 5 - Big Bang Nucleosynthesis

Why are only very light elements (H, D, He, and traces of Li) synthesized in the first three minutes of the Big Bang?

(I haven't figured out how to convert latex tables into google doc yet)

Time after BB	Description	Chemical Reaction
< 1 second	The neutron:proton ratio is maintained at thermal equilibrium	$p + e^{-1} \rightleftharpoons n + \nu$
≈ 1 second	Temperature cools, is slightly less than the neutron:proton mass difference, these weak reactions become slower than expansion and the ratio freezes out at about 1:6	$n + e^+ \rightleftharpoons p + \bar{\nu}$
> 1 second	Only neutron decay changes number of neutrons. Half life of a neutron is 615 seconds, ~ 10 minutes. Without further reactions to preserve neutron the Universe would be pure Hydrogen.	$n \rightarrow p + e^{-1} + \bar{\nu}$
> 100 seconds	Deuterons preserve the neutron. The reaction releases 2.2 MeV but since photons are a billion times more numerous than protons the reaction doesn't proceed until $T < 0.1$ MeV. The neutron proton ratio is about 1:7.	$p + n \rightleftharpoons d + \gamma$
	Further reactions proceed to make helium nuclei including He_3 , He_4 , and radioactive Hydrogen H_3 (triatomic). Because the binding energy of helium is 28 MeV more bound than deuterons and the temperature has fallen to 0.1 MeV, the reactions go one way.	$d + n \rightarrow \text{H}_3 + \gamma$ $\text{H}_3 + p \rightarrow \text{He}_4 + \gamma$ $d + p \rightarrow \text{He}_3 + \gamma$ $\text{He}_3 + n \rightarrow \text{He}_4 + \gamma$
	Reactions that don't produce a photon occur and can happen even faster.	$d + d \rightarrow \text{He}_3 + n$ $d + d \rightarrow \text{H}_3 + p$ $\text{H}_3 + d \rightarrow \text{He}_4 + n$ $\text{He}_3 + d \rightarrow \text{He}_4 + p$
	H_3 has a 12 year half life and decays into He_3 so none survive to the present.	$\text{H}_3 \rightarrow \text{He}_3 + e^- + \bar{\nu}_e$
	Be_7 decays into Li_7 with a 53 day half life and does not survive.	$\text{He}_3 + \text{He}_4 \rightarrow \text{Be}_7 + \gamma$ $\text{Be}_7 + e^- \rightarrow \text{Li}_7$

Table 1: BBN Timeline

Deuteron is the nucleus of deuterium.

Deuterium is the heavy form of hydrogen H_2 .

Deuteron is just a nucleus with a proton where hydrogen has only one proton and no neutron.

Deuterium peaks around 100 seconds and is rapidly swept up into helium nuclei.

Very few helium nuclei combine into heavier nuclei giving an abundance of lithium.

Q5) Herman Cosmo Q6

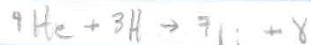
C6

- Deuterium Bottleneck - low binding energy and limited neutrons
- A=5 roadblock - no stable nuclei and little D, ^3H , ^3He available
- A=8 roadblock - no stable nuclei; expansion rate exceeds reaction rates

Synthesis of Light Elements during BBN

• Deuterium acted as a bottleneck for further synthesis. Basically all of the neutrons were bound up in D, but D has a relatively low binding energy so it's easy to photo dissociate. The temperature had to cool past $\sim 10^9 \text{ K}$ ($\sim 2 \text{ min.}?$) before a significant fraction of D could be built up, and this was a race against neutron decay (w/ a half life of $\sim 10 \text{ min.}$).

• After this ^4He forms rapidly, w/ a high binding energy making it very stable. Almost all D is converted to ^4He . Building heavier elements than this is difficult, since there is a road block at A=5. No stable nuclei exist w/ 5 nucleons. You can only make small amounts of Li via:



• There is limited D, ^3H , and ^3He available, so you only get trace amounts of Li. Another roadblock occurs for A=8, since no stable nuclei exist here. Fusion higher elements is nearly impossible, since getting over this bump is tough and the reaction rates are quickly being exceeded by the rate of expansion until further synthesis is not possible.

• 75% H, 25% He, trace Li

Q5) Campbell Cosmo Q6

1.7 Question 6

Why are only very light elements (H, D, He, and traces of Li) synthesized in the first three minutes of the Big Bang?

1.7.1 Short answer

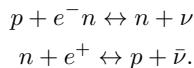
Deuterium synthesis acted as the first bottleneck for nucleosynthesis of heavier elements. The deuteron (i.e., deuterium nucleus) has a binding energy of 2.225 MeV which is only 4.3 times larger than $m_e c^2$ and 1.7 times larger than the neutron-proton mass-energy difference. Therefore, typical photons are energetic enough to easily photoionize deuterium. The Universe must therefore cool to $T \sim 10^9$ K, just below the binding energy of deuterium for the equilibrium to swing in favour of deuterium synthesis. This occurs at $t \sim 3$ min. All the while, free neutrons are decaying with a half life of $t_n \sim 15$ min. Since the simplest way to synthesize helium is by fusing deuterium, helium synthesis must await significant quantities of deuterium which happens at a temperature roughly one third that at which helium would be expected to dominate. Once significant nucleosynthesis begins, deuterium is rapidly converted into helium owing to its greater binding energy per nucleon (7 MeV as opposed to 1.1 MeV for deuterium). This occurs at a temperature of roughly 0.1 MeV, by which point the density and temperature of the Universe are too low for significant synthesis of heavier nuclei to proceed.

1.7.2 Additional context

At sufficiently early times, the temperature of the Universe was that of the centre of the Sun ($T = 1.55 \times 10^7$ K), where we know that nuclear reactions occur. Starting in the 1940's, Gamow considered the fascinating question of whether nuclear reactions were possible in the early Universe. He noted that the abundances of some elements in stars showed great regularities, especially a universal proportion of about 25% helium by mass. This led to the vision that a chain of nuclear reactions in the early Universe could generate not only helium, but all elements. In 1957, the Burbidges, Fowler & Hoyle showed that almost all elements could in fact be generated in stars, but the problem of helium remained. Gamow showed that its existence could be used to predict the present radiation temperature (as argued below). It will be convenient to refer to particle masses and temperatures in nuclear physics units, which are MeV. Some useful conversions are:

$$\begin{aligned} 1 \text{ MeV} &= 10^{10.065} \text{ K} \\ m_e c^2 &= 0.511 \text{ MeV} \\ m_p c^2 &= 939 \text{ MeV} \\ (m_n - m_p) c^2 &= 1.3 \text{ MeV}. \end{aligned}$$

Neutron freezeout: After the annihilation between matter and anti-matter (below $\sim 10^{13}$ K), the balance between neutrons and protons are maintained in equilibrium by weak interactions:



While this persists, the relative number densities of neutrons and protons should vary according to a Boltzmann factor based on their rest energy difference of 1.3 MeV:

$$\frac{n_n}{n_p} = e^{-\Delta mc^2/k_B T} \text{ [dimensionless].}$$

The reason that neutrons exist today is that the timescale for the weak interactions needed to keep this equilibrium set up eventually becomes longer than the expansion timescale. The reactions thus rapidly cease, and the proton-neutron ratio undergoes freezeout at some characteristic value which determined the He abundance. Since most He is ${}^4\text{He}$ (i.e., two neutrons and two protons), the Helium mass fraction (denoted Y) is given by

$$Y = \frac{4 \times n_n/2}{n_n + n_p} = \frac{2}{1 + n_n/n_p} \text{ [dimensionless]}$$

neglecting neutrons in other elements. So, $Y = 0.25$ requires a neutron-proton freezeout $n_n/n_p \simeq 1/7$. To calculate when the neutron-proton ratio undergoes freezeout, we need to know the rates of weak nuclear reactions. Fermi discovered how to calculate the relevant cross sections in the 1930s. Remember that, at $T \sim 10^{10}$ K, we are above the electron-proton energy threshold, so there exist thermal populations of both electrons and neutrinos to make the reaction $p + e^- \leftrightarrow n + \nu$ go equally well in either direction.

Q5) Campbell Cosmo Q6

All that is needed, therefore, is to consider either the reaction timescale for one proton immersed in a thermal bath of electrons or of one neutron immersed in a bath of neutrinos (the rates are the same). When this timescale equals the local Hubble time, $a(t)/\dot{a}(t)$, we get freezeout of the proton-neutron ratio. Taking the known weak interaction rates, this happens at

$$T(n \text{ freezeout}) \simeq 10^{10.14} \text{ K} \Rightarrow \frac{n_n}{n_p} \simeq 0.34.$$

This number is not a precisely correct result because nucleosynthesis is a process that contains a number of interesting (but potentially confusing) coincidences:

1. The freezeout condition was calculated assuming a temperature well above the electron mass threshold, but freezeout actually happens only a very little above this critical temperature.
2. Neutrons are not stable: they decay spontaneously with the e -folding lifetime of $\tau_n = 887 \pm 2$ s. Unless the frozen-out neutrons can be locked away in nuclei before $t = 887$ s, the relic abundance will decay freely to zero. The freezeout point occurs at an age of a few seconds, so there are only a few e -foldings of expansion available in which to salvage some neutrinos.

Locking up the neutrons: It may seem implausible that we can add one more coincidence – i.e., that nuclear reactions will become important at about the same time – this is just what happens. The Deuteron binding energy of 2.225 MeV is only 4.3 times larger than $m_e c^2$ and only 1.7 times larger than the neutron-proton mass difference. At higher temperatures, the strong interaction $n + p \leftrightarrow D + \gamma$ is fast enough to produce deuterium, but thermal equilibrium favours a small deuterium fraction – i.e., typical photons are energetic enough to disrupt deuterium nuclei very easily. The second key temperature in nucleosynthesis is therefore where the Universe has cooled sufficiently for the equilibrium to swing in favour of deuterium. In practice, this happens at a temperature a little below the deuteron binding energy. This is because the large photon to baryon ratio: even if photons lack sufficient energy to disintegrate deuterons, the rare ones in the tail of the distribution can still do the job.

Nevertheless, the temperature at which deuterium switches from being rare to dominating the equilibrium is still at $k_B T$ of order the deuteron binding energy:

$$T(\text{Deuterium formation}) \simeq 10^{8.9} \text{ [K]},$$

or at a time of about 3 minutes.

Notice that we have not needed to know the nuclear reaction rates that form deuterium, since the argument is an equilibrium one. However, if the matter density is too low, the nuclear reactions will freeze out before much deuterium has formed. Gamow took the known nuclear cross-sections and argued that the typical reaction time for deuterium formation had to be the cosmological age at that time (i.e., 3 minutes). This let him conclude that the matter density must have been about $10^{-3} \text{ kg m}^{-3}$ at that time. This gives a ratio of number densities of photons to nucleons, which is preserved as the Universe expands. Therefore, the present-day matter density allows a prediction of the present-day photon density, and hence its temperature. Alpher & Herman used Gamow's argument to predict a current photon temperature of 4 K to 5 K, which is impressively accurate. On the other hand, this prediction was based on a figure for the $z = 0$ matter density that is probably too low by at least a factor of 100, raising the temperature estimate by a factor of about 5. Actually, Gamow's argument is an inequality: there is a minimum matter density at 10⁹ K, but it could have been higher. The prediction for the current temperature is therefore really an upper limit. It works because the nuclear reactions are not too far from freeze out when deuterium forms.

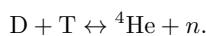
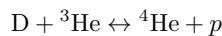
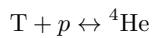
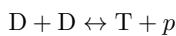
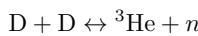
The deuterium measurements (Burles and Tytler, 1998) are the new developments in the field. These measurements are so exciting because they explore the deuterium abundance at redshifts of order 3-4, well before much processing could have altered the primordial abundances. Figure 11 shows one such detection. The basic idea is that light from distant QSOs is absorbed by intervening neutral hydrogen systems. The key absorption feature arises from transition from the ground state ($n = 1$) of hydrogen to the first excited state ($n = 2$), requiring a photon with wavelength $\lambda = 1215.7 \text{ \AA}$. Since photons are absorbed when exciting hydrogen in this fashion, there is a trough in the spectrum at \AA , redshifted by a factor of $(1+z)$. The corresponding line from deuterium should be (i) shifted over by $0.33(1+z) \text{ \AA}$ and (ii) much less damped since there is much less deuterium. Figure 11 shows just such a system; there are now half a dozen with detections precisely in the neighborhood shown in Figure 10. Note that the steep decline in deuterium as a function of baryon density helps here: even relatively large errors in deuterium measurements translate into small errors on the baryon density.

Formation of helium: The argument so far has produced a Universe consisting of just hydrogen and deuterium, but this is not realistic as one would expect ⁴He to be preferred on thermodynamic grounds, owing to its greater binding energy per nucleon (7 MeV as opposed to 1.1 MeV for deuterium). In

Q5) Campbell Cosmo Q6

equilibrium, the first nuclei to come into existence in significant numbers should be ${}^4\text{He}$: the abundance of ${}^4\text{He}$ relative to protons should reach unity at an energy of about 0.3 MeV, at which point the relative abundance of deuterium is only $\sim 10^{-12}$.

Since the simplest way to synthesize helium is by fusing deuterium, it is no surprise that the equilibrium prediction fails miserably in the expanding Universe: the production of helium must await the synthesis of significant quantities of deuterium which we have seen happens at a temperature roughly one third that at which helium would be expected to dominate. What the thermodynamic argument does show, however, is that it is expected that deuterium will be rapidly converted to helium once significant nucleosynthesis begins. This argument is what allows us to expect that the helium abundance can be calculated from the final n_n/n_p ratio. The main reactions of importance are of course 2-body ones, rather than the improbable coincidence of 2 protons and 2 neutrons all arriving at the same place simultaneously to make ${}^4\text{He}$ in one go. The process starts by fusing deuterium to make either tritium and ${}^3\text{He}$, following which there are four main ways of reaching ${}^4\text{He}$ (leaving aside rarer reactions involving residual free neutrons):



The same thermodynamic arguments that say helium should be favoured at temperatures around 0.1 MeV say that more massive nuclei would be preferred in equilibrium at lower temperatures still. A Universe that stayed in nuclear equilibrium as it cooled would eventually consist entirely of iron since this has the largest binding energy per nucleon. However, by the time helium synthesis is accomplished, the density and temperature are too low for significant synthesis of heavier nuclei to proceed. Apart from helium, the main nuclear residue of the Big Bang is therefore those deuterium nuclei that escape being mopped up into helium, plus a trace of ${}^3\text{He}$. The other intermediate produce, tritium, is not so strongly bound and thus leaves no significant relic. There also exists extremely small fractions of other elements: ${}^7\text{Li}$ ($\sim 10^{-9}$ by mass) and ${}^7\text{Be}$ ($\sim 10^{-11}$ by mass). The helium content in the Universe changes later by nuclear fusion in stars, which also forms heavier nuclei ('metals'). However, the total amount of helium produced in stars is expected to be smaller by about one order of magnitude compared to that in BBN. In summary, nucleosynthesis starts at about 10^{10} K when the Universe was about 1 s old, and effectively ends when it has cooled by a factor of 10, and is about 100 times older.

1.7.3 Follow-up Questions

- How does nucleosynthesis scale with cosmological parameters?
- How do we determine primordial abundances?
- Why is there more matter than anti-matter?

Q5) Zhu Cosmo Q5

19

1.7.2. Why can't we explain the Hubble flow through the physical motion of galaxies through space?

If space were not expanding, but galaxies are moving away from us isotropically, then

$$v = c \frac{(1+z)^2 - 1}{(1+z)^2 + 1}. \quad (31)$$

If we assume $v = H_0 d_c$ applies to find the co-moving distance (in SR we have no way of accommodating further redshifting after photon emission, so we assume the galaxy still has the same velocity today, and follows the Hubble flow), we can use Eqn. 13 to determine the luminosity distance. We also use Eqns. 11 and 13 to determine the luminosity distance in GR. We compare this to the calculated luminosity distance using SNe Ia (any standardizable candle allows one to properly calculate the luminosity distance). The result is plotted in Fig. 11, and shows a clear bias against the special relativistic model. The reason why in the figure SR does even worse than Newtonian is simply because as $v \rightarrow c$, $d_c \rightarrow c/H_0$, resulting in a linear relationship between luminosity distance and redshift. This is not the case in either Newtonian or GR.

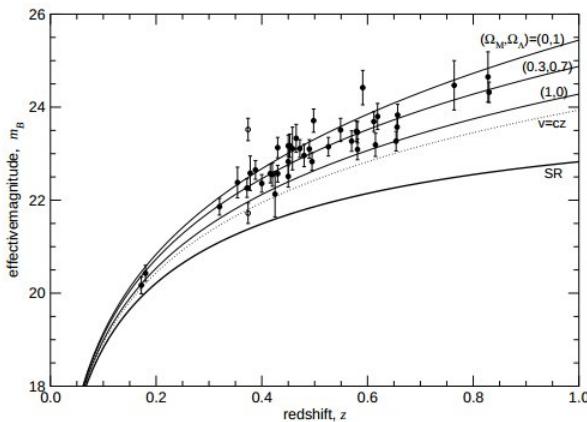


FIG. 11.— A plot of the magnitude-redshift relation, with a comparison between SR, Newtonian ($v = cz$) and several Λ CDM universes. Magnitude is calculated from luminosity distance. From Davis & Lineweaver (2004), their Fig. 5.

1.7.3. Can galaxies with recession velocities $v > c$ slow down until $v < c$?

Certainly! $v = H(t)d_p(t) = \dot{a}(t)d_c$, and therefore $v/c = \dot{a}(t)d_c/c$. For a matter-dominated universe $\dot{a}(t) \propto t^{-1/3}$ and therefore over time objects “slow down” (we cannot observe this, of course; light from these objects has yet to reach us!). This can be seen by the Hubble sphere expanding in Fig. 10.

1.8. Question 7

QUESTION: What happened in the first 3 minutes after the Big Bang? Why is only He (and tiny traces of Li) synthesized in the Big Bang?

A whole bunch of things happened in the first few minutes after the Big Bang, including inflation, CP symmetry breaking, neutrino decoupling. These features are summarized in Sec. 1.12. This question speaks mainly, however, of Big Bang nucleosynthesis (BBN).

The energy scale of BBN is set by the binding energy of nuclei - deuterium binding is about 10^5 times greater than the ionization energy of a hydrogen atom, and as a result BBN occurred when $T \approx 4 \times 10^8$ K. The universe grew too cold to maintain such temperatures when it was only several minutes old.

The basic building blocks of matter are protons and neutrons. A free neutron has 1.29 MeV more energy than a proton, and 0.78 MeV more than a proton and electron. $n \rightarrow p + e^- + \bar{n}u_e$, then, is energetically (and entropically) highly favourable, and the half-life of a neutron is about 890 seconds.

At age $t = 0.1$ s, $T \approx 3 \times 10^{10}$ K, and the mean energy per photon was about $E \approx 10$ MeV, high enough to easily begin pair production. Neutrons and protons will be at equilibrium with each other via $n + \nu_e \rightleftharpoons p + e^-$ and $n + e^+ \rightleftharpoons p + \bar{n}u_e$, and given LTE, their densities will be given by the Maxwell-Boltzmann equation,

$$n = g \left(\frac{mkT}{2\pi\hbar^2} \right)^{3/2} \exp \left(-\frac{E}{k_B T} \right), \quad (32)$$

where the energy scale we consider is the rest mass of a proton vs. a neutron. The relative balance of neutrons and protons, then, is given by

$$\frac{n_n}{n_p} = \left(\frac{m_n}{m_p} \right)^{3/2} \exp \left(-\frac{(m_n - m_p)c^2}{k_B T} \right) \approx \exp \left(-\frac{Q}{k_B T} \right), \quad (33)$$

where $Q = 1.29$ MeV, which corresponds to $\sim 1.5 \times 10^{10}$ K. This shows a high preference for protons at low temperatures. In truth, however, $n + \nu_e \rightleftharpoons p + e^-$ is a weak reaction and the cross-sectional dependence of a weak reaction is $\sigma_w \propto T^2$. Since in a radiation-dominated universe $T \propto t^{-1/2}$, $\omega_w \propto t^{-1}$, and the neutron density is greater than $\propto t^{-3/2}$ (from $\rho \propto a^{-3}$ and the fact that neutron numbers are decreasing with temperature). As a result, Γ_w falls dramatically. When $\Gamma \approx H$, the neutrinos decouple from the neutrons and protons. This occurs (empirically) at about 0.8 MeV, or $T_{\text{freeze}} = 9 \times 10^9$ K. Using Eqn. 33, we obtain 1 neutron for 5 protons.

The lack of neutrons prevented BBN from fusing to nickel. Proton-proton fusion is difficult due to Coulombic repulsion, and in the Sun the pp-chain has a timescale of several Gyr. This means that in the several minutes when the temperature of the universe was sufficiently high for nuclear fusion to occur, $p + n \rightleftharpoons D + \gamma$ dominated (neutron-neutron fusion has a very small cross-section). if every neutron binded to a proton, and the only nucleosynthetic product was ${}^4\text{He}$, the fraction of ${}^4\text{He}/\text{H}$ would be $(2 \text{ neutrons} + 2 \text{ protons})/(6 \text{ free protons}) = 1/3$.

This fusion happened in several stages. The time of deuterium fusion (when $n_D/n_n = 1$) occurred at $T \approx 7.6 \times 10^8$ K, or $t \approx 200$ s - this can be derived from the Saha equation ($g_D = 3$). Deuterium can then be fused into tritium (${}^3\text{H}$, half-life 18 years) or ${}^3\text{He}$, and from there quickly fused into ${}^4\text{He}$. ${}^4\text{He}$ is very tightly bound (hence α -decay), and there are no stable nuclei with atomic weight 5 or 8. Small amounts of ${}^6\text{Li}$ and ${}^7\text{Li}$ can be made via ${}^4\text{He} + D \rightleftharpoons {}^6\text{Li} + \gamma$ and ${}^4\text{He} + {}^3\text{H} \rightleftharpoons {}^7\text{Li} + \gamma$, which are fairly slow reactions. By the time the temperature has dropped to $T \approx 4 \times 10^8$ K at $t = 10$ min, BBN is over, and neutrons are locked up in ${}^4\text{He}$ and a small amount of Li.

1.8.1. How does nucleosynthesis scale with cosmological parameters?

Nucleosynthesis depends critically on η , the baryon-to-photon ratio. A high ratio increases the temperature at which deuterium synthesis occurs, and hence gives an earlier start to BBN, allowing a greater conversion of D to ${}^4\text{He}$. ${}^7\text{Li}$ is produced both by fusing ${}^4\text{He}$ and ${}^3\text{He}$ (decreases with increased baryon fraction) and by electron capture of ${}^7\text{Be}$ (increases). Fig. 12 shows the nucleosynthetic products of BBN as a function of baryon density.

1.8.2. How do we determine primordial densities if D is easily destroyed in stars?

One way is to look at Ly- α transitions in the neutral, high-redshift ISM. The greater mass of the D nucleus shifts slightly downward (i.e. more negative energy) the energy levels of the electron, creating a slightly shorter Ly- α transition.

1.8.3. Why is there more matter than antimatter?

When the temperature of the universe was greater than 150 MeV, quarks would roam free, and photons could pair-produce quarks. The various flavours of quarks were in LTE with each other, and very nearly equal. CP violation, however, produced a $\sim 10^{-9}$ bias in favour of quarks, and when the temperature cooled enough that quark pair production was no longer favourable, the quarks and antiquarks annihilated, producing an enormous photon to baryon ratio, and leaving only quarks. A similar situation occurred for leptons.

1.8.4. What are WIMPs?

WIMPs (weakly interacting massive particles) are dark matter particle candidates that, due to their small cross-sections, would have stopped interacting with baryons at about the same time as neutrino decoupling. If WIMPs have masses < 1 MeV, they would be ultrarelativistic today and would have the same number density as neutrinos. This gives $\Omega_{\text{WIMP}} h^2 \approx \frac{m_{\text{WIMP}}}{91.5 \text{ eV}}$ (this also applies to neutrinos). The mass of an individual WIMP, then must be < 100 eV. If instead the WIMP is massive, then it is not relativistic, and, then $\Omega_{\text{WIMP}} h^2 \approx \left(\frac{m_{\text{WIMP}}}{1 \text{ TeV}} \right)^2$.

1.9. Question 8

QUESTION: Explain how Supernovae (SNe of Type Ia in particular) are used in the measurements of cosmological parameters.

This is adopted from my own qual notes.

Suppose a very bright standard candle exists throughout the history of the universe; since the luminosity of the candle is known, we would be able to use it to measure the luminosity distance (Eqn. 13, using $R = \frac{c}{H_0} \sqrt{|\Omega_\kappa|}$ and sinn to represent sin, sinh, etc.):

$$d_L = (1+z) \frac{c}{H_0} \frac{1}{\sqrt{|\Omega_\kappa|}} \text{sinn} \left(\sqrt{|\Omega_\kappa|} H_0 \int_0^z \frac{dz}{H} \right) \quad (34)$$

Q6) TYPE 1A SUPERNOVAE FOR COSMOLOGY

Explain how and why Type Ia Supernovae are used in the measurements of cosmological parameters.

Q6) Ludwig Cosmo Q6

Question 6 - Type IA Supernova

Explain how and why Type Ia Supernovae are used in the measurements of cosmological parameters.

Relevant Equations:

- $m_{\text{apparent}} - M_{\text{absolute}} = 5 \log_{10}(\frac{d}{10})$ Distance Modulus
- $v = cz$
- $v = H_0 d$
- $d = \frac{c}{H_0} z$
- $H = H_0 \sqrt{\Omega_m a^{-3} + \Omega_r a^{-4} + \Omega_k a^{-2} + \Omega_\Lambda}$

For supernova, we can measure a) the apparent magnitude and b) the light curve.

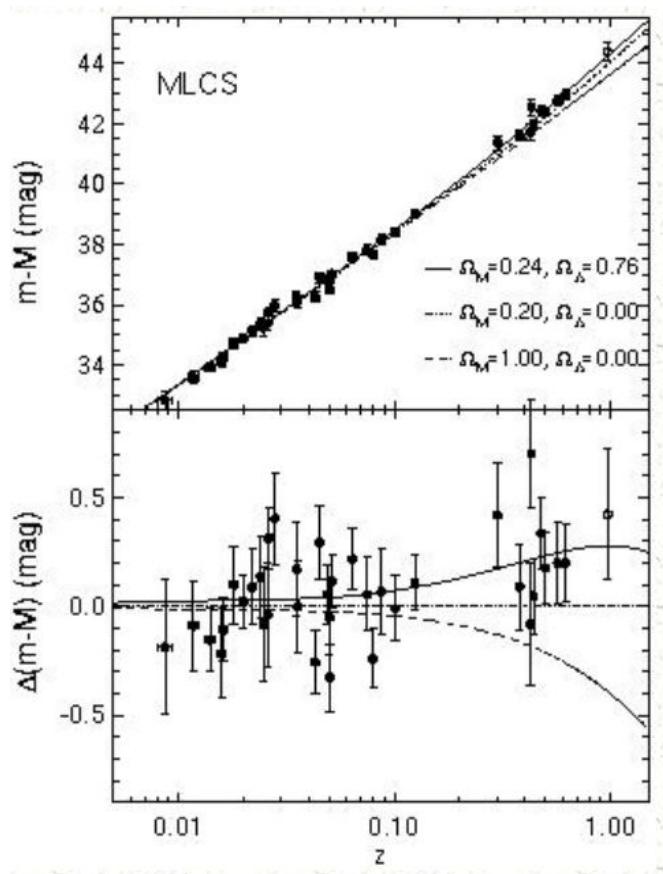
From the light curve we can get the absolute magnitude.

Plugging the absolute and apparent magnitudes into the distance modulus gives us the distance to these supernova.

Follow Up

Why are type Ia SN standardizable?

- Type Ia SN occur when a companion binary star accretes mass onto a white dwarf until the Chandrasekhar limit is reached, in which case it explodes. The Chandrasekhar limit is pretty universal so you get a consistent progenitor mass.



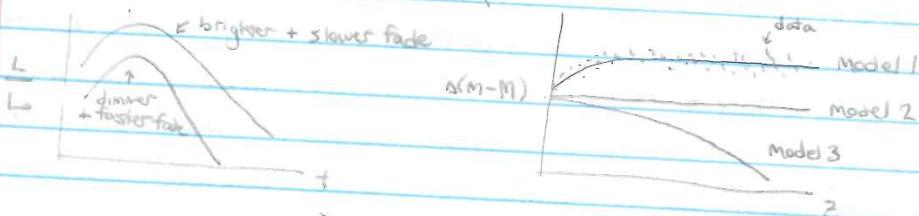
Q6) Herman Cosmo Q7

C7

- M related to light curve shape
- get $M, m \rightarrow d \rightarrow z$
- shape of $\Delta(m-M)$ v. z varies w/ choice of cosmological parameter values

Type Ia Supernovae

- The absolute magnitude of a Type Ia SN is directly related to the shape of its light curve. By measuring the light curve shape we can determine the absolute magnitude and use this (plus the apparent mag) with the distance modulus to measure their distance. This allows us to trace the expansion of the universe, because we can plot their $\Delta(m-M)$ as a function of z and compare to models w/ different cosmological parameters, like Ω_m and Ω_Λ .



- The luminosity of Type Ia SNe is almost uniform, but does have some variance as SN explosions aren't all the exact same. To get them to be useful as standard candles, we have to account for their 'stretch factor', which standardizes the peak luminosity based on the stretched-ness of the light curve relative to some template. The Multicolor light curve shape (MLCS, how the light curve changes for $\sim 15d$ post peak brightness) allows us to compare the observed light curve to a parametrized one to get the peak L . It allows the reddening and dimming/extinction effect of ISM dust to be detected and removed.

- I believe the difference in decay times for different light curves is due to the abundance of ^{56}Ni : This decays into ^{58}Co & ^{56}Fe , which leads to a slower decline as \uparrow opacity $\rightarrow \uparrow T \rightarrow \uparrow$ peak L ...? ^{56}Ni : $T \sim 9$ days, releases energy during decay

- Apparent SNe at $\uparrow z$ look brighter... $D_L \sim \frac{c}{H_0} z \sim \sqrt{\frac{L}{4\pi F}}$

Hilary

Q6) Campbell Cosmo Q7

1.8 Question 7

Explain how and why Type Ia supernovae are used in the measurements of cosmological parameters.

1.8.1 Short answer

Type 1a are good standard candles – they are very luminous and standardized. The average Type 1a has a peak luminosity of $L = 4 \times 10^9 L_\odot$ which is 100,000 times brighter than even the brightest Cepheid variable. When $z \ll 1$, the luminosity distance to the light source is

$$d_L \approx \frac{c}{H_0} z \left(1 + \frac{1 - q_0}{2} z \right) [\text{Mpc}]$$

The recipe for finding the Hubble constant and deceleration parameters is a simple one:

- Measure the redshift z and flux F for each SN.
- Compute the luminosity distance $d_L = \sqrt{L/4\pi F}$ for each SN.
- Plot the recessional velocity cz versus luminosity distance d_L .
- Measure the slope of the cz versus d_L relation when $z \ll 1$; this gives H_0 .
- At slightly larger values of z (but still < 1), the deviation of the plot from a straight line tells you the value of q_0 .

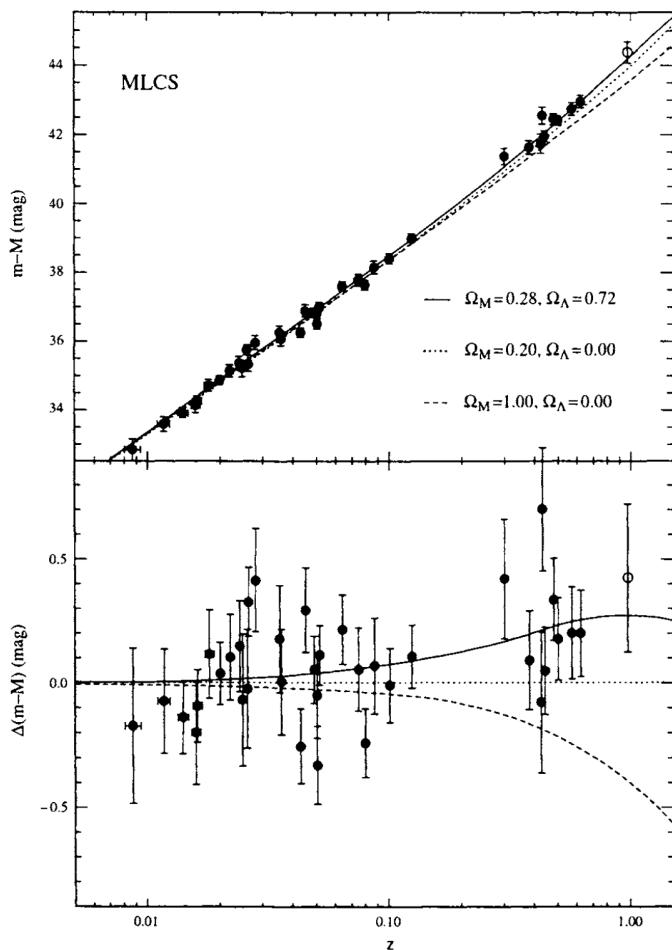


Figure 14: Hubble diagram from distant Type Ia supernovae. Top panel shows apparent magnitude (an indicator of the distance) vs redshift. Lines show the predictions for different energy contents in the Universe, with Ω_M the ratio of energy density today in matter compared to the critical density and Ω_Λ the ratio of energy density in a cosmological constant to the critical density. Bottom panel plots the residuals, making it clear that the high-redshift supernovae favor a Λ -dominated Universe over a matter-dominated one. Figure taken from Dodelson (2003).

1.8.2 Additional context

In recent years, the standard candle of choice among cosmologists has been Type Ia supernovae (SNe). A supernova may be loosely defined as an exploding star. Early in the history of supernova studies, when little was known about their underlying physics, supernovae were divided into two classes, on the basis of their spectra. Type I supernovae contain no hydrogen absorption lines in their spectra; Type II supernovae contain strong hydrogen absorption lines. Gradually, it was realized that all Type II supernovae are the same species of beast; they are massive stars ($M > 8 M_{\odot}$) whose cores collapse to form a black hole or neutron star when their nuclear fuel is exhausted. During the rapid collapse of the core, the outer layers of the star are thrown off into space. Type I supernovae are actually two separate species, which are called Type Ia and Type Ib. Type Ib supernovae, it is thought, are massive stars whose cores collapse after the hydrogen-rich outer layers of the star have been blown away in strong stellar winds. Thus, Type Ib and Type II supernovae are driven by very similar mechanisms – their differences are superficial, in the most literal sense. Type Ia supernovae, however, are something completely different. They occur in close binary systems where one of the two stars in the system is a white dwarf; that is, a stellar remnant which is supported against gravity by electron degeneracy pressure. The transfer of mass from the companion star to the white dwarf eventually nudges the white dwarf over the Chandrasekhar limit of $1.4 M_{\odot}$; this is the maximum mass at which the electron degeneracy pressure can support a white dwarf against its own self-gravity. When the Chandrasekhar limit is exceeded, the white dwarf starts to collapse until its increased density triggers a runaway nuclear fusion reaction. The entire white dwarf becomes a fusion bomb, blowing itself to smithereens; unlike Type II supernovae, Type Ia supernovae do not leave a condensed stellar remnant behind.

Within our Galaxy, Type Ia supernovae occur roughly once per century, on average. Although Type Ia supernovae are not frequent occurrences locally, they are extraordinarily luminous, and hence can be seen to large distances. The luminosity of an average Type Ia supernova, at peak brightness, is $L = 4 \times 10^9 L_{\odot}$; that's 100,000 times more luminous than even the brightest Cepheid. For a few days, a Type Ia supernova in a moderately bright galaxy can outshine all the other stars in the galaxy combined. Since moderately bright galaxies can be seen at $z \sim 1$, this means that Type Ia supernovae can also be seen at $z \sim 1$. Not only are Type Ia supernovae bright standard candles, they are also reasonably standardized standard candles. Consider Type Ia supernovae in the Virgo cluster. Although there's only one Type Ia supernova per century in our own Galaxy, the total luminosity of the Virgo cluster is a few hundred times that of our Galaxy. Thus, every year you can expect a few type Ia supernovae to go off in the Virgo cluster. Several type Ia supernovae have been observed in the Virgo cluster in the recent past, and have been found to have similar fluxes at maximum brightness.

So far, Type Ia supernovae sound like ideal standard candles; very luminous and very standardized. There's one complication, however. Observation of supernovae in galaxies whose distances have been well determined by Cepheids reveal that Type Ia supernovae do not have identical luminosities. Instead of all having $L = 4 \times 10^9 L_{\odot}$, their peak luminosities lie in the fairly broad range $L = 3 \rightarrow 5 \times 10^9 L_{\odot}$. However, it has also been noted that the peak luminosity of a Type Ia supernova is tightly correlated with the shape of its light curve. Type Ia supernovae with luminosities that shoot up rapidly and decline rapidly are less luminous than average at their peak; supernovae with luminosities that rise and fall in a more leisurely manner are more luminous than average at their peak. Thus, just as the period of a Cepheid tells you its luminosity, the rise and fall time of a Type Ia supernova tells you its peak luminosity.

Recently, two research teams, the “Supernova Cosmology Project” and the “High-z Supernova Search Team”, have been conducting searches for supernovae in distant galaxies. They have used the observed light curves and redshifts of Type Ia supernovae to measure cosmological parameters. First, by observing Type Ia supernovae at $z \sim 0.1$, the value of H_0 can be determined. The results of the different groups are in reasonable agreement with each other. If the distance to the Virgo cluster is pegged at $d_L = 15$ Mpc, as indicated by the Cepheid results, then the observed supernovae fluxes and redshifts are consistent with $H_0 = 70 \pm 7 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

In addition, the supernova groups have been attempting to measure the acceleration (or deceleration) of the Universe by observing Type Ia supernovae at higher redshift. Before discussing these results, let's introduce the “magnitude” system used by astronomers to express fluxes and luminosities. The magnitude system, like much else in astronomy, has its roots in ancient Greece. The Greek astronomer Hipparchus, in the second century BC, divided the stars into six classes, according to their apparent brightness. The brightest stars were of “first magnitude”, the faintest stars visible to the naked eye were of “sixth magnitude”, and intermediate stars were ranked as second, third, fourth, and fifth magnitude. Long after the time of Hipparchus, it was realized that the response of the human eye is roughly logarithmic, and that stars of the first magnitude have fluxes (at visible wavelengths) about 100 times greater than stars of the sixth magnitude. On the basis of this realization, the magnitude system was placed on a more rigorous mathematical basis.

Q6) Campbell Cosmo Q7

Nowadays, the bolometric apparent magnitude of a light source is defined in terms of the sources **bolometric flux** as

$$m \equiv -2.5 \log_{10} \left(\frac{F}{F_0} \right) [\text{mag}],$$

where the reference flux F_0 is set at the value $F_0 = 2.53 \times 10^{-8} [\text{W m}^{-2}]$. Thanks to the negative sign in the definition, a small value of m corresponds to a large flux F . For instance, the flux of sunlight at the Earth's location is $F = 1367 [\text{W m}^{-2}]$; the Sun thus has a bolometric apparent magnitude of $m = 26.8$. The choice of reference flux F_0 constitutes a tip of the hat to Hipparchus, since for stars visible to the naked eye it typically yields $0 < m < 6$.

The bolometric absolute magnitude of a light source is defined as the apparent magnitude that it would have if it were at a luminosity distance of $d_L = 10 \text{ pc}$. Thus, a light source with luminosity L has a bolometric **absolute magnitude** of

$$M \equiv -2.5 \log_{10} \left(\frac{L}{L_0} \right) [\text{mag}],$$

where the reference luminosity is $L_0 = 78.7 L_\odot$, since that is the luminosity of an object which produces a flux $F_0 = 2.53 \times 10^{-8} [\text{W m}^{-2}]$ when viewed from a distance of 10 pc. The bolometric absolute magnitude of the Sun is thus $M = 4.74$. Although the system of apparent and absolute magnitudes seems strange to the uninitiated, the apparent magnitude is really nothing more than a logarithmic measure of the flux, and the absolute magnitude is a logarithmic measure of the luminosity.

Given the definitions of apparent and absolute magnitude, the relation between an object's apparent magnitude and its absolute magnitude can be written in the form

$$M = m - 5 \log_{10} \left(\frac{d_L}{10 \text{ pc}} \right) [\text{mag}],$$

where d_L is the luminosity distance to the light source. If the luminosity distance is given in units of megaparsecs, this relation becomes

$$M = m - 5 \log_{10} \left(\frac{d_L}{1 \text{ Mpc}} \right) - 25 [\text{mag}].$$

Since astronomers frequently quote fluxes and luminosities in terms of apparent and absolute magnitudes, they find it convenient to quote luminosity distances in terms of the distance modulus to a light source. The **distance modulus** is defined as $m - M$, and is related to the luminosity distance by the relation

$$m - M = 5 \log_{10} \left(\frac{d_L}{1 \text{ Mpc}} \right) - 25 [\text{mag}].$$

The distance modulus of the Large Magellanic Cloud (LMC), for instance, at $d_L = 0.05 \text{ Mpc}$, is $m - M = 18.5$. The distance modulus of the Virgo cluster, at $d_L = 15 \text{ Mpc}$, is $m - M = 30.9$.

Unfortunately, the current proper distance to a Type 1a supernova is not a measurable property. If you tried to measure the distance with a tape measure, for instance, the distance would be continuously increasing as you extended the tape. To measure the proper distance at time t_0 , you would need a tape measure which could be extended with infinite speed; alternatively, you would need to stop the expansion of the Universe at its current scale factor while you measured the distance at your leisure. Neither of these alternatives is physically possible.

Since cosmology is ultimately based on observations, if we want to find the distance to a galaxy, we need some way of computing a distance from that galaxy's observed properties. For objects at cosmological distances, we can measure the flux of light, F , from the object. The complete flux, integrated over all wavelengths of light, is called the bolometric flux. (A bolometer is an extremely sensitive thermometer capable of detecting electromagnetic radiation over a wide range of wavelengths; it was invented in 1881 by the astronomer Samuel Langley, who used it to measure solar radiation.) More frequently, given the difficulties of measuring the true bolometric flux, the flux over a limited range of wavelengths is measured. If the light from the object has emission or absorption lines, we can measure the redshift, z .

Since Type 1a supernovae are standard candles in the sense that their luminosities can be measured, then you can use its measured flux F to define a function called the **luminosity distance**:

$$d_L \equiv \sqrt{\frac{L}{4\pi F}} [\text{Mpc}].$$

The function d_L is called a “distance” because its dimensionality is that of a distance, and because it is what the proper distance to the standard candle would be if the Universe were static and Euclidean. In a static Euclidean Universe, the propagation of light follows the inverse square law $F = L/(4\pi d^2)$.

Q6) Campbell Cosmo Q7

Suppose, though, that you are in a Universe described by a **Robertson-Walker metric**:

$$ds^2 = -c^2 dt^2 + a(t)^2 (dr^2 + S_\kappa(r)^2 d\Omega^2),$$

with

$$S_\kappa(r) = \begin{cases} R \sin(r/R), & (\kappa = +1) \\ r, & (\kappa = 0) \\ R \sinh(r/R), & (\kappa = -1) \end{cases}.$$

You are at the origin. At the present moment $t = t_0$, you see light that was emitted by a standard candle at comoving coordinate location (r, θ, ϕ) at a time t_e . The photons which were emitted at time t_e are, at the present moment, spread over a sphere of proper radius $d_p(t_0) = r$ and proper surface area $A_p(t_0)$. If space is flat ($\kappa = 0$), then the proper area of the sphere is given by the Euclidean relation $A_p(t_0) = 4\pi d_p(t_0)^2 = 4\pi r^2$. More generally, however,

$$A_p(t_0) = 4\pi S_\kappa(r)^2 [\text{Mpc}^2].$$

When space is positively curved, $A_p(t_0) < 4\pi r^2$, and the photons are spread over a smaller area than they would be in flat space. When space is negatively curved, $A_p(t_0) > 4\pi r^2$, and photons are spread over a larger area than they would be in flat space.

In addition to these geometric effects, which would apply even in a static Universe, the expansion of the Universe causes the observed flux of light from a standard candle of redshift z to be decreased by a factor of $(1+z)^2$. First, the expansion of the Universe causes the energy of each photon from the standard candle to decrease. If a photon starts with an emitted energy $E = hc/\lambda$ when the scale factor is $a(t)$, by the time we observe it when the scale factor is $Aa(t_0) = 1$, the wavelength will have grown to

$$\lambda_0 = \frac{1}{a(t)} \lambda = (1+z)\lambda [\text{m}],$$

and the energy will have fallen to

$$E_0 = E(1+z)^{-1} [\text{eV}].$$

Second, thanks to the expansion of the Universe, the time between photon detections will be greater. If two photons are emitted in the same direction separated by a time interval δt , the proper distance between them will initially be $c\delta t$; by the time we detect the photons at time t_0 , the proper distance between them will be stretched to $c\delta t(1+z)$, and we will detect them separated by a time interval $\delta t_0 = \delta t(1+z)$.

The net result is that in an expanding, spatially curved Universe, the relation between the observed flux F and the luminosity L of a distant light source is

$$F = \frac{L}{4\pi S_\kappa(r)^2 (1+z)^2} [\text{W m}^{-2}],$$

and the luminosity distance is

$$d_L = S_\kappa(r)(1+z) [\text{Mpc}].$$

The available evidence indicates that our Universe is nearly flat, with a radius of curvature R_0 which is larger than the current horizon distance $d_{\text{hor}}(t_0)$. Objects with finite redshift are at proper distances smaller than the horizon distance, and hence smaller than the radius of curvature. Thus, it is safe to make the approximation $r \ll R_0$, implying $S_\kappa(r) \approx r$. With our assumption that space is very close to being flat, the relation between the luminosity distance and the current proper distance becomes very simple:

$$d_L(\kappa=0) = r(1+z) = d_p(t_0)(1+z) [\text{Mpc}].$$

Thus, even if space is perfectly flat, if you estimate the distance to a standard candle by using a naive inverse square law, you will overestimate the actual proper distance by a factor $(1+z)$.

Figure 15 shows the luminosity distance d_L as a function of redshift z for the Benchmark Model, and for two other flat Universes, one dominated by matter and one dominated by a cosmological constant Λ . When $z \ll 1$, the current proper distance may be approximated as

$$d_p(t_0) \approx \frac{c}{H_0} z \left(1 - \frac{1+q_0}{2} z\right) [\text{Mpc}].$$

In a Universe which is nearly flat, the luminosity distance may thus be approximated as

Q6) Campbell Cosmo Q7

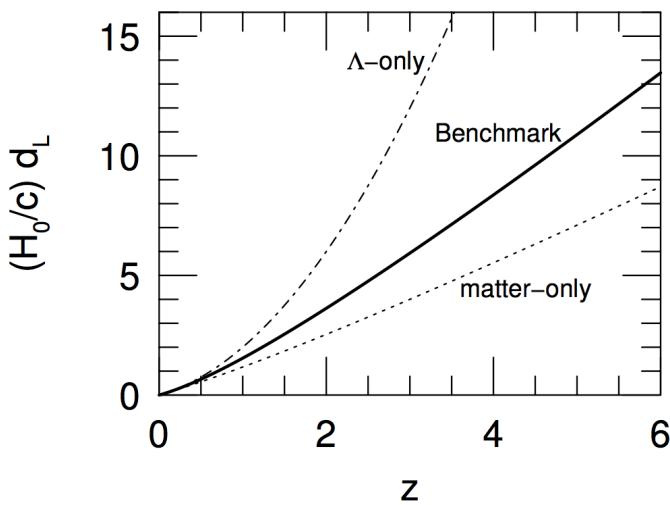


Figure 15: The luminosity distance of a standard candle with observed redshift z . The bold solid line gives the result for the Benchmark Model, the dot-dash line for a flat, Λ -only Universe, and the dotted line for a flat, matter-only Universe. Figure taken from Ryden (2006).

$$\begin{aligned} d_L &\approx \frac{c}{H_0} z \left(1 - \frac{1 + q_0}{2} z \right) (1 + z) \\ &\approx \frac{c}{H_0} z \left(1 + \frac{1 - q_0}{2} z \right). \end{aligned}$$

Note that in the limit $z \rightarrow 0$,

$$d_p(t_0) \approx d_L \approx \frac{c}{H_0} z \text{ [Mpc].}$$

In a Universe described by the Robertson-Walker metric, the luminosity distance is a good approximation to the current proper distance for objects with small redshifts.

Substituting the luminosity distance approximation into the expression for the distance modulus gives us the relation between distance modulus and redshift:

$$m - M \approx 43.17 - 5 \log_{10} \left(\frac{H_0}{70 \text{ km s}^{-1} \text{ Mpc}} \right) + 5 \log_{10} z + 1.086(1 - q_0)z \text{ [mag].}$$

For a population of standard candles with known luminosity L (and hence of known bolometric absolute magnitude M), you measure the flux F (or equivalently, the bolometric apparent magnitude m) and the redshift z . In the limit $z \rightarrow 0$, a plot of $m - M$ versus $\log_{10}(z)$ gives a straight line whose amplitude at a given value of z tells you the value of H_0 . At slightly larger values of z , the deviation of the plot from a straight line tells you the value of q_0 . At a given value of z , an accelerating Universe (with $q_0 < 0$) yields standard candles with a smaller flux than would a decelerating Universe (with $q_0 > 0$).

The upper panel of Figure 16 shows the plot of distance modulus versus redshift for the combined supernova samples of the High-z Supernova Search Team (given by the filled circles) and the Supernova Cosmology Project (given by the open circles). The observational results are compared to the expected results for three model Universes. One Universe is flat, and contains nothing but matter ($\Omega_{m,0} = 1, q_0 = 0.5$). The second is negatively curved, and contains nothing but matter ($\Omega_{m,0} = 0.3, q_0 = 0.15$). The third is flat, and contains both matter and a cosmological constant ($\Omega_{m,0} = 0.3, \Omega_{\Lambda,0} = 0.7, q_0 = 0.55$). The data are best fitted by the third of the models – which is, in fact, our Benchmark Model. The bottom panel of Figure 16 shows this result more clearly. It shows the difference between the data and the predictions of the negatively curved, matter-only model. The conclusion that the Universe is accelerating derives from the observation that the supernovae seen at $z \sim 0.5$ are, on average, about 0.25 mag fainter than they would in a decelerating Universe with $\Omega_{m,0} = 0.3$ and no cosmological constant.

The supernova data extend out to $z \sim 1$; this is beyond the range where an expansion in terms of H_0 and q_0 is adequate to describe the scale factor $a(t)$. Thus, the two supernova teams customarily describe their results in terms of a model Universe which contains both matter and a cosmological constant. After choosing values of $\Omega_{m,0}$ and $\Omega_{\Lambda,0}$, they compute the expected relation between $m - M$ and z , and compare it to the observed data. The results of fitting these model Universes are given in Figure 17. The ovals drawn on Figure 17 enclose those values of $\Omega_{m,0}$ and $\Omega_{\Lambda,0}$ which give the best fit to the supernova data. The results of the two teams (the solid ovals and dotted ovals) give very similar results. Three concentric ovals are shown for each team's result; they correspond to 1σ , 2σ , and 3σ confidence

Q6) Campbell Cosmo Q7

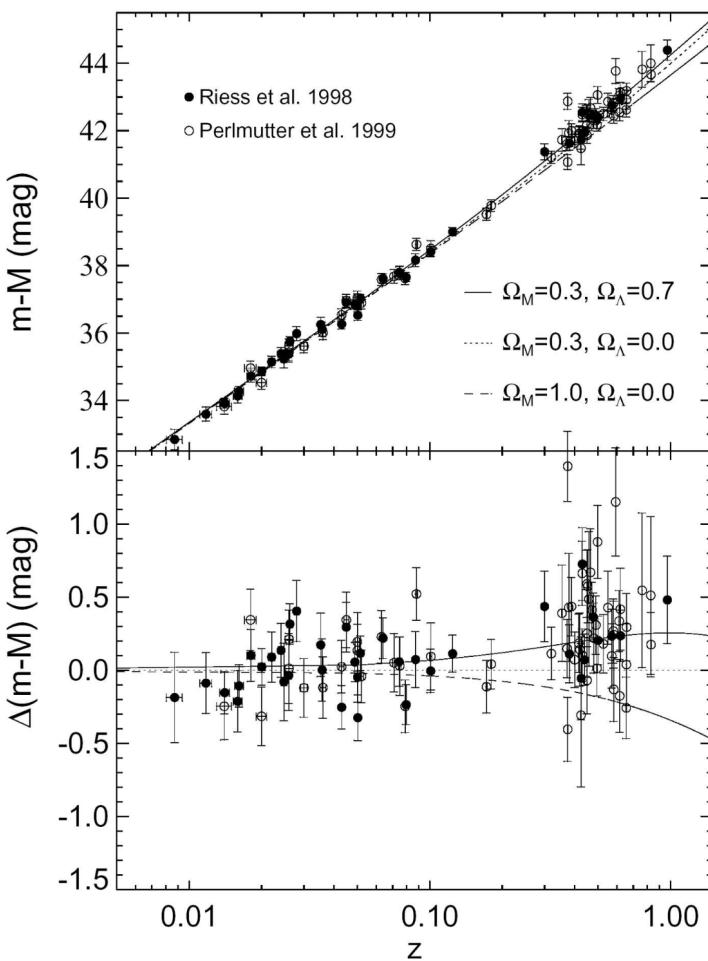


Figure 16: Distance modulus versus redshift for Type Ia supernovae from the Supernova Cosmology Project (Perlmutter et al. 1999, ApJ, 517, 565) and the High-z Supernova Search Team (Riess et al. 1998, AJ, 116, 1009). The bottom panel shows the difference between the data and the predictions of a negatively curved $\Omega_{m,0} = 0.3$ model (from Riess 2000, PASP, 112, 1284). Figure taken from Ryden (2006).

intervals, with the inner oval representing the highest probability. The best fitting models lie along the line $0.8\Omega_m, 0 - 0.6\Omega_{\Lambda,0} \approx -0.2$. Note that decelerating Universes (with $q_0 > 0$) can be strongly excluded by the data, as can Big Crunch Universes (labeled ‘Recollapses’ in Figure 17), and Big Bounce Universes (labeled ‘No Big Bang’ in Figure 17). The supernova data are consistent with negative curvature (labeled ‘Open’ in Figure 17), positive curvature (labeled ‘Closed’ in Figure 17), or with a Universe which is spatially flat.

The results of the supernova teams made headlines when they were first announced; the discovery of the accelerating Universe was named by Science magazine as the ‘Scientific Breakthrough of the Year’ for 1998. It is prudent to remember, however, that all the hoopla about the accelerating Universe is based on the observation that Type Ia supernova at $z \sim 0.5$ and beyond have somewhat lower fluxes (by about 25%) than they would have in a decelerating Universe. There are other reasons why their fluxes might be low. For instance, if Type Ia supernovae were intrinsically less luminous at $z \sim 0.5$ than at $z \sim 0$, that could explain their low fluxes. (If a typical supernova at $z \sim 0.5$ had $L = 3 \times 10^9 L_\odot$ rather than $L = 4 \times 10^9 L_\odot$, that would explain their observed dimness, without the need to invoke a cosmological constant. Conversely, if the typical supernova at $z \sim 0.5$ had $L = 5 \times 10^9 L_\odot$ rather than $L = 4 \times 10^9 L_\odot$, that would require an even larger cosmological constant to explain their observed dimness.) However, the other properties of Type Ia supernovae, such as their spectra, don’t seem to evolve with time, so why should their luminosity? Perhaps the fluxes of supernovae at $z \sim 0.5$ are low because some of their light is scattered or absorbed by intervening dust. However, dust tends to scatter some wavelengths of light more than others. This would change the shape of the spectrum of distant Type Ia supernovae, but no dependence of spectral shape on redshift is observed.

In sum, the supernova results of Figure 17 provide persuasive evidence for a nearly flat accelerating Universe with $\Omega_{m,0} \approx 0.3$ and $\Omega_{\Lambda,0} \approx 0.7$.

Q6) Campbell Cosmo Q7

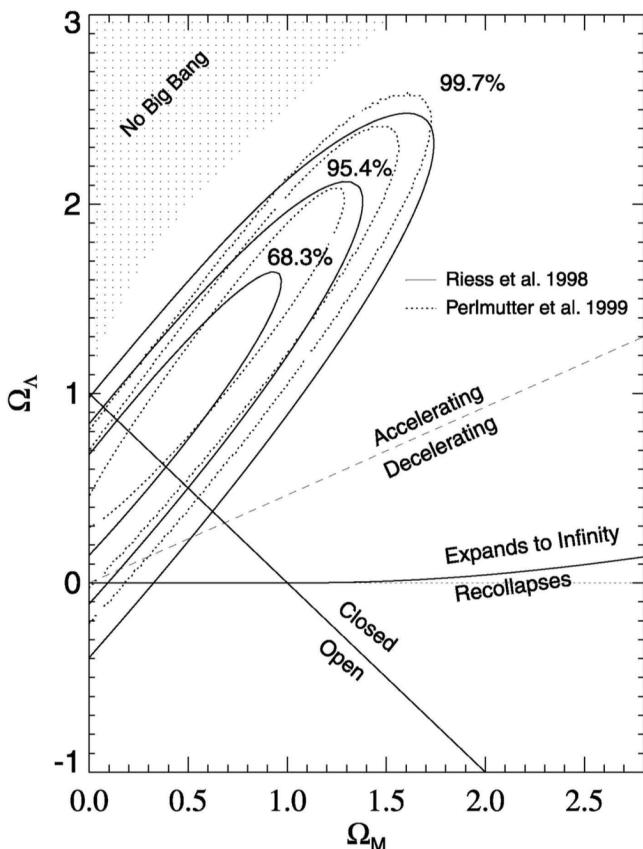


Figure 17: The values of $\Omega_{m,0}$ (horizontal axis) and $\Omega_{\Lambda,0}$ (vertical axis) which best fit the data shown in Figure 16. The solid ovals show the best-fitting values for the High-z Supernova Search Team data; the dotted ovals show the best-fitting values for the Supernova Cosmology Project data (from Riess 2000, PASP, 112, 1284) Figure taken from Ryden (2006).

1.8.3 Follow-up Questions

- What are Type Ia supernovae (i.e., what is physically happening)?
- Can Type Ia constrain all cosmological parameters?
- What are some sources of error? (Maximum luminosity is not perfectly consistent but the light curve can be used to calibrate this due to their tight correlation.)
- How do we know that redshift can be used as a measurement of distance? (Wanted to hear about Hubble's Law.)
- Can other types of supernovae be used as standard candles? Why or why not?

Q6) Zhu Cosmo Q8

20

1.9. Question 8

QUESTION: Explain how Supernovae (SNe of Type Ia in particular) are used in the measurements of cosmological parameters.

This is adopted from my own qual notes.

Suppose a very bright standard candle exists throughout the history of the universe; since the luminosity of the candle is known, we would be able to use it to measure the luminosity distance (Eqn. 13, using $R = \frac{c}{H_0} \sqrt{|\Omega_\kappa|}$ and sinn to represent sin, sinh, etc.):

$$d_L = (1+z) \frac{c}{H_0} \frac{1}{\sqrt{|\Omega_\kappa|}} \text{sinn} \left(\sqrt{|\Omega_\kappa|} H_0 \int_0^z \frac{dz}{H} \right) \quad (34)$$

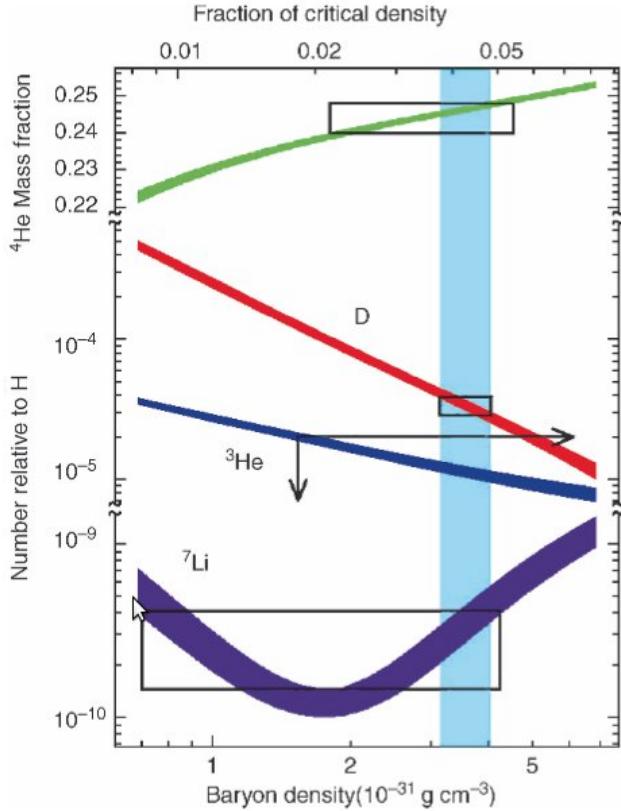


FIG. 12.— BBN predictions of the primordial abundances of light elements as a function of todays baryon density. If a photon density is known, then the x-axis scales with η . Rectangles indicate measured abundance values. Ω_b can be obtained by determining the baryon fraction consistent with measured abundances. From Schneider (2006), his Fig. 4.14.

This allows us to measure the expansion history of the universe.

As a (particularly appropriate) series of examples, let us consider the following. If the universe were empty, $\Omega_\kappa = -1$ and $H = H_0(1+z)$, giving us $d_L = (1+z) \frac{c}{H_0} \sinh \left(\frac{1}{H_0^2} \int_0^z \frac{dz}{1+z} \right)$. Assuming the universe is flat (Sec. 1.3), the luminosity distance reduces to

$$d_L = (1+z)d_c = (1+z)c \int_0^z \frac{dz}{H} \quad (35)$$

If $\Omega_\Lambda = 0$, then $\Omega_m = 1$, which gives $H = 2/3t = H_0(1+z)^{3/2}$. A critical matter-dominated universe has on average smaller d_L than an empty universe. If $\Omega_\Lambda = 1$, then $H = H_0$ and while the comoving distance is much smaller than the curvature radius, d_L will be larger than either the empty or the critical matter-dominated universe.

SNe Ia are standardizable candles: they have a maximum luminosity spread of $\sigma \approx 0.4$ mag in B-band, which is empirically correlated with the rise and decay times of the light curve. This “Phillips Relation” allows for the determination of the true peak absolute magnitude of any SN Ia. (It should be noted that colour is almost just as important as duration; see below.) Since redshift can easily be determined from SNe Ia spectra (Si II P Cygni profile, intermediate element absorption lines, etc.), and SNe Ia are visible out to enormous distances, a plot of luminosity distance vs. redshift, or luminosity distance modulus vs. redshift covering a large portion of the universe’s history is feasible.

Fig. 13 shows the result, from Riess et al. (2004). The results strongly rule out the empty and critical universes discussed earlier, in favour of a universe with a non-zero Ω_Λ . These results are in agreement with the ΛCDM concordance cosmology values $\Omega_m = 0.27$ and $\Omega_\Lambda = 0.73$.

1.9.1. Describe systematic errors.

The two common methods used to determine the luminosity distance moduli to SNe are the stretch and multi-light curve method (MLCS). The stretch method fits a dataset of SNe Ia to a “stretch parameter” α (which adjusts the light curve based on the fall-off timescale equivalent to δm_{15}) and a luminosity/colour slope β . The MLCS method uses a low- z set of SNe to determine the Phillips and colour relations, and then uses these values for high- z SNe. MLCS also treats intrinsic SNe reddening with extinction from the galaxy independently. Both features can add systematics into the analysis. The two methods, however, give remarkably similar results.

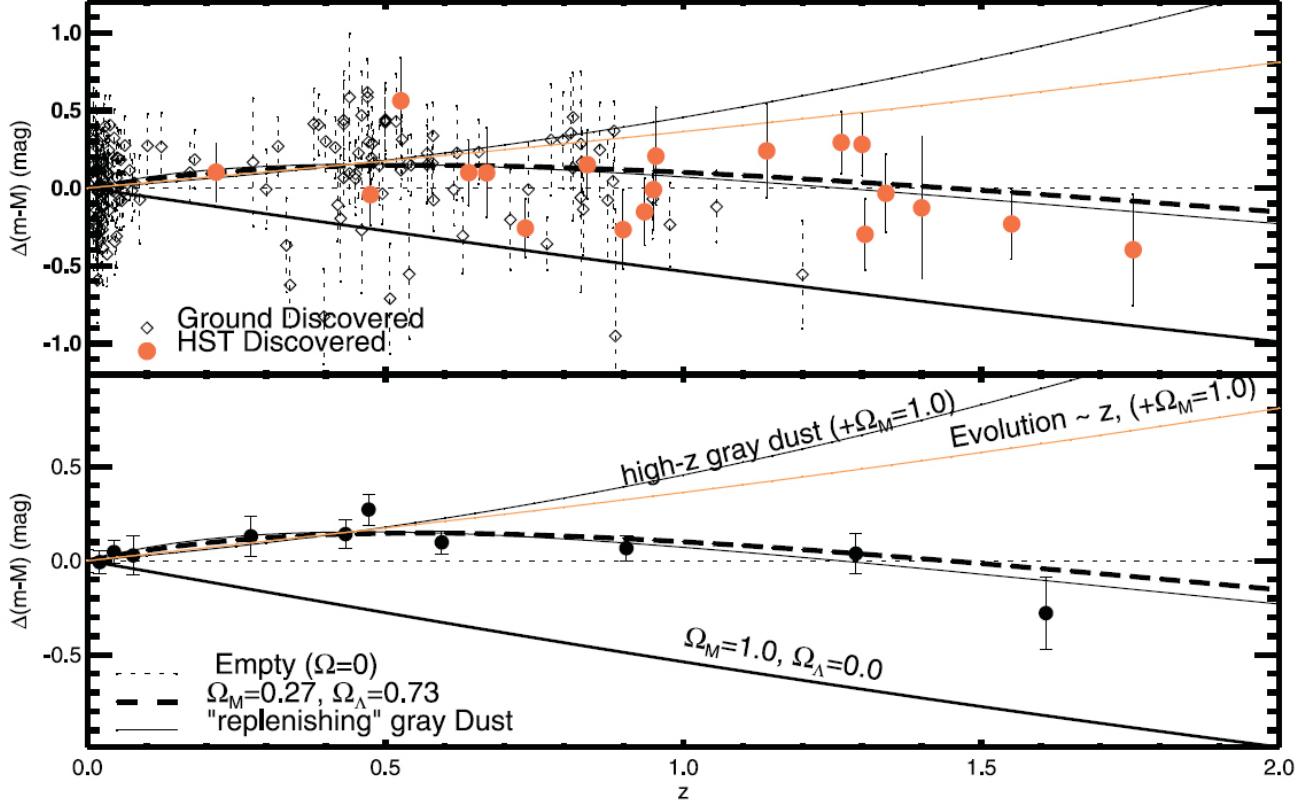


FIG. 13.— Above: the difference in luminosity distance modulus (positive means more distant, since $\Delta(m - M) = 5 \log(d_L/10 \text{ pc})$) between observed values from SNe Ia and theoretical values in an empty universe, as a function of redshift. Various cosmological scenarios are also depicted. Below: the same data, except averaged in redshift bins for a cleaner fit. A best fit $\Omega_m = 0.27$, $\Omega_\Lambda = 0.73$ is also drawn. Note that at lower z the observed deviation is positive, indicating a greater luminosity distance (equivalent to reduced flux) than expected for an empty universe. Making Ω_m non-zero only increases $E(z)$ which decreases the luminosity distance, so this increase must be accounted for by a non-zero Ω_Λ . At large z the curve approaches the same slope as the $\Omega_m = 1$ line, indicating matter dominance in the early universe. The “grey dust” lines are for a uniform extinction medium (dust that provides dimming without reddening) - the “replenishing grey dust” assumes a constant dust density even though the universe is expanding. The orange line assumes SNe Ia become dimmer by a percentage monotonically increasing with z (i.e. early universe SNe Ia are much dimmer than modern-day SNe Ia). From Riess et al. (2004), their Fig. 7.

There are a large number of systematic errors. They include:

1. Calibration: k-corrections require good knowledge of both the SED of a typical SN Ia as well as the filter system used.
2. UV spread: SNe Ia have a higher spread in the UV than in optical or infrared (in the IR SNe Ia actually have an uncalibrated σ of 0.15!), which becomes problematic when redshift pulls the U band into the B band.
3. Reddening: intrinsic reddening of the SNe themselves and reddening due to dust should be handled separately, but in practice they are hard to deconvolve. Intrinsically fainter SNe Ia are redder than normal Ias, an effect almost as important as the luminosity/falloff time relationship, but the same effect occurs with dust extinction.
4. Galactic evolution: fainter SNe Ia tend to be embedded in older stellar populations. This translates to a $\sim 12\%$ brightness increase for $z = 1$ SNe Ia due to increased star formation. While this is not a problem if the stretch factor corrects for light curves, it turns out that current fitting methods do not bring SNe Ia in high and low-mass galaxies to the same absolute magnitude (the difference is ~ 0.08 mag).
5. Since the progenitors of SNe Ia are unknown, it is not known if the physical nature of SNe Ia changes with redshift.

1.9.2. Describe alternate explanations to the SNe luminosity distance data, and why they can be ruled out?

Alternate scenarios for why the luminosity distance appears to increase for mid- z values include a 20% drop in SNe flux due to a “grey dust” (uniform opacity dust). This does not work, since at higher redshift the obscuration would be more significant, and this trend is not seen in the dataset. The idea that SNe Ia have fundamentally changed from the young universe to today is also difficult to support: the measured luminosity distance approaches what we would expect for a critical matter dominated universe at high redshift (a consequence of Λ becoming prominent only recently). It is not obvious how a change in the nature of SNe Ia could produce the same trend.

1.9.3. *Can SNe II be used as standard candles?*

Yes; in particular the recombination bump of SNe II-P can be used as a standardizable candle (Kasen & Woosley 2009). In these SNe, there is a tight relationship between luminosity and expansion velocity (as measured from ~ 5000 ÅFe II absorption lines), explained by the simple behavior of hydrogen recombination in the supernova envelope (Kasen & Woosley 2009). There is sensitivity to progenitor metallicity and mass that could lead to systematic errors, and overall SNe II are dimmer than SNe Ia, however (Kasen & Woosley 2009).

Q7) THE DISTANCE LADDER

Describe as many steps of the distance ladder and the involved techniques as you can.

Q7) Herman ExtraGal Q3

EG 3

- distances to objects
- Radar, parallax, spec. parallax, W-B effect, cepheids, TF / FJ relation, Type Ia SNe

• Distance Ladder

LMC - 50 kpc

M31 - 700 kpc ($1 M_{\odot}$)

SMC - 60 kpc

Virgo Cluster - ~10 Mpc

• Radar

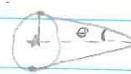
- not very far; Solar system

- send a signal and wait for it to bounce back at c

• Parallax

- 20 pc

- measure angle but need to resolve source



Spectroscopic Parallax

- ~7 Mpc b/w ~ 100 pc
- use spectrum to get T or sp. t., use TIR diagram to get M , then dist. \propto

Wilson Bappu Effect

- width of narrow em. line w/in Ca absorption line is strongly correlated w/ M_V
- for late-type stars

• Cepheids / Variable stars

- 30 Mpc

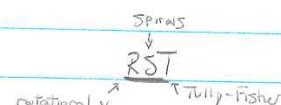
- relation b/w star's pulsation period and peak luminosity, use this plus apparent mag. & distance modulus to get distance $M \sim \log_{10} P$

$$m - M = -5 \log_{10} (d/10\text{pc})$$

• Tully-Fisher Relation

- > 100 Mpc

- relation b/w Max. rotational v and luminosity for spiral galaxies



• Faber-Jackson Relation

- > 100 Mpc

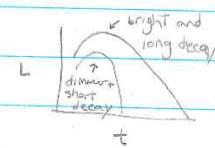
- relation b/w velocity dispersion and luminosity for elliptical galaxies



• Type Ia Supernovae

- > 1000 Mpc

- relation b/w decay time of light curve and peak luminosity
over 1st 2 mag: $M \sim \log_{10} m$



* Radar \rightarrow Parallax \rightarrow Spectroscopic Parallax \rightarrow Wilson Bappu effect \rightarrow cepheids \rightarrow 20pc 100 kpc 30 Mpc

Tully-Fisher \rightarrow Faber Jackson \rightarrow Type Ia SNe

100 Mpc

100 Mpc

1000 Mpc

Q7) Campbell ExtraGal Q3

1.3 Question 3

Describe as many steps of the distance ladder and the involved techniques as you can. What are the rough distances to the Magellanic Clouds, Andromeda, and the Virgo Cluster?

1.3.1 Short answer

The rough distances to the Magellanic Clouds, Andromeda, and the Virgo Cluster are as follows:

$$\begin{aligned}d_{\text{LMC}} &\approx 50 \text{ [kpc]} \\d_{\text{SMC}} &\approx 60 \text{ [kpc]} \\d_{\text{M31}} &\approx 700 \text{ [kpc]} \\d_{\text{Virgo}} &\approx 10 \text{ [Mpc].}\end{aligned}$$

1.3.2 Additional context

Trigonometric parallax ($< 30 \text{ pc}$): The most important method of distance determination is the trigonometric parallax, not only from a historical point-of-view. This method is based on a purely geometric effect and is therefore independent of any physical assumptions. Due to the motion of the Earth around the Sun the positions of nearby stars on the sphere change relative to those of very distant sources (e.g., extragalactic objects such as quasars). The latter therefore define a fixed reference frame on the sphere. In the course of a year the apparent position of a nearby star follows an ellipse on the sphere, the semi-major axis of which is called the parallax. The axis ratio of this ellipse depends on the direction of the star relative to the ecliptic (the plane that is defined by the orbits of the Earth and the other planets). The parallax depends on the radius r of the Earth's orbit, hence on the Earth-Sun distance which is, by definition, one astronomical unit (AU). Furthermore, the parallax depends on the distance d of the star,

$$\frac{r}{d} = \tan(p) \approx p \text{ [rad].}$$

where we used $p \ll 1$ in the last step, and p is measured in radians as usual. The trigonometric parallax is also used to define the common unit of distance in astronomy: one parsec (pc) is the distance of a hypothetical source for which the parallax is exactly $p = 1''$. With the conversion of arcseconds to radians ($1'' \approx 4.848 \times 10^{-6} \text{ rad}$) one gets $p/1'' = 206265p$, which for a parsec yields

$$1 \text{ pc} = 206265 \text{ AU} = 3.086 \times 10^{16} \text{ m.}$$

The distance corresponding to a measured parallax is then calculated as

$$d = \left(\frac{p}{1''} \right)^{-1} \text{ [pc].}$$

To determine the parallax p , precise measurements of the position of an object at different times are needed, spread over a year, allowing us to measure the ellipse drawn on the sphere by the object's apparent position. For ground-based observations the accuracy of this method is limited by the atmosphere. The seeing causes a blurring of the images of astronomical sources and thus limits the accuracy of position measurements. From the ground this method is therefore limited to parallaxes larger than $\approx 0.1''$, implying that the trigonometric parallax yields distances to stars only within $\approx 30 \text{ pc}$.

Proper motions: Stars are moving relative to us or, more precisely, relative to the Sun. To study the kinematics of the Milky Way we need to be able to measure the velocities of stars. The radial component v_r of the velocity is easily obtained from the Doppler shift of spectral lines,

$$v_r = \left(\frac{\Delta\lambda}{\lambda_0} \right) c \text{ [m s}^{-1}]$$

where λ_0 is the rest-frame wavelength of an atomic transition and $\Delta\lambda = \lambda - \lambda_0$ is the Doppler shift of the wavelength due to the radial velocity of the source. The sign of the radial velocity is defined such that $v_r > 0$ corresponds to a motion away from us, i.e., to a redshift of spectral lines.

In contrast, the determination of the other two velocity components is much more difficult. The tangential component, v_t , of the velocity can be obtained from the proper motion of an object. In addition to the motion caused by the parallax, stars also change their positions on the sphere as a function of time because of the transverse component of their velocity relative to the Sun. The proper motion μ is thus an angular velocity, e.g., measured in milliarcseconds per year (mas/yr or ''/year). This angular velocity is linked to the tangential velocity component via

Q7) Campbell ExtraGal Q3

$$v_t = d\mu = 4.74 \left(\frac{d}{1 \text{ pc}} \right) \left(\frac{\mu}{1'' \text{ year}^{-1}} \right) [\text{m s}^{-1}].$$

Therefore, one can calculate the tangential velocity from the proper motion and the distance. If the latter is derived from the trigonometric parallax, these can be combined to yield

$$v_t = d\mu = 4.74 \left(\frac{\mu}{1'' \text{ year}^{-1}} \right) \left(\frac{p}{1''} \right)^{-1} [\text{m s}^{-1}].$$

Of course, the proper motion has two components, corresponding to the absolute value of the angular velocity and its direction on the sphere. Together with v_r this determines the three-dimensional velocity vector. Correcting for the known velocity of the Earth around the Sun, one can then compute the velocity vector \mathbf{v} of the star relative to the Sun, called the heliocentric velocity.

Moving cluster parallax (< 200 pc): The stars in an (open) star cluster all have a very similar spatial velocity. This implies that their proper motion vectors should be similar. To what accuracy the proper motions are aligned depends on the angular extent of the star cluster on the sphere. Like two railway tracks that run parallel but do not appear parallel to us, the vectors of proper motions in a star cluster also do not appear parallel. They are directed towards a convergence point, as depicted in Figure 4.

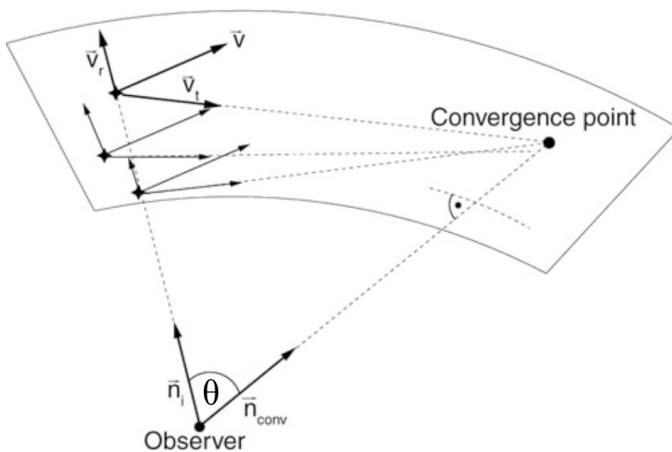


Figure 4: The moving cluster parallax is a projection effect, similar to that known from viewing railway tracks. The directions of velocity vectors pointing away from us seem to converge and intersect at the convergence point. The connecting line from the observer to the convergence point is parallel to the velocity vector of the star cluster. Image taken from Schneider (2006).

We consider a star cluster and assume that all stars have the same spatial velocity \mathbf{v} . The position of the i -th star as a function of time is then described by

$$\mathbf{r}_i(t) = \mathbf{r}_i(t_0) + \mathbf{v}t [\text{m}],$$

where \mathbf{r}_i is the current position if we identify the origin of time, $t = 0$, with ‘today’. The direction of a star relative to us is described by the unit vector

$$\mathbf{n}_i(t) \equiv \frac{\mathbf{r}_i}{|\mathbf{r}_i|} \text{ [dimensionless].}$$

From this, one infers that for large times, $t \rightarrow \infty$, the direction vectors of the convergence point are identical for all stars in the cluster,

$$\mathbf{n}_i(t) \rightarrow \frac{\mathbf{v}}{|\mathbf{v}|} \equiv \mathbf{n}_{\text{conv}} \text{ [dimensionless].}$$

Hence for large times all stars will appear at the same point \mathbf{n}_{conv} : the convergence point. This only depends on the direction of the velocity vector of the star cluster. In other words, the direction vector of the stars is such that they are all moving towards the convergence point. Thus, \mathbf{n}_{conv} (and hence $\mathbf{v}/|\mathbf{v}|$) can be measured from the direction of the proper motions of the stars in the cluster. On the other hand, one component of \mathbf{v} can be determined from the (easily measured) radial velocity v_r . With these two observables the three-dimensional velocity vector \mathbf{v} is completely determined, as is easily demonstrated: let θ be the angle between the line-of-sight \mathbf{n} towards a star in the cluster and \mathbf{v} . The angle is directly read off from the direction vector \mathbf{n} and the convergence point, $\cos \theta = \mathbf{n} \cdot \mathbf{v}/|\mathbf{v}| = \mathbf{n}_{\text{conv}} \cdot \mathbf{n}$. With $v \equiv |\mathbf{v}|$ one then obtains

$$v_r = v \cos \theta [\text{m s}^{-1}], \quad v_t = v \sin \theta [\text{m s}^{-1}],$$

and so

Q7) Campbell ExtraGal Q3

$$v_t = v_r \tan \theta \text{ [m s}^{-1}\text{].}$$

This means that the tangential velocity \mathbf{v}_t can be measured without determining the distance to the stars in the cluster. On the other hand, we have a relation between the proper motion, the distance, and \mathbf{v} . Hence, a distance determination for the star is now possible with

$$\mu = \frac{v_t}{d} = \frac{v_r \tan \theta}{d} \rightarrow d = \frac{v_r \tan \theta}{\mu} \text{ [pc].}$$

This method yields accurate distance estimates of star clusters within ~ 200 pc. The accuracy depends on the measurability of the proper motions. Furthermore, the cluster should cover a sufficiently large area on the sky for the convergence point to be well defined. For the distance estimate, one can then take the average over a large number of stars in the cluster if one assumes that the spatial extent of the cluster is much smaller than its distance to us.

Main sequence fitting: Most stars in the color-magnitude diagram (CMD) are located along the main sequence. This enables us to compile a calibrated main sequence of those stars whose trigonometric parallaxes are measured, thus with known distances. Utilizing photo-metric methods, it is then possible to derive the distance to a star cluster.

The stars of a star cluster define their own main sequence; since they are all located at the same distance, their main sequence is already defined in a CMD in which only apparent magnitudes are plotted. This cluster main sequence can then be fitted to a calibrated main sequence by a suitable choice of the distance, i.e., by adjusting the **distance modulus** $m - M$,

$$m - M = 5 \log(d/\text{pc}) - 5 \text{ [mag],}$$

where m and M denote the apparent and absolute magnitude, respectively.

In reality this method cannot be applied so easily since the position of a star on the main sequence does not only depend on its mass but also on its age and metallicity. Furthermore, only stars of luminosity class V (i.e., dwarf stars) define the main sequence, but without spectroscopic data it is not possible to determine the luminosity class.

Extinction and reddening: Another major problem is extinction. Absorption and scattering of light by dust affect the relation of absolute to apparent magnitude: for a given M , the apparent magnitude m becomes larger (fainter) in the case of absorption, making the source appear dimmer. Also, since extinction depends on wavelength, the spectral energy distribution of the source is modified and the observed colour of the star changes. Because extinction by dust is always associated with such a change in colour, one can estimate the absorption – provided one has sufficient information on the intrinsic colour of a source or of an ensemble of sources.

We consider the **equation of radiative transfer** for pure absorption or scattering:

$$\frac{dI_\nu}{ds} = -\alpha_\nu I_\nu \text{ [erg s}^{-1} \text{ cm}^{-3} \text{ ster}^{-1} \text{ Hz}^{-1}],$$

where I_ν is the specific intensity at frequency ν , α_ν is the absorption coefficient, and s is the distance along the light beam.

This says that the amount by which the intensity of a light beam is diminished on a path of length ds is ds . The absorption coefficient α_ν is thus defined as the constant of proportionality. In other words, on the distance interval ds , a fraction $\alpha_\nu ds$ of all photons at frequency ν is absorbed or scattered out of the beam. The solution of the transport equation is obtained by writing it in the form $d \ln I_\nu = dI_\nu/I_\nu = -\alpha_\nu ds$ and integrating from 0 to s ,

$$\ln I_\nu(s) - \ln I_\nu(0) = - \int_0^s \alpha_\nu(s') ds \equiv \tau_\nu(s) \text{ [erg s}^{-1} \text{ cm}^{-2} \text{ ster}^{-1} \text{ Hz}^{-1}],$$

where in the last step we defined the optical depth, τ_ν , which depends on frequency. This yields

$$I_\nu(s) = I_\nu(0) e^{-\tau_\nu(s)} \text{ [erg s}^{-1} \text{ cm}^{-2} \text{ ster}^{-1} \text{ Hz}^{-1}\text{].}$$

The specific intensity is thus reduced by a factor $e^{-\tau}$ compared to the case of no absorption taking place. Because of the relation between flux and magnitude $m = -2.5 \log S + \text{const}$ or $S \propto 10^{-0.4m}$, one has

$$\frac{I_\nu}{I_{\nu,0}} = 10^{-0.4(m-m_0)} = e^{-\tau_\nu} = 10^{-\log(e)\tau_\nu} \text{ [dimensionless],}$$

or,

Q7) Campbell ExtraGal Q3

$$\begin{aligned} A_\nu &\equiv m - m_0 = -2.5 \log(I_\nu/I_{\nu,0}) \text{ [mag]} \\ &= 2.5 \log(e)\tau_\nu \text{ [mag]} \\ &= 1.086\tau_\nu \text{ [mag].} \end{aligned}$$

Here, A_ν is the **extinction coefficient** describing the change of apparent magnitude m compared to that without absorption, m_0 . Since the absorption coefficient α_ν depends on frequency, absorption is always linked to a change in colour. This is described by the **colour excess** which is defined as follows:

$$E(X - Y) \equiv A_X - A_Y = (X - X_0) - (Y - Y_0) = (X - Y) - (X_0 - Y_0) \text{ [mag].}$$

The color excess describes the change of the color index $(X - Y)$, measured in two filters X and Y that define the corresponding spectral windows by their transmission curves. The ratio $A_X/A_Y = \tau_{\nu(X)}/\tau_{\nu(Y)}$ depends only on the optical properties of the dust or, more specifically, on the ratio of the absorption coefficients in the two frequency bands X and Y considered here. Thus, the color excess is proportional to the extinction coefficient,

$$E(X - Y) = A_X - A_Y = A_X \left(1 - \frac{A_Y}{A_X}\right) \equiv A_X R_X^{-1} \text{ [mag].}$$

where in the last step we introduced the factor of proportionality R_X between the extinction coefficient and the colour excess, which depends only on the properties of the dust and the choice of the filters. Usually, one considers a blue and a visual filter and writes

$$A_V = R_V E(B - V) \text{ [mag].}$$

For example, for dust in our Milky Way we have the characteristic relation

$$A_{V,\text{MWG}} = (3.1 \pm 0.1)E(B - V) \text{ [mag].}$$

This relation is not a universal law, but the factor of proportionality depends on the properties of the dust. They are determined, e.g., by the chemical composition and the size distribution of the dust grains. Figure 5 shows the wavelength dependence of the extinction coefficient for different kinds of dust, corresponding to different values of R_V . In the optical part of the spectrum we have approximately $\tau_\nu \propto \nu$, i.e., blue light is absorbed (or scattered) more strongly than red light. The extinction therefore always causes a reddening.

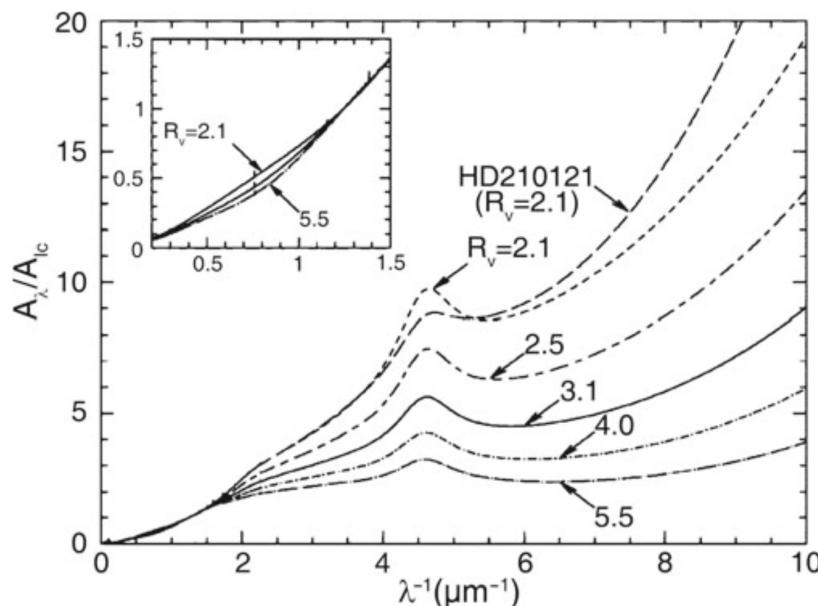


Figure 5: Wavelength dependence of the extinction coefficient A_ν , normalized to the extinction coefficient A_f at $\lambda = 9000 \text{ \AA} = 0.9 \mu\text{m}$. Different kinds of clouds, characterized by the value of R_V , i.e., by the reddening law, are shown. The solid curve specifies the mean Galactic extinction curve. The extinction coefficient, as determined from the observation of an individual star, is also shown. The figure insert shows a detailed plot at relatively large wavelengths in the NIR range of the spectrum; at these wavelengths the extinction depends only weakly on the value of R_V . Source: B. Draine 2003, Interstellar Dust Grains, ARA&A 41, 241. Image taken from Schneider (2006).

The extinction coefficient A_V is proportional to the optical depth towards a source and so is the colour excess. Since the extinction is due to dust along the line-of-sight, the colour excess is proportional to the column density of dust towards the source. If we assume that the dust-to-gas ratio in the interstellar medium does not vary greatly, we expect that the column density of neutral hydrogen N_H is proportional to the colour excess. The former can be measured from the Lyman- α absorption in the spectra of stars,

Q7) Campbell ExtraGal Q3

whereas the latter is obtained by comparing the observed colour of these stars with the colour expected for the type of star, given its spectrum (and thus, its spectral classification). One finds indeed that the color excess is proportional to the HI column density with

$$E(B-V) = 1.7 \left(\frac{N_{\text{H}}}{10^{22} \text{ atoms cm}^{-2}} \right) [\text{mag}],$$

and a scatter of about 30% around this relation. The fact that this scatter is so small indicates that the assumption of a constant dust-to-gas ratio is reasonable.

In the Solar neighborhood the extinction coefficient for sources in the disk is about

$$A_V \approx \frac{d}{1 \text{ kpc}} [\text{mag}],$$

but this relation is at best a rough approximation, since the absorption coefficient can show strong local deviations from this law, for instance in the direction of molecular clouds.

Colour-colour diagram: As a first step in this measurement, it is necessary to determine the degree of extinction, which can only be done by analyzing the reddening. The stars of the cluster are plotted in a color-color diagram, for example by plotting the colors $(U - B)$ versus $(B - V)$. A colour-colour diagram also shows a main sequence along which the majority of the stars are aligned. The wavelength-dependent extinction causes a reddening in both colors. This shifts the positions of the stars in the diagram. The direction of the reddening vector depends only on the properties of the dust and is here assumed to be known, whereas the amplitude of the shift depends on the extinction coefficient. In a similar way to the CMD, this amplitude can now be determined if one has access to a calibrated, unreddened main sequence for the colour-colour diagram which can be obtained from the examination of nearby stars. From the relative shift of the main sequence in the two diagrams one can then derive the reddening and thus the extinction. The essential point here is the fact that the colour-colour diagram is *independent of the distance*.

This then defines the procedure for the distance determination of a star cluster using photometry: in the first step we determine the reddening $E(B-V)$, and thus also A_V via $A_V = (3.1 \pm 0.1)E(B-V)$ for the Galactic medium, by shifting the main sequence in a colour-colour diagram along the reddening vector until it matches a calibrated main sequence. In the second step the distance modulus is determined by vertically (i.e., in the direction of M) shifting the main sequence in the CMD until it matches a calibrated main sequence. From this, the distance is finally obtained according to

$$m - M = 5 \log \left(\frac{d}{1 \text{ pc}} \right) - 5 + A [\text{mag}].$$

Spectroscopic distance: From the spectrum of a star, the spectral type as well as its luminosity class can be obtained. The former is determined from the strength of various absorption lines in the spectrum, while the latter is obtained from the width of the lines. From the line width the surface gravity of the star can be derived, and from that its radius (more precisely, M/R^2). The spectral type and the luminosity class specify the position of the star in the HRD unambiguously. By means of stellar evolution models, the absolute magnitude M_V can then be determined. Furthermore, the comparison of the observed colour with that expected from theory yields the color excess $E(B-V)$, and from that we obtain A_V . With this information we are then able to determine the distance using

$$m_V - A_V - M_V = 5 \log \left(\frac{d}{1 \text{ pc}} \right) - 5 [\text{mag}].$$

Visual binaries: Keplers third law for a two-body problem,

$$P = \sqrt{\frac{4\pi^2}{G(m_1 + m_2)} a^3} [\text{yr}]$$

relates the orbital period P of a binary star to the masses m_i of the two components and the semi-major axis a of the ellipse. The latter is defined by the separation vector between the two stars in the course of one period. This law can be used to determine the distance to a visual binary star. For such a system, the period P and the angular diameter 2θ of the orbit are direct observables. If one additionally knows the mass of the two stars, for instance from their spectral classification, a can be determined according to Kepler's third law, and from this the distance follows with the small angle approximation $d = a/\theta$.

Variable stars: Several types of pulsating stars show periodic changes in their brightnesses, where the period of a star is related to its mass, and thus to its luminosity. This period-luminosity (PL) relation is ideally suited for distance measurements: since the determination of the period is independent of distance, one can obtain the luminosity directly from the period if the calibrated PL-relation is known.

Q7) Campbell ExtraGal Q3

The distance is thus directly derived from the measured magnitude using the photometric distance, if the extinction can be determined from color measurements.

The existence of a relation between the luminosity and the pulsation period can be expected from simple physical considerations. Pulsations are essentially radial density waves inside a star that propagate with the speed of sound, c_s . Thus, one can expect that the period is comparable to the sound crossing time through the star, $P \sim R/c_s$. The speed of sound c_s in a gas is of the same order of magnitude as the thermal velocity of the gas particles, so that $k_B T \sim m_p c_s^2$, where m_p is the proton mass (and thus a characteristic mass of particles in the stellar plasma) and k_B is Boltzmann's constant. According to the virial theorem, one expects that the gravitational binding energy of the star is about twice the kinetic (i.e., thermal) energy, so that for a proton,

$$\frac{GMm_p}{R} \sim k_B T.$$

Combining these relations, we obtain for the pulsation period

$$P \sim \frac{R}{c_s} \sim \frac{R\sqrt{m_p}}{\sqrt{k_B T}} \sim \frac{R^{3/2}}{\sqrt{GM}} \propto \langle \rho \rangle^{-1/2},$$

where $\langle \rho \rangle$ is the mean density of the star. This is a remarkable result – the pulsation period depends only on the mean density. Furthermore, the stellar luminosity is related to its mass by approximately L/M^3 . If we now consider stars of equal effective temperature T_{eff} (where $L \propto R^2 T_{\text{eff}}^4$), we find that

$$P \propto \frac{R^{3/2}}{\sqrt{M}} \propto L^{7/12},$$

which is the relation between period and luminosity that we were aiming for.

One finds that a well-defined period-luminosity relation exists for three types of pulsating stars:

- **δ Cepheid stars** (classical Cepheids): These are young stars found in the disk population (close to the Galactic plane) and in young star clusters.
- **W Virginis stars**: Also called population II Cepheids. These are low-mass, metal-poor stars located in the halo of the Galaxy, in globular clusters, and near the Galactic center.
- **RR Lyrae stars**: These are likewise population II stars and thus metal-poor. They are found in the halo, in globular clusters, and in the Galactic bulge.

Globular clusters:

Tully-Fisher relation: Using 21 cm observations of spiral galaxies, in 1977 R. Brent Tully and J. Richard Fisher found that the maximum rotation velocity of spirals is closely related to their luminosity, following the relation

$$L_{\text{TF}} \propto v_{\max}^\alpha [\text{erg s}^{-1}],$$

where the power-law index (i.e., the slope) of the Tully-Fisher relation is about $\alpha \sim 4$. The larger the wavelength of the filter in which the luminosity is measured, the smaller the dispersion of the Tully-Fisher relation (see Figure 6). This is to be expected because radiation at larger wavelengths is less affected by dust absorption and by the current star formation rate, which may vary to some extent between individual spirals. Furthermore, it is found that the value of α increases with the wavelength of the filter: The Tully-Fisher relation is steeper in the red, which follows from the fact that more massive, or more luminous galaxies (i.e., those with larger v_{\max}) are redder. The dispersion of galaxies around this relation in the NIR (e.g., in the H-band) is about 10%.

Because of this close correlation, the luminosity of spirals can be estimated quite precisely by measuring the rotational velocity. The determination of the (maximum) rotational velocity is independent of the galaxy's distance. By comparing the luminosity, as determined from the Tully-Fisher relation, with the measured flux, one can then estimate the distance of the galaxy – without utilizing the Hubble relation! The measurement of v_{\max} is obtained either from a spatially resolved rotation curve, by measuring v_{rot} , which can be done with optical spectroscopy or, for relatively nearby galaxies, also with spatially resolved 21 cm spectroscopy. Alternatively, one can observe an integrated spectrum of the 21 cm line of HI that has a Doppler width corresponding to about $2v_{\max}$ (see Fig.3.28). The Tully-Fisher relation shown in Figure 6 was determined by measuring the width of the 21 cm line.

The shapes of the rotation curves of spirals are very similar to each other, in particular with regard to their flat behavior in the outer part. The flat rotation curve implies

$$M = \frac{v_{\max}^2 R}{G} [\text{M}_\odot],$$

Q7) Campbell ExtraGal Q3

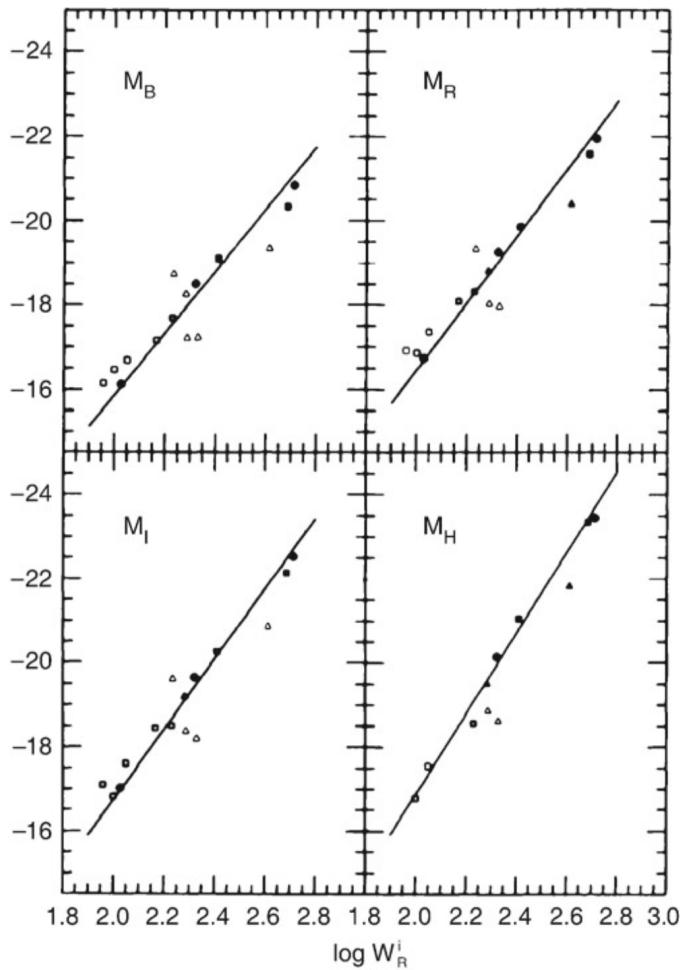


Figure 6: The Tully-Fisher relation for galaxies in the Local Group (dots), in the Sculptor group (triangles), and in the M81 group (squares). The absolute magnitude is plotted as a function of the width of the 21 cm profile which indicates the maximum rotation velocity. Filled symbols represent galaxies for which independent distance estimates were obtained, either from RR Lyrae stars, Cepheids, or planetary nebulae. For galaxies represented by open symbols, the average distance of the respective group is used. The solid line is a fit to similar data for the Ursa-Major cluster, together with data of those galaxies for which individual distance estimates are available (filled symbols). The larger dispersion around the mean relation for the Sculptor group galaxies is due to the groups extent along the line-of-sight. Source: M.J. Pierce & R.B. Tully 1992, Luminosity-line width relations and the extragalactic distance scale. I-Absolute calibration, ApJ 387, 47, p. 51, Fig. 1. Image taken from Schneider (2006).

where the value of the distance R from the center of the galaxy is chosen to be in the range of the flat part of the rotation curve (i.e., where $v_{\text{rot}}(R) \approx v_{\text{max}}$). We note that the exact value of R is not important; of course, $M = M(R)$. By re-writing this,

$$L = \left(\frac{M}{L}\right)^{-1} \frac{v_{\text{max}}^2 R}{G} [\text{erg s}^{-1}],$$

and by replacing R by the mean surface brightness $\langle I \rangle = L/R^2$, we obtain

$$L = \left(\frac{M}{L}\right)^{-2} \left(\frac{1}{G^2 \langle I \rangle}\right) v_{\text{max}}^4 [\text{erg s}^{-1}].$$

This is the Tully-Fisher relation if M/L and $\langle I \rangle$ are the same for all spirals. The latter is in fact suggested by Freemans law. Since the shapes of rotation curves for spirals seem to be very similar, the radial dependence of the ratio of luminous to dark matter may also be quite similar among spirals. Furthermore, since the mass-to-light ratios of a stellar population as measured from the red or infrared emission do not depend strongly on its age, independently of their Hubble type. Within R_{25} one finds $M/L_B = 6.2$ for Sa's, 4.5 for Sb's, and 2.6 for Sc's. This trend does not come as a surprise because late types of spirals contain more young, blue and luminous stars.

Faber-Jackson relation: A relation for elliptical galaxies, analogous to the Tully-Fisher relation, was found by Sandra Faber and Roger Jackson. They discovered that the velocity dispersion in the center of ellipticals, σ_v , scales with luminosity,

$$L_{\text{FJ}} \propto \sigma_v^4 [\text{erg s}^{-1}]$$

'Deriving' the Faber-Jackson scaling relation is possible under the same assumptions as for the Tully-Fisher relation. However, the dispersion of ellipticals about this relation is larger than that of spirals about the Tully-Fisher relation.

Q7) Campbell ExtraGal Q3

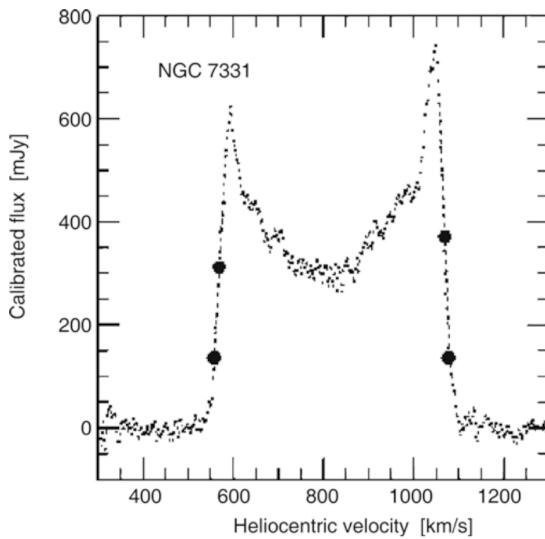


Figure 7: 21 cm profile of the galaxy NGC7331. The bold dots indicate 20 and 50% of the maximum flux; these are of relevance for the determination of the line width from which the rotational velocity is derived. Source: L.M. Macri et al. 2000, A Database of Tully-Fisher Calibrator Galaxies, ApJS 128, 461, p. 467, Fig. 5. Image taken from Schneider (2006).

$D_n - \sigma$ relation: Another scaling relation for ellipticals which is of substantial importance in practical applications is the $D_n - \sigma$ relation. D_n is defined as the mean diameter of an ellipse within which the average surface brightness. It corresponds to a value of $20.75 \text{ mag arcsec}^{-2}$ in the B-band. If we now assume that all ellipticals have a self-similar brightness profile, $I(R) = I_e f(R/R_e)$, with $f(1) = 1$, then the luminosity within D_n can be written as

$$\begin{aligned} I_n \left(\frac{D_n}{2} \right)^2 \pi &= 2\pi I_e \int_0^{D_n/2} R f(R/R_e) dR \\ &= 2\pi I_e R_e^2 \int_0^{D_n/(2R_e)} [\text{erg s}^{-1} \text{ m}^{-2}] x f(x) dx, \end{aligned}$$

where in the last step we changed the integration variable to $x = R/R_e$. For a de Vaucouleurs profile we have approximately $f(x) \propto x^{-1.2}$ in the relevant range of radius. Computing the integral with this expression, we obtain

$$D_n \propto R_e I_e^{0.8}.$$

Empirically, we find that ellipticals follow the normalized $D_n - \sigma$ relation

$$D_n = 2.05 \times \left(\frac{\sigma_v}{100 \text{ km s}^{-1}} \right) [\text{kpc}]$$

and they scatter around this relation with a relative width of about 15%.

Type Ia supernovae: The **Phillips relation** is the empirical correlation between the peak luminosity of a Type Ia supernova and the speed of luminosity evolution after maximum light; the faster the supernova fades from maximum light, the fainter its peak magnitude was (i.e., slower-evolving light curves are found to have brighter maximum luminosities).

Q7) Zhu ExtraGal Q4

49

2.3.2. What are the properties of these neutral H clouds? Can they form galaxies?

We deduce the size of the intergalactic clouds by comparing the Ly- α forest in the spectra of pairs of lensed quasars. Many of the absorption lines are seen in both spectra, but some are not. This indicates that the clouds are, on average, about the size of the lensing galaxy. From the total calculated column density of hydrogen (ionized plus neutral), the mass of atypical cloud is somewhere around $10^7 M_{\odot}$. At the temperate estimated for a typical cloud ($T \sim 3 \times 10^4$ K), its self-gravity would be too weak to keep it from dispersing. It may be held together by the pressure of less dense (but hotter) external IGM or by the presence of DM within the cloud.

The clouds are placed into three categories. The low column density Ly- α forest absorbers ($\Sigma_{HI} < 10^{16} \text{ cm}^{-2}$) are associated with the diffuse IGM. These systems probe low-density, highly ionized gas and are thought to trace the dark matter distribution throughout the IGM as well as contain the bulk of the baryons at high redshift and a significant amount of the baryons even today. At the other end, the high column density damped Ly- α absorbers (DLAs, $\Sigma_{HI} > 10^{20} \text{ cm}^{-2}$) appear associated with the main bodies of galaxies. These high-density, predominantly neutral systems serve as neutral gas reservoirs for high redshift star formation. The intermediate column density systems, known as Lyman Limit Systems, mark the transition from the optically thin Ly- α forest to the optically thick absorbers found in and around the extended regions of galaxies. Typically these absorbers are easy to identify in QSO spectra due to the characteristic attenuation of QSO flux by the Lyman limit at $\sim 912 \text{ \AA}$ in the rest frame. In addition, they are optically thick enough to be harbouring neutral hydrogen cores.

2.4. Question 4

QUESTION: Describe as many steps of the distance ladder and the involved techniques as you can. What are the rough distances to the Magellanic Clouds, Andromeda, and the Virgo Cluster?

This is mostly from my own notes.

The cosmic distance ladder consists of a large number of means to determine (luminosity, for the most part) distances to objects. They are listed below, and extragalactic ones in Fig. 32. The differing techniques are generally calibrated to one another (to provide consistency), and therefore it is highly advantageous for differing rungs of the ladders to overlap.

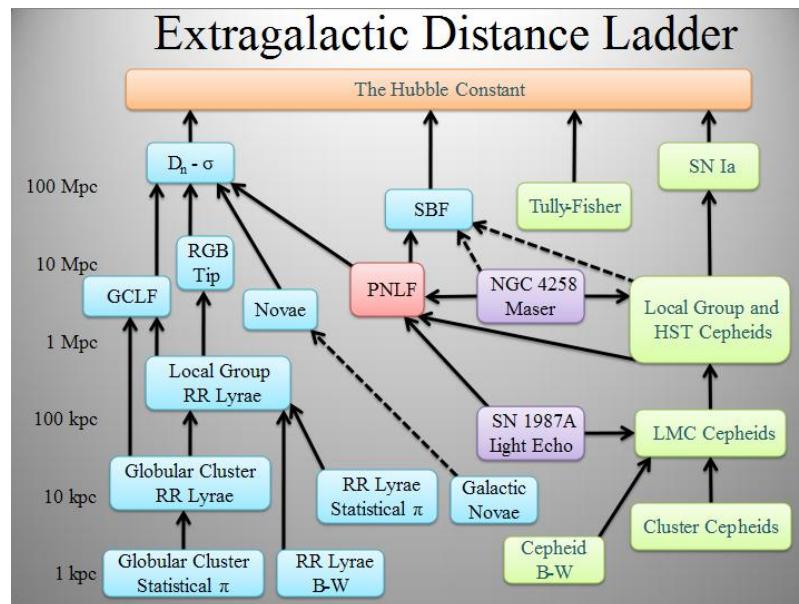


FIG. 32.— Plot of various extragalactic distance-determining techniques, with their maximum effective luminosity distance measures located to the left. The various colours represent applicability to different classes of object (but can effectively be ignored). Solid black lines indicate good calibration between two steps on the ladder; dashed lines indicate poor calibration. “GCLF” is the globular cluster luminosity function, “PNLF” is planetary nebula luminosity function, and “SBF” is surface brightness function. From Wikipedia (2011a).

Solely Galactic distance determining techniques:

- **Geometric parallax** is the measurement of the position of a distant objects from two different observation points separated by some physical distance called a “baseline”. The baseline divided by the angle by which the objects shifts when moving between the two points of observation gives a distance estimate. The technique is usually only applicable to nearby stars (the Gaia satellite will change this). Commonly the baseline used is the

diameter of the Earth's orbit around the Sun, but the proper motion of the Sun can be used to create longer baselines.

- **The moving cluster method** uses the proper motion of cluster members, as well as cluster member radial velocities, to determine distance to the cluster.
- **Dynamical parallax** uses measurements of the period (or a related orbital value) and angular semi-major axis of binary stars to determine their distance. By plugging in the period, angular semi-major axis and a mass-luminosity relation into Kepler's Third Law, one obtains an equation solvable for the distance to the binary.
- **Main sequence fitting*** of stars in a cluster onto an HR diagram, and comparison of that fit with a similar fit to the Hyades cluster, can be used to determine open cluster distances up to 7 kpc.
- **Spectroscopic parallax** is the determination of the luminosity of a star from a combination of spectral class and line-broadening (the 2-dimensional Morgan-Keenan luminosity class). Combined with the apparent magnitude, this can be used to determine a distance modulus to the star. Technically spectroscopic parallax is useful up to 10 Mpc, but in practice it is only used up to 10^5 pc.
- **Expansion velocities of SNe ejecta** proper motion measurements, combined with measurements of radial velocity from Doppler shifts and the assumption that expansion is spherically symmetric can be combined to obtain distances.

Extragalactic distance determining techniques (some can also be used in our own Galaxy):

- **The Wilson-Bappu effect** (up to 0.1 Mpc) is a spectroscopic parallax effect where the width of a small emission line housed in the K absorption line of calcium is correlated to absolute magnitude for G, K and M stars. Calibrated to geometric parallax, the Wilson-Bappu effect is mostly used for distances up to the LMC.
- **Bright red supergiants** (up to 7 Mpc) which appear to have the same absolute V and bolometric magnitude. Requires distinguishing individual stars, giving it the same range as spectroscopic parallax.
- **Tip of the RGB** (up to 7 Mpc) is a similar method that uses the tip of the red giant branch star luminosity function. Stars travelling up the RGB will eventually experience an He flash and transition off the RGB to the zero-age horizontal branch. An He flash occurs when the He core of an RGB reaches $\sim 0.5 M_\odot$, and the luminosity prior is dependent on properties of the H-burning shell, which in turn is dependent on the He core; this means that the most luminous a red giant can get is an almost constant value. A distance, then, can be estimated from the brightest RGB stars in a galaxy (since they will be very near the RGB tip).
- **Novae** (up to 20 Mpc) have, like their SNe Ia cousins, a luminosity vs. light curve decline time relationship that allows them to be calibrated into standard candles using $M_V^{max} = -9.96 - 2.31 \log_{10} \dot{m}$, where \dot{m} is the average rate of decline over the first 2 mag in magnitudes per day. This relationship comes from the fact that more massive WDs tend to accrete smaller amounts of matter before a nuclear runaway occurs (i.e. producing less bright an explosion), and this thinner layer is ejected more easily.
- **Variables** (up to 29 Mpc) have a correlation between pulsation and luminosity⁸, allowing determination of a distance modulus (assuming extinction is also known). Classical Cepheids have been seen up to 29 Mpc away, while RR Lyrae and W Virginis stars are generally fainter and therefore can only be used for closer distances.
- **Globular clusters*** (up to 50 Mpc) follow an empirical luminosity function. Because the function has a turnover, fitting a sampling of globular clusters around a distant galaxy with the luminosity function can be used to derive a relative distance.
- **Planetary nebulae*** (up to 50 Mpc) also appear to follow a luminosity function⁹ with a rapid cutoff at around -4.6 mag. Fitting, and comparing to a fit with a known distance, provides a distance measure; alternatively, finding the brightest PN in a galaxy and assuming they reside near the cutoff also gives a distance. As the cutoff method uses brighter PN, it can be to larger distances (the fitting method only goes up to 20 Mpc).
- **The surface brightness method*** (up to 50 Mpc) is based on the fact that the number of bright stars (which contribute the majority of the brightness) per area element in a galaxy fluctuates by Poisson noise. Since a galaxy further away will subtend a smaller area on the sky, there will be more bright stars per angular area, and therefore less fluctuation in the surface brightness (since the fluctuation goes like \sqrt{N}/N).

⁸ $M_V = -3.53 \log_{10}(P_d) - 2.13 + 2.13(B - V)$.

⁹ Here the measured luminosity is L_λ at $\lambda = 5007$ Å.

- **The Tully-Fisher relation*** (> 100 Mpc) is a relation between a spiral galaxy's luminosity and its maximum rotational velocity described in Sec. 2.1. The analogous relation for ellipticals is the Faber-Jackson relation, and is much noisier and more difficult to use as a standard ruler. Physically, the Tully-Fisher relation means that the mass-to-light ratio and mean surface brightness of spirals is fairly constant. In fact, due to the changing fraction of baryons in gas instead of stars for lower mass spirals, the mass-to-light ratio does change - adding a correction term to the Tully-Fisher relation results in a much tighter relationship.
- **The $D - \sigma$ relation*** (> 100 Mpc) is a relation between an elliptical galaxy's angular diameter D out to a constant surface brightness (20.75 B-mag per arcsec) and its velocity dispersion. Since surface brightness is independent of distance, D is inversely proportional to the distance to the elliptical galaxy. Physically, this relation is a natural outcome of the fundamental plane and the (fairly good) assumption that all ellipticals are self-similar.
- **Masers** (> 1000 Mpc) are amplified microwave emissions coming from regions of interstellar media where populations (due to very low densities) can become inverted. Cosmic masers exist around some galactic nuclear regions with line luminosities up to $10^4 L_\odot$; these maser sources orbit the supermassive black hole, and observations of both the radial and proper motions of these sources can be made. If we assume circular orbits, these two values can be combined to give a distance. Source radial velocities can also be combined with an independent measure of the black hole mass to determine distances.
- **Supernovae Ia** (> 1000 Mpc) have (or at least a sub-class of them have) a strong relation between light curve decay time and peak luminosity, the Phillips relation (Sec. 1.9). While most optical and infrared observations can be used, this relation has the smallest spread in the infrared. Other supernovae have similar relationships that could be used to determine luminosity (ex. SNe IIP plateau longevity is correlated with maximum luminosity), though these relationships are generally not as well defined and/or studied, and SNe Ia are brighter in the optical and IR and other SNe.
- **Brightest galaxies in clusters*** (> 1000 Mpc) fits a luminosity function to the galaxy cluster, and determines a distance modulus from the fit. It assumes that the luminosity function of galaxies remains fixed over time.
- **Cosmological redshift** (> 1000 Mpc) requires finding characteristic spectral features (lines, bumps, etc.). Redshift can be made to correspond to a physical distance through Hubble's Law ($v = H_0 d$) at low redshifts, and to d_c (assuming a cosmological model) at large redshift. Use of this measure must take into account peculiar velocities of objects.

All techniques denoted with a * are secondary distance indicators, which only give distance scaling, and therefore require a calibration galaxy with a distance known by other means (primary distance indicators can be used on their own because they give absolute luminosity distances).

The distance to the LMC can be determined from the Wilson-Bappu, RGB tip/supergiants, and variable star methods, giving a distance of ~ 48 kpc (the same answer can be obtained by "going backwards" by taking the NGC 4258 maser distance as a baseline for other distance measures). The most accurate distance for an object, however, was found using the light echo of SN1987A off of a nearly perfectly circular ring of material (ejected during the SN progenitor's AGB phase) to determine a distance. The ring is inclined to us, and so while light from the SN should impact all sections of the ring at the same time, due to the inclination and light delay we see different sections of the ring light up at different times. The light delay can be used to determine the physical size of the ring, while its angular size is easily observed, and from these two measurements we can obtain a distance of 51.8 ± 3 kpc to the SNR.

The distance to Andromeda has been measured with variable stars, surface brightness fluctuations, RGB tip/supergiants, the Wilson-Bappu effect and measurements of eclipsing binaries; the current accepted distance is ~ 770 kpc.

The distance to the Virgo Cluster can be determined by using a large number of these methods, including variable stars, novae, luminosity function methods, surface brightness fluctuations, the Tully-Fisher and $D - \sigma$ relations and Type Ia supernovae, applied to its component galaxies; all methods agree that it is ~ 17 Mpc away.

2.5. Question 5

QUESTION: What evidence is there that most galaxies contain nuclear black holes? How do those black holes interact with their host galaxies?

For most observational purposes, a black hole (BH) is defined as a mass-concentration whose radius is smaller than the Schwarzschild radius r_S of its corresponding mass. This value is about $r_S \sim 10^7$ km $\sim 15 R_\odot$ for the SMBH in the Galactic centre (GC). Since we cannot even resolve the supermassive black hole (SMBH) in our own galactic centre (which spans 10^{-5} ") in order to show that other galaxies contain SMBHs we must find indirect evidence of a massive compact object at the centres of galaxies.

Q8) OTHER METHODS FOR PARAMETER CONSTRAINTS

Describe a method, other than Type Ia supernovae and CMB foregrounds, by which the cosmological parameters can be determined by astronomical observations, and describe the current status of constraints from this method.

Q8) Ludwig Cosmo Q7

Question 7 - Observational Cosmology

Describe two methods, other than Type Ia supernovae, by which the cosmological parameters can be determined by astronomical observations.

Baryon Acoustic Oscillations - BAO - Can be used as a standard ruler.

Relevant Equations:

- $\Delta\theta = \frac{\Delta\chi}{d_A(z)}$
- $\Delta\theta \rightarrow$ subtended angle. $\Delta\chi \rightarrow$ length of ruler. $d_A(z) \rightarrow$ angular diameter distance.
- $d_A(z) \propto \int_0^z \frac{dz}{H(z)}$
- $c\Delta z = H(z)\Delta\chi \rightarrow$ the redshift interval can be measured from data thus determining the Hubble Parameter.

In the early universe, things are clumpy. Sound waves can propagate because things are also very dense. The initial sound wave mechanism is set off by high amounts of pressure in particular regions which compresses this hot photon-baryon fluid which then decompresses like a sound wave does.

The electrons follow the photons because they're charged. The protons follow the electrons because otherwise there would be this huge electrical field.

These ripples travel onward for about 150 Mpc until recombination where the expansion of the universe drops the temperature, the electrons and protons combine to form neutral hydrogen, which no longer has to follow along with the photons. This leaves a matter ripple/shell frozen in place.

We can predict where this should occur so SDSS went and looked to see if there was a bump in the correlation function between galaxies and how clustered they are. This correlation function just describes how far galaxies should be from each other if everything just got pushed homogeneously and the bump in it shows that there is actual structure to how the matter got pushed.

Since that BAO signal is confirmed, and we know its comoving distance and angular size, we can place constraints on cosmological parameters.

Angular Power Spectrum of CMB.

The CMB map is a superposition of BAO shells with the temperature anisotropies resulting from the level of compression and rarefaction of the baryon photon fluid. We can express the temperature fluctuations in terms of spherical harmonics and get the amplitude of the fluctuations at different scales. The first peak in the CMB power spectrum is from the BAO waves reaching maximum compression when the universe became transparent at recombination. Its position tells us about how curved the universe is. The position of the first peak would shift left or right depending on a positively or negatively curved universe, respectively. Comparing the heights of the first two peaks tells us about Ω_m and the heights of the second and third peaks tells us about dark matter since the dark matter pulls the baryon fluid in more during compression.

Other methods: Weak Gravitational Lensing, CMB foreground like SZ and ISW effects.

Q8) Herman Cosmo Q8

C8

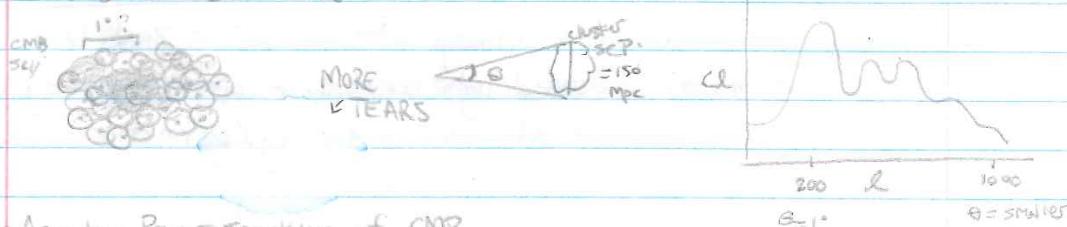
- BAOs result from density fluctuations and produce structure on 150 Mpc scale \rightarrow can be used as standard rulers
- Use BAOs to probe density of baryons, DM, Y_s, and use angular size to get flatness $\rightarrow f_{\Omega_0}$
- CL spectrum describes amplitude of T fluctuations on different scales, can be used to determine density of baryons, DM

Determining Cosmological Parameters Observationally

• BAOs

- initial quantum fluctuations in the early universe resulted in slight over- and under-densities ^{growing} as the gravitationally coupled DM & photon-baryon fluid fell inward. The resulting compression produced a bit of heating until radiation pressure pushed the baryon-photon fluid outward. This produced sound waves that oscillated due to the interaction b/w pressure and gravity. The characteristic distance these sound waves could travel before recombination, when the photons decoupled and the baryons were left in a shell around the DM over-densities, was 150 Mpc. We thus expect an increase in the amount of structure we see in the universe at this separation.

- this makes BAOs useful as a standard ruler. We can measure the average separation b/w galaxy clusters, which would've formed at the over-density locations, and use that to probe things like the baryon to photon to DM ratio. By looking at the angular size of the large scale structure we probe the flatness of the universe, which tells us about f_{Ω_0} .



• Angular Power Spectrum of CMB

- The CMB map is a superposition of BAO shells, w/ the temperature anisotropies resulting from the level of compression / expansion that the baryon-photon fluid underwent before decoupling. Since the oscillations arise from spherical harmonics, we can express the temperature fluctuations in terms of them, and determine their average amplitude at different scales (i.e. the power spectrum, $C_l \propto l^2$). The temperature ^{variance} will be highest at maximum compression at the time of decoupling, which explains the 1st peak. There will also be a large temperature anisotropy if the BAO reaches

Q8) Herman Cosmo Q8

Maximum expansion at decoupling, producing the 2nd peak. Because the DM pulls the baryon-photon fluid in more during compression, the 2nd peak is lower than the 1st. By comparing their relative heights we can learn about Ω_m , and by comparing 2 and 3 we can learn about DM. The location of the 1st peak (i.e. its angular scale) tells us about curvature. It would shift left/right if the universe was positively/negatively curved, respectively.

longer angular scales

A d

- Also weak gravitational lensing but fck that.
- Also CMB foregrounds like SZ and ISW effects

Hilary

Q8) Campbell Cosmo Q8

1.9 Question 8

Describe two methods, other than Type Ia supernovae, by which the cosmological parameters can be determined by astronomical observations.

1.9.1 Short answer

CMB Angular Power Spectrum: Comparisons of the relative strengths of the first few peaks can inform you about Ω_m and Ω_Λ since these are the components that caused the oscillations. In addition to this, the smallest multipole moments (i.e., the largest of scales) in the power spectrum can tell you about the modes that were never affected by the sound horizon and should tell you about inflation and the initial conditions that remain unperturbed. In contrast, the largest multipole moments (i.e., the smallest scales) in the power spectrum can tell you about Silk damping.

BAOs: The peaks in the CMB angular power spectrum are the result of compressions and expansions of the baryonic acoustic oscillations: the first and second peaks being the first compression and first decompression, respectively, etc. Given that the first compression happened on scales of the sound horizon at the time of last scattering, the first peak in the CMB angular power spectrum should tell you at what that scale was. This scale (i.e., the sound horizon) can be compared to what would be expected for a flat Universe – if it is smaller, the Universe is open whereas if it is larger, the Universe is closed (if they’re equal, the Universe is of course flat).

1.9.2 Additional context

CMB Angular Power Spectrum: The cosmic microwave background consists of photons that last interacted with matter at $z \sim 1, 100$. Since the Universe must already have been inhomogeneous at this time, in order for the structures present in the current Universe to be able to form, it is expected that these spatial inhomogeneities are visible as a (small) anisotropy of the CMB: *the angular distribution of the CMB temperature reflects the matter inhomogeneities at the redshift of decoupling of radiation and matter*. The CMB anisotropies reflect the conditions in the Universe at the epoch of recombination, thus at $z \sim 1, 100$. Temperature fluctuations originating at this time are called **primary anisotropies**. Later, as the CMB photons propagate through the Universe, they may experience a number of distortions along their way which, again, may change their temperature distribution on the sky. These effects then lead to **secondary anisotropies**.

Of the primary anisotropies (which directly pertain to inhomogeneities at the time of last scattering) the most basic mechanisms can be divided into those which occur on scales larger than the horizon size at recombination, i.e., which can not have been affected by physical interactions up to the time of last scattering, and those on smaller scales. The effects on superhorizon scales are the following:

- **Sachs-Wolfe effect:** Inhomogeneities in the gravitational potential cause photons which originate in regions of higher density to climb out of a potential well. As a result of this, they lose energy and are redshifted (**gravitational redshift**). This effect is partly compensated for by the fact that, besides the gravitational redshift, a gravitational time delay also occurs: a photon that originates in an overdense region will be scattered at a slightly earlier time, and thus at a slightly higher temperature of the Universe, compared to a photon from a region of average density. Both effects always occur side by side. This results in photons that are cooler in baryon overdensities. (This is the dominant mechanism of primary anisotropies at superhorizon scales.)
- **Peculiar velocities:** The electrons that scatter the CMB photons for the last time do not follow exactly the Hubble expansion, but have an additional velocity that is closely linked to the density fluctuations. This results in a Doppler effect: if photons are scattered by gas receding from us with a speed larger than that corresponding to the Hubble expansion, these photons experience an additional redshift which reduces the temperature measured in that direction.
- **Enhanced baryon density:** The distribution of baryons follows that of the dark matter, so that in regions of a higher dark matter density, the baryon density is also enhanced. This leads to an increased temperature of the baryons in overdense regions.

These three effects are relevant on scales larger than the (sound) horizon scale at the epoch of recombination. Obviously, they are closely coupled to each other. In particular, on scales $> r_{\text{H,com}}(z_{\text{rec}})$, the first two effects can partially compensate each other, though the Sachs-Wolfe effect is the dominant one at superhorizon scales. Inside the (sound) horizon, two other effects dominate the primary anisotropy signal:

Q8) Campbell Cosmo Q8

1. **Baryon acoustic oscillations:** On subhorizon scales, the pressure of the baryon-photon fluid is effective because, prior to recombination, these two components had been closely coupled by Compton scattering. This leads to sound waves in the baryon-photon fluid, called the baryonic acoustic oscillations (BAOs). In the density peaks of these sound waves, the baryon-photon fluid is adiabatically compressed and thus hotter than the average. The CMB sky yields a two-dimensional cut through this three-dimensional density (and temperature) field of these sound waves, and thus reflect these fluctuations, yielding temperature anisotropies with characteristic length (or angular) scales. (More on this later.)
2. **Silk damping:** The coupling of baryons and photons is not perfect since, owing to the finite mean free path of photons, the two components are decoupled on small spatial scales. This implies that on small length-scales, the temperature fluctuations can be smeared out by the diffusion of photons. This process is known as **Silk damping**, and it implies that on angular scales below about $\sim 50'$, only very small primary fluctuations exist.

BAOs: On subhorizon scales, the pressure of the baryon-photon fluid is effective because, prior to recombination, these two components had been closely coupled by Compton scattering. This leads to sound waves in the baryon-photon fluid, called the baryonic acoustic oscillations (BAOs). In the density peaks of these sound waves, the baryon-photon fluid is adiabatically compressed and thus hotter than the average. The CMB sky yields a two-dimensional cut through this three-dimensional density (and temperature) field of these sound waves, and thus reflect these fluctuations, yielding temperature anisotropies with characteristic length (or angular) scales. The length-scale of the resulting density perturbation is given by the sound horizon at recombination; it depends only on the baryon-to-photon and the matter-to-radiation density ratios. The former is proportional to $\Omega_b h^2$ and is determined, e.g., from BBN; the latter is proportional to $\Omega_m h^2$ which is well determined from measurements of the CMB. Together, one finds that the acoustic scale has a comoving value of $r_s \approx 150$ Mpc.

We expect that these frozen sound waves of the baryons leave an imprint on the overall matter correlation function, best seen in Figure 7. This unique feature in the correlation function, if it can be detected in the galaxy correlation function, would provide a well-defined ‘standard rod’ in the observable Universe. Since we can observe only angular scales on the sky, the relation between the (comoving) length of the standard rod and the associated angular scale provides a measure of distance. Therefore, a measurement of the baryonic acoustic oscillations (BAOs) in the correlation of galaxies at a given redshift z can be used to determine the (comoving) angular diameter distance $d_A(z)$ – which depends on the density parameters Ω_m and Ω_Λ .

In fact, the three-dimensional correlation function does not only depend on the transverse length scale which is related to the angular scale via the angular-diameter distance, but also on the separation of galaxies along the line-of-sight. The comoving distance interval corresponding to a redshift interval Δz is given by $\Delta x = c\Delta z/H(z)$. Since there are two transversal dimensions, and one along the line-of-sight, the distance measure that is determined best from BAOs is the geometric mean

$$D(z) = \left(d_A^2(z) \frac{cz}{H(z)} \right)^{1/3}.$$

The large redshift surveys 2dFGRS and SDSS allowed the first detection of these BAOs in the galaxy distribution in 2005. Figure 18 shows the discovery of BAOs from the SDSS, where a clear feature in the galaxy correlation function is seen at the expected length scale. The mean redshift of the galaxies from which the correlation function was determined is $z \sim 0.35$; thus, this measurement yields an estimate of the angular diameter distance to that redshift, with about a 5% accuracy. In particular, we point out that the sound horizon at recombination is visible in the current Universe!

The power of the method depends on whether the galaxies trace the underlying matter distribution sufficiently well, so that the measured galaxy correlation function reflects the correlation function of matter. Given the large spatial scale on which BAOs are observed, the proportionality between the galaxy and matter fluctuation fields, assumed by the simple bias model, is expected to hold very well. This then turns BAOs into a straightforward, almost purely geometrical tool for measuring the geometry of our Universe.

For this reason, several surveys are underway to measure the acoustic scale as a function of redshift. Figure 19 shows recent measurements of BAOs over a range of redshifts. Amazingly, the measurements are in perfect agreement with the cosmological parameters as determined by CMB anisotropy measurements. In particular, the spatial flatness of our Universe is confirmed, and any curvature is constrained to be very small, $|\Omega_m + \Omega_\Lambda| \lesssim 0.01$.

Q8) Campbell Cosmo Q8

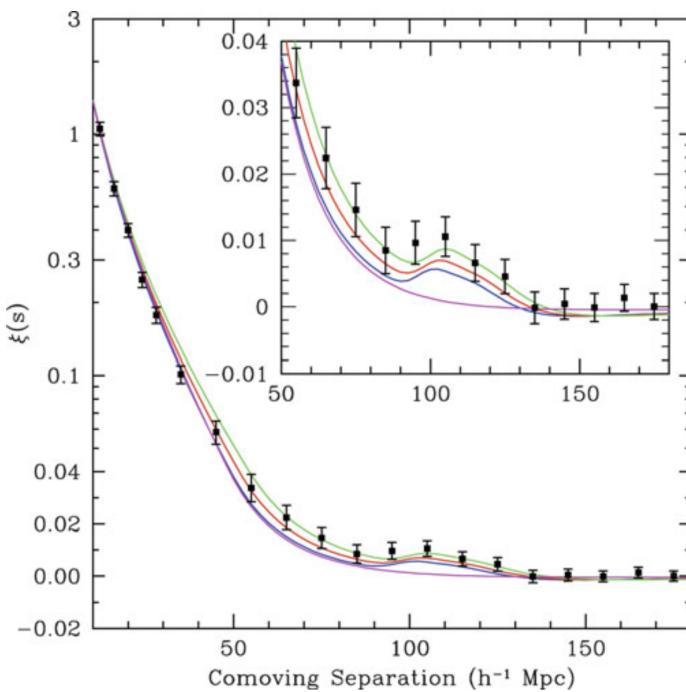


Figure 18: The correlation function of galaxies, as observed in the SDSS, shows a clear indication of a secondary peak on a comoving scale of about $100h^{-1}\text{Mpc} \sim 150\text{Mpc}$. Curves show models with slightly different density parameter $\Omega_m h^2 = 0.12; 0.13; 0.14$, with fixed baryon density of $\Omega_b h^2 = 0.024$. The lowest, smooth curve is a CDM model without baryons which thus shows no features due to baryonic oscillations. Source: D. Eisenstein et al. 2005, Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies, ApJ 633, 560, p. 563, Fig. 2. Figure taken from Schneider (2006).

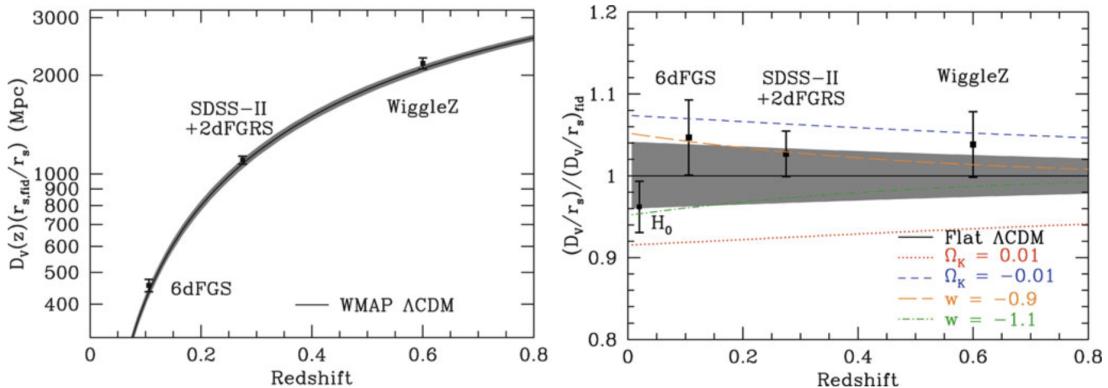


Figure 19: Recent results on the distance-redshift relation from measurements of BAOs. Left panel: The black line and the grey band shows the best-fit cosmological model obtained from the WMAP CMB anisotropy measurements and its 1σ uncertainty range. The BAO distance determination from three surveys are indicated, which are located right on this best-fit model. Right panel: The same data are shown, now divided by the prediction of a flat ΛCDM model with parameters as determined from WMAP. Also shown are curves of this distance ratio for models with a small positive or negative curvature, or for different equations-of-state of dark energy. In particular, the BAO measurements exclude any appreciable curvature parameter of our Universe. Source: D.H. Weinberg et al. 2012, Observational Probes of Cosmic Acceleration, arXiv:1201.2434, Fig. 8. Figure taken from Schneider (2006).

1.9.3 Follow-up Questions

- How does the CMB angular power spectrum change with cosmological parameters?
- How does the CMB angular power spectrum support inflation?
- Do we see super-horizon modes in the CMB angular power spectrum?
- How do we measure the CMB angular power spectrum?
- What about CMB polarization?

Q9) CMB POLARIZATION

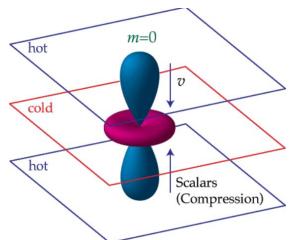
Why is the cosmic microwave background expected to be weakly polarized, and what is practically required to observe this signal?

Q9) Ludwig Cosmo Q8

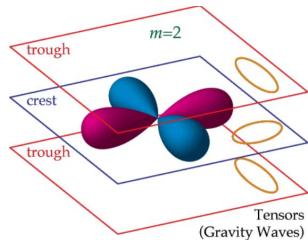
Question 8 - CMB Polarization

Why is the cosmic microwave background expected to be weakly polarized, and what is practically required to observe this signal?

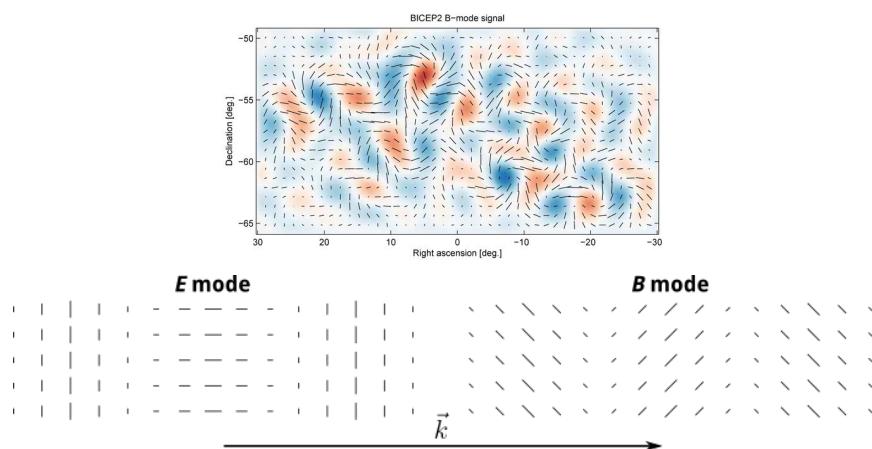
- The cosmic background radiation is blackbody radiation and should therefore be unpolarized. However, measurements have confirmed a finite polarization exists.
 - Since BB's absorb photons at all polarizations it should emit at all polarizations. No preferred direction means unpolarized.
- Temperature/density fluctuations produce a local quadrupole moment. This results in radiation coming toward an electron along a preferred axis.
- Thomson scattering produces linearly polarized light when the electron interacts with this radiation.
- The quadrupole can come about in two ways.
 - An electron falls into a gravitational potential well.



- As primordial gravitational waves stretch and squeeze space they affect the density of the photon baryon fluid. This is cool because inflation predicts gravitational waves and finding a polarization in the cmb supports that.



- Measuring polarization requires a detector with polarized sensors to measure the E and B field components of the CMB. Bicep2 detected CMB B modes (but it turned out to only be dust -- this is why understanding foregrounds is super important)



Q9) Herman Cosmo Q9

C9

CMB Polarization

- Density perturbations \rightarrow development of quadrupole moment \rightarrow higher intensity on one axis \rightarrow linear polarization
- Can be due to infalling matter or grav. waves
- Need instrument w/ polarized antennae to measure E field direction of photons.

The density (and thus T) fluctuations produce a local quadrupole moment that results in more radiation coming toward an electron along one axis than another (higher intensity). Thomson scattering produces linearly polarized light when the electron interacts with this radiation. The quadrupole can come about 2 ways:

- as an electron falls into a grav. potential well (a hot spot), the neighboring fluid looks like this:

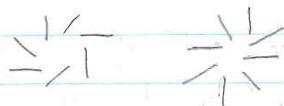


So a quadrupole moment develops. You can also think in terms of a T quadrupole: cold $\begin{matrix} \text{Hot} \\ \text{Hot} \end{matrix}$ cold

This results in scalar E modes, which produces a polarization pattern directed radially away from cold regions and tangent to hot regions.



- as primordial gravitational waves stretch and squeeze space, they affect the density of the photon-baryon fluid as well. These tensor B-modes produce a polarization pattern with a 'handedness' that spirals out from the temperature fluctuations



- Measuring polarization requires a detector with polarized sensors to measure the electric-field components of the CMB. BICEP2 had 2 antennae (?) polarized in different directions, then thought they were seeing CMB B-modes but were actually seeing polarization due to galactic dust.

Q9) Campbell Cosmo Q9

1.10 Question 9

Why is the cosmic microwave background expected to be weakly polarized, and what is practically required to observe this signal?

1.10.1 Short answer

The CMB is linearly polarized at the 10% level due to **Thomson scattering** of photons off of free electrons at the surface of last scattering. The degree of linear polarization is directly related to the quadrupole anisotropy in the photons when they last scatter. If the incoming radiation field were isotropic, orthogonal polarization states from incident directions would balance so that the outgoing radiation would remain unpolarized. Conversely, if the incident radiation field possesses a quadrupolar variation in intensity or temperature (which possess intensity peaks at $90^\circ = \pi/2$ separations), the result is a linear polarization of the scattered radiation. No other multipoles other than the quadrupole contributes to the CMB polarization.

1.10.2 Additional context

Why should we be concerned with the polarization of the cosmic microwave background (CMB) anisotropies? That the CMB anisotropies are polarized is a fundamental prediction of the gravitational instability paradigm. Under this paradigm, small fluctuations in the early Universe grow into the large scale structure we see today. If the temperature anisotropies we observe are indeed the result of primordial fluctuations, their presence at last scattering would polarize the CMB anisotropies themselves. The verification of the (partial) polarization of the CMB on small scales would thus represent a fundamental check on our basic assumptions about the behavior of fluctuations in the Universe, in much the same way that the redshift dependence of the CMB temperature is a test of our assumptions about the background cosmology. Furthermore, observations of polarization provide an important tool for reconstructing the model of the fluctuations from the observed power spectrum (as distinct from fitting an a priori model prediction from observations). *The polarization probes the epoch of last scattering directly as opposed to the temperature fluctuations which may evolve between last scattering and the present.* This localization in time is a very powerful constraint for reconstructing the sources of anisotropy. Moreover, different sources of temperature anisotropies (scalar, vector, and tensor) give different patterns in the polarization: both in its intrinsic structure and in its correlation with the temperature fluctuations themselves. Thus by including polarization information, one can distinguish the ingredients which go to make up the temperature power spectrum and so the cosmological model. Finally, the polarization power spectrum provides information complementary to the temperature power spectrum even for ordinary (scalar or density) perturbations. This can be of use in breaking parameter degeneracies and thus constraining cosmological parameters more accurately. The prime example of this is the degeneracy, within the limitations of cosmic variance, between a change in the normalization and an epoch of “late” reionization.

Yet how polarized are the fluctuations? The degree of linear polarization is directly related to the quadrupole anisotropy in the photons when they last scatter. While the exact properties of the polarization depend on the mechanism for producing the anisotropy, several general properties arise. The polarization peaks at angular scales smaller than the horizon at last scattering due to causality. Furthermore, the polarized fraction of the temperature anisotropy is small since only those photons that last scattered in an optically thin region could have possessed a quadrupole anisotropy. The fraction depends on the duration of last scattering. For the standard thermal history, it is 10% on a characteristic scale of tens of arcminutes. Since temperature anisotropies are at the 10^{-5} level, the polarized signal is at (or below) the 10^{-6} level, or several μK representing a significant experimental challenge.

The Thomson scattering cross section σ_T depends on polarization as (see, e.g., Chandrasekhar, 1960) as

$$\frac{d\sigma_T}{d\Omega} \propto |\epsilon \cdot \epsilon'|^2,$$

where ϵ and ϵ' are the incident and scattered polarization directions, and Ω is the solid angle. Heuristically, the incident light sets up oscillations of the target electron in the direction of the electric field vector E (i.e., the polarization). The scattered radiation intensity thus peaks in the direction normal to, with polarization parallel to, the incident polarization. More formally, the polarization dependence of the cross section is dictated by electromagnetic gauge invariance and thus follows from very basic principles of fundamental physics.

If the incoming radiation field were isotropic, orthogonal polarization states from incident directions separated by 90° would balance so that the outgoing radiation would remain unpolarized. Conversely, if the incident radiation field possesses a quadrupolar variation in intensity or temperature (which possess intensity peaks at $90^\circ = \pi/2$ separations), the result is a linear polarization of the scattered radiation

Q9) Campbell Cosmo Q9

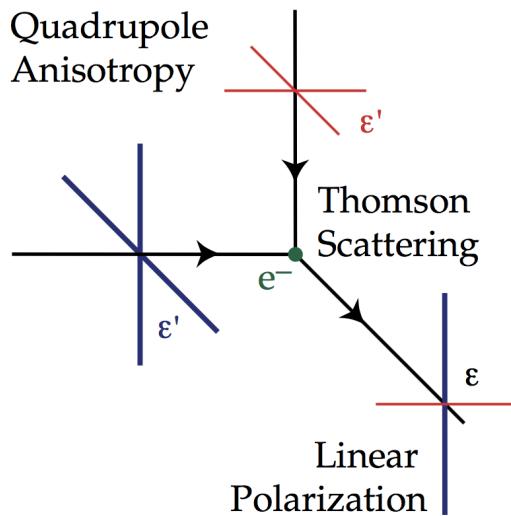


Figure 20:

Thomson scattering of radiation with a quadrupole anisotropy generates linear polarization. Blue colors (thick lines) represent hot and red colors (thin lines) cold radiation. Figure taken from Hu & White (1997).

(see Figure 20). A reversal in sign of the temperature fluctuation corresponds to a 90° rotation of the polarization, which reflects the spin-2 nature of polarization.

In terms of a multipole decomposition of the radiation field into spherical harmonics, $Y_\ell^m(\theta, \phi)$, the five quadrupole moments are represented by $\ell = 2, m = 0, \pm 1, \pm 2$. The orthogonality of the spherical harmonics guarantees that no other moment can generate polarization from Thomson scattering. In these spherical coordinates, with the north pole at $\theta = 0$, we call a N-S (E-W) polarization component $Q > 0$ ($Q < 0$) and a NE-SW (NW-SE) component $U > 0$ ($U < 0$). The polarization amplitude and angle clockwise from north are

$$P = \sqrt{Q^2 + U^2}$$

and

$$\chi = \frac{1}{2} \arctan \left(\frac{U}{Q} \right).$$

where P is the polarized intensity and Q and U are the linear polarization Stokes vectors.

If Thomson scattering is rapid, then the randomization of photon directions that results destroys any quadrupole anisotropy and polarization. The problem of understanding the polarization pattern of the CMB thus reduces to understanding the quadrupolar temperature fluctuations at last scattering. Temperature perturbations have 3 geometrically distinct sources: the scalar (compressional), vector (vertical) and tensor (gravitational wave) perturbations. Formally, they form the irreducible basis of the symmetric metric tensor. We shall consider each of these in Question 17 and show that the scalar, vector, and tensor quadrupole anisotropy correspond to $m = 0, \pm 1, \pm 2$ respectively. This leads to different patterns of polarization for the three sources.

1.10.3 Follow-up Questions

- What are some foregrounds or challenges to observing this signal?
- Why does dust polarize the CMB?
- Why is dust not randomly oriented?
- How did we first determine that dust is not randomly oriented?

Q10) CMB SECONDARY ANISOTROPIES

Our view of the cosmic microwave background is affected by what is along the line of sight. Give two examples of CMB secondary anisotropies that also provide information about the cosmic parameters.

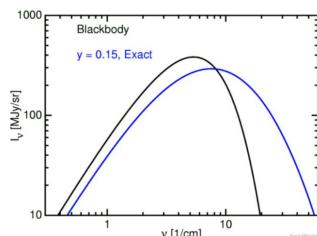
TSZ gives tau, amplifies Omega M
ISW give H0 and Omega Lambda

Q10) Ludwig Cosmo Q9

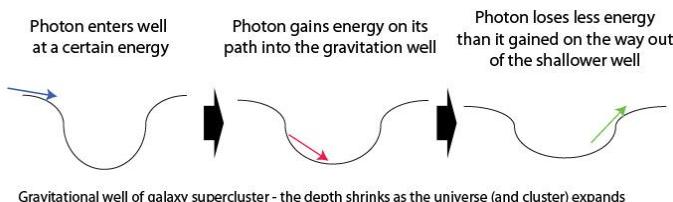
Question 9 - CMB Foregrounds

Our view of the cosmic microwave background is affected by what is along the line of sight. Give two examples of CMB foregrounds that also provide information about the cosmic parameters.

- It's a long journey from the CMB to us and a lot of opportunity for things to get in the way of CMB photons. Most of it is galactic stuff like like synchrotron radiation and dust but some can tell us about cosmological parameters.
- Thermal Sunyaev Zel'dovich \Rightarrow **ICM scatters photons, shifts BB by optical depth * thermal energy/ rest mass energy. Can measure distance and Ω_m**
 - Involves scattering of CMB photons by rapidly moving electrons in the hot gas in clusters of galaxies.
 - It is possible to use a combination of the SZ effect and the X-ray emission from the hot gas to derive a distance to the cluster.
 - This effect is proportional to (1) the number density of electrons, (2) the thickness of the cluster along our line of sight, and (3) the electron temperature. The parameter that combines these factors is called the Kompaneets y parameter, with $y = \tau(kT/mc^2)$. Tau (τ) is the optical depth or the fraction of photons scattered, while (kT/mc^2) is the electron temperature in units of the rest mass of the electron.
 - The SZ effect also amplifies the first peak in the power spectrum which can tell us about Ω_m
 - The usual order of magnitude for y is about 0.0001, which is very small.



- Integrated Sachs-Wolfe Effect \Rightarrow **a particle falls into well and expansion makes it easier to get out. Particle gets to keep some energy. Tells us about expansion**
 - Light travelling through a supercluster picks up energy in the form of speed and heat like a particle rolling down a valley.
 - Normally it would give all that energy back when rolling back up the valley, but dark energy changes the shape of the valley as the particles rolls through it to make it shallower and the particle gets to keep some of that heat.
 - It's a gravitational redshift that occurs between the surface of last scattering and the earth but happens when the universe is still dominated by radiation. If the universe were matter dominated than large scale gravitational potential energy wells and hills don't evolve significantly.



Q10) Herman Cosmo Q10

C10

CMB Foregrounds

- TSZ effect increases CMB T on large scales due to hot ICM, w/ typical separation of 150 Mpc. Will increase amplitude of 1st peak and inform you about Ω_m
- ISW effect increases CMB T on large scales due to gravitational red shift, expanding universe. Tells you about Ω_m

• A lot of stuff messes w/ the CMB photons on their journey to us. A lot of those things are just various forms of galactic stuff, like synchrotron radiation and dust, but some can tell us about cosmological parameters.

• Thermal Sunyaev-Zeldovich effect

- As CMB photons travel through the hot ICM of a galaxy cluster, they undergo inverse Compton scattering and gain energy from the relativistic electrons they encounter. This shifts the Planck spectrum of the CMB so there are fewer photons at low frequencies and more at higher ν . This will lead to regions of slightly higher temperature wherever the cluster is located in the CMB sky. Since clusters are generally 150 Mpc apart, the SZ effect will increase the amplitude of the 1st peak of the C_T spectrum, informing us about Ω_m .

• Integrated Sachs-Wolfe effect

- As CMB photons travel toward us they can encounter regions over over- (and under)-densities in the universe, and will fall into their grav. potential wells. When they fall in they gain a little bit of energy, but as they try to climb out, the universe's expansion reduces the strength of gravitational well, so it requires less energy to escape. The photons will thus gain energy on their passage through overdense regions, appearing as hotter spots in the CMB (similar to SZ). Underdense regions will result in the opposite effect. The ISW effect thus tells you about large scale structure (i.e. what causes the density region [OR VICE VERSA B/C FML]) as well as the expansion of the universe.

• Also polarization due to epoch of reionization, which I guess informs you about the optical depth at EoR which is a cosm. parameter.

Hilroy

Q10) Campbell Cosmo Q10

1.11 Question 10

Our view of the cosmic microwave background is affected by what is along the line of sight. Give two examples of CMB foregrounds that also provide information about the cosmic parameters.

1.11.1 Short answer

- **Thermal Sunyaev-Zeldovich effect:** Electrons in the hot gas of the intracluster medium can inverse Compton scatter photons of the CMB. Since galaxy clusters are separated by $\sim 150 \text{ Mpc}$, this will increase the power of the first peak in the CMB power spectrum and can inform us about Ω_m .
- **Integrated Sachs-Wolfe effect:** Inhomogeneities in the gravitational potential cause photons which originate in regions of higher density to climb out of a potential well. As a result of this, they lose energy and are gravitationally redshifted. This is a primary anisotropy effect which dominates on superhorizon scales.

1.11.2 Additional context

The CMB consists of photons that last interacted with matter at $z \sim 1100$. Since the Universe must already have been inhomogeneous at this time, in order for the structures present in the current Universe to be able to form, it is expected that these spatial inhomogeneities are visible as a (small) anisotropy of the CMB: the angular distribution of the CMB temperature reflects the matter inhomogeneities at the redshift of decoupling of radiation and matter.

Since the discovery of the CMB in 1965, such anisotropies have been searched for. Under the assumption that the matter in the Universe solely consists of baryons, the expectation was that we would find relative fluctuations in the CMB temperature of amplitude $\Delta T/T 10^{-3}$ on scales of a few arcminutes. This expectation is based on the theory of gravitational instability for structure growth: to account for the density fluctuations observed today where the density contrast $\delta \sim 1$ on scales of $\sim 10 h^{-1} \text{ Mpc}$, one needs relative density fluctuations at $z \sim 1000$ of order $D_+(z=1000) \sim 10^{-3}$. Despite increasingly more sensitive observations, such fluctuations were not detected. The upper limits resulting from these searches for anisotropies provided one of the arguments that, in the mid-1980s, caused the idea of the existence of dark matter on cosmic scales to increasingly enter the minds of cosmologists. In a Universe which is dominated by dark matter the expected CMB fluctuations on small angular scales are considerably smaller than in a purely baryonic Universe. It was the COBE satellite with which temperature fluctuations in the CMB were finally observed in 1992 (Figure 21), a discovery which was awarded the Physics Nobel Prize in 2006. Over the following years, sensitive and significant measurements of the CMB anisotropy were carried out using balloons and ground-based telescopes. Two more satellites have observed the full microwave sky, WMAP and Planck; their results, together with ground-based measurements at smaller angular scales, have yielded the most stringent constraints on cosmological parameters yet.

The CMB anisotropies depend on nearly all cosmological parameters, such as Ω_m , Ω_b , Ω_Λ , Ω_{HDM} , H_0 , the normalization σ_8 , the primordial slope n_s , and the shape parameter Γ of the power spectrum. Therefore, from an accurate mapping of the angular distribution of the CMB and by comparison with theoretical expectations, all these parameters can, in principle, be determined.

Temperature fluctuations originating at the time of last scattering are called primary anisotropies. Later, as the CMB photons propagate through the Universe, they may experience a number of distortions along their way which, again, may change their temperature distribution on the sky. These effects then lead to secondary anisotropies.

Given that the amplitude of the polarization is so small the question of foregrounds is even more important than for the temperature anisotropy. Unfortunately, the level and structure of the various foreground polarization in the CMB frequency bands is currently not well known. Atmospheric emission is believed to be negligibly polarized, leaving the main astrophysical foregrounds: free-free, synchrotron, dust, and point source emissions. Of these the most important foreground is synchrotron emission.

Free-free emission (bremsstrahlung) is intrinsically unpolarized but can be partially polarized by Thomson scattering within the HII region. This small effect is not expected to polarize the emission by more than 10%. The emission is larger at low frequencies but is not expected to dominate the polarization at any frequency.

The polarization of dust is not well known. In principle, emission from dust particles could be highly polarized, however it's been found that the majority of dust is polarized at the $\approx 2\%$ level at $100 \mu\text{m}$ with a small fraction of regions approaching 10% polarization. It's also been shown that even at 100% polarization, extrapolation of the IRAS $100 \mu\text{m}$ map with the COBE FIRAS index shows that dust emission is negligible below 80 GHz. At higher frequencies it will become the dominant foreground.

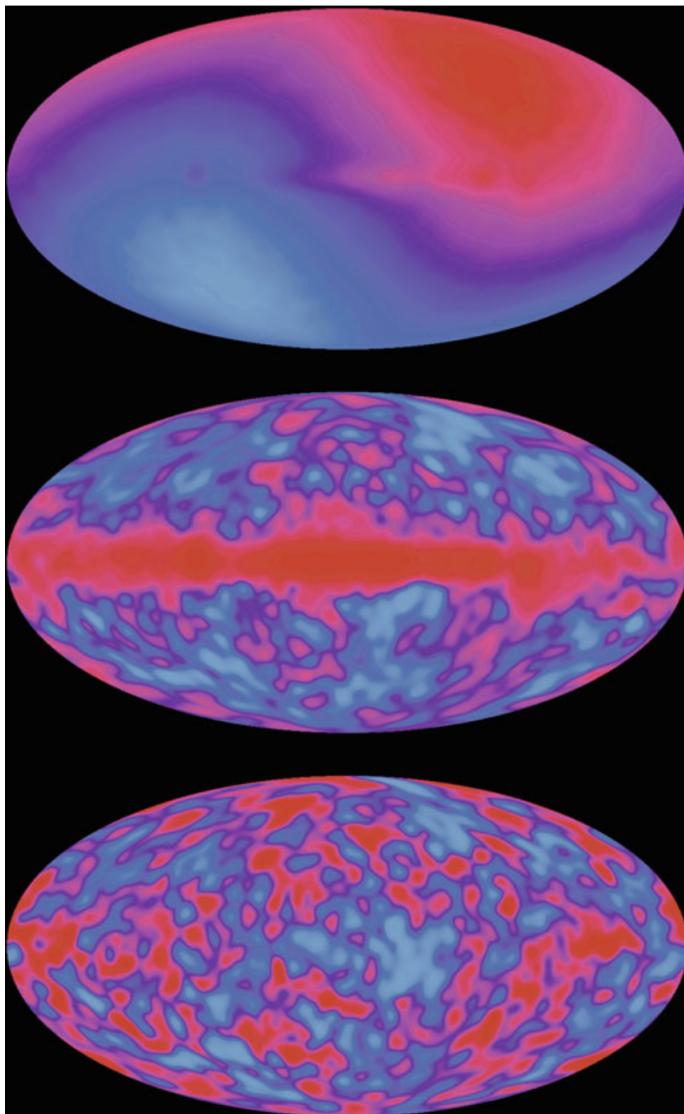


Figure 21: Temperature distribution of the CMB on the sky as measured by the COBE satellite. The top image shows a dipole distribution; it originates from the Earth's motion relative to the restframe of the CMB. Our Solar System moves at a speed of 369 km s^{-1} relative to that system, which leads to a dipole anisotropy with an amplitude of $\Delta T/T \sim v/c \sim 1.2 \times 10^{-3}$ due to the Doppler effect. If this dipole contribution is subtracted, we get the map in the middle which clearly shows the emission from the Galactic disk. Since this emission has a different spectral energy distribution (it is not a blackbody of $T \sim 3 \text{ K}$), it can also be subtracted to get the temperature map at the bottom. These are the primordial fluctuations of the CMB, with an amplitude of about $\Delta T/T \sim 2 \times 10^{-5}$. Credit: COBE/DRM team, NASA. Figure taken from Schneider (2006).

Radio point sources are polarized due to synchrotron emission at $< 20\%$ level. For large angle experiments, the random contribution from point sources will contribute negligibly, but may be of more concern for the upcoming satellite missions.

Galactic synchrotron emission is the major concern. It is potentially highly polarized with the fraction dependent on the spectral index and depolarization from Faraday rotation and non-uniform magnetic fields. The level of polarization is expected to lie between 10%-75% of a total intensity which itself is approximately $5 \mu\text{K}$ at 30 GHz. This estimate follows from extrapolating measurements at 1411 MHz with an index of $T \propto \nu^{-3}$.

Due to their different spectral indices, the minimum in the foreground polarization, like the temperature, lies near 100 GHz. For full sky measurements, since synchrotron emission is more highly polarized than dust, the optimum frequency at which to measure intrinsic (CMB) polarization is slightly higher than for the anisotropy. Over small regions of the sky where one or the other of the foregrounds is known a priori to be absent the optimum frequency would clearly be different. However as with anisotropy measurements, with multi-frequency coverage, polarized foregrounds can be removed.

It is also interesting to consider whether the spatial as well as frequency signature of the polarization can be used to separate foregrounds. Using angular power spectra for the spatial properties of the foregrounds is a simple generalization of methods already used in anisotropy work. For instance, in the case of synchrotron emission, if the spatial correlation in the polarization follows that of the temperature itself, the relative contamination will decrease on smaller angular scales due to its diffuse nature. Furthermore the peak of the cosmic signal in polarization occurs at even smaller angular scales than for the anisotropy. Of those that can provide information about the cosmic parameters...

The thermal Sunyaev-Zeldovich effect: Electrons in the hot gas of the intracluster medium (ICM) can inverse Compton scatter photons of the CMB. The optical depth and thus the scattering probability for this inverse Compton scattering is relatively low, but the effect is nevertheless observable and, in addition, is of great importance for the analysis of clusters. A photon moving through a cluster of galaxies towards us will change its direction through scattering and thus will not reach us. But since the cosmic background radiation is isotropic, for any CMB photon that is scattered out of the line-of-sight, another photon exists – statistically – that is scattered into it, so that the total number of photons reaching us is preserved. However, the energy of the photons changes slightly through scattering by the hot electrons, in a way that they have an (on average) higher frequency after scattering. Hence, by this inverse Compton scattering, energy is on average transferred from the electrons to the photons, as can be seen in Figure 22.

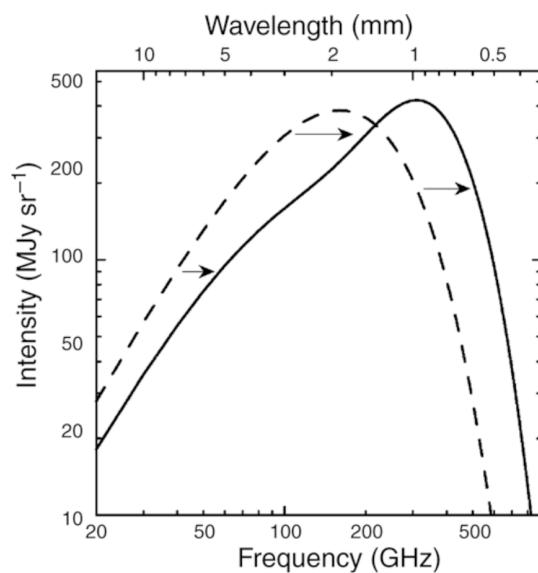


Figure 22: The influence of the Sunyaev-Zeldovich effect on the cosmic background radiation. The dashed curve represents the Planck distribution of the unperturbed CMB spectrum, the solid curve shows the spectrum after the radiation has passed through a cloud of hot electrons. The magnitude of this effect, for clarity, has been very much exaggerated in this sketch. Source: Carlstrom et al. 2002, ARA&A 40, 643. Image taken from Schneider (2006).

In the Rayleigh-Jeans (RJ) domain of the CMB spectrum, at wavelengths larger than about 2 mm, the intensity of the CMB is decreased by the SZ-effect. For the change in specific intensity in the RJ part, one obtains

$$\frac{\Delta I_\nu^{\text{RJ}}}{I_\nu^{\text{RJ}}} = -2y,$$

where y is the the Compton y-parameter

$$y = \int \frac{k_B T}{m_e c^2} \sigma_T n_e dl$$

and σ_T is the Compton cross-section for electron scattering

$$\sigma_T = \frac{8\pi}{3} \left(\frac{e^2}{m_e c^2} \right)^2 = \frac{8\pi}{3} e_{e,0}^2 [\text{m}^2].$$

Obviously, y is proportional to the optical depth with respect to Compton scattering, given as an integral over $n_e \sigma_T$ along the line-of-sight. Furthermore, y is proportional to the gas temperature, because that defines the average energy transfer per scattering event. Overall, y is proportional to the integral over the gas pressure $PD = nk_B T$ along the line-of-sight through the cluster.

The SZ effect affects the temperature distribution of the CMB. Some of the photons propagating along lines-of-sight passing through clusters of galaxies or other regions of dense and hot gas are scattered by the hot electrons, resulting in a temperature change in these directions. We recall that in the direction of clusters the measured intensity of the CMB radiation is reduced at low frequencies, whereas it is increased at high frequencies. Hence, the SZ effect can be identified in the CMB data if measurements are conducted over a sufficiently large frequency range.

The SZE spectral distortion of the CMB expressed as a temperature change ΔT_{SZE} at dimensionless frequency $x \equiv h\nu k_B T_{\text{CMB}}$ is given by

$$\frac{\Delta T_{\text{SZE}}}{T_{\text{CMB}}} = f(x)y = f(x) \int n_e \frac{k_B T_e}{m_e c^2} \sigma_T dl \text{ [dimensionless]},$$

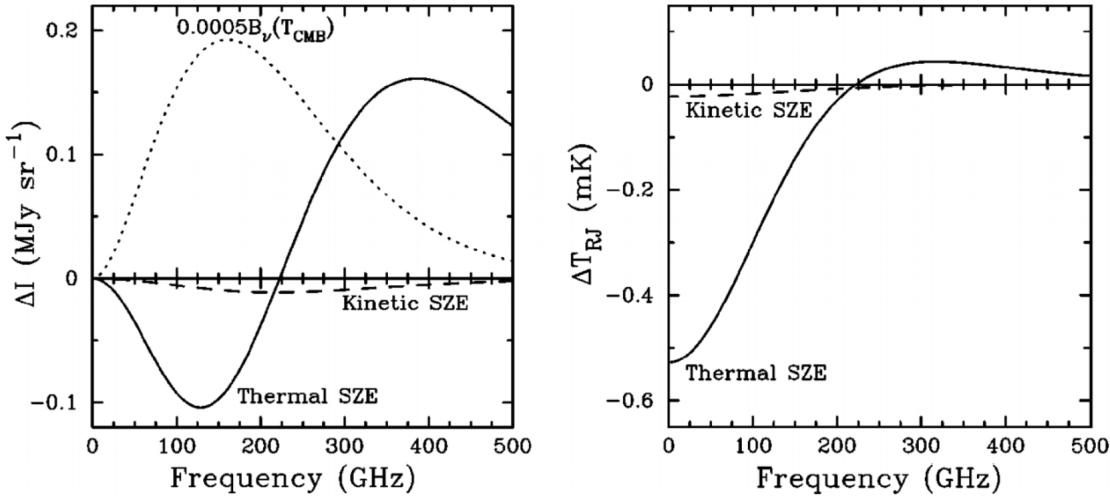


Figure 23: Spectral distortion of the CMB radiation due to the SZE. The left panel shows the intensity and the right panel shows the RJ brightness temperature. The thick solid line is the thermal SZE and the dashed line is the kinetic SZE. For reference the 2.7 K thermal spectrum for the CMB intensity scaled by 0.0005 is shown by the dotted line in the left panel. The cluster properties used to calculate the spectra are an electron temperature of 10 keV, a Compton y parameter of 10^{-4} , and a peculiar velocity of 500 km s^{-1} . Image taken from Carlstrom (2002).

where y is the Compton y -parameter, which for an isothermal cluster equals the optical depth, τ_e , times the fractional energy gain per scattering, σ_T is the Thomson cross-section, n_e is the electron number density, T_e is the electron temperature, k_B is the Boltzmann constant, $m_e c^2$ is the electron rest mass energy, and the integration is along the line of sight. The frequency dependence of the SZE is

$$f(x) = x \left(\frac{e^x + 1}{e^x - 1} \right) (1 + \delta_{\text{SZE}}(bx, T_e)) \text{ [dimensionless]},$$

where $\delta_{\text{SZE}}(x, T_e)$ is the relativistic correction to the frequency dependence. Note that $f(x) \rightarrow -2$ in the nonrelativistic and Rayleigh-Jeans (RJ) limits.

It is worth noting that $\Delta T_{\text{SZE}}/T_{\text{CMB}}$ is independent of redshift. This unique feature of the SZE makes it a potentially powerful tool for investigating the high-redshift Universe.

The spectral distortion of the CMB spectrum by the thermal SZE is shown in Figure 23 (solid line) for a realistic massive cluster ($y = 10^{-4}$) in units of intensity (left panel) and RJ brightness temperature (right panel). The RJ brightness is shown because the sensitivity of a radio telescope is calibrated in these units. It is defined simply by $I_{\nu} = (2k_B\nu^2/c^2)T_{RJ}$, where I_{ν} is the intensity at frequency ν , k_B is Boltzmann's constant, and c is the speed of light. The CMB blackbody spectrum, $B_{\nu}(T_{\text{CMB}})$, multiplied by 0.0005 (dotted line) is also shown for comparison. Note that the spectral signature of the thermal effect is distinguished readily from a simple temperature fluctuation of the CMB. The kinetic SZE distortion is shown by the dashed curve. In the nonrelativistic regime it is indistinguishable from a CMB temperature fluctuation.

The most important features of the thermal SZE are that (a) it is a small spectral distortion of the CMB of order $\sim 1 \text{ mK}$, which is proportional to the cluster pressure integrated along the line of sight; (b) it is independent of redshift; (c) it has a unique spectral signature with a decrease in the CMB intensity at frequencies $\lesssim 218 \text{ GHz}$ and an increase at higher frequencies; and (d) the integrated SZE flux is proportional to the temperature-weighted mass (total thermal energy) of the cluster, implying that SZE surveys will have a mass threshold nearly independent of redshift.

The kinetic Sunyaev-Zeldovich effect: If the cluster is moving with respect to the CMB rest frame, there will be an additional spectral distortion due to the Doppler effect of the cluster bulk velocity on the scattered CMB photons. If a component of the cluster velocity, v_{pec} , is projected along the line of sight to the cluster, then the Doppler effect will lead to an observed distortion of the CMB spectrum referred to as the kinetic SZE. In the nonrelativistic limit the spectral signature of the kinetic SZE is a pure thermal distortion of magnitude

$$\frac{\Delta T_{\text{SZ}}}{T_{\text{CMB}}} = -\tau_e \left(\frac{v_{\text{pec}}}{c} \right) \text{ [dimensionless]},$$

where v_{pec} is along the line of sight; i.e., the emergent spectrum is still described completely by a Planck

Campbell Cosmo Q10

spectrum, but at a slightly different temperature, lower (higher) for positive (negative) peculiar velocities (see Figure 23).

The scattering of the CMB photons by the hot intracluster medium (ICM) electrons can result in polarization at levels proportional to powers of (v_{pec}/c) and τ_e . The largest polarization is expected from the anisotropic optical depth to a given location in the cluster. For example, toward the outskirts of a cluster one expects to see a concentric (radial) pattern of the linear polarization at frequencies at which the thermal SZE is positive (negative). Nonspherical morphology for the electron distributions will lead to considerably complicated polarization patterns. The peak polarization of this signal will be of order τ_e times the SZE signal (i.e., of order $0.025(k_B T_e/m_e c^2)\tau_e^2$ times the CMB intensity). For a massive cluster with $\tau_e = 0.01$, the effect would be at the $0.1 \mu\text{K}$ level toward the edge of the cluster. In principle, this effect could be used to measure the optical depth of the cluster and therefore separate T_e and τ_e from a measurement of the thermal SZE.

It can be shown that polarization of the SZE comes entirely from the quadrupole component of the local radiation field experienced by the scattering electron. In the case above, the quadrupole component at the outskirts of the cluster is caused by the anisotropy in the radiation field in the direction of the cluster center due to the SZE. Sunyaev and Zel'dovich discussed polarization due to the motion of the cluster with respect to the CMB and transverse to our line of sight. In this case, the quadrupole comes from the Doppler shift. They found the largest terms to be of order $0.1\tau_e(v_{\text{pec}}/c)^2$ and $0.025\tau_e^2(v_{\text{pec}}/c)$ of the CMB intensity. The latter term, second order in τ_e , can be thought of as imperfect cancellation of the dipole term due to the anisotropic optical depth. Using $\tau_e = 0.01$ and a bulk motion of 500 km s^{-1} results in polarization levels of order $10 \mu\text{K}$, far beyond the sensitivity of current instrumentation.

The CMB as seen by the cluster electrons will have a quadrupole component and therefore the electron scattering will lead to linear polarization. This mechanism could possibly be used to trace the evolution of the CMB quadrupole if polarization measurements could be obtained for a large number of clusters binned in direction and redshift. Sazonov and Sunyaev calculated the expected polarization level and found that the maximum CMB quadrupole-induced polarization is $50(\tau_e/0.01) \mu\text{K}$, somewhat higher than the expected velocity-induced terms discussed above. The effect is again too small to expect detection in the near future. However, by averaging over many clusters, detecting this polarization might be possible with future satellite missions.

The integrated Sachs-Wolfe effect: Inhomogeneities in the gravitational potential cause photons which originate in regions of higher density to climb out of a potential well. As a result of this, they lose energy and are redshifted (gravitational redshift). This effect is partly compensated for by the fact that, besides the gravitational redshift, a gravitational time delay also occurs: a photon that originates in an overdense region will be scattered at a slightly earlier time, and thus at a slightly higher temperature of the Universe, compared to a photon from a region of average density. Both effects always occur side by side. They are combined under the term Sachs-Wolfe effect. Its separation into two processes is necessary only in a simplified description; a general relativistic treatment of the Sachs-Wolfe effect jointly yields both processes.

1.11.3 Follow-up Questions

- How does the thermal Sunyaev-Zeldovich effect compare to the kinetic Sunyaev-Zeldovich effect?
- How do we actually measure the SZ effect?
- What is the redshift dependence of the SZ effect? (There is none!)
- How do we actually measure the integrated Sachs-Wolfe effect?

Q11) COSMOLOGICAL INFLATION

Describe cosmological inflation. List at least three important observations it is intended to explain.

Ludwig Cosmo Q10

Question 10 - Inflation

Describe cosmological inflation. List at least three important observations it is intended to explain.

Solution

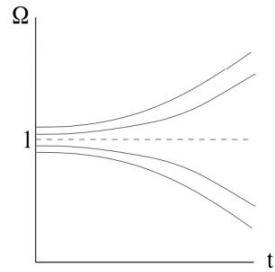
Inflation occurred 10^{-36} to 10^{-34} seconds after the big bang when quantum fluctuations allowed a small region of space to enter a true vacuum state in a universe otherwise filled with false vacuum. The conceptual part of this is tricky. A simple approach is that the universe was temporarily dominated by a positive cosmological constant Λ_i with $w = -1$.

Relevant Equations

- $a(t) = e^{H_it}$
- $w = p/\rho, \rho \propto a^{-3(1+w)}$ with $w = -1$

Flatness Problem \Rightarrow The Universe is nearly flat today, and even flatter in the past. Seems fine tuned.

- We can write the Friedmann equation as
$$1 - \Omega(t) = \frac{H_0^2(1 - \Omega_0)}{H(t)^2 a(t)^2}$$
- $\Omega(t) = 1$ is unstable equilibrium. Any deviation from this would quickly make things less flat.
- Inflation reduces the dependence of the current flatness on the initial conditions.



Horizon Problem \Rightarrow How is the Universe so homogeneous if not all particles are in causal contact?

- Distance to last scatter $d_p(t_0) = c \int_{t_{ls}}^{t_0} \frac{dt}{a(t)}$, but $t_{ls} \ll t_0$ only slightly smaller than current horizon distance.
- If you look at two points separated by 180 degrees on the sky, their current proper distance is $\approx 2d_{hor}(t_0)$, hence they are causally disconnected.
- CMB measurements show that those particles are the same temperature to within one part in 10^5 . How are they in thermal equilibrium?
- Universe is matter dominated at last scattering, $a = (3/2H_0t)^{2/3}$, you can get the angular diameter distance $d_a = \frac{c}{1+z} \int \frac{dz}{H(z)}$, which shows you that anything not in a 2 degrees radius is not in causal contact and $d_{hor}(t_{ls}) \approx 0.4$ Mpc.
 - If you factor in inflation and allow 100 e-foldings $d_{hor}(t_{ls}) \approx 10^{43}$ Mpc allowing the last scattering surface to be in causal contact.

Monopole Problem

- It would be beautiful if all of the forces were originally unified in the beginning of the universe. We know that the EM and weak forces become the same force at super high temperatures after all. However one of the predictions of this is that at one point the universe would have been dominated by magnetic monopoles since they occur at every defect in the quantum field and we can't find them.
- Inflation dilutes the number of magnetic monopoles practically out of existence.

Herman Cosmo Q11

C11

- Expansion of true vacuum bubbles in the false vacuum Universe was driven by false vacuum's $\propto E$, so expanded exponentially with time
- Solves magnetic monopole, horizon, and flatness problem

• Inflation

- At $\sim 10^{-34}$ s the universe underwent a period of exponential expansion. The idea is that prior to inflation, the universe was in a weird false vacuum state where it had supercooled while remaining in a high energy density state (ie not the preferred energy state w/ lowest energy). Due to tiny quantum fluctuations, small regions of space would actually be in the true vacuum state. These regions would've started expanding rapidly, since the false vacuum had a negative pressure and the true vacuum had zero pressure ($0 > \text{negative} \Rightarrow \text{expansion}$). So the negative pressure (and thus const. energy density) of the false vacuum dominated exponential expansion.

* like dark energy in today's universe, and

by like 10^{30} .

• Magnetic Monopoles

- Monopoles are expected to be very prevalent in the universe, since they would occur any time there was a defect in the quantum field following the break up of GUT Force. With inflation, these monopoles can be described as remnants of the false vacuum, still w/ high energy density and thus a lot of mass. Such defects would exist on the boundaries of the true vacuum bubbles, so when inflation caused exponential expansion these monopoles became far removed from what we can observe today.

• Flatness Problem

- our universe is incredibly flat, so much so that it must have been even flatter in the past otherwise the universe would have deviated significantly away from flatness by this time. Inflation solves this by exponential expansion. The curvature of the universe was flattened out by the expansion so that the initial curvature is irrelevant.

$$\Omega = 1 + \left(\frac{Kc^t}{a}\right)^2 \rightarrow a \propto e^t \text{ so } K \text{ doesn't matter, second term} \rightarrow 0, \Omega = 1.$$

• Horizon Problem

- the CMB appears to be homogeneous & isotropic on large scales, implying every part of it was in thermodynamic equilibrium at some

earlier time. However, with the original Big Bang model, there would be no way the entire universe could have been in causal contact at the time of BB? \rightarrow recombination. Inflation suggests every part of our universe was in causal contact prior to the expansion, and thus in thermal equilibrium, which is why the CMB looks smooth on large scales - inflation spread everything out so that any inhomogeneities lie far outside our observable universe.

Hilross

Q12) Two-POINT CORRELATION FUNCTION (2PCF)

Define the two-point correlation function of a Gaussian random field. How is it related to the power spectrum of that field? Describe how the above two concepts are used in cosmology.

Short Summary

The two point correlation function (2PCF) – $\xi(R)$ – describes the excess likelihood over a random distribution of finding 2 galaxies separated by a distance R . Equivalently, the 2PCF is the probability of finding a galaxy at some point x given another galaxy R away¹.

A Gaussian Random Field (field) is a stochastic process in spatial coordinates defined by at least one Gaussian kernel. The power spectrum describes the power per scale-length from each kernel.

The initial conditions of the universe (QM fluctuations during Inflation) are thought to be a GRF with a scale-invariant spectrum. This means that the amplitudes are independent of the scale-length of the Gaussians. Now, the universe is not scale-independent – The BAO ($\sim 150\text{Mpc}$) scale is the peak of galaxy clustering.

Equations

From empirical fits, the **2PCF** is given by

$$\xi_g(r) = \left(\frac{r}{r_0}\right)^{-\gamma} \quad [\text{dimensionless}],$$

where $r_0 \sim 5 h^{-1} \text{ Mpc}$ is the correlation length and $\gamma \simeq 1.7$ is the slope. This relation is reasonable over a range of separations: $0.2h^{-1}\text{Mpc} \lesssim r \lesssim 30h^{-1}\text{Mpc}$. The 2PCF is related to the **power spectrum** $P(k)$ via the Fourier transform,

$$P(k) = 2\pi \int_0^\infty \frac{\sin(kx)}{kx} \xi_g(x) x^2 dx \quad [\text{dimensionless}].$$

Gaussian Random Field

“One way of constructing a GRF is by assuming that the field is the sum of a large number... of plane waves with uniformly distributed random phase... At any point, the sum of these individual plane-wave contributions will exhibit a Gaussian distribution. This type of GRF is completely described by its power spectral density.

Suppose $f(x)$ is the value of a GRF at a point x in some D-dimensional space. If we make a vector of the values of f at N points, x_1, \dots, x_N , in the D-dimensional space, then the vector $(f(x_1), \dots, f(x_N))$ will always be distributed as a multivariate Gaussian”

(Modified from Wikipedia).

¹this definition is useful for actually evaluating the 2PCF

Q12) Herman Cosmo Q13

C13

- 2PCF describes prob of finding galaxies w/ some separation

- Related to Power spectrum via Fourier transform

- Ccl's relate to low z galaxy clustering via secondary anisotropies, like ISW, SZ, grav. lensing.

- Two-Point Correlation Function

- The 2PCF describes the likelihood of finding a galaxy at point X given that another galaxy exists with some characteristic separation. Worded differently, given some distance, what is the probability of finding 2 galaxies separated by that distance?

$$\xi(r) = (r/r_0)^{-\delta}, \quad \delta \approx 1.7$$

- If galaxies are more concentrated, $\xi(r) > 0$, and v.v.

- The 2PCF is related to the power spectrum via Fourier transform:

$$P(k) = 2\pi \int_0^\infty dx x^2 \frac{\sin(kx)}{kx} \xi(x)$$

- We can measure the correlation/separation b/w galaxies and determine the functional form of the 2PCF, which we can then use to describe the level of structure we see on different scales in the form of the power spectrum. So we can use one to determine the other.

- The relevance of low z galaxy clustering to the CMB spectrum of the CMB is through secondary anisotropies. Low z galaxies can produce changes to the spectrum via the ISW effect, the SZ effect, and grav. lensing (that one effects higher multipoles, small angular scales).

Hilroy

Q12) Campbell Cosmo Q13

1.14 Question 13

Define the two-point correlation function. How is it related to the power spectrum? How is the C_ℓ spectrum of the CMB related to low redshift galaxy clustering?

1.14.1 Short answer

The **two-point correlation function** is given by

$$\xi_g(r) = \left(\frac{r}{r_0}\right)^{-\gamma} \text{ [dimensionless]},$$

where $r_0 \sim 5 h^{-1}$ Mpc denotes the correlation length, and the slope is about $\gamma \simeq 1.7$. It is related to the **power spectrum** $P(k)$ via the Fourier transform

$$P(k) = 2\pi \int_0^\infty x^2 \frac{\sin(kx)}{kx} \xi(x) dx \text{ [dimensionless].}$$

1.14.2 Additional context

Galaxies are not randomly distributed in space, but rather they gather in groups, clusters, or even larger structures. Phrased differently, this means that the probability of finding a galaxy at location \mathbf{x} is not independent of whether there is a galaxy at a neighboring point \mathbf{y} . It is more probable to find a galaxy in the vicinity of another one than at an arbitrary location. This phenomenon is described such that one considers two points \mathbf{x} and \mathbf{y} , and two volume elements dV around these points. If \bar{n} is the average number density of galaxies, the probability of finding a galaxy in the volume element dV around x is then

$$P_1 = \bar{n}dV \text{ [dimensionless]},$$

independent of \mathbf{x} if we assume that the Universe is statistically homogeneous. We choose dV such that $P_1 \ll 1$, so that the probability of finding two or more galaxies in this volume element is negligible.

The probability of finding a galaxy in the volume element dV at location \mathbf{x} and at the same time finding a galaxy in the volume element dV at location \mathbf{y} is then

$$P_2 = (\bar{n}dV)^2 [1 + \xi_g(\mathbf{x}, \mathbf{y})] \text{ [dimensionless].}$$

If the distribution of galaxies was uncorrelated, the probability P_2 would simply be the product of the probabilities of finding a galaxy at each of the locations \mathbf{x} and \mathbf{y} in a volume element dV , so $P_2 = P_1^2$. But since the distribution is correlated, the relation does not apply in this simple form; rather, it needs to be modified, as was done. This equation defines the **two-point correlation function** (or simply '**correlation function**') of galaxies $\xi_g(\mathbf{x}, \mathbf{y})$.

By analogy to this, the correlation function for the total matter density can be defined as

$$\begin{aligned} \langle \rho(\mathbf{x})\rho(\mathbf{y}) \rangle &= \bar{\rho}^2 \langle [1 + \delta(\mathbf{x})][1 + \delta(\mathbf{y})] \rangle \\ &= \bar{\rho}^2 (1 + \delta(\mathbf{x})\delta(\mathbf{y})) \\ &= \bar{\rho}^2 (1 + \xi(\mathbf{x}, \mathbf{y})) \text{ [dimensionless]}, \end{aligned}$$

because the mean (or expectation) value $\langle \delta(\mathbf{x}) \rangle$ for all locations \mathbf{x} .

In the above equations, angular brackets denote averaging over an ensemble of distributions that all have identical statistical properties. In our example of the lake, the correlation function of the wave amplitudes at positions \mathbf{x} and \mathbf{y} , for instance, would be determined by taking a large number of snapshots of its surface and then averaging the product of the amplitudes at these two locations over all these realizations. Since the Universe is considered statistically homogeneous, ξ can only depend on the difference $\mathbf{x} - \mathbf{y}$ and not on \mathbf{x} and \mathbf{y} individually. Furthermore, ξ can only depend on the separation $r = |\mathbf{x} - \mathbf{y}|$, and not on the direction of the separation vector $\mathbf{x} - \mathbf{y}$ because of the assumed statistical isotropy of the Universe. Therefore, $\xi = \xi(r)$ is simply a function of the separation between two points.

For a homogeneous random field, the ensemble average can be replaced by spatial averaging (i.e., the correlation function can be determined by averaging over the products of densities at pairs of points) for a large number of pairs of points with given separation r . For determining the correlation function of galaxies, we note that $\xi_g(r)$ is the excess probability to find a galaxy at a separation r from another galaxy, relative to that of a random distribution. Therefore, $\xi_g(r)$ can be determined by first counting the number of galaxy pairs with separation in the interval Δr around r . Then one creates a random

Q12) Campbell Cosmo Q13

distribution of the same number of objects in the same volume, and again counts the pairs in the same distance interval. The ratio of these two pair counts then yields an estimate for $\xi_g(r)$.

The equivalence of ensemble average and spatial average is called the ergodicity of the random field. Only by this can the correlation function (and all other statistical properties) in our Universe be measured at all, because we are able to observe only a single (namely our) realization of the hypothetical ensemble. From the measured correlations between galaxy positions, as determined from spectroscopic redshift surveys of galaxies, one finds the approximate relation

$$\xi_g(r) = \left(\frac{r}{r_0}\right)^{-\gamma} \text{ [dimensionless]},$$

for galaxies of luminosity $\sim L^*$, where $r_0 \sim 5 h^{-1}$ Mpc denotes the correlation length, and where the slope is about $\gamma \simeq 1.7$. This relation is approximately valid over a range of separations $0.2h^{-1}$ Mpc $\lesssim r \lesssim 30h^{-1}$ Mpc.

Hence, the correlation function provides a means to characterize the structure of the cosmological matter distribution. Besides this two-point correlation function, correlations of higher order may also be defined, leading to general n-point correlation functions. These are more difficult to determine from observation, though. It can be shown that the statistical properties of a random field are fully specified by the set of all n-point correlations.

An alternative (and equivalent) description of the statistical properties of a random field, and thus of the matter distribution in a Universe, is the power spectrum $P(k)$. Roughly speaking, the power spectrum $P(k)$ describes the level of structure as a function of the length-scale $L \simeq 2\pi/k$; the larger $P(k)$, the larger the amplitude of the fluctuations on a length-scale $2\pi/k$. Here, k is a comoving wave number. Phrased differently, the density fluctuations are decomposed into a sum of plane waves of the form $\delta(\mathbf{x}) = \sum a_k \cos(\mathbf{x} \cdot \mathbf{k})$, with a wave vector \mathbf{k} and an amplitude a_k . The power spectrum $P(k)$ then describes the mean of the squares, $|a_k|^2$, of the amplitudes, averaged over all wave vectors with equal length $k = |\mathbf{k}|$. Technically speaking, this is a Fourier decomposition. Referring back to the example of waves on the surface of a lake, one finds that a characteristic wavelength L_c exists, which depends, among other factors, on the wind speed. In this case, the power spectrum will have a prominent maximum at $k = 2\pi/L_c$.

The power spectrum $P(k)$ and the correlation function are related through a Fourier transform; formally, one has

$$P(k) = 2\pi \int_0^\infty x^2 \frac{\sin(kx)}{kx} \xi(x) dx \text{ [dimensionless]},$$

i.e., the integral over the correlation function with a weight factor depending on $k \sim 2\pi/L$. This relation can also be inverted, and thus $\xi(x)$ can be computed from $P(k)$.

Since the CMB temperature fluctuations $\delta T/T$ is defined on the surface of a sphere – the celestial sphere, in this case – it is useful to expand it in spherical harmonics:

$$\frac{\delta T}{T}(\theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=-1}^{\ell} a_{\ell m} Y_{\ell m}(\theta, \phi)$$

where $Y_{\ell m}(\theta, \phi)$ are the usual **spherical harmonic functions**⁴. What concerns cosmologists is not the exact pattern of hot spots and cold spots on the sky, but their statistical properties. The most important statistical property of $\delta T/T$ is the correlation function $C(\theta)$. Consider two points on the last scattering surface. Relative to an observer, they are in the directions \hat{n} and \hat{n}' , and are separated by an angle θ given by the relation $\cos\theta = \hat{n} \cdot \hat{n}'$. To find the correlation function $C(\theta)$, multiply together the values of $\delta T/T$ at the two points, then average the product over all points separated by the angle θ :

$$C(\theta) = \left\langle \frac{\delta T}{T}(\hat{n}) \frac{\delta T}{T}(\hat{n}') \right\rangle_{\hat{n} \cdot \hat{n}' = \cos\theta}$$

If cosmologists knew the precise value of $C(\theta)$ for all angles from $\theta = 0$ to $\theta = 180^\circ$, they would have a complete statistical description of the temperature fluctuations over all angular scales. Unfortunately, the CMB measurements which tell us about $C(\theta)$ contain information over only a limited range of angular scales.

The limited angular resolution of available observations is what makes the spherical harmonic expansion of $\delta T/T$ so useful. Using the expansion of $\delta T/T$ in spherical harmonics, the correlation function can be written in the form

⁴https://en.wikipedia.org/wiki/Table_of_spherical_harmonics

Q12) Campbell Cosmo Q13

$$C(\theta) = \frac{1}{4\pi} \sum_{\ell=0}^{\infty} (2\ell + \ell)(C_\ell) P_\ell(\cos\theta),$$

where P_ℓ are the usual **Legendre polynomials**:

$$\begin{aligned} P_0(x) &= 1 \\ P_1(x) &= x \\ P_2(x) &= \frac{1}{2}(3x^2 - 1) \end{aligned}$$

and so forth. In this way, a measured correlation function $C(\theta)$ can be broken down into its multipole moments C_ℓ .

Generally speaking, a term C_ℓ is a measure of temperature fluctuations on the angular scale $\theta \sim 180^\circ/\ell$. Thus, the multipole ℓ is interchangeable, for all practical purposes, with the angular scale θ . The $\ell = 0$ (monopole) term of the correlation function vanishes if you've defined the mean temperature correctly. The $\ell = 1$ (dipole) term results primarily from the Doppler shift due to our motion through space. It is the moments with $\ell \geq 2$ which are of the most interest to cosmologists, since they tell us about the fluctuations present at the time of last scattering.

In presenting the results of CMB observations, it is customary to plot the function

$$\Delta T \equiv \sqrt{\frac{\ell(\ell+1)}{2\pi}} \langle T \rangle [\text{K}],$$

since this function tells us the contribution per logarithmic interval in ℓ to the total temperature fluctuation δT of the CMB. The detailed shape of the ΔT versus ℓ curve contains a wealth of information about the Universe at the time of photon decoupling.

At the time of last scattering, a particularly interesting length scale, cosmologically speaking, is the Hubble distance,

$$d_H = \frac{c}{H(z_{ls})} \approx \frac{3 \times 10^8 \text{ m s}^{-1}}{1.24 \times 10^{-18} \text{ s}^{-1} (1101)^{3/2}} \approx 6.6 \times 10^{21} \text{ m} \approx 0.2 \text{ [Mpc]},$$

where the redshift of last scattering is $z_{ls} \approx 1100$. A patch of the last scattering surface with this physical size will have an angular size, as seen from Earth, of

$$\theta_H = \frac{c/H(z_{ls})}{d_A} \approx \frac{0.2 \text{ Mpc}}{13 \text{ Mpc}} \approx 0.015 \text{ rad} \approx 1^\circ.$$

It is no coincidence that the peak in the ΔT versus ℓ curve occurs at an angular scale $\sim \theta_H$. The origin of temperature fluctuations with $\theta > \theta_H$ ($\ell < 180$) is different from those with $\theta < \theta_H$ ($\ell > 180$).

Q13) ANGULAR POWER SPECTRUM

How is the angular (C_l) power spectrum of the CMB related to the matter power spectrum and how can we use the C_{ls} s to learn about the initial conditions of the universe?

Short Summary

The angular power spectrum of the CMB is *sourced* by the density fluctuations that the matter power spectrum describes. That is, hot and cold spots in the CMB temperature arise from photons freely streaming out of over- and under-dense regions (respectively). The acoustic peaks in the CMB are so-named because they result from the Baryon Acoustic Oscillations (BAOs) in the matter-radiation plasma that existed prior to recombination.

Campbell

CMB angular power spectrum: The comparison between the predicted acoustic peak scale and its angular extent provides a measurement of the angular diameter distance to recombination. The angular diameter distance in turn depends on the spatial curvature and expansion history of the Universe. Assuming the size of the Universe's horizon at the time of recombination and the distance to the last scattering surface, the geometry (or curvature) of the Universe can be measured using the angular size of the first peak of the angular power spectrum. If the first peak is at $\ell \sim 220$, the Universe is flat, whereas if the first peak is at $\ell < 220$ or $\ell > 220$, the Universe is open or closed, respectively.

BAOs: The peaks in the CMB angular power spectrum are the result of compressions and expansions of the baryonic acoustic oscillations: the first and second peaks being the first compression and first decompression, respectively, etc. Given that the first compression happened on scales of the sound horizon at the time of last scattering, the first peak in the CMB angular power spectrum should tell you at what that scale was. This scale (i.e., the sound horizon) can be compared to what would be expected for a flat Universe – if it is smaller, the Universe is open whereas if it is larger, the Universe is closed (if they're equal, the Universe is of course flat).

Q14) POSITIVE COSMOLOGICAL CONSTANT

Consider a cosmological model including a positive cosmological constant. Show that, in such a model, the expansion factor eventually expands at an exponential rate. Sketch the time dependence of the expansion factor in the currently favoured cosmological model.

Short Summary

Q14) Ludwig Cosmo Q13

Question 13 - Cosmological Constant

Consider a cosmological model including a positive cosmological constant. Show that, in such a model, the expansion factor eventually expands at an exponential rate. Sketch the time dependence of the expansion factor in the currently favoured cosmological model.

Solution

- The scale factor eventually increases at an exponential rate if the cosmological constant is positive.

The Friedmann equation is written as:

$$\left(\frac{\dot{a}}{a}\right)^2 = \Omega_r a^{-4} + \Omega_m a^{-3} + \Omega_k a^{-2} + \Omega_\Lambda$$

For large a , only Ω_Λ is left

$$\left(\frac{\dot{a}}{a}\right)^2 = \Omega_\Lambda$$

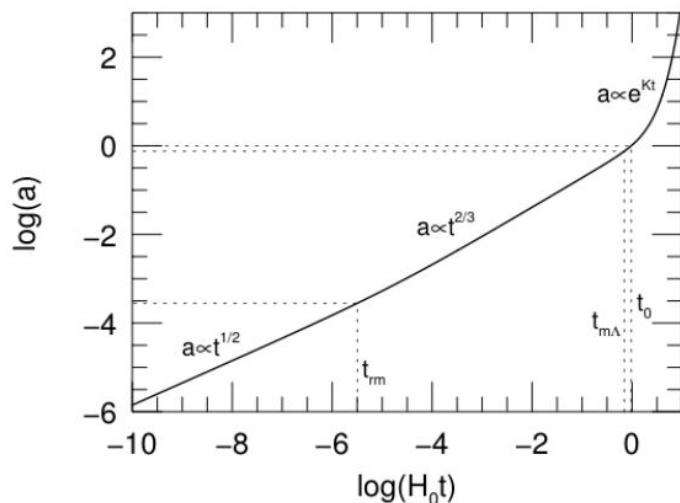
$$\frac{1}{a} da = \sqrt{\Omega_\Lambda} dt$$

$$\ln(a) = \sqrt{\Omega_\Lambda} t$$

$$a = e^{\sqrt{\Omega_\Lambda} t}$$

$$a \propto e^t$$

- Show the time dependence of the expansion factor in the currently favoured cosmological model.



Q14) Herman Cosmo Q14

C14

- Friedmann eqn describes \dot{a} w/ Ω
- Eventually Ω_m dominates as $a \uparrow$
- Draw current model

- Expansion with a positive cosmological constant Λ

- Starting from the Friedmann equation:

$$\left(\frac{\dot{a}}{a}\right)^2 = \left(\frac{H}{H_0}\right)^2 = \Omega_r a^{-4} + \Omega_m a^{-3} + \Omega_k a^{-2} + \Omega_\Lambda$$

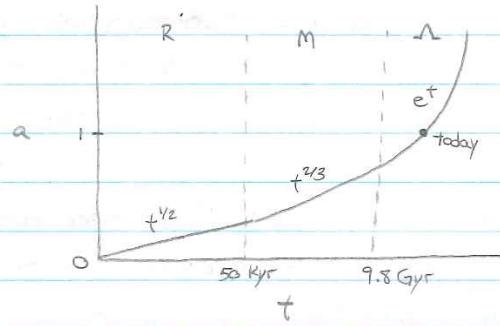
- Eventually, the scale factor will grow large enough that the 1st three parameters will approach zero and only Ω_Λ will matter. In this case:

$$\frac{\dot{a}}{a} = \sqrt{\Omega_\Lambda}$$

$$a(t) = \exp^{\sqrt{\Omega_\Lambda} t}$$

$$\therefore a(t) \sim e^t$$

- Current model of the universe:



$$\Omega_m \approx 0.3$$

$$\Omega_\Lambda \approx 0.7$$

$$\Omega_r \approx 10^{-5}$$

- Scale factor when radiation = matter:

$$\frac{\Omega_r}{a^4} = \left(\frac{\dot{a}}{a}\right)^2 = \frac{\Omega_m}{a^3}$$

$$a = \frac{\Omega_r / \Omega_m}{10^{-5} / 0.3}$$

\therefore very small!

Q14) Campbell Cosmo Q14

1.15 Question 14

Consider a cosmological model including a positive cosmological constant. Show that, in such a model, the expansion factor eventually expands at an exponential rate. Sketch the time dependence of the expansion factor in the currently favoured cosmological model.

1.15.1 Short answer

Friedmann equation can be written in terms of $\Omega_{r,0}$, $\Omega_{m,0}$, and $\Omega_{\Lambda,0}$ via

$$\left(\frac{H}{H_0}\right)^2 = \frac{\Omega_{r,0}}{a^4} + \frac{\Omega_{m,0}}{a^3} + \Omega_{\Lambda,0} + \frac{1 - \Omega_0}{a^2}$$

where $\Omega_0 = \Omega_{r,0} + \Omega_{m,0} + \Omega_{\Lambda,0}$. Since $(H/H_0)^2 = (\dot{a}/a)^2$, this can be re-written as

$$\begin{aligned} \left(\frac{\dot{a}(t)}{a(t)}\right)^2 &= \frac{\Omega_{r,0}}{a^4} + \frac{\Omega_{m,0}}{a^3} + \Omega_{\Lambda,0} + \frac{1 - \Omega_0}{a^2} \\ \dot{a}(t) &= a(t) \sqrt{\frac{\Omega_{r,0}}{a^4} + \frac{\Omega_{m,0}}{a^3} + \Omega_{\Lambda,0} + \frac{1 - \Omega_0}{a^2}}. \end{aligned}$$

For $a \gg 1$, this simplifies to

$$\begin{aligned} \frac{da}{dt} &\approx a(t) \sqrt{\Omega_{\Lambda,0}} \\ a(t) &\approx e^{\sqrt{\Omega_{\Lambda,0}} t}. \end{aligned}$$

Therefore, at large scale factors, the expansion proceeds at the exponential rate as follows:

$$a(t) \propto e^{\sqrt{\Omega_{\Lambda,0}} t}.$$

Figure 27 shows the time-dependence of the scale factor $a(t)$ in the Benchmark Model.

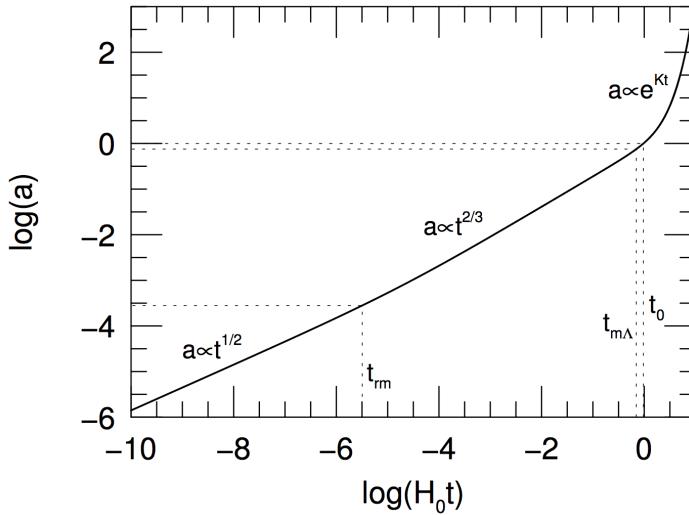


Figure 27: The scale factor a as a function of time t (measured in units of the Hubble time), computed for the Benchmark Model. The dotted lines indicate the time of radiation-matter equality, $a_{rm} = 2.8 \times 10^{-4}$, the time of matter-lambda equality, $a_{m\Lambda} = 0.75$, and the present moment, $a_0 = 1$. Figure taken from Ryden (2006).

1.15.2 Additional context

The Benchmark Model, is adopted as the best fit to the currently available observational data, is spatially flat, and contains radiation, matter, and a cosmological constant. The Hubble constant of the Benchmark Model is assumed to be $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$. The radiation in the Benchmark Model consists of photons and neutrinos. The photons are assumed to be provided solely by a CMB with current temperature $T_0 = 2.725 \text{ K}$ and density parameter $\Omega_{\gamma,0} = 5.0 \times 10^{-5}$. The energy density of the CMB is theoretically calculated to be 68% of that of the CMB, as long as neutrinos are relativistic. If a neutrino has a non-zero mass m_ν , it deflects from the “radiation” column to the “matter” column when the scale factor is $a \sim 5 \times 10^{-4} \text{ eV}/(m_\nu c^2)$. The matter content of the Benchmark Model consists partly of baryonic matter (that is, matter composed of protons and neutrons, with associated electrons), and partly of nonbaryonic dark matter; the evidence indicates that most of the matter in the Universe is nonbaryonic dark matter. The baryonic material that we are familiar with from our everyday existence has a density

Q14) Campbell Cosmo Q14

parameter of $\Omega_{b,0} \approx 0.04$ today. The density parameter of the nonbaryonic dark matter is roughly six times greater: $\Omega_{c,0} \approx 0.26$. The bulk of the energy density in the Benchmark Model, however, is not provided by radiation or matter, but by a cosmological constant, with $\Omega_{\Lambda,0} = 1 - \Omega_{m,0} - \Omega_{r,0} \approx 0.70$. The Benchmark Model was first radiation-dominated, then matter-dominated, and is now entering into its lambda-dominated phase. Radiation gave way to matter at a scale factor $a_{rm} = \Omega_{r,0}/\Omega_m, 0 = 2.8 \times 10^{-4}$, corresponding to a time $t_{rm} = 4.7 \times 10^4$ yr. Matter, in turn, gave way to the cosmological constant at $a_{m\Lambda} = (\Omega_{m,0}/\Omega_{\Lambda,0})^{1/3}$ corresponding to $t_{m\Lambda} = 9.8$ Gyr. The current age of the Universe, in the Benchmark Model, is $t_0 = 13.5$ Gyr.

Figure 27 shows the scale factor, thus computed, for the Benchmark Model. Note that the transition from the $a \propto t^{1/2}$ radiation-dominated phase to the $a \propto t^{2/3}$ matter-dominated phase is not an abrupt one; neither is the later transition from the matter-dominated phase to the exponentially growing lambda-dominated phase. One curious feature of the Benchmark Model which Figure 27 illustrates vividly is that we are living very close to the time of matter-lambda equality.

Radiation-Matter equality: To find the redshift of radiation-matter equality, obtain the Friedmann equation for each component using the scale factor relations above along with the definition of the density parameter $\Omega \equiv \rho/\rho_c = (8\pi G/3H_0)\rho$. For the radiation component, this gives us:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho_r(a) \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho_r a^{-4} \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3} \left(\frac{3H_0}{8\pi G}\Omega_r\right) a^{-4} \left(\frac{\dot{a}}{a}\right)^2 = H_0\Omega_r a^{-4}$$

along with the following for the matter component:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho_m(a) \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho_m a^{-3} \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3} \left(\frac{3H_0}{8\pi G}\Omega_m\right) a^{-3} \left(\frac{\dot{a}}{a}\right)^2 = H_0\Omega_m a^{-3}.$$

Equating the two Friedmann equations and solving for the radiation-matter equality scale factor a_{rm} using the relationship $a_{rm} = (1 + z_{rm})^{-1}$ for redshift:

$$\begin{aligned} H_0\Omega_r a_{rm}^{-4} &= H_0\Omega_m a_{rm}^{-3} \\ \Omega_r a_{rm}^{-1} &= \Omega_m a_{rm}^{-1} \\ a_{rm} &= \frac{\Omega_r}{\Omega_m} \\ (z_{rm} + 1)^{-1} &= \frac{\Omega_r}{\Omega_m} \\ (z_{rm} + 1) &= \frac{\Omega_m}{\Omega_r} \\ z_{rm} &= \frac{\Omega_m}{\Omega_r} - 1 \\ z_{rm} &= \frac{0.27}{5 \times 10^{-5}} - 1 \\ z_{rm} &= 5399. \end{aligned}$$

Matter- Λ equality: To find the redshift of matter- Λ equality, obtain the Friedmann equation for each component using the scale factor relations above along with the definition of the density parameter $\Omega \equiv \rho/\rho_c = (8\pi G/3H_0)\rho$. We already have the following for the matter component:

$$\left(\frac{\dot{a}}{a}\right)^2 = H_0\Omega_m a^{-3}$$

so let's work it out for the Λ component:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho_\Lambda(a) \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho_\Lambda \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3} \left(\frac{3H_0}{8\pi G}\Omega_\Lambda\right) \left(\frac{\dot{a}}{a}\right)^2 = H_0\Omega_\Lambda.$$

Equating the two Friedmann equations and solving for the matter- Λ equality scale factor $a_{m\Lambda}$ using the relationship $a_{m\Lambda} = (1 + z_{m\Lambda})^{-1}$ for redshift:

Q14) Campbell Cosmo Q14

$$\begin{aligned}H_0\Omega_m a_{m\Lambda}^{-3} &= H_0\Omega_\Lambda \\a_{m\Lambda}^{-3} &= \frac{\Omega_\Lambda}{\Omega_m} \\a_{m\Lambda}^3 &= \frac{\Omega_m}{\Omega_\Lambda} \\[(1+z_{m\Lambda})^{-1}]^3 &= \frac{\Omega_m}{\Omega_\Lambda} \\(1+z_{m\Lambda})^{-3} &= \frac{\Omega_m}{\Omega_\Lambda} \\(1+z_{m\Lambda}) &= \left(\frac{\Omega_m}{\Omega_\Lambda}\right)^{1/3} \\z_{m\Lambda} &= \left(\frac{\Omega_m}{\Omega_\Lambda}\right)^{1/3} - 1 \\z_{m\Lambda} &= \left(\frac{0.27}{0.73}\right)^{1/3} - 1 \\z_{m\Lambda} &= 0.39.\end{aligned}$$

1.15.3 Follow-up Questions

- At what time/scalefactor/redshift did the radiation-matter equality take place?
- At what time/scalefactor/redshift did the matter- Λ equality take place?
- What are the possible fates of the Universe?

Q15) EPOCH OF REIONIZATION

Define and describe the epoch of reionization. What are the observational constraints on it?

Short Summary

The Universe was largely neutral after recombination. The earliest galaxies/quasars emit photo-ionizing radiation which ionizes all of the neutral gas around it. This results in the Gunn-Peterson trough.

Q15) Ludwig Cosmo Q14 Question 14 - Reionization

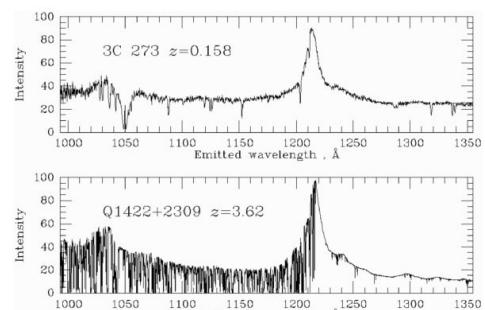
Define and describe the epoch of reionization. What are the observational constraints on it?

Solution

The reionization of the universe was due almost exclusively to photo-ionization. Collisional ionization is ruled out since the ICM wasn't hot enough for efficient collisional ionization. Reionization took place from $z \sim 12-6$

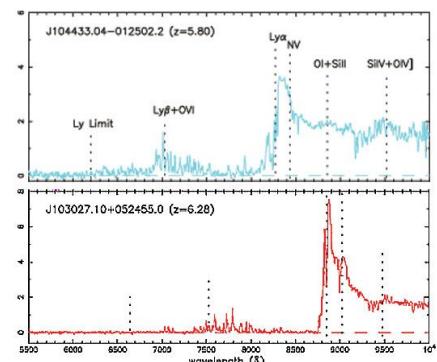
Lyman Alpha Forest

The presence of neutral hydrogen absorbs much more of the Lyman-alpha series. Direct observational evidence we have of the existence and properties of the general IGM.



Gunn-Peterson Trough of Quasars

Can constrain the end of reionization. The Gunn Peterson trough is a feature of the spectra of quasars due to the presence of neutral hydrogen in the Intergalactic Medium (IGM). The trough is characterized by suppression of electromagnetic emission from the quasar at wavelengths less than that of the Lyman-alpha line at the redshift of the emitted light. The discovery of the trough in a $z = 6.28$ quasar, and the absence of the trough in quasars detected at redshifts just below $z = 6$ presented strong evidence for the hydrogen in the universe having undergone a transition from neutral to ionized around $z = 6$.



The top is a close quasar, the bottom is very distant. Note the height of the spectral lines on the left side of the spectrum.

CMB Polarization

Can constrain the beginning of reionization. The degree of polarization from thomson scattering is related to the optical depth and therefore the abundance of free electrons.

Kinetic SZ Effect

Can tell us how long reionization lasted. The kinetic SZ effect looks at the temperature fluctuations in the CMB. The peculiar velocities of the ionized bubbles produce a doppler shift of CMB photons and the strength of this effect scales with the number of ionized bubbles present.

21 cm Spin Flip

Can trace the ionized fraction of hydrogen. This is the only way to probe the dark ages from recombination to reionization. It can provide a picture of the matter power spectrum after recombination as well as providing the picture for how the universe was reionized. In the late ($z < 9$) reionization epoch the 21 cm line is proportional to ionized fraction of hydrogen.

Q15) Herman Cosmo Q15

C15

- First stars and galaxies/AGN reionize Universe by photo ionization
- See evidence in metal lines of quasar spectra, GP trough, 21 cm line strength, increased CMB polarization

Epoch of Reionization

- The reionization of the universe was due (almost exclusively) to photo-ionization processes. Collisional ionization can be ruled out, since the IGM wasn't hot enough for efficient collisional ionization. Reionization took place from $z \approx 6$.
- As baryons fell into the DM wells, the gas needed to cool to form the 1st stars. This was accomplished mostly by H₂ cooling. The 1st stars (popn III) would have been a lot hotter and brighter due to their lack of metals, as metals increase opacity. The higher L meant they were efficient sources of ionizing photons, but also had shorter lifetimes. These massive stars ionized the H in bubbles around them, ironically destroying the H₂ capable of further cooling and thus continued star formation. The explosions of these stars as SNe enriched the IGM with metals, which we can observe in high z quasar spectra, since the Ly- α forest of the IGM shows metal absorption lines.
- Further ionization occurred on larger scales, due to the 1st proto-galaxies and AGN. Cooling by atomic H works best for more massive halos, allowing the 1st galaxies to form and continue reionization. The regions of ionized material gradually expand and grow until they overlap, and eventually the entire universe is ionized by $z \approx 6$.
- We see evidence of this EoR in the Ly- α forest & Gunn-Peterson Trough of quasars, which are very sensitive to even small amounts of HI capable of producing absorption lines at $\lambda_{rest} = 1216 \text{ \AA}$. The GP Trough allows us to place constraints on the end of reionization.
- We also see evidence of the EoR in secondary anisotropies of CMB + its polarization. The increase in free e⁻s from this time resulted in additional Thomson scattering, producing further polarization. The increase in scattering events also increased the optical depth, affecting the Hilbert

amplitude of T fluctuations (maybe? DK). At low multipoles in the polarization spectrum we observe a reionization bump, since this took place on the horizon scale.

- The Kinetic SZ-effect can also be seen in the temperature fluctuations of the CMB. The peculiar velocities of the ionized bubbles produce a Doppler shift of CMB photons, and the strength of this effect scales w/ the # of ionized bubbles present. This anisotropy tells us the duration of reionization.
- Also consider 21 cm spin flip of H I during EoR.

Q15) Campbell Cosmo Q15

1.16 Question 15

Define and describe the epoch of reionization. What are the observational constraints on it?

1.16.1 Short answer

After recombination occurring at a redshift $z \sim 1,100$, the Universe entered the cosmic **Dark Ages** where it became predominantly neutral until $z \sim 16$. Around this time, gravitational Jeans instabilities allowed for the formation of the first generation of stars. While they have yet to be directly observed, these Population III (Pop III) stars were particularly massive ($\sim 100 M_{\odot}$), implying that they were extremely hot and capable of radiating significant amounts of ionizing UV photons. These Pop III stars triggered a significant phase transition in the IGM as they began ionizing the predominantly neutral gas during a time called the *Epoch of Reionization* (EoR). There are several observational constraints on this period, including Ly α forest lines, Gunn-Peterson absorption, and CMB polarization.

1.16.2 Additional context

The baryonic pre-galactic medium (PGM) evolves in three distinct phases. At high redshifts ($z > 1100$) the PGM is hot, fully ionized, and optically thick to **Thomson scattering**, and hence coupled to the photon field. As the Universe expands, the PGM cools, and eventually recombines, leaving a surface of last scattering (the CMB), plus a neutral PGM. This neutral phase lasts from $z = 1100$ to $z \sim 14$. At some point between $z \sim 14$ and 6, hydrogen in the PGM is “reionized,” due to UV radiation from the first luminous objects, leaving the fully reionized IGM seen during the “realm of the galaxies” ($0 < z < 6$). The ionized, dense PGM at very high redshift has been well studied through extensive observations of the CMB. Likewise, the reionized, rarified IGM at low redshift has been well characterized through QSO absorption line studies. The middle phase (the existence of a neutral IGM during the so-called dark ages and the process of reionization of this medium) is the last directly observable phase of cosmic evolution that remains to be verified and explored. The epoch of reionization (EoR) is crucial in cosmic structure formation studies, because it sets a fundamental benchmark indicating the formation of the first luminous objects, either star-forming galaxies or active galactic nuclei (AGN).

After recombination at $z \sim 1100$, the intergalactic gas became neutral, with a residual ionization fraction of only 10^{-4} . Had the Universe remained neutral we would not be able to receive any photons that were emitted bluewards of the Ly α line of a source, because the absorption cross section for Ly α photons is too large. Since such photons are observed from QSOs, as can be seen for instance in the spectra of the $z > 5.7$ QSOs in Figure 28, and since an appreciable fraction of homogeneously distributed neutral gas in the intergalactic medium can be excluded for $z \gtrsim 5$, from the tight upper bounds on the strength of the Gunn-Peterson effect, the Universe must have been reionized between the recombination epoch and the redshift $z \sim 7$ of the most distant known QSOs.

This raises the question of how this reionization proceeded, in particular which process was responsible for it. The latter question is easy to answer – reionization must have happened by photoionization. Collisional ionization can be ruled out because for it to be efficient the intergalactic medium (IGM) would need to be very hot, a scenario which can be excluded due to the perfect Planck spectrum of the CMB. Hence, the next question is what produced the energetic photons that caused the photoionization of the IGM.

Two kinds of sources may in principle account for them – hot stars or AGNs. Currently, it is not unambiguously clear which of these is the predominant source of energetic photons causing reionization since our current understanding of the formation of supermassive black holes is still insufficient. However, it is currently thought that the main source of photoionization photons is the first generation of hot stars. Following on from the above arguments, understanding reionization is thus directly linked to studying the first generation of stars. In the present Universe star formation occurs in galaxies; thus, one needs to examine when the first galaxies could have formed. From the theory of structure formation, the mass spectrum of dark matter halos at a given redshift can be computed by means of, e.g., the Press-Schechter model. Two conditions need to be fulfilled for stars to form in these halos. First, gas needs to be able to fall into the dark halos. Since the gas has a finite temperature, pressure forces may impede the infall into the potential well. Second, this gas also needs to be able to cool, condensing into clouds in which stars can then be formed.

By means of a simple argument, we can estimate under which conditions pressure forces are unable to prevent the infall of gas into a potential well. To do this, we consider a slightly overdense spherical region of radius R whose density is only a little larger than the mean cosmic matter density $\bar{\rho}$. If this sphere is homogeneously filled with baryons, the gravitational binding energy of the gas is about

Q15) Campbell Cosmo Q15

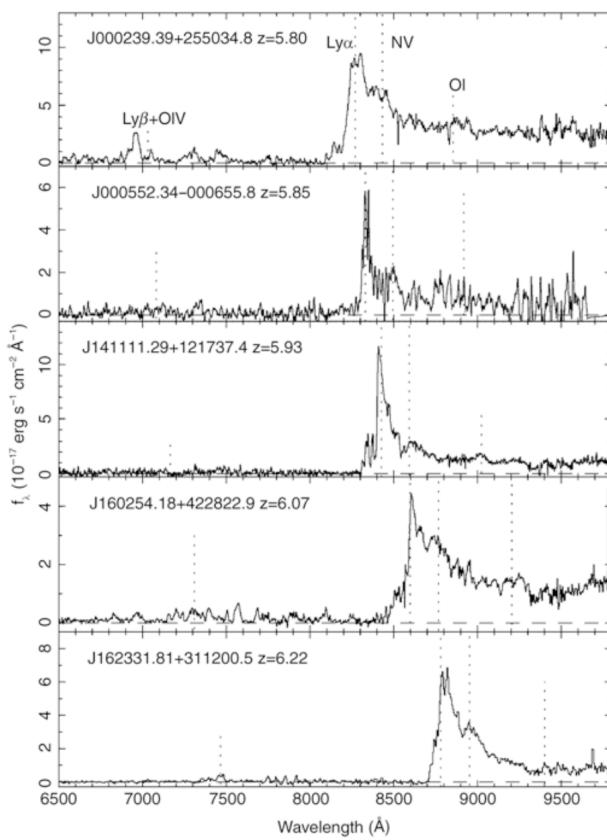


Figure 28:

Spectra of five QSOs at redshifts $z > 5.7$, discovered in multi-color data from the Sloan Digital Sky Survey. The positions of the most important emission lines are marked. Particularly remarkable is the almost complete lack of flux bluewards of the Ly α emission line in some of the QSOs, indicating a strong Gunn-Peterson effect. However, this absorption is not complete in all QSOs, which points at strong variations in the density of neutral hydrogen in the IGM at these high redshifts. Either the hydrogen density varies strongly for different lines-of-sight, or the degree of ionization is very inhomogeneous. Source: X. Fan et al. 2004, A Survey of $z > 5.7$ Quasars in the Sloan Digital Sky Survey. III. Discovery of Five Additional Quasars, AJ 128, 515, p. 517, Fig. 1. Figure taken from Ryden (2006).

$$|E_{\text{pot}}| \sim \frac{GMM_g}{R} [\text{J}],$$

where M and M_g denote the total mass and the gas mass of the sphere, respectively. The thermal energy of the gas can be computed from the kinetic energy per particle, multiplied by the number of particles in the gas, or

$$E_{\text{th}} \sim c_s^2 M_g [\text{J}],$$

where the sound speed is given by

$$c_s \approx \sqrt{\frac{k_B T}{\mu m}} [\text{m s}^{-1}]$$

which is about the average velocity of the gas particles, and m denotes the average particle mass in the gas. For the gas to be bound in the gravitational field, its gravitational binding energy needs to be larger than its thermal energy, $E_{\text{pot}} > E_{\text{th}}$, which yields the condition $GM > c_s^2 R$. Since we have assumed an only slightly overdense region, the relation $M \sim \bar{\rho} R^3$ between mass and radius of the sphere applies. From the two latter equations, the radius can be eliminated, yielding the condition

$$M > M_J \equiv \frac{\pi^{5/2}}{6} \left(\frac{c_s^2}{G} \right)^{3/2} \frac{1}{\sqrt{\bar{\rho}}} [\text{M}_\odot],$$

where M_J is the **Jeans mass**. The Jeans mass depends on the temperature of the gas, expressed through the sound speed c_s , and on the mean cosmic matter density $\bar{\rho}$. The latter can easily be expressed as a function of redshift, $\bar{\rho}(z) = \bar{\rho}_0(1+z)^3$.

The baryon temperature T_b has a more complicated dependence on redshift. For sufficiently high redshifts, the small fraction of free electrons that remains after recombination provides a thermal coupling of the baryons to the cosmic background radiation, by means of Compton scattering. This is the case for redshifts $z > z_t$, where

Q15) Campbell Cosmo Q15

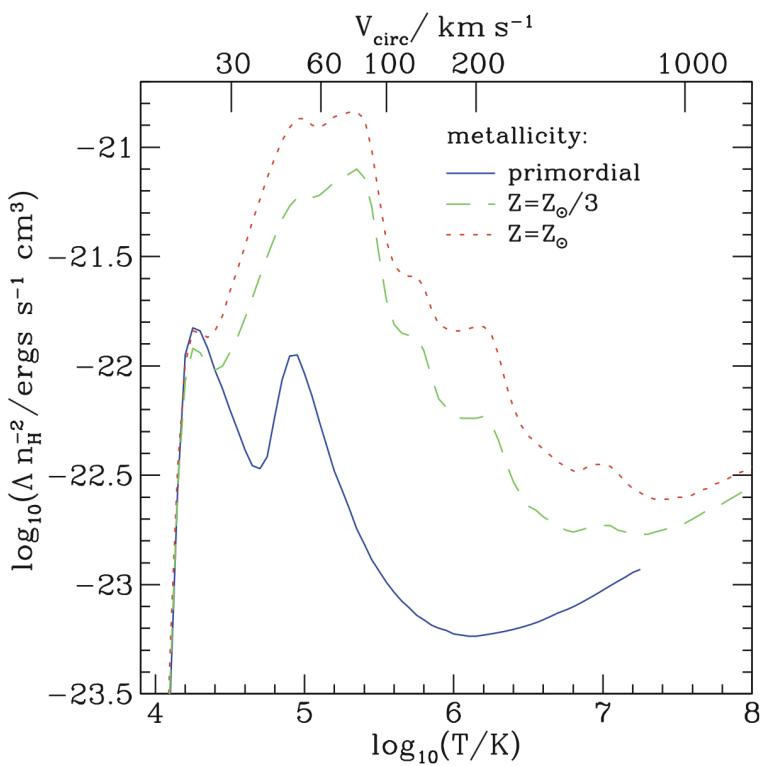


Figure 29: The cooling function for gas with primordial composition (blue solid curve), $1/3$ of the Solar metallicity (green dashed curve) and Solar metallicity (red dotted curve). On the top axis, the temperature is converted into a circular velocity. To obtain such a cooling function, one needs to assume an equilibrium state of the gas. Here it is assumed that the gas is in thermodynamical equilibrium, where the fraction of ionization states of any atom depends just on T . The total cooling function shown here is a superposition of different gas cooling processes, including atomic processes and bremsstrahlung, the latter of which dominating at high T where the gas is fully ionized. Source: C.M. Baugh 2006, A primer on hierarchical galaxy formation: the semi-analytical approach, arXiv:astro-ph/0610031, Fig.9. Figure taken from Ryden (2006).

$$z_t \approx 140 \left(\frac{\Omega_b h^2}{0.022} \right)^{2/5} \text{ [dimensionless]}$$

hence, $T_b(z) \approx T(z) = T_0(1+z)$ for $z > z_t$. For smaller redshifts, the density of photons gets too small to maintain this coupling, and baryons start to adiabatically cool down by the expansion, so that for $z \lesssim z_t$ we obtain approximately $T_b \propto \rho_b^{2/3} \propto (1+z)^2$.

From these temperature dependencies, the Jeans mass then be calculated as a function of redshift. For $z_t \gtrsim z \gtrsim 1000$, M_J is independent of z because $c_s \propto T^{1/2} \propto (1+z)^{1/2}$ and $\bar{\rho} \propto (1+z)^3$ and its value is

$$M_J = 1.35 \times 10^5 \left(\frac{\Omega_m h^2}{0.15} \right)^{-1/2} [\text{M}_\odot],$$

whereas for $z \gtrsim z_t$ we obtain, with $T_b \simeq 1.7 \times 10^{-2}(1+z)^2 \text{ K}$,

$$M_J = 5.7 \times 10^3 \left(\frac{\Omega_m h^2}{0.15} \right)^{-1/2} \left(\frac{\Omega_b h^2}{0.022} \right)^{-3/5} \left(\frac{1+z}{10} \right)^{3/2} [\text{M}_\odot].$$

Hence, gas can not fall into halos with mass lower than these values.

The Jeans criterion is a necessary condition for the formation of proto-galaxies (i.e., dark matter halos which contain baryons). In order to form stars, the gas in the halos needs to be able to cool further. Here, we are dealing with the particular situation of the first galaxies, whose gas is metal-free, so metal lines cannot contribute to the cooling. The cooling function of primordial gas is much smaller than that of enriched material; in particular, the absence of metals means that even slow cooling though excitation of fine-structure lines cannot occur, as there are no atoms with such transitions present. Thus, cooling by the primordial gas is efficient only above $T \gtrsim 2 \times 10^4 \text{ K}$. However, the halos formed at high redshift have low mass. The abundance of dark matter halos depends on the parameter ν which is the ratio between the density threshold required for collapse and the dispersion of fluctuations on a given mass scale. At high redshift, the growth factor $D_+(a)$ is small, and thus to have a noticeable abundance of halos of mass M , the dispersion $\sigma(M)$ must be correspondingly large. At redshift $z \sim 10$, the parameter ν is about unity for halos of mass $\sim 10^3 \text{ M}_\odot$. Hence, at that time, substantially more massive halos than that were (exponentially) rare (i.e., only low-mass halos were around) and their virial temperature

$$T_{\text{vir}} \approx 2 \times 10^2 \left(\frac{M}{10^5 h^{-1} \text{ M}_\odot} \right)^{2/3} \left(\frac{1+z}{10} \right) [\text{K}]$$

Q15) Campbell Cosmo Q15

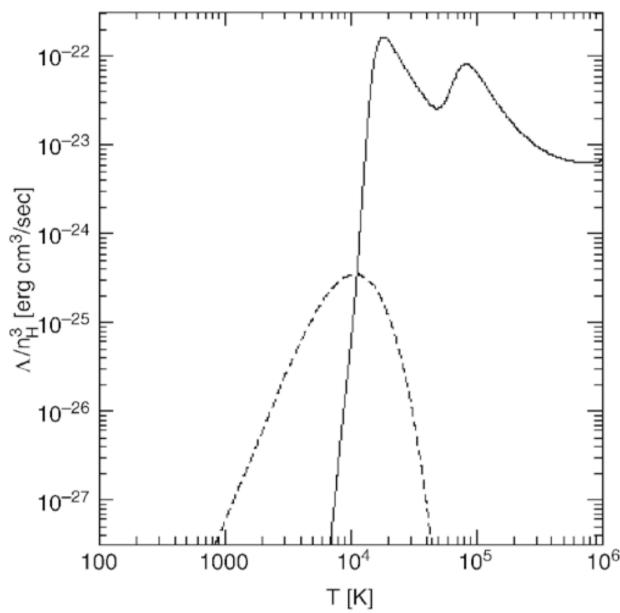


Figure 30: Cooling rate as a function of the temperature for a gas consisting of atomic and molecular hydrogen (with 0.1% abundance) and of helium. The solid curve describes the cooling by atomic gas, the dashed curve that by molecular hydrogen; thus, the latter is extremely important at temperatures below $\sim 10^4$ K. At considerably lower temperatures the gas cannot cool, hence no star formation can take place. Source: R. Barkana & A. Loeb 2000, In the Beginning: The First Sources of Light and the Reionization of the Universe, astro-ph/0010468, Fig. 12. Figure taken from Ryden (2006).

is considerably below the energy scale where atomic hydrogen can efficiently cool. Here, we used the fact that the mean matter density of a halo inside its virial radius is 200 times the critical density at a given redshift. Therefore, atomic hydrogen is a very inefficient coolant for these first halos, insufficient to initiate the formation of stars. Furthermore, helium is of no help in this context, since its excitation temperature is even higher than that of hydrogen.

Besides atomic hydrogen and helium, the primordial gas contains a small fraction of molecular hydrogen which represents an extremely important component in cooling processes. Whereas in enriched gas, molecular hydrogen is formed on dust particles, the primordial gas had no dust, and so H_2 must form in the gas phase itself, rendering its abundance very small. However, despite its very small density and transition probability, H_2 dominates the cooling rate of primordial gas at temperatures below $T \sim 10^4$ K (see Figure 30) where the precise value of this temperature depends on the abundance of H_2 .

By means of H_2 , the gas can cool in halos with a temperature exceeding about $T_{vir} \gtrsim 1000$ K, corresponding to a halo mass of $M \gtrsim 5 \times 10^4 M_\odot$ at $z \sim 20$. In these halos, stars may then be able to form. These stars will certainly be different from those known to us, because they do not contain any metals. Therefore, the opacity of the stellar plasma is much lower. Such stars, which at the same mass presumably have a much higher temperature and luminosity (and thus a shorter lifetime), are called population III stars. Due to their high temperature they are much more efficient sources of ionizing photons than stars with ‘normal’ metallicity.

The energetic photons from these population III stars are now capable of ionizing hydrogen in their vicinity. More important still is another effect: photons with energy above 11.26 eV can destroy H_2 . Since the Universe is transparent for photons with energies below 13.6 eV, photons with $11.26 \text{ eV} \geq E \geq 13.6 \text{ eV}$ can propagate very long distances and dissociate molecular hydrogen. This means that as soon as the first stars have formed in a region of the Universe, molecular hydrogen in their vicinities will be destroyed and further gas cooling and star formation will then be prevented. At this point, the Universe contains a low number density of isolated bubbles of ionized hydrogen, centered on those halos in which population III stars were able to form early, but this constitutes only a tiny fraction of the volume; most of the baryons remain neutral.

Soon after population III stars have formed, they will explode as supernovae. Through this process, the metals produced by them are ejected into the IGM, by which the initial metal enrichment occurs. The kinetic energy transferred by SNe to the gas within the halo can exceed its binding energy, so that the baryons of the halo can be blown away and further star formation is prevented. Whether this effect may indeed lead to gas-free halos, or whether the released energy can instead be radiated away, depends on the geometry of the star-formation regions. In any case, it can be assumed that in those halos where the first generation of stars was born, further star formation was considerably suppressed, particularly since all molecular hydrogen was destroyed.

We can assume that the metals produced in these first SN explosions are, at least partially, ejected from the halos into the IGM, thus enriching the latter. The existence of metal formation in the very early Universe is concluded from the fact that even sources at very high redshift (like QSOs at $z \sim 6$) have

Q15) Campbell Cosmo Q15

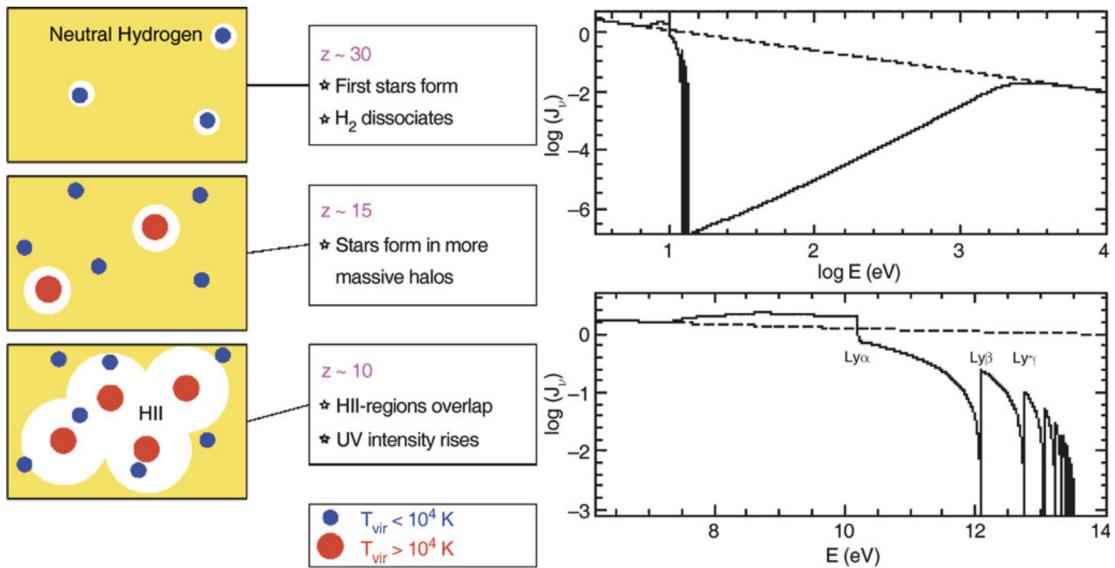


Figure 31: *On the left*, a sketch of the geometry of reionization is shown: initially, relatively low-mass halos collapse, a first generation of stars ionizes and heats the gas in and around these halos. By heating, the temperature increases so strongly (to about $T \sim 10^4$ K) that gas can escape from the potential wells; these halos may never again form stars efficiently. Only when more massive halos have collapsed will continuous star formation set in. Ionizing photons from this first generation of hot stars produce HII-regions around their halos, which is the onset of reionization. The regions in which hydrogen is ionized will grow until they start to overlap; at that time, the flux of ionizing photons will strongly increase. *On the right*, the average spectrum of photons at the beginning of the reionization epoch is shown; here, it has been assumed that the flux from the radiation source follows a power law (dashed curve). Photons with an energy higher than that of the Ly α transition are strongly suppressed because they are efficiently absorbed. The spectrum near the Lyman limit shows features which are produced by the combination of breaks corresponding to the various Lyman lines, and the redshifting of the photons. Source: R. Barkana & A. Loeb 2000, In the Beginning: The First Sources of Light and the Reionization of the Universe, astro-ph/0010468, Figs. 4, 11. Image taken from Ryden (2006).

a metallicity of about one tenth the Solar value. Furthermore, the Ly α forest also contains gas with non-vanishing metallicity. Since the Ly α forest is produced by the IGM, this therefore must have been enriched.

For gas to cool in halos without molecular hydrogen, their virial temperature needs to exceed about 10^4 K (see Figure 30). Halos of this virial temperature form with appreciable abundance at redshifts of $z \sim 10$, corresponding to a halo mass of $\sim 10^7 M_\odot$, as can be estimated from the Press-Schechter model. In these halos, efficient star formation can then take place and the first proto-galaxies form. These then ionize the surrounding IGM in the form of HII-regions, as sketched in Figure 31. The corresponding HII-regions expand because increasingly more photons are produced. If the halo density is sufficiently high, these HII-regions start to overlap and soon after fill the whole volume. Once this occurs, the IGM is ionized, and reionization is completed.

We therefore conclude that reionization is a two-stage process. In a first phase, population III stars form through cooling of gas by molecular hydrogen, which is then destroyed by these very stars. Only in a later epoch and in more massive halos cooling is provided by atomic hydrogen, leading to reionization. The increase of temperature causes an increase of the Jeans mass, due to its dependence on the sound speed. Once the IGM is heated to $\sim 10^4$ K by intergalactic UV radiation, the gas pressure prevents gas inflow into low-mass halos, corresponding to circular velocities $\lesssim 30 \text{ km s}^{-1}$. For this reason, one expects that halos of lower mass have a lower baryon fraction than that of the cosmic mixture, $f_b = \Omega_b/\Omega_m$. The actual value of the baryon fraction depends on the details of the merger history of a halo. Quantitative studies yield an average baryon mass of

$$\bar{M}_b = \frac{f_b M}{[1 + (2^{\alpha/3} - 1)(M_C/M)^\alpha]^{3/\alpha}} [\text{kg}],$$

where $M_C \sim 10^9 M_\odot$ is a characteristic mass, defined such that for a halo with mass M_C , $\bar{M}_b/M = f_b/2$. For halos of mass smaller than M_C , the baryon fraction is suppressed, decreasing as $(M/M_C)^3$ for small masses, whereas for halo masses $\gg M_C$, the baryon fraction corresponds to the cosmic average. The index $\alpha \sim 2$ determines the sharpness of the transition between these two cases. The characteristic mass M_C depends on redshift, being much smaller at high z due to the stronger ionizing background.

Q15) Campbell Cosmo Q15

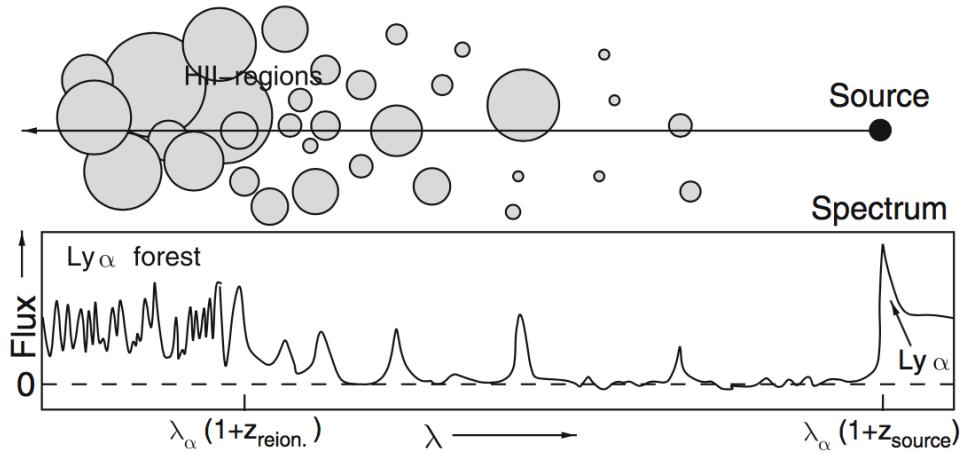


Figure 32: Sketch of a potential observation of reionization: light from a very distant QSO propagates through a partially ionized Universe; at locations where it passes through HII-regions, radiation will get through – flux will be visible at the corresponding wavelengths. When the HII-regions start to overlap, the normal Ly α forest will be produced. Adapted from: R. Barkana & A. Loeb 2000, In the Beginning: The First Sources of Light and the Reionization of the Universe, astro-ph/0010468. Figure taken from Ryden (2006).

The ionizing flux has two additional effects on the gas that resides in halos: it provides a source of heating due to photoionization and it leads to a higher degree of ionization in the gas, reducing the number density of atoms which can be excited by collisions and cool through de-excitation. Both effects act in the same direction, by impeding an efficient cooling of the gas and hence the formation of stars. For halos of larger mass, intergalactic radiation is of fairly little importance because the corresponding heating rate is substantially smaller than that occurring by the dissipation of the gas which is needed to concentrate the baryons towards the halo center. For low-mass halos, however, this effect is important. Together, these two effects reduce the cooling rate of the gas, which is a dominant effect for low-mass halos. Thus, the gas in low-mass halos cannot cool efficiently, suppressing star formation – unless star formation occurred before the reionization was completed. We hence found one of the elements for the second part of the answer to the question about the different mass-to-light ratios in halos: star formation in low mass halos is strongly suppressed due to the ionizing background radiation. This also provides an explanation of the ‘missing satellite problem’.

To singly ionize helium, photons of energy ≥ 24.6 eV are required, and the ionization energy of He II is four times that of hydrogen. In addition, the recombination rate of fully ionized helium is about five times higher than that of hydrogen. Therefore, the reionization of helium is expected to be completed at a later epoch when the density of photons with $\lambda < 304$ Å was high enough. Since even massive stars do not generate photons exceeding this energy in large quantities, the photons leading to helium reionization presumably are emitted by quasars; therefore, the ionization of helium has to wait for the ‘quasar epoch’ of the Universe, at $z \lesssim 4$. From the statistical analysis of the Ly α forest and from the analysis of helium absorption lines and the helium Gunn-Peterson effect in high-redshift QSOs, a reionization redshift of $z \sim 3$ for helium is obtained.

One of the challenges of current observational cosmology is to link the history of reionization, as outlined above, to the observation of the highest redshift sources (i.e., to see whether we can observe the sources which are responsible for cosmic reionization). Are the galaxy populations that we can find at very high redshifts sufficient to understand the reionization process? Here we shall mention some of the major obstacles for a direct observation probe of these ionizing sources.

If reionization was caused by the energetic photons emitted during star formation, the remnants of this first generation of stars must be present in the post-reionization Universe, and thus be observable. Galaxies at redshift $z > 6$ are observed, either as Lyman-break galaxies (LBGs), Lyman-alpha emitters (LAEs) or as sub-millimeter galaxies (SMGs). Their stellar masses can be estimated from their observed light. However, most of the LBGs are observed only in the near-IR, which means that we see their restframe UV-emission. Converting the UV-light into a stellar mass is highly uncertain, since it depends strongly on the instantaneous star-formation rate. For LAEs, it is even more challenging to determine a stellar mass, since they are typically fainter in their broad-band (i.e., continuum) emission, which renders the determination of the stellar mass even more challenging.

Nevertheless, galaxies at very high redshift were found which appear to have high stellar masses, including a LAE at $z = 6.6$ with an estimated stellar mass $M_* \gtrsim 10^{11} M_\odot$. The high-redshift QSOs require a SMBH

Q15) Campbell Cosmo Q15

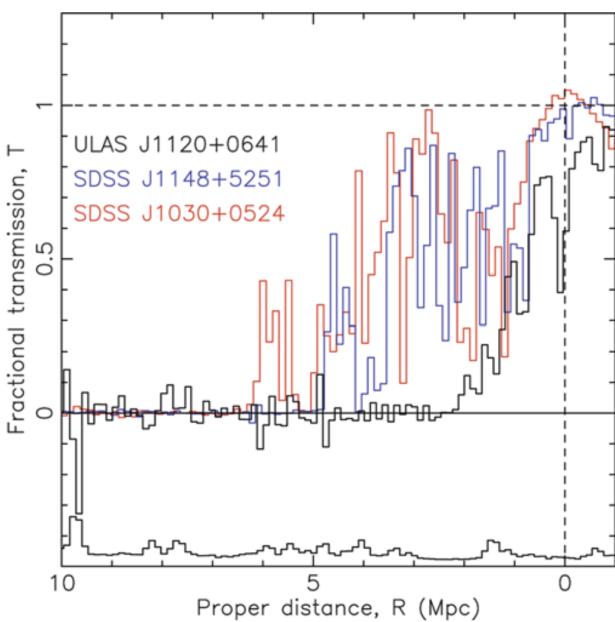


Figure 33: The spectra of three high-redshift QSOs (SDSSJ1148+5251 at $z = 6.42$, SDSS J1030+0524 at $z = 6.31$, and the $z = 7.085$ QSO ULAS J1120+0641) at the Lyman α emission line. For this figure, the wavelength difference to the Lyman α transition is expressed in proper distance away from the QSOs. The spectra are normalized, dividing them by the extrapolation of the continuum on the red side of the emission line, yielding the transmission. Source: D.J. Mortlock et al. 2011, A luminous quasar at a redshift of $z = 7.085$, Nature 474, 616, Fig. 3. Figure taken from Ryden (2006).

with $M \gtrsim 10^9 M_\odot$ to power their energy output, and these must be hosted in galaxies with very large stellar masses. Therefore, massive galaxies have formed very early on, delivering ionizing photons. However, these highest mass objects are very rare and, by themselves, by far not able to explain reionization. This fact can be clearly seen by considering the spectral shape of the Ly α emission line of high-redshift QSOs. Figure 32 describes qualitatively the result of observing the emission towards bright background QSOs to study (re)ionization. Figure 33 shows the spectrum of three very high redshift QSOs near to the Lyman α emission line. Whereas all three QSO show essentially no flux blueward of Lyman α , once the wavelength difference exceeds ~ 20 Å in the restframe, there is some transmitted flux very close to the Ly α transition. This near-zone transmission is understood as a region around the QSO where the intergalactic gas is fully ionized by the QSO, so it becomes transparent. The figure shows a clear trend that the size of this near zone decreases for higher redshifts, as would be expected due to the higher gas density and probably larger mean neutral fraction in the Universe. Thus, these very luminous objects are able to reionize the IGM in their immediate surroundings, but their effect is constrained to a rather limited volume. Most of the ionizing photons must come from the far more numerous lower-mass galaxies (i.e., far less luminous sources).

Gunn & Peterson (1965) first proposed using Ly α resonance absorption in the spectrum of distant quasars as a direct probe to the neutral hydrogen density in the IGM at high redshift. For objects beyond reionization, neutral hydrogen in the IGM creates complete GP absorption troughs in the quasar spectrum blueward of Ly α emission. Observations of the GP effect directly constrain the evolution of neutral hydrogen fraction and the ionization state of the IGM.

The Gunn-Peterson optical depth to Ly α photons is

$$\tau_{\text{GP}} = \frac{\pi e^2}{m_e c^2} f_\alpha \lambda_\alpha H^{-1}(z) n_{\text{HI}} \text{ [dimensionless]},$$

where f_α is the oscillator strength of the Ly α transition, $\lambda_\alpha = 1216$ Å, $H(z)$ is the Hubble constant at redshift z , and n_{HI} is the density of neutral hydrogen in the IGM.

At high redshifts,

$$\tau_{\text{GP}}(z) = 4.9 \times 10^5 \left(\frac{\Omega_m h^2}{0.13} \right)^{-1/2} \left(\frac{\Omega_b h^2}{0.02} \right) \left(\frac{1+z}{7} \right)^{3/2} \left(\frac{n_{\text{HI}}}{n_{\text{H}}} \right) \text{ [dimensionless]}$$

for a uniform IGM. Even a tiny neutral fraction, $X_{\text{HI}} \sim 10^{-4}$, gives rise to complete GP absorption. This test is only sensitive at the end of the reionization when the IGM is already mostly ionized, and the absorption saturates for the higher neutral fraction in the earlier stage.

At $z < 5$, the IGM absorption is resolved into individual Ly α forest lines; their number density increases strongly with redshift: $N_\alpha(z) \propto (1+z)$. Earlier attempts to study GP absorption concentrated on measuring the amount of flux between individual Ly α forest lines using high-resolution spectroscopy to place limits on the diffuse neutral IGM. However, at $z > 5$, even with a moderately high resolution

Q15) Campbell Cosmo Q15

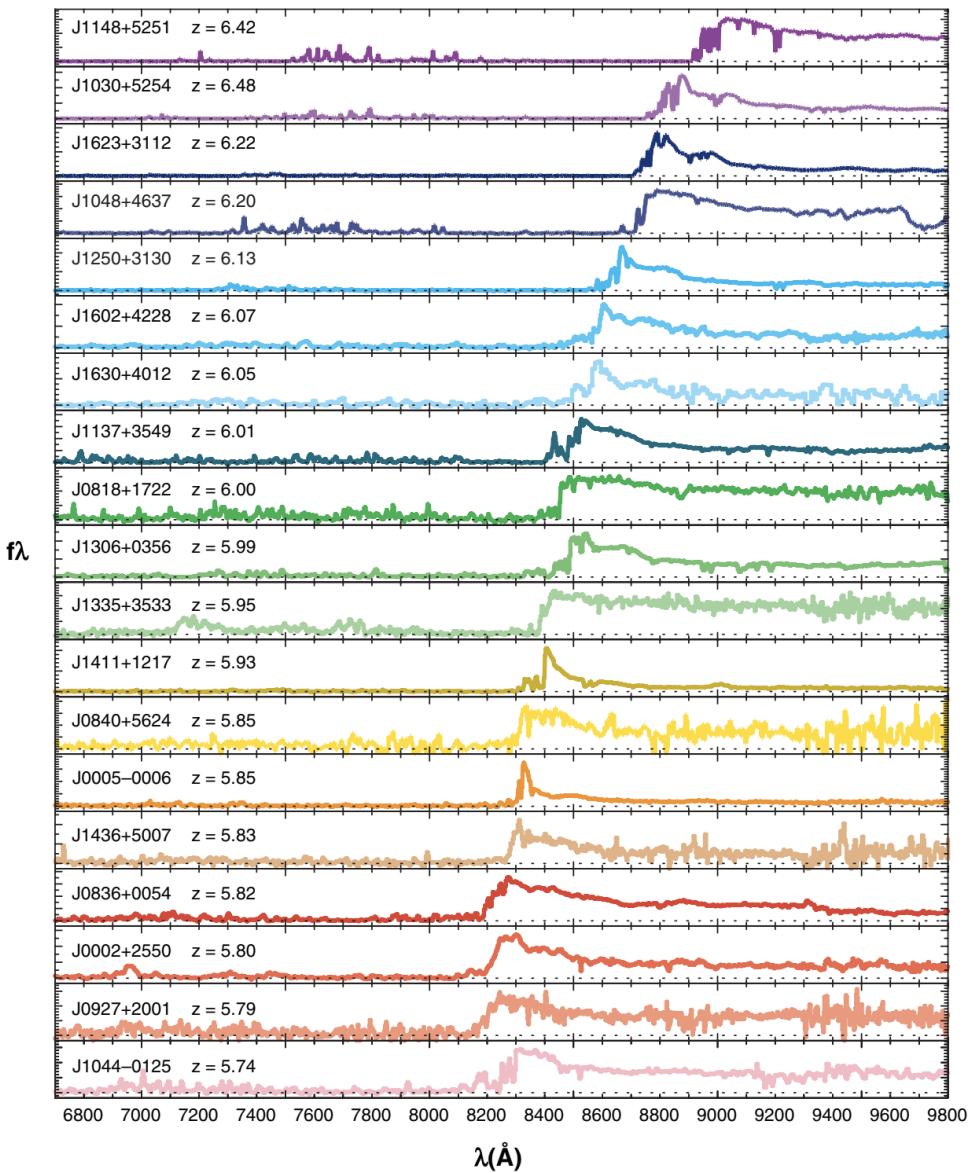


Figure 34: Moderate resolution spectra of nineteen SDSS quasars at $5.74 < z < 6.42$. Adapted from Fan et al. (2006b). Image taken from Ryden (2006).

spectrum, Ly α forest lines overlap severely, making it impossible to find a truly “line-free” region. A more accurate picture of the IGM evolution interprets the Ly α forest as a fluctuating GP effect: absorption arises from low-density gas in the IGM that is in approximate thermal equilibrium between photoionization heating by the UV background and adiabatic cooling due to the Hubble expansion, rather than as discrete Ly α forest clouds. The neutral hydrogen fraction and therefore the GP optical depth depend on the local density of the IGM. By studying the evolution of the average transmitted flux or effective optical depth, one can trace the evolution of the UV ionizing background and neutral fraction of the IGM. At high-redshift, the IGM is highly clumpy, which must be taken into account in order to estimate the IGM ionization from observations.

The SDSS provides large samples of luminous quasars over $0 < z < 6.5$. Fan et al. (2000, 2001a,b, 2003, 2004, 2006a) carried out a survey of i-dropout quasars ($z > 5.7$) using the SDSS imaging data, resulting in the discovery of 19 luminous quasars in this redshift regime (Figure 34). Other multicolour survey projects are also searching for quasars at similar redshifts. They provide by far the best probes of IGM evolution toward the end of the EoR.

The observed spectrum of high-redshift QSOs bluewards of the Ly α emission line shows that an increasing fraction of the radiation is absorbed by neutral hydrogen on the line-of-sight. We have seen that the density of the Ly α forest increases with redshift (see Figures 28 and 34) in such a way that only a tiny

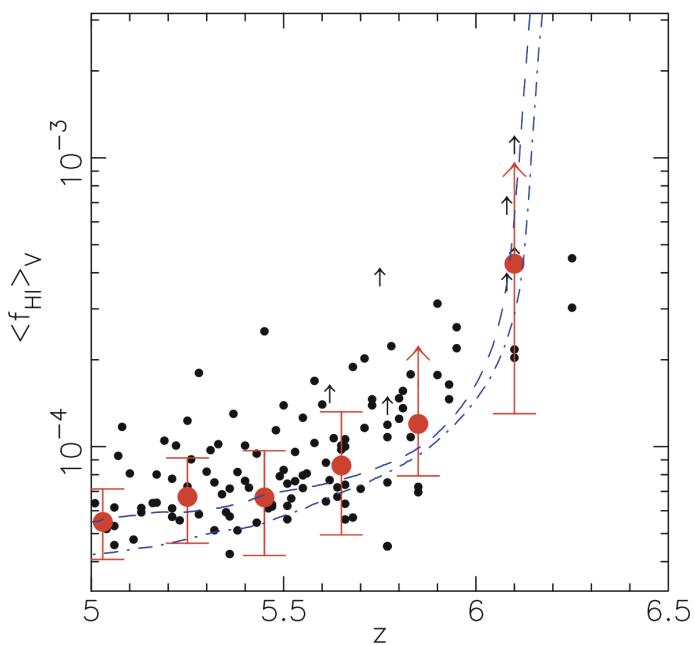


Figure 35: Redshift evolution of the mean neutral fraction of hydrogen in the IGM, as obtained from the absorption of ionizing radiation from high-redshift QSOs (Gunn-Peterson effect). Individual measurements are shown as small dots, whereas the large circles with error bars represent averages over redshift bins. The two curves show results from numerical simulations. Source: X. Fan et al. 2006, Constraining the Evolution of the Ionizing Background and the Epoch of Reionization with $z \sim 6$ Quasars. II. A Sample of 19 Quasars, AJ 132, 117, p. 126, Fig. 7. Figure taken from Ryden (2006).

fraction of ionizing photons manage to escape absorption. This observation may be seen as an indication that we approach the EoR as the QSO redshift increases beyond $z \sim 6$. However, as shown in Figure 35, the mean neutral fraction of intergalactic hydrogen needed to cause this strong absorption of ionizing photons is still very small – a neutral fraction of much less than 1% is sufficient to entirely block the light of QSOs blueward of the Ly α emission. Hence, the strong absorption implied by QSO spectra cannot be taken as evidence for $z \sim 6$ signalling the end of the EoR. Nevertheless, the trend of the data shown in Figure 35 may suggest that beyond $z \sim 6$, we may approach a phase where the neutral hydrogen fraction indeed starts to increase significantly.

Observing reionization directly may in principle be possible if a very high-redshift QSO could be identified whose absorption spectrum could reveal a tomographic view through the ionized ‘bubbles’ of the IGM, as sketched in Figure 32. But we point out again that the very dense Ly α forest seen towards QSOs at high redshift, is no unambiguous sign for approaching the redshift of reionization, because a very small fraction of neutral atoms (about 1%) is already sufficient to produce a large optical depth for Ly α photons.

1.16.3 Follow-up Questions

- Besides the first stars and galaxies, what else could have ionized the Universe?
- How does the recent paper by Bowman et al. relate to this?
- What about Helium ionization?

Cosmology Class Fall 2019: Martine's Notes

Define and describe:

Today, the IGM is mainly ionized. The ISM is mostly ionized. Two main constraints on reionization: CMB spectrum and the spectra of quasars.

Spectra of quasars

If you can find quasars with the Gunn-Peterson trough, then you know there was a period of time where the universe was neutral. What this would look like: Lyman alpha peak, then at redder frequencies an extended trough, then the Lyman-alpha forest at redder frequencies from clumps of neutral hydrogen.

From this evidence, reionization happens at $z \approx 6$. Galaxies at higher redshifts have been seen with this trough.

CMB evidence

CMB scatters off of ionized universe. Different electrons which scatter CMB photons see different quadropoles and thus the scattered light is polarized. The variations in the polarization are largest at large scales because far-apart electrons see the most different version of the CMB. Nearby electrons see similar CMB quadropole. There is a peak at low ℓ in the $(\ell)(\ell+1)C_{ell}^{EE}$ power spectrum. This will be higher in power if reionization happened earlier, because the universe was denser at that point so there was more scattering and thus more polarization. It will also be at lower ell if reionization happened earlier, because the distance between an electron which sees one CMB quadropole and another electron which sees an entirely different CMB quadropole is smaller at earlier times (I think; don't quote me on that last bit).

Q16) 21 CM HYDROGEN

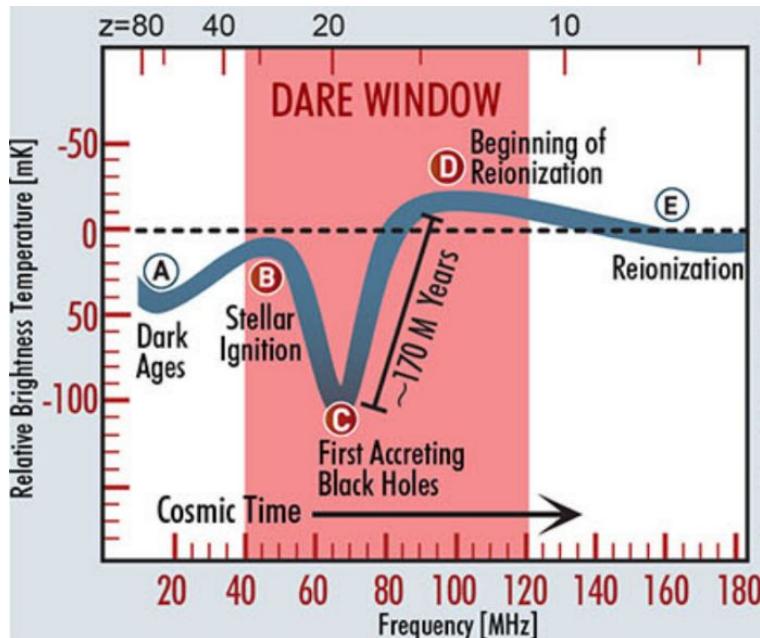
The 21 cm line of hydrogen is expected to show up in absorption against the cosmic microwave background at some redshifts, and in emission at other redshifts. What physical processes lead to this behaviour?

Short Summary

Q16) Ludwig Cosmo Q15

Question 15 - 21 cm Line

The 21 cm line of hydrogen is expected to show up in absorption against the cosmic microwave background at some redshifts, and in emission at other redshifts. What physical processes lead to this behaviour?



21 cm spin flip can be a spontaneous emission or induced through collisions.

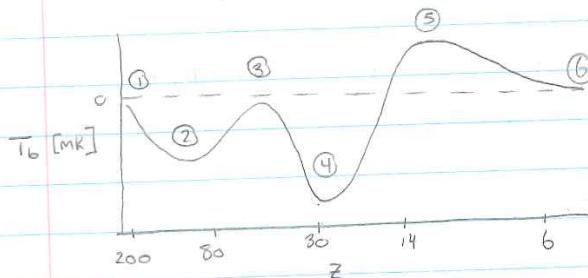
0. Prior to Dark Ages, Redshift $z \sim 200$ - Post Recombination
 - The intergalactic medium is still dense enough to couple photons to baryons via collisions so we don't see any significant absorption or deviation from the CMB temperature. No 21 cm signal.
- A. Dark Ages, $z \sim 80$, -
 - As the universe expands, collisions can't keep baryons and photons coupled anymore and the temperature drops so neutral hydrogen with electrons in a parallel configuration will seek the lower energy state with electrons in anti parallel configurations. When one of the electrons flips its spin it emits a 21 cm line which can be detected even when everything else around it is dark.
- B. Stellar Ignition, $z \sim 30$
 - The first stars and dwarf galaxies form in the universe and their collective luminosity puts an end to the cosmological dark ages.
 - They excite hydrogen to produce the Lyman series photons
- C. First Accreting Black Holes, $z \sim 20$
 - Galaxies begin to emit X-rays which heat gas the 21 cm line counteracts the Lyman gun Peterson trough.
- D. Beginning of Reionization, $z \sim 15$
 - Stars produce UV radiation which ionizes the ISM. No 21 cm emission is seen in bubbles around these stars.
- E. Reionization, $z \sim 6$
 - No more 21 cm line

Q16) Herman Cosmo Q16

C16

- collisional coupling that can't be maintained as universe expands, so get absorption
- galaxies form, Ly- α coupling, but then more absorption induced by Ly- α interactions
- IGM is heated by x-rays, produces emission
- Reionization decreases HI, so less emission

21 cm absorption and emission against CMB



↑ Emission of 21 cm

↓ Absorption of 21 cm

$T_b \approx \text{Measure of radio intensity}$

$$I_{\nu} = 2kT_b v^2/c^2$$

Spin-flip of e^- in ground state of neutral H.

$p^+ + e^- \xrightarrow{\text{mag. mom}} \text{Aligned } \uparrow\uparrow \downarrow\downarrow$
IS \downarrow 21cm \downarrow 0.068K
Antialigned $\uparrow\downarrow$

- ① Post recombination, the IGM is still dense enough to couple baryons and photons via collisional coupling, so we don't see any significant absorption or deviation from the CMB temperature. No 21 cm signal.
- ② As the universe expands, the gas cools adiabatically, and the excitation T of the 21 cm H transition decouples from T_{gas} . The Dark ages take over. The diffusion of the IGM by expansion means collisions can't keep baryons + photons coupled, so HI absorbs & at 21 cm. I think...
- ③ The first galaxies form, and they produce Ly- α photons that couple the excitation T of 21 cm spin states to T_{gas} by scattering. This results in strong absorption, with spatial variance since galaxies aren't evenly distributed.
- ④ X-ray heating of the IGM begins as galaxies emit x-ray emission which heats gas, producing a 21 cm emission signal that counteracts the Ly- α absorption trough
- ⑤ UV photoionization begins as first stars produce ionizing UV photons. The 21 cm signal develops spatial holes where no 21 cm emission is seen in regions of ionized bubbles surrounding groups of galaxies, until the entire universe is ionized by $z \approx 6$.

Q16) Campbell Cosmo Q16

1.17 Question 16

The 21 cm line of hydrogen is expected to show up in absorption against the cosmic microwave background at some redshifts, and in emission at other redshifts. What physical processes lead to this behaviour?

1.17.1 Short answer

The key to the detectability of the 21 cm signal hinges on the spin temperature T_s . Only if this temperature deviates from the background temperature, will a signal be observable.

- $200 \lesssim z \lesssim 1100$: The residual free electron fraction left after recombination allows Compton scattering to maintain thermal coupling of the gas to the CMB, setting $T_k = T_\gamma$. The high gas density leads to effective collisional coupling so that $T_s = T_\gamma$ and we expect $\bar{T}_b = 0$ and no detectable 21 cm signal.
- $40 \lesssim z \lesssim 200$: In this regime, the gas cools adiabatically so that $T_k \propto (1+z)^2$ leading to $T_k < T_\gamma$ and collisional coupling sets $T_s < T_\gamma$, leading to $\bar{T}_b < 0$ and an early absorption signal. At this time, T_b fluctuations are sourced by density fluctuations, potentially allowing the initial conditions to be probed.
- $z_* \lesssim z \lesssim 40$: As the expansion continues, decreasing the gas density, collisional coupling becomes ineffective and radiative coupling to the CMB sets $T_s = T_\gamma$, and there is no detectable 21 cm signal.
- $z_\alpha \lesssim z \lesssim z_*$: Once the first sources switch on at z_* , they emit both Ly α photons and X-rays. In general, the emissivity required for Ly α coupling is significantly less than that for heating T_k above T_γ . We therefore expect a regime where the spin temperature is coupled to cold gas so that $T_s \sim T_k < T_\gamma$ and there is an absorption signal. Fluctuations are dominated by density fluctuations and variation in the Ly α flux. As further star formation occurs the Ly α coupling will eventually saturate ($x_\alpha \gg 1$), so that by a redshift z_α the gas will everywhere be strongly coupled.
- $z_h \lesssim z \lesssim z_\alpha$: After Ly α coupling saturates, fluctuations in the Ly α flux no longer affect the 21 cm signal. By this point, heating becomes significant and gas temperature fluctuations source T_b fluctuations. While T_k remains below T_γ we see a 21 cm signal in absorption, but as T_k approaches T_γ hotter regions may begin to be seen in emission. Eventually by a redshift z_h the gas will be heated everywhere so that $\bar{T}_k = T_\gamma$.
- $z_T \lesssim z \lesssim z_h$: After the heating transition, $T_k > T_\gamma$ and we expect to see a 21 cm signal in emission. The 21 cm brightness temperature is not yet saturated, which occurs at z_T , when $T_s \sim T_k \gg T_\gamma$. By this time, the ionization fraction has likely risen above the percent level. Brightness temperature fluctuations are sourced by a mixture of fluctuations in ionization, density and gas temperature.
- $z_r \lesssim z \lesssim z_T$: Continued heating drives $T_k \gg T_\gamma$ at z_T and temperature fluctuations become unimportant. $T_s \sim T_k \gg T_\gamma$ and the dependence on T_s may be neglected which greatly simplifies analysis of the 21 cm power spectrum. By this point, the filling fraction of HII regions probably becomes significant and ionization fluctuations begin to dominate the 21 cm signal.
- $z \lesssim z_r$: After reionization, any remaining 21 cm signal originates primarily from collapsed islands of neutral hydrogen (damped Ly α systems).

These events are summarized in Figure 36.

1.17.2 Additional context

As the most common atomic species present in the Universe, hydrogen is a useful tracer of local properties of the gas. The simplicity of its structure (a proton and electron) belies the richness of the associated physics. The 21 cm line of hydrogen arises from the hyperfine splitting of the $1S$ ground state due to the interaction of the magnetic moments of the proton and the electron. This splitting leads to two distinct energy levels separated by $\Delta E_{21\text{ cm}} = 5.9 \times 10^{-6}$ eV, corresponding to a wavelength of 21 cm and a frequency of 1420 MHz. This frequency is one of the most precisely known quantities in astrophysics having been measured to great accuracy from studies of hydrogen masers.

Radio telescopes look for emission by warm hydrogen gas within galaxies. Since the line is narrow with a well measured rest frame frequency it can be used in the local Universe as a probe of the velocity distribution of gas within our galaxy and other nearby galaxies. 21 cm rotation curves are often used to trace galactic dynamics. Traditional techniques for observing 21 cm emission have only detected the line in relatively local galaxies, although it has been seen in absorption against radio loud background sources

Q16) Campbell Cosmo Q16

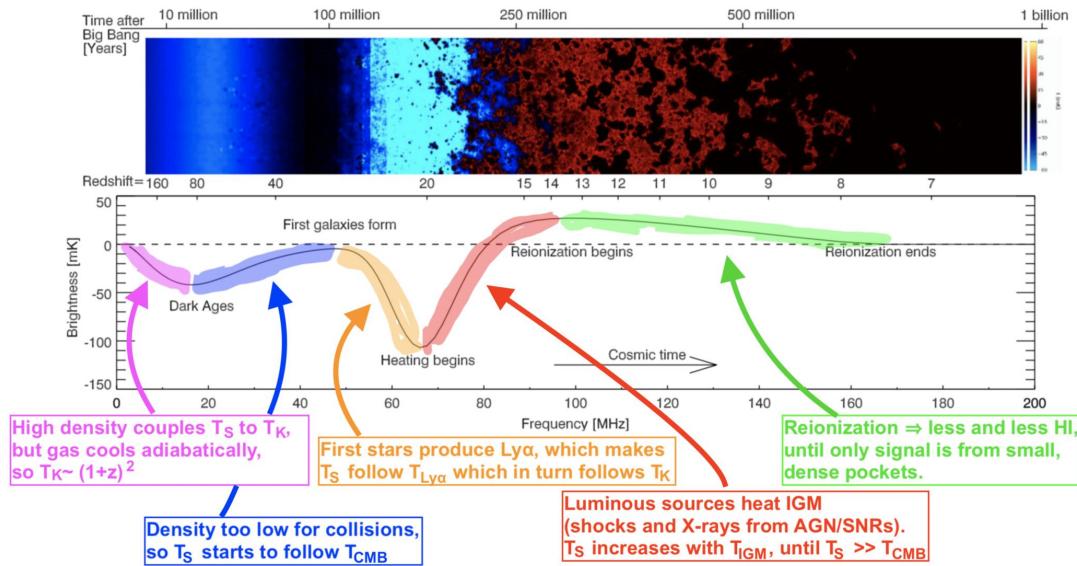


Figure 36: The 21-centimeter cosmic hydrogen signal. (a) Time evolution of fluctuations in the 21 cm brightness from just before the first stars formed through to the end of the EoR. This evolution is pieced together from redshift slices through a simulated cosmic volume. Colouration indicates the strength of the 21 cm brightness as it evolves through two absorption phases (purple and blue), separated by a period (black) where the excitation temperature of the 21 cm hydrogen transition decouples from the temperature of the hydrogen gas, before it transitions to emission (red) and finally disappears (black) owing to the ionization of the hydrogen gas. (b) Expected evolution of the sky-averaged 21 cm brightness from the “dark ages” at redshift 200 to the end of reionization, sometime before redshift 6 (solid curve indicates the signal; dashed curve indicates $T_b = 0$). The frequency structure within this redshift range is driven by several physical processes, including the formation of the first galaxies and the heating and ionization of the hydrogen gas. There is considerable uncertainty in the exact form of this signal, arising from the unknown properties of the first galaxies. Figure taken from Ryden (2006) with annotations by Matt Young.

from individual systems at redshifts $z \lesssim 3$. A new generation of radio telescopes offers the exciting prospect of using the 21 cm line as a probe of cosmology.

In cosmological contexts the 21 cm line has been used as a probe of gas along the line of sight to some background radio source. The detailed signal depends upon the radiative transfer through gas along the line of sight. We recall the basic equation of radiative transfer (RT) for the specific intensity I_ν (per unit frequency ν) in the absence of scattering along a path described by coordinate s

$$\frac{dI_\nu}{ds} = j_\nu - \alpha_\nu I_\nu \quad [\text{erg s}^{-1} \text{cm}^{-2} \text{Hz}^{-1} \text{sr}^{-1} \text{pc}^{-1}],$$

where absorption and emission by gas along the path are described by the coefficients α_ν and j_ν , respectively.

To simplify the discussion, we will work in the Rayleigh-Jeans (RJ) limit, appropriate here since the relevant photon frequencies ν are much smaller than the peak frequency of the CMB blackbody. This allows us to relate the intensity I_ν to a brightness temperature T_b by the relation $I_\nu = 2k_B T_b \nu^2 / c^2$, where c is the speed of light and k_B is Boltzmann’s constant. We will also make use of the standard definition of the optical depth $\tau = \int \alpha_\nu ds$. With this we may rewrite the equation of RT to give the radiative transfer for light from a background radio source of brightness temperature T_r (primarily the CMB) along the line of sight through a cloud of optical depth τ_ν and uniform excitation temperature T_{ex} so that the observed brightness temperature T_b^{obs} at a frequency ν is given by

$$T_b^{\text{obs}} = T_{ex}(1 - e^{-\tau_\nu}) + T_r(\nu)e^{-\tau_\nu} \quad [\text{K}].$$

Figure 36 provides a summary of the 21 cm signal showing the key features of the signal with the relevant cosmic time, frequency, and redshift scales indicated. The earliest period of the signal arises in the period after thermal decoupling of the ordinary matter (baryons) from the CMB, so that the gas is able to cool adiabatically with the expansion of the Universe. In these cosmic “Dark Ages”, before the first stars have formed, the first structures begin to grow from the seed inhomogeneities thought to be produced by quantum fluctuations during inflation. The cold gas can be seen in a 21 cm absorption signal, which has both a mean value (shown in the bottom panel) and fluctuations arising from variation in density (shown in the top panel). Once the first stars and galaxies form, their light radically alters the properties of the gas. Scattering of Ly α photons leads to a strong coupling between the excitation of the 21 cm line

Q16) Campbell Cosmo Q16

spin states and the gas temperature. Initially, this leads to a strong absorption signal that is spatially varying due to the strong clustering of the rare first generation of galaxies. Next, the X-ray emission from these galaxies heats the gas leading to a 21 cm emission signal. Finally, ultraviolet photons ionize the gas producing dark holes in the 21 cm signal within regions of ionized bubbles surrounding groups of galaxies. Eventually all of the hydrogen gas, except for that in a few dense pockets, is ionized.

The excitation temperature of the 21 cm line is known as the spin temperature T_s . It is defined through the ratio between the number densities n_i of hydrogen atoms in the two hyperfine levels (which we label with a subscript 0 and 1 for the $1S$ singlet and $1S$ triplet levels, respectively)

$$\frac{n_1}{n_2} = \frac{g_1}{g_2} \exp\left(-\frac{T_*}{T_s}\right) \text{ [dimensionless]},$$

where $(g_1/g_0) = 3$ is the ratio of the statistical degeneracy factors of the two levels, and $T_* \equiv hc/k_B\lambda_{21\text{ cm}} = 0.068\text{ K}$.

With this definition, the optical depth of a cloud of hydrogen is then

$$\tau_\nu = \int \left[1 - \exp\left(-\frac{E_{10}}{k_B T_s}\right) \right] \sigma_0 \phi(\nu) n_0 ds \text{ [dimensionless]},$$

where $n_0 = n_{\text{H}}/4$ with n_{H} being the hydrogen density, and we have denoted the 21 cm cross-section as $\sigma(\nu) = \sigma_0 \phi(\nu)$, with $\sigma_0 \equiv 3c^2 2A_{10}/8\pi\nu^2$, where $A_{10} = 2.85 \times 10^{-15}\text{ s}^{-1}$ is the spontaneous decay rate of the spin-flip transition, and the line profile is normalized so that $\int \phi(\nu)d\nu = 1$. To evaluate this expression we need to find the column length as a function of frequency $s(\nu)$ to determine the range of frequencies $d\nu$ over the path ds that correspond to a fixed observed frequency ν_{obs} . This can be done in one of two ways: by relating the path length to the cosmological expansion $ds = cdz/(1+z)H(z)$ and the redshifting of light to relate the observed and emitted frequencies $\nu_{\text{obs}} = \nu_{\text{em}}/(1+z)$ or assuming a linear velocity profile locally $v = (d\nu/ds)s$ (the well known Sobolev approximation) and using the Doppler law $\nu_{\text{obs}} = \nu_{\text{em}}(1-v/c)$ self-consistently to $\mathcal{O}(v/c)$. Since the latter case describes the well known Hubble law in the absence of peculiar velocities these two approaches give identical results for the optical depth. The latter picture brings out the effect of peculiar velocities that modify the local velocity-frequency conversion.

The optical depth of this transition is small at all relevant redshifts, yielding a differential brightness temperature

$$\begin{aligned} \delta T_b &= \frac{T_s - T_r}{1+z} (1 - e^{-\tau_\nu}) \approx \frac{T_s - T_r}{1+z} \tau \\ &\approx 27x_{\text{HI}}(1+f_b) \left(\frac{\Omega_b h^2}{0.023} \right) \left(\frac{0.15}{\Omega_m h^2} \frac{1+z}{10} \right)^{1/2} \left(\frac{T_s - T_r}{T_s} \right) \left(\frac{\partial_r v_r}{(1+z)H(z)} \right) [\text{mK}], \end{aligned}$$

Here x_{HI} is the neutral fraction of hydrogen, f_b is the fractional overdensity in baryons, and the final term arises from the velocity gradient along the line of sight $\partial_r v_r$.

The key to the detectability of the 21 cm signal hinges on the spin temperature. Only if this temperature deviates from the background temperature, will a signal be observable; the spin temperature and spatial variation in the spin temperature can be used to convey information about astrophysical sources.

Three processes determine the spin temperature: (i) absorption/emission of 21 cm photons from/to the radio background, primarily the CMB; (ii) collisions with other hydrogen atoms and with electrons; and (iii) resonant scattering of Ly α photons that cause a spin flip via an intermediate excited state. The rate of these processes is fast compared to the de-excitation time of the line, so that to a very good approximation the spin temperature is given by the equilibrium balance of these effects. In this limit, the spin temperature is given by

$$T_s^{-1} = \frac{T_\gamma^{-1} + x_\alpha T_\alpha^{-1} + x_c T_K^{-1}}{1 + x_\alpha + x_c} [\text{K}^{-1}],$$

where T_γ is the temperature of the surrounding bath of radio photons, typically set by the CMB so that $T_\gamma = T_{\text{CMB}}$; T_α is the colour temperature of the Ly α radiation field at the Ly α frequency and is closely coupled to the gas kinetic temperature T_k by recoil during repeated scattering, and x_c , x_α are the coupling coefficients due to atomic collisions and scattering of Ly α photons, respectively. The spin temperature becomes strongly coupled to the gas temperature when $x_{\text{tot}} \equiv x_c + x_\alpha \gtrsim 1$ and relaxes to T_γ when $x_{\text{tot}} \ll 1$.

Two types of background radio sources are important for the 21 cm line as a probe of astrophysics. Firstly, we may use the CMB as a radio background source. In this case, $T_r = T_{\text{CMB}}$ and the 21 cm feature is seen as a spectral distortion to the CMB blackbody at appropriate radio frequencies (since fluctuations in the CMB temperature are small $\Delta T_{\text{CMB}} \sim 10^{-5}$ the CMB is effectively a source of uniform brightness).

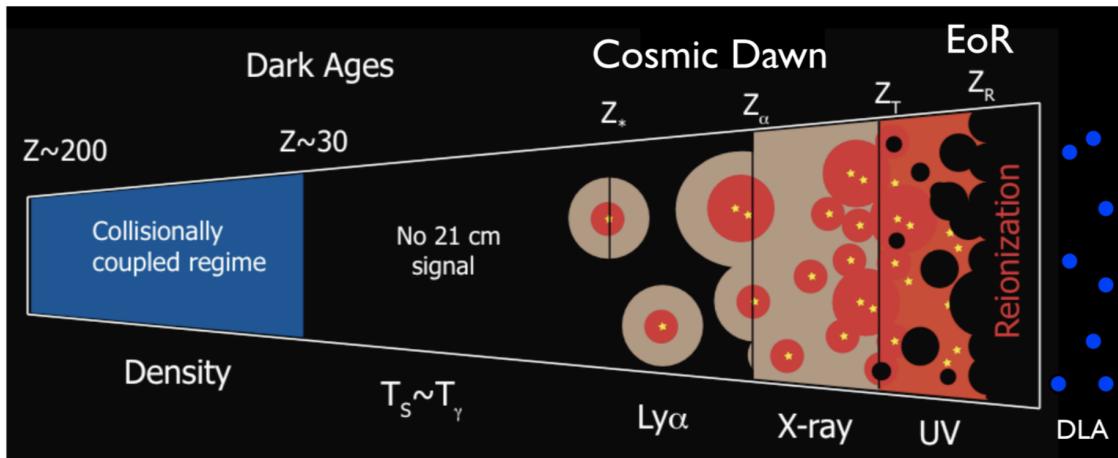


Figure 37: Cartoon of the different phases of the 21 cm signal. The signal transitions from an early phase of collisional coupling to a later phase of Ly α coupling through a short period where there is little signal. Fluctuations after this phase are dominated successively by spatial variation in the Ly α , X-ray, and ionizing UV radiation backgrounds. After reionization is complete there is a residual signal from neutral hydrogen in galaxies. Figure taken from Ryden (2006).

The distortion forms a diffuse background that can be studied across the whole sky in a similar way to CMB anisotropies. Observations at different frequencies probe different spherical shells of the observable Universe, so that a 3D map can be constructed.

The second situation uses a radio loud point source, for example a radio loud quasar, as the background. In this case, the source will always be much brighter than the weak emission from diffuse hydrogen gas, $T_r \gg T_s$, so that the gas is seen in absorption against the source. The appearance of lines from regions of neutral gas at different distances to the source leads to a “forest” of lines known as the “21 cm forest” in analogy to the Ly α forest. The high brightness of the background source allows the 21 cm forest to be studied with high frequency resolution so probing small scale structures (\sim kpc) in the IGM. For useful statistics, many lines of sight to different radio sources are required, making the discovery of high redshift radio sources a priority.

Collisions between different particles may induce spin-flips in a hydrogen atom and dominate the coupling in the early Universe where the gas density is high. Three main channels are available: collisions between two hydrogen atoms and collisions between a hydrogen atom and an electron or a proton. The collisional coupling for a species is

$$x_c^i \equiv \frac{C_{10}}{A_{10}} \frac{T_*}{T_\gamma} = \frac{n_i k_{10}^i}{A_{10}} \frac{T_*}{T_\gamma},$$

where C_{10} is the collisional excitation rate, k_{10}^i is the specific rate coefficient for spin de-excitation by collisions with species i (in units of $\text{cm}^3 \text{s}^{-1}$).

The total collisional coupling coefficient can be written as

$$\begin{aligned} x_c &= x_c^{\text{HH}} + x_c^{\text{eH}} + x_c^{\text{pH}} \\ &= \frac{T_*}{A_{10} T_\gamma} \left[k_{1-0}^{\text{HH}}(T_k) n_{\text{H}} + k_{1-0}^{\text{eH}}(T_k) n_e + k_{1-0}^{\text{pH}}(T_k) n_p \right] \text{ [dimensionless]}, \end{aligned}$$

where k_{1-0}^{HH} is the scattering rate between hydrogen atoms, k_{1-0}^{eH} is the scattering rate between hydrogen atoms and electrons, and k_{1-0}^{pH} is the scattering rate between hydrogen atoms and protons.

We may express the 21 cm brightness temperature as a function of four variables $T_b = T_b(T_k, x_i, J_\alpha, n_{\text{H}})$, where x_i is the volume-averaged ionized fraction of hydrogen and J_α is the Ly α specific flux. In calculating the 21 cm signal, we require a model for the global evolution of and fluctuations in these quantities. Before looking at the evolution of the signal quantitatively, we will first outline the basic picture to delineate the most important phases.

An important feature of T_b is that its dependence on each of these quantities saturates at some point, for example once the Ly α flux is high enough the spin and kinetic gas temperatures become tightly coupled and further variation in J_α becomes irrelevant to the details of the signal. This leads to conceptually separate regimes where variation in only one of the variables dominating fluctuations in the signal. These different regimes can be seen in Figure 36 and are shown in schematic form in Figure 37.

Q17) COSMIC NEUTRINO BACKGROUND

What are the similarities and differences between the cosmic neutrino background and the cosmic microwave background?

Q17) Ludwig Cosmo Q17

Question 17 - Cosmic Neutrino Background

What are the similarities and differences between the cosmic neutrino background and the cosmic microwave background?

The CNB

- Produced when neutrinos decoupled from the photon-baryon fluid at $t \sim 1$ second.
- The expansion of the universe exceeded the reaction rate of neutrino interactions.
- The temperature dependent cross section became too small as the universe cooled
- Neutrinos then streamed freely through the universe.
- Peak temperature of 1.9 K
- Very difficult to observe neutrinos from our own sun, much less 14 billion light years away.
- We look for indirect evidence of it.
 - Has an effect on the small scale part of the C_l spectrum.
 - Neutrinos diffusing outward damps the structure at small angles, just like Silk damping.
 - Element abundance.

	CNB	CMB
Particle	Neutrinos	Photons
Decoupling of	Neutrinos and Photon-Baryon Fluid	Photons and Baryons
Time	1 second	380,000 years
Temp Then	10^{10} K	3000 K
Temp Now	1.9 K	2.73 K

If the universe cools uniformly for all its constituents why is $T_{cnb} < T_{cmb}$?

- Electron-Positron annihilation.
 - This event produced an in flux of photons with more energy, giving the overall CMB progenitor photon distribution a slightly higher temperature than the neutrinos.
 - This annihilation period occurred around $t = 6s$ when it became more energetically favorable to produce photons rather than electrons.

Q17) Herman Cosmo Q18

C18

Cosmic Neutrino Background

- Neutrinos decoupled b/c universe's T decreased below interaction cross section; γ 's decoupled b/c no more free e^\pm 's for Thomson scattering
- Both have Planck distribution(?), were in T.E. before
- Different T, t, particle type

The CNB was produced when neutrinos decoupled from the photon-baryon fluid at $t \approx 1s$. This happened because the expansion rate of the universe exceeded the reaction rate of neutrino interactions, or in other words, the temperature-dependent interaction cross section became too small as the universe cooled. Neutrinos were then able to stream freely through the universe, and by now they would have a peak temperature of $\sim 1.9\text{ K}$. It's incredibly difficult to detect energetic neutrinos from our own sun, so detecting the CNB seems impossible. Instead, we can look for indirect evidence of it, such as its effect on the small scale part of the CMB spectrum (neutrinos diffusing outward damps the structure at small angular scales, like Silk damping), or the impact the neutrinos had on element abundance due to their part in neutron production/decay, I think...

	CNB	CMB
Particle	Neutrinos	Photons
Decoupling of	Neutrinos + photon-baryon fluid	Photons + baryons
Time	1s	380,000 yr
Temp. then	10^{10} K	3000 K
Temp. now	1.9 K	2.73 K

If the universe cools uniformly for all its constituents, why is $T_{\text{obs}} < T_{\text{CMB}}$? Electron-positron annihilation. This event produced an influx of photons w/ more energy, giving the overall CMB progenitor photon distribution a slightly higher temperature than the neutrinos. This annihilation period occurred around $t \approx 6s$, when it became more energetically favorable to produce γ 's rather than e^\pm 's.

Q17) Campbell Cosmo Q18

1.19 Question 18

What are the similarities and differences between the cosmic neutrino background and the cosmic microwave background?

1.19.1 Short answer

- The CMB is from $\sim 380,000$ years after the BB (or a redshift of $z \sim 1100$), while the CNB is from ~ 1 second after the BB.
- The CMB occurred when the Universe was $\sim 1,000$ K (or ~ 0.1 eV) while that of the CNB was $\sim 10^{10}$ K (or ~ 1 MeV).
- The expansion of the Universe has cooled the temperature of the CMB to ~ 2.7 K and the CNB to ~ 1.95 K.

1.19.2 Additional context

Just as the CMB is a relic of the time when the Universe was opaque to photons, the CNB is a relic of the time when the Universe was hot and dense enough to be opaque to neutrinos. The number density of each of the three flavors of neutrinos (ν_e , ν_μ , and ν_τ) has been calculated to be 3/11 times the number density of CMB photons. This means that at any moment, about twenty million cosmic neutrinos are zipping through your body.

The CMB yields information about our Universe at around 380,000 years after the BB. Due to the weak interaction of the neutrinos with matter the CNB should give information about a much earlier time of our Universe, around one second after the Big Bang.

The CNB (often also called ‘relic neutrinos’) decouples from matter about 1 second after the BB in the radiation dominated era at a temperature of 10^{10} K $\equiv 1$ MeV. Today due to the expansion and cooling of the Universe the relic neutrino temperature is 1.95 K and the average neutrino density in the Universe is 340 cm^{-3} or $56 \nu_e \text{ cm}^{-3}$.

The existence of a CNB (the analogue of the CMB) is a fundamental prediction of standard big bang cosmology. Up to now, the observational evidence for its existence is rather indirect and rests entirely on cosmological observations of, e.g., the light elemental abundances, the anisotropies in the CMB, and the large scale distribution of matter. All these measurements, however, probe the presence of relic neutrinos only at early stages in the cosmological evolution and, moreover, in a rather indirect way. The most promising laboratory search, based on neutrino capture on beta decaying nuclei, may be done in future experiments designed to measure the neutrino mass through decay kinematics.

Since the neutrinos interact with matter only by the weak interaction they decouple much earlier than the photons, which interact by the electromagnetic force. The decoupling of the neutrinos from matter is determined by the competition of the expansion rate of the Universe (Hubble parameter) and the interaction rate of the neutrinos with matter, occurring about 1 second after the BB.

Despite its successes, the standard model is known to be incomplete, and in at least one aspect, it directly conflicts with observations: according to the standard model, neutrinos should be massless. However, the Solar neutrino problem and its solution has shown this to be not the case: The (electron) neutrinos generated through nuclear fusion in the center of the Sun can escape, due to their small interaction cross section. These Solar neutrinos can be detected in (big!) terrestrial detectors. However, the measured rate of electron neutrinos from the Sun is only half as large as expected from Solar models. This Solar neutrino problem kept physicists and astrophysicists busy for decades. Its solution consists of a purely quantum-mechanical process, that of neutrino oscillations. It is possible that during its propagation, an electron neutrino transforms into a muon or tau neutrino, and back again. The existence of such oscillations was speculated for a long time, and provides a possible explanation for the missing rate of Solar electron neutrinos. Indeed, in 2001 the Sudbury Neutrino Observatory showed that neutrinos of all three flavors are received from the Sun, and that the sum of the rates of these neutrinos is compatible with the predicted rate from Solar models. In the meantime, these neutrino oscillations have also been observed in terrestrial experiments with neutrino beams.

Whereas neutrino oscillations are therefore well established today, they are in conflict with the Standard Model, according to which neutrinos have zero rest mass. From Special Relativity one can easily show that massless particles can not change their ‘identity’ while propagating through space. The existence of neutrino oscillations necessarily requires that neutrinos have a finite rest mass. Indeed, the oscillation experiments were able to constrain these masses, since they determine the length-scale over which the flavor of neutrinos changes – more precisely, it depends on the difference of their squared mass m_i^2 . One finds that $|m_2^2 - m_1^2| = (7.6 \pm 0.2) \times 10^{-5}$ eV and $|m_3^2 - m_2^2| \approx |m_3^2 - m_1^2| = (2.4 \pm 0.2) \times 10^{-3}$ eV.

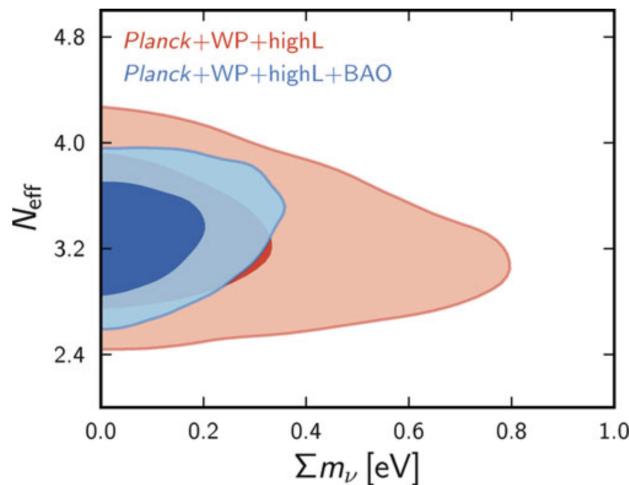


Figure 44: Allowed region in the parameter plane spanned by the sum of neutrino masses and the effective number of neutrino families N_{eff} . Constraints are given from CMB data alone (red) and in combination with the BAO results (blue). Source: Planck Collaboration 2013, Planck 2013 results. XVI. Cosmological parameters, arXiv:1303.5076, Fig. 28; original source: ESA and the Planck Collaboration. Figure taken from Schneider (2006).

These squared-mass differences thus do not provide us with an absolute mass scale of neutrinos, but their mass is non-zero. That means that neutrinos contribute to the cosmic matter density today, giving a contribution of

$$\Omega_\nu h^2 = \frac{\Sigma m_\nu}{91.5 \text{ eV}},$$

which depends only on the sum of the three neutrino masses – since the number density of neutrinos is known from the thermal history after the Big Bang. If the neutrino masses take the lowest values allowed by the squared-mass differences given above, this contribution is about 0.1%. Observations of the large-scale structure in the Universe show that neutrinos cannot contribute a substantial fraction to the matter density. Indeed, these observations yield a constraint of $\Sigma m_{\nu,i} = 0.1 \text{ eV}$, and thus the upper bound on neutrino masses from cosmology are much stricter than those obtained from laboratory experiments. For the electron neutrino, an upper limit on its mass was determined from decay experiments of tritium, yielding $m_{\nu_e} \lesssim 2 \text{ eV}$, which together with the results from neutrino oscillations implies a maximum density of $\Omega_\nu < 0.12$.

If the sum of the neutrino masses is significantly larger than the minimum mass required from neutrino oscillations, then they would contribute to the current matter density in the Universe in the form of hot dark matter, and thus affect the shape of the power spectrum. If the number of neutrino families is larger than three, there would be a larger radiation component in the early Universe, changing the expansion law and the epoch of matter-radiation equality. All of this affects the CMB fluctuations.

Significant constraints on the neutrino masses and the number of neutrino families can thus be obtained from the CMB results. As shown in Figure 44, the CMB alone yields a 2σ upper limit on the neutrino mass of $\Sigma m_\nu < 0.60 \text{ eV}$ and limits the effective number of neutrinos to $N_{\text{eff}} = 3.29^{+0.67}_{-0.64}$. These constraints get even tighter when the results from the BAOs are included in the analysis, yielding $\Sigma m_\nu < 0.28 \text{ eV}$ and $N_{\text{eff}} = 3.29^{+0.54}_{-0.652}$. The former results presents the strongest bound on the neutrino masses yet available, it is a mere factor of ~ 4 larger than the lower bound derived from neutrino oscillations. The latter result confirms our picture of particle physics, according to which there are three families of leptons, and thus three kinds of neutrinos.

In addition to the sun and the CNB, various other astrophysical objects also emit neutrinos. For example, the major fraction of the binding energy released in the formation of compact objects is emitted in the form of neutrinos: about $3 \times 10^{53} \text{ erg}$.

1.19.3 Follow-up Questions

- What is a neutrino?
- How do we determine their mass?
- Besides the sun and the CNB, what other sources of neutrinos are there?
- What is the neutrino contribution to the density of the Universe?
- Why do neutrinos decouple at 1 second (i.e., before photons)?

Cosmology Class Fall 2019: Martine's Notes

$$a + b \leftrightarrow c + d \quad (1.3)$$

$$f = \frac{1}{e^{E-\mu}/kT \pm 1} \quad (1.4)$$

Early universe, at very high temperatures, every particle has equal likelihood. μ is 1 when T is high.

At $kT > 511\text{keV}$:

$$\nu_e + \bar{\nu}_e \leftrightarrow e^+ + e^- \quad (1.5)$$

Define the interaction rate, Γ . Define the expansion rate, H . When $\Gamma \gg H$, lots of interactions. When $\Gamma \leq H$, you get freeze out.

Let's talk about

$$p^+ + e^- \leftrightarrow n + \nu_e \quad (1.6)$$

$$\frac{n_n}{n_p} = e^{-Q/T} \quad (1.7)$$

$$Q = \Delta M = 1.3\text{MeV} \quad (1.8)$$

(Q is the difference in the masses of the two particles.)

At around $T \approx 1\text{ MeV}$, or around 1 second, the interaction rate of Equation 1.5 goes below H and the numbers of neutrinos freezes out.

The temperature of the CνB is $T_\nu = 1.95K \approx 0.17\text{meV}$.

Around the same time, another process stops:

$$e^+ + e^- \leftrightarrow \gamma + \gamma \quad (1.9)$$

The likelihood of this reaction (cross section) is higher than of 1.5. But then after $T < 511\text{ keV}$, the gammas can no longer interact to form electrons because they don't have enough energy. But the electron-positron interaction keeps going.

Cosmic neutrino background decoupled earlier, and didn't get the extra energy from electron-positron annihilation added to it. But the CMB did.

Neutrinos have mass. The CνB contributes to the mass density of the universe.

$$\Omega_\nu h^2 = \Sigma_i \frac{M_{\nu,i}}{94\text{eV}} \quad (1.10)$$

We know: $\Sigma_i M_{\nu,i} > 60\text{ meV}$: Neutrino oscillation experiments. $\Sigma_i M_{\nu,i} < 120\text{ meV}$: Fits to the CMB. **Barth's follow up question**

What makes detection of the Cosmic Neutrino background challenging? If it could be measured at the same resolution and precision as the cosmic microwave background, what would its angular power spectrum look like?

Q18) DARK MATTER CANDIDATES

Give three examples of possible dark matter candidates (current or historical). What is their status observationally?

C20

- MACHOs should result in more lensing and be more spread out (less pt like) to explain rotation curve
- Neutrinos would diffuse out from & damp small scale structure, require top down formation model
- WIMPs are theoretical new elementary particles; massive and non-relativistic so no damping.

Dark Matter Candidates

• Dark matter is expected to account for $\sim 80\%$ of all matter in the universe and is invoked to explain things like the M/L ratio of galaxy clusters and the rotational velocity curves of our own galaxy. It's thought to be non-baryonic (mostly) and to only really interact gravitationally, not electromagnetically, i.e cold DM.

MACHOs - Massive Compact Halo Objects

- To explain the flat rotation curve of the MW at large radii, there must be a lot of mass in the halo which we can't see. Perhaps this non-luminous objects (or just weakly luminous) are black holes, neutron stars, brown dwarfs, and rogue planets. If this were the case, we'd expect to detect them via gravitational lensing events. In fact we do see such events, but only enough to account for $\sim 20\%$ of the mass we expect. ^{plus,} these MACHOs are large point sources, which wouldn't quite produce the smooth rotation curve - we need a more diffuse source.

Neutrinos as HDM

- because neutrinos only interact via gravity and the weak force, they don't really interact w/ normal baryonic matter very much. This makes them very difficult to detect and thus potential DM candidates, but because they're so light there must be a huge number of them to account for the mass we know exists in the universe. We can rule out HDM models for a couple reasons. An increase in neutrinos (to account for the DM) in the early universe would have resulted in less structure on small scales, as neutrinos leaking out of density perturbations would've suppressed this structure growth. This means structure would only really be able to form on large scales, and the rest of the structure we see today would've formed via fragmentation. We know this top-down method is incorrect because we see small structures in the early universe.

- also neutrinos are relativistic and the best guess for their mass is $< 2\text{eV}$.

• WIMPS - Weakly Interacting Massive Particles

- these are hypothetical, new kinds of elementary particles that could have been produced thermally in the early universe alongside all the other particles we see today. Like neutrinos, they would interact gravitationally, but they would be a lot heavier and not relativistic (otherwise they'd diffuse out from over densities on small scales like neutrinos do). We don't have any evidence for these particles yet, but supersymmetric extensions of the standard model of particle physics does predict particles with such properties.

Q18) Campbell Cosmo Q20

1.21 Question 20

Give three examples of possible dark matter candidates (current or historical). What is their status observationally?

1.21.1 Short answer

There is no shortage of candidates for non-baryonic dark matter. The primary candidates include:

- WIMPs (weakly interacting massive particles). These are currently purely theoretical and have yet to be detected.
- MACHOs (massive compact halo objects). These include white dwarfs, brown dwarfs, black holes etc. These have been detected, of course, but not in large enough quantities to describe the amount of dark matter necessary.
- Standard Model neutrinos. These have also been detected but would only work in a HDM cosmology which has already been ruled out in favour of CDM.

1.21.2 Additional context

There are even more particle candidates...

- Sterile neutrinos.
- Axions.
- Supersymmetric candidates.
- Neutralinos.
- Sneutrinos.
- Gravitinos.
- Axinos.
- Light scalar dark matter.
- Dark matter from little Higgs models.
- Kaluza-Klein states.
- Superheavy dark matter.
- Q-balls.

1.21.3 Follow-up Questions

- What does the “hot” in HDM refer to?
- What distinguishes between HDM and CDM?
- What is MOND and how does it fit in with DM?
- What happens if we don’t find WIMPs? How is this search going?

Q19) GALAXY CLUSTERS

What are galaxy clusters? List and explain three methods for detecting them or determining their basic properties.

Herman ExtraGal Q13

EG 13

- clusters are grav. bound collections of galaxies, and are in virial equilibrium
- List properties
- Look for extended x-ray sources, SZ effect on inverse compton scattered CMB, lensing signals due to mass of cluster warping space

• Galaxy Cluster Detection

• Galaxies aren't uniformly distributed in space, but instead are found in grav. bound clusters (which have a characteristic separation of ~ 150 Mpc due to growth of density perturbations in early universe). Clusters are dynamically relaxed (in virial equilibrium). They're classified as regular if they're spherical & centrally condensed

• Properties:

Galaxies $\gtrsim 50$ to 1000 (groups have < 50)

Diameter $\sim 2\text{-}10$ Mpc

Mass $\sim 10^{14} M_\odot$

Velocity dispersion ~ 1000 Km/s

M/L ratio $\sim 10^2 M_\odot / L_\odot$ (indicates a lot of DM)

• Detection Methods

- X-ray emission

- ↳ extended x-ray sources are more likely to be clusters than quasars, which have point-like x-ray emission. The extended source implies the radiation doesn't come from just a single galaxy, so we can search for those sources.
- ↳ the x-ray emission is due to hot optically thin gas where the e⁻s experience thermal Bremsstrahlung during collisions.

of all x-ray sources

(15%)

(85%)

- Sunyaev-Zeldovich Effect

- ↳ e⁻s in the hot ICM gas can inverse Compton scatter CMB photons, giving them an energy boost. This causes the CMB spectrum to deviate from a perfect blackbody, as low- ν γ 's are shifted to higher ν 's. We will therefore see an increase in the temperature fluctuation of the CMB in the direction of the cluster.

- Weak Gravitational Lensing

- ↳ The high mass concentration produces a tangentially-oriented shear field, bending light around the cluster and distorting more distant galaxies behind the appearance of

The cluster into arcs and multiple images. By searching for such lensed signals you can infer the presence of a cluster.

↳ the advantage of this method is that it depends only on cluster mass, not its EM radiation

- Red Cluster Sequence

↳ Color-Mag diagrams of cluster members show a horizontal sequence (RCS) of early type galaxies, meaning these galaxies have the same color, only weakly depending on L or metallicity. The RCS gets redder w/ redshift, so you can constrain the z based on the RCS color, and determine if the galaxies are spatially associated in 3D rather than just 2D (on the sky).

Hilroy

STAR FORMATION QUENCHING

What is star formation quenching in galaxies? What is the evidence for it, and why is it thought to happen?

Herman ExtraGal Q14

EG 14

- SFG is a process through which feedback slows down or halts SF
- Methods: SNe, AGN, Strangulation. All remove gas
- Evidence: bimodal galaxy distribution in color-mag; galaxy evolution simulations

• Star Formation Quenching

• SFG is generally when some feedback process causes SF to decrease or halt entirely, or some event removes gas from a galaxy, limiting SF.

• SNe Feedback

- high SFR leads to high rate of SNe explosions, which can do 2 things to the ISM:
 - ↳ heat the gas due to its release of KE into the ISM, resulting in gas expansion and decreased density, making cooling less efficient and further SFR harder
 - ↳ in low mass galaxies, the ISM can be blown out of the disk and into the halo, removing gas and limiting SF as the gas is no longer available

• AGN Feedback

- as gas accretes onto the SMBH / its accretion disk, the grav. PE is converted to KE/Thermal E, so the AGN heats and expels the gas from the galaxy. The removal of gas limits SF.

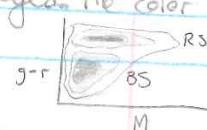
• Strangulation

- as a galaxy approaches the inner cluster, the high density of ICM gas and other galaxies can strip all the gas from the galaxy, limiting SF.

• Generally these processes serve to slow down the SFR rate such that galaxies can still be seen with SF after billions of years, not just when they're very young. The quenching has to happen quickly.

- We see evidence of this SFG in the bimodal distribution of galaxies in color-mag diagrams. Red-sequence galaxies are more luminous^{high mass, spherical} and narrowly distributed in color, since the color of an old stellar pop'n depends little on its exact age. The galaxies have an old stellar pop'n w/ no/little recent SF. Blue-sequence galaxies are star-forming, low mass galaxies that are disk dominated, and their spread in color reflects different levels of SFR, leading to different mean ages. The color of B.S. galaxies is more correlated w/ luminosity than R.S. galaxies

Fittery



• We can also find 'evidence' of SFQ if we look at galaxy evolution simulations. If we don't include feedback processes (like SNe), SF proceeds way too quickly to explain the galaxy types and ages we see today.

• Also consider the epoch of reionization. At first H₂ cooling allowed the first stars to form from condensed gas, but then the stars produced a bunch of UV photons capable of photo-dissociating H₂, preventing further cooling + SF.

OLD COSMOLOGY QUESTIONS, NOT IN THIS QUAL

Herman Cosmo Q5

C5

- Universe isn't expanding \rightarrow photons travel some distance \rightarrow lose energy \rightarrow change in $P \rightarrow$ blurring of distant sources \rightarrow we don't see this

- Universe is expanding, but infinite in size and age \rightarrow requires constant matter creation \rightarrow violates conservation laws

- can't explain quasar dist. in z , CMB, metal abundance

Tired Light and Steady State Universes

Tired Light

- rather than cosmological redshift (due to expanding space), photons lose energy as they travel, with an exponential dependence on distance

$$E = E_0 \exp(-d/R_0)$$

- photons can't lose energy w/out a change in their momenta, which would lead to blurring of distance galaxies. This is observationally disproven.

Steady State

- the universe is infinite in both size and age, and is homogeneous and isotropic on large scales. To keep the universe in a steady state requires constant creation of matter to account for the expanding universe (const. ρ)

- this violates conservation laws and can't explain the presence of quasars (more prevalent at $z \sim 3$), the CMB (light from ancient stars scattered by galactic dust...that was their solution), or the abundance of elements and lower metallicity stars at higher z .

YUP
TEARS

Hilary

Herman Cosmo Q12

C12

- The universe has to have very specific values for us to observe it / exist
- We exist because the universe was tuned specifically for us
- The universe is too flat

Fine Tuning and the Anthropic Principle

• The fine tuning problem is that the universe appears to require very specific values for its fundamental constants in order to allow us to exist and observe it to have such values. For instance, the flatness of the universe. Right now it's $\Omega_k = 0 \pm 0.02$, and scaling that back to Planck time implies flatness to 1 part in 10^{60} . That's WEIRD.

$$1 - \Omega_0 = \Omega_k$$

• The anthropic principle attempts to resolve this by saying that we can only exist in a universe with such specific parameters, otherwise we wouldn't be able to observe it, so it shouldn't be that surprising that the parameters are finely tuned. Perhaps we're but one possibility of many. G, stated even more strongly, perhaps the universe is made for us, so it's inevitable that intelligent life should arise to observe the universe with such parameter values.

I think this is stupid and implies that humans are special in some way when really we all suck and it's just chance that we developed in the first place.

C 17

- Scalar modes have only a magnitude, no direction, and will produce E mode polarization
- Tensor modes have magnitude and multiple dimensions (they're distortions of the metric), and will produce B mode polarization
- we can detect them in the CMB polarization map

Scalar and Tensor Modes of Perturbation

- Scalar perturbations are just single order perturbations, with some magnitude but no direction. They come in the form of density perturbations which result in temperature fluctuations and polarized variations - but only E-mode polarization. The temperature fluctuations produce quadrupole moments which polarize light ^{inertial}_{non-inertial}. We can detect their presence in the CMB by looking at the polarization power spectrum and C_l power spectrum for temperature fluctuations.
- Tensor perturbations are a distortion of space time (in the form of grav. waves). They're an extension along one axis and a compression on another, propagating forward at the speed of light, and which don't damp out because there's nothing to do so. They're thought to be very faint but we haven't detected them yet. ^(at least not primordial ones.) Tensor perturbations should seed not only temperature and E-mode variations, but B-modes as well since tensors can have a handedness, while scalars cannot as they're just a magnitude.
- To detect such modes we can observe the CMB and determine its linear polarization. That polarization map can then be decomposed into E- and B-modes. BICEP2, for instance had linearly polarized antennae and just stared at the sky.
- E-Modes can be sheared into B-modes if the CMB photons are gravitationally squished by ~~massive~~ massive galaxy clusters. Grav. lensing mixes E-modes and B-modes, but only on small scales since it's hard to shear large scale modes. B-modes can be transformed into E-modes too.
- Vector Modes have a magnitude and a single direction, so they're basically travelling through space. These are easily damped out by the expansion of the universe since velocities are difficult to maintain. We don't observe these and don't expect to.

Hilroy

C19

- adiabatic = proportional density fluctuations resulting in net change + curvature
- isocurvature = balanced fluctuations, not net change or curvature

• Compare C_l spectrum to models; isocurvature smears things out.

Isocurvature and Adiabatic Modes.

- Adiabatic Modes are when ^{the} density fluctuations of all the universe's components are proportional to each other - if the density of baryons goes up in a certain region, the densities of photons, neutrinos, and DM will go up as well, so the total density goes up in that region. This results in a change in the local curvature of spacetime.
- Isocurvature Modes are when the density fluctuations are all balanced such that there's no net over/under density. If the density of one component goes up, the densities of others go down to compensate. Thus there's no effect on the local curvature, and hence 'isocurvature'.
- We know the initial conditions of the universe must have mostly been in the form of adiabatic modes based on the shape of the C_l power spectrum of the CMB. If we had more isocurvature modes, the power spectrum would be more smeared out, as over-densities in baryons and photons wouldn't necessarily follow each other, and the resulting temperature fluctuations would have different amplitudes, probably being smoothed out among angular scales. Comparing observations of the C_l spectrum with models containing adiabatic v. isocurvature perturbations tells us that the former must dominate.

Hilary

OLD Q3) CDM Spectrum evolution

Outline the development of the Cold Dark Matter spectrum of density fluctuations from the early universe to the current epoch.

Q3) Ludwig Cosmo Q3

Question 3 - CDM Power Spectrum

Outline the development of the Cold Dark Matter spectrum of density fluctuations from the early universe to the current epoch.

Rephrased: Describe how fluctuations in the matter density grow with time in different cosmological epochs and how this leads to the matter power spectrum we see.

Solution

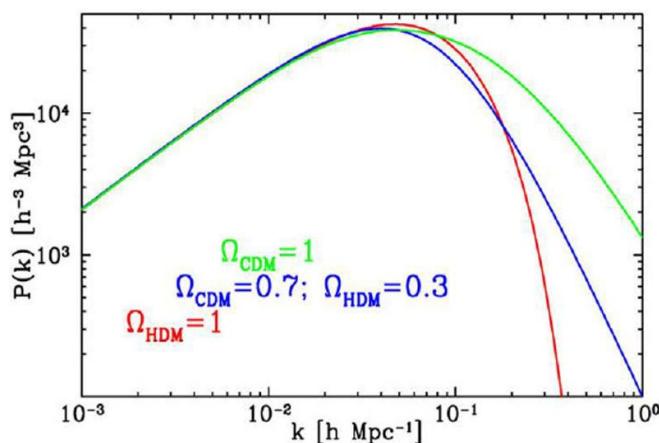
Gravitational instability increases density fluctuations over time.

- On large scales, the universe is homogeneous.
 - But on small scales you see anisotropies with small amplitudes in the cmb that can be described with a gaussian random field.
 - These become bigger over time.
- To talk about this we define a fractional overdensity, or relative density contrast

$$\delta = \frac{p(\vec{r}, t)}{\bar{p}(t)} - 1$$
- In over dense regions, $\Delta p > 0$, $\delta > 0$, there is a stronger gravitational field than the mean.
- Expansion is related to gravity so these over dense regions expand more slowly, while conversely, underdense regions expand much quicker than the mean. There's some positive feedback between density creating a gravitational field and the gravitational field making the region more dense.

The Matter-Power Spectrum

- The power spectrum is parameterized by a power law $P(k) = k^n$
- Harrison and Zeldovich argue that n has to equal 1 or else you're introducing a preferred scale for mass scale for fluctuations. $P(k) = k$
- A transfer function is introduced which is dependent on the cosmological model. This is where it matters if the universe is matter or radiation dominated. The assumption we make here is that we can use linear perturbation theory, so fluctuation amplitudes have to be small. Often this just sets the initial condition.
- For small k , $P(k) = k$.
- For large k , $P(k) = k^{-3}$,
- Peaks at $k_{peak} \sim 2 \times 10^{-2} h Mpc^{-1}$ corresponding to $\lambda_{peak} \sim 350 h^{-1} Mpc$

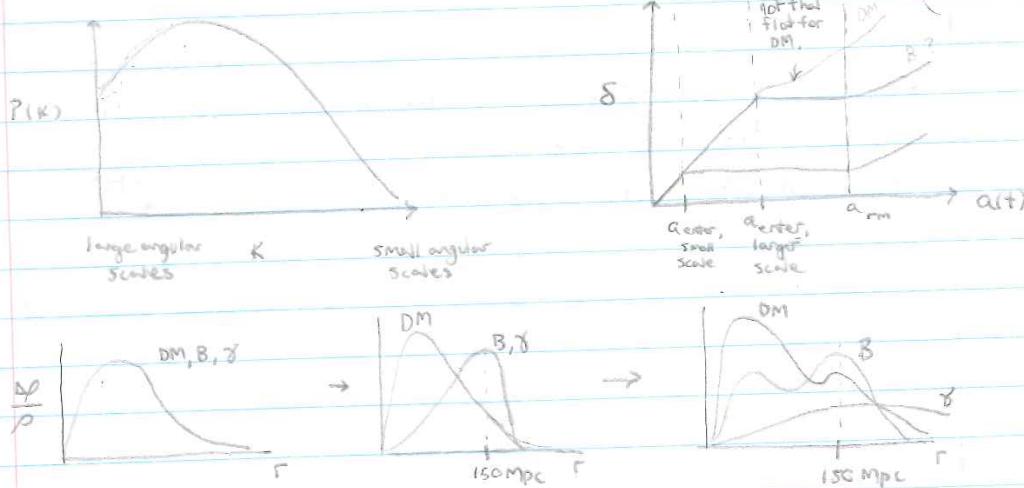


Q3) Herman Cosmo Q3

C3

- density perturbations start to grow on all scales
- as the horizon expands to encompass them, grav. coupling to γ -B fluid damps out DM fluctuation
- peak corresponds to max growth in rad. era before being encompassed & slowed in matter era.

CDM spectrum of density fluctuations



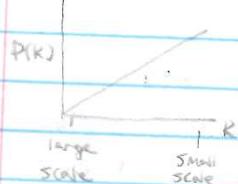
• The 1st part of the power spectrum follows basically a linear power law (Harrison-Zeldovich spectrum), where density fluctuations on large scales have just increased in power over time, starting from $P(k) = 0$ and $a \approx 10^{-3}$. The peak corresponds to the angular scale at which the density perturbation was encompassed by the horizon at the same time as the matter-radiation energy density equality (50 kyr), so these modes (?) had the longest possible time to grow. At smaller angular scales, the power is lower, because these have been damped out by the diffusion of γ s and stuff.

• I think that the smaller scales are lower Power because these modes entered the horizon sooner, and were grav. coupled to the baryons and photons, slowing their density growth as the baryons & photons pushed outward and tried to pull the DM into their potential well. The angular scale at which the most power is output corresponds to a size that would only just have reached the horizon when the energy density (and thus expansion) of the universe switched from radiation to matter dominated, and would have had maximal growth in the radiation era. Perturbation theory tells us that the density perturbations grew as a^2 during this time. In

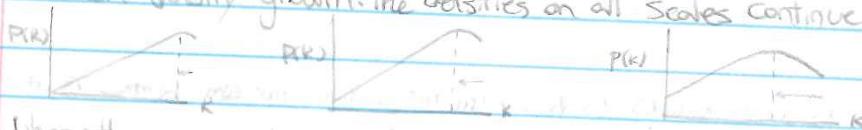
In the matter dominated era, perturbations grew linearly with $a(t)$. So any fluctuations that entered the horizon at later times (post a_m) would not grow as quickly. The initial part of the power spectrum will just look like the initial power spectrum (linear power law) with higher amplitude as the density perturbations grow over time.

IDFK I HATE THIS LEMME TRY AGAIN.

Initial power spectrum = linear; Harrison-Zeldovich:



As the horizon grows to encompass progressively larger scales, the DM is gravitationally coupled to the baryon-photon plasma, which damps out the DM density growth. The densities on all scales continue to grow.



When the energy density of the universe flips over from radiation dominated to matter dominated, there's a specific angular scale associated with the size of the horizon.



As time continues, eventually the baryons have fallen as far into the DM over densities as they ever will, so as the horizon expands there isn't a lot of damping happening on the angular scales being encompassed by the horizon, probably, but IDFK.



Screw this question

Hilroy

Q3) Campbell Cosmo Q3

1.4 Question 3

Outline the development of the Cold Dark Matter spectrum of density fluctuations from the early Universe to the current epoch.

1.4.1 Short answer

The primordial CDM spectrum is that of the scale-invariant **Harrison-Zeldovich power spectrum** (i.e., a flat power spectrum favouring neither large nor small scales). As the sound horizon grows, continuously larger modes are encompassed and become in causal contact which allows them to begin collapsing under their own self-gravity. The CDM power spectrum therefore begins to diverge from a flat spectrum at small modes which continues until the Universe turns over from being radiation-dominated to matter-dominated. At this point, the CDM power spectrum is frozen-in and there is a characteristic peak at ~ 250 Mpc which is defined by the sound horizon.

1.4.2 Additional context

The generally accepted theoretical framework for the formation of structure is that of gravitational instability. The gravitational instability scenario assumes that the early Universe was almost perfectly smooth, with the exception of tiny density perturbations with respect to the global cosmic background density and the accompanying tiny velocity perturbations from the general Hubble expansion. The observed fluctuations in the CMB temperature are a reflection of these density perturbations, so we know that the primordial density perturbations must have been on the order of $\Delta T/T \sim 10^{-5}$. The origin of this density perturbation field has yet to be fully understood, but the most plausible theory currently is that they are a result of **quantum fluctuations** which, during the inflationary phase, expanded to macroscopic scales.

Originally minute deviations from the average density of the Universe, and the corresponding deviations from the global cosmic expansion velocity (i.e., the **Hubble expansion**), will start to grow under the influence of gravity where the density perturbations will have induced local differences in gravity. During its early evolution, an overdensity will experience a gradually stronger deceleration of its expansion velocity such that its expansion velocity will continue to slow down with respect to the Hubble expansion. Since matter is gravitationally attracted to regions of higher density, it will also have the tendency to move towards that region. When the region has become sufficiently overdense, the mass of the density perturbation will have grown so much that its expansion would come to a halt. The region then completely decouples from the Hubble expansion and begins to collapse under its own gravity. The newly formed gravitationally bound object will reach virial equilibrium at which point it becomes a recognizable cosmic object; its precise nature (e.g., galaxy, cluster of galaxies etc.) being determined by the properties of the initial density perturbation and its surroundings.

The opposite tendency will have occurred for underdensities in the matter density field. Since they contain less matter than the average density field, its expansion deceleration is less than the Hubble expansion which results in matter becoming even more dispersed and underdensities continuing to expand. As this process continues and becomes more pronounced, such underdensities result in the gradual emergence of voids in the matter distribution of the Universe. The most dramatic evidence for these spatial nonuniformities (i.e., overdensities and voids) on the largest of scales comes from redshift surveys such as the 2dF Galactic Redshift Survey² shown in Figure 3.

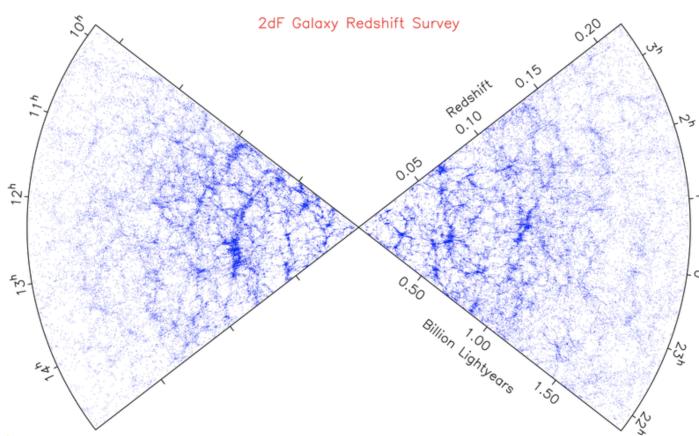


Figure 3: The spatial distribution of galaxies in two four-degree strips on the sky, according to the 2dF Galaxy Redshift Survey. Note the 100 Mpc filamentary features and the prominent voids. One of the principal challenges in cosmology is to explain this pattern, which is most probably a relic of the very earliest stages of the expanding Universe. Figure taken from Ryden (2017).

²2dFGRS: <http://www.2dfgrs.net/>

Q3) Campbell Cosmo Q3

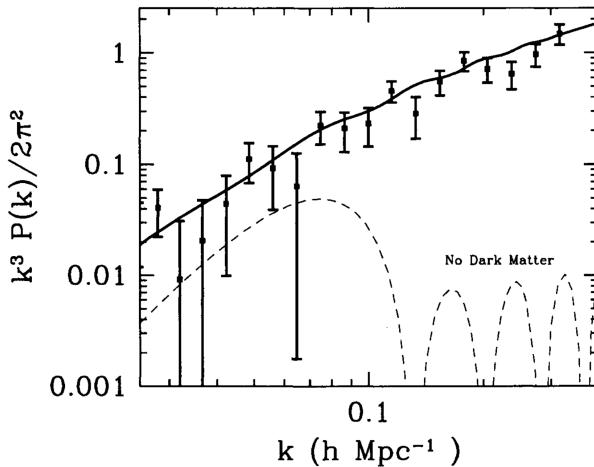


Figure 4: The variance $\Delta^2 \equiv k^3 P(k)/2\pi^2$ of the Fourier transform of the galaxy distribution as a function of scale. On large scales, the variance is smaller than unity so the distribution is smooth. The solid line is the theoretical model in which the Universe contains dark matter and a cosmological constant with perturbations generated by inflation. The dashed line is a model with only baryons and no dark matter. Data came from the PSCz survey (Saunders et al, 2000) as analyzed by Hamilton and Tegmark (2001). Figure from Ryden (2017).

Galaxies in Figure 3 are clearly not distributed randomly: the Universe has structure on large scales. To understand this structure, we must go beyond the Standard Model not only by including dark matter, but also by allowing for deviations from smoothness. We must develop the tools to study perturbations around the smooth background of the Standard Model. This is straightforward in theory, as long as the perturbations remain small (i.e., linear). Indeed, understanding the development of structure in the Universe has become a major goal of most cosmologists today. Dark matter is needed not only to explain rotation curves of galaxies but to explain structure in the Universe at large!

The best ways to learn about the evolution of structure and to compare theory with observations are to look at anisotropies in the CMB and at how matter is distributed on large scales. Anisotropies in the CMB today tell us what the Universe looked like when it was several hundred thousand years old, so they are wonderful probes of the perturbations. To compare theory with observations, we must at first try to avoid scales dominated by nonlinearities. As an extreme example, we can never hope to understand cosmology by carefully examining rock formations on Earth. The intermediate steps – collapse of matter into a galaxy; molecular cooling; star formation; planetary formation; etc. – are much too complicated to allow comparison between linear theory and observations. While perturbations to matter on small scales (less than about 100 Mpc) have grown nonlinear, large-scale perturbations are still small so they have been processed much less than the corresponding small-scale structure. Similarly, anisotropies in the CMB have remained small because the photons that make up the CMB do not clump.

Identifying large-scale structure and the CMB as the two most promising areas of study solves just one issue. Another very important challenge is to understand how to characterize these distributions so that theory can be compared to experiment. It is one thing to look at a map and quite another to quantitatively test cosmological models. To make such tests, it is often useful to take the Fourier transform of the distribution in question; as we will see, working in Fourier space makes it easier to separate large from small scales. The most important statistic in the cases of both the CMB and large-scale structure is the **two-point correlation function**, called the **power spectrum** in Fourier space.

If the mean density of galaxies at some time t is $\langle \rho(t) \rangle$, then we can characterize the inhomogeneities with the dimensionless **density perturbation field** (also known as the **density contrast**) which is used to relate the energy density $\rho(\mathbf{x}, t)$ at some co-moving spatial coordinate \mathbf{x} and time t to the average energy density at that time $\langle \rho(t) \rangle$:

$$\delta(\mathbf{x}, t) \equiv \frac{\rho(\mathbf{x}, t) - \langle \rho(t) \rangle}{\langle \rho(t) \rangle} \text{ [dimensionless].}$$

Evidently, in an unperturbed Universe with $\rho(\mathbf{x}, t) = \langle \rho(t) \rangle$ everywhere, $\delta(\mathbf{x}, t) = 0$. Note that positive density fluctuations (i.e., overdensities) may in principle grow limitless whereas negative density fluctuations (i.e., underdensities) have a strict lower-limit of $\delta \geq -1$ (emptier than empty simply doesn't exist). In a complete description of density perturbations, the total energy density of the Universe includes components of baryonic matter (ρ_b), cold dark matter (ρ_c), radiation (ρ_{rad}), and dark energy (ρ_Λ):

$$\rho(\mathbf{x}, t) = \rho_b(\mathbf{x}, t) + \rho_c(\mathbf{x}, t) + \rho_{\text{rad}}(\mathbf{x}, t) + \rho_\Lambda(\mathbf{x}, t) \text{ [g cm}^{-3}\text{].}$$

In terms of their global gravitational influence, dark matter and baryonic matter contribute and evolve equivalently, therefore on cosmological scales they can be combined into a total matter energy density $\rho_m = \rho_b + \rho_c$. Each component may have its own (primordial) perturbation character. Dark energy does not have any density fluctuations, so $\delta_\Lambda(\mathbf{x}, t) = 0$. Since most of the structure formation happened during

Q3) Campbell Cosmo Q3

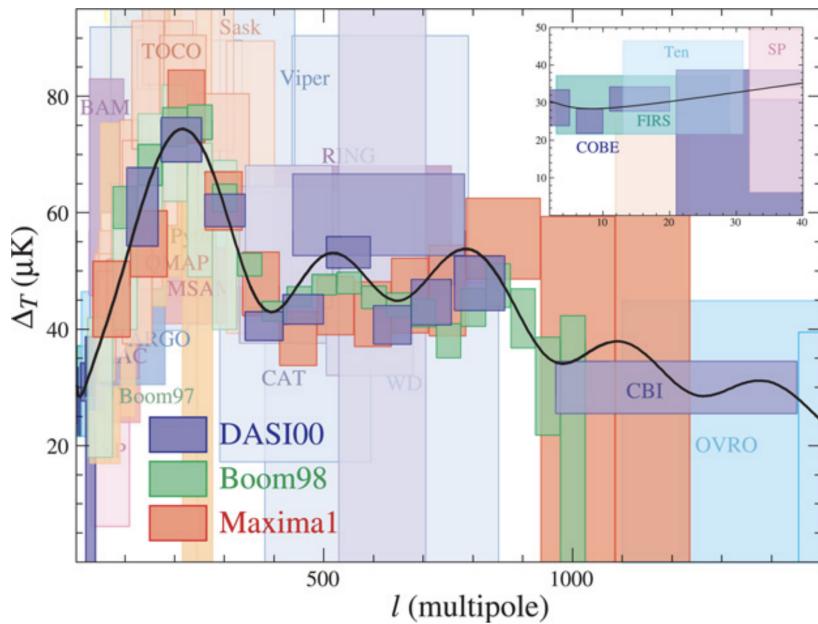


Figure 5: Anisotropies in the CMB predicted by the theory of inflation compared with observations, x-axis is multipole moment (e.g., $\ell = 1$ is the dipole, $\ell = 2$ the quadrupole) so that large angular scales correspond to low ℓ ; y-axis is the root mean square anisotropy (the square root of the two-point correlation function) as a function of scale. The characteristic signature of inflation is the series of peaks and troughs, a signature which has been verified by experiment. Figure from Ryden (2017).

the *matter-dominated era*, this mainly involves the evolution of matter linear perturbations $\delta_m(\mathbf{x}, t)$. In addition to the density contrast, an alternative (and equivalent) description of the statistical properties of the matter distribution in the Universe is the **power spectrum** $P(k)$. Roughly speaking, the power spectrum describes the level of structure as a function of the length scale $L \simeq 2\pi/k$ where k is a co-moving wavenumber. Phrased differently, the density fluctuations are decomposed into a set of plane waves of the form $\delta(\mathbf{x}) = \sum a_k \cos(\mathbf{x} \cdot \mathbf{k})$ with wave vector \mathbf{k} and amplitude a_k . The power spectrum $P(k)$ then describes the mean of the squares of the amplitudes $|a_k|^2$ averaged over all wave vectors with equal length $|\mathbf{k}| = k$. Technically speaking, this is a Fourier decomposition. The Fourier transform of $\delta(\mathbf{x}, t)$ is $\tilde{\delta}(\mathbf{x}, t)$, which allows the power spectrum $P(k)$ to be defined via

$$\langle \tilde{\delta}(\mathbf{k}) \tilde{\delta}(\mathbf{k}') \rangle = (2\pi)^3 P(k) \delta^3(\mathbf{k} - \mathbf{k}') [\mu\text{K}^2],$$

where the angular brackets denote averaging over the entire distribution, k is the wavenumber, and $\delta^3()$ is the delta Dirac function which constrains $\mathbf{k} = \mathbf{k}'$. This indicates that the power spectrum is the spread, or the variance, in the distribution. Figure 4 shows the combination $k^3 P(k)/2\pi^2$, a dimensionless number which is a good indication of the clumpiness on scale k .

The power spectrum $P(k)$ and the two-point correlation function $\xi(x)$ are related through the Fourier transform as

$$P(k) = 2\pi \int_0^\infty \frac{\sin(kx)}{kx} x^2 \xi(x) dx [\mu\text{K}^2],$$

i.e., the integral over the correlation function with a weight factor depending on the wave number provides the power spectrum. This relation can also be inverted to obtain the inverse Fourier transform such that the correlation function can be computed from the power spectrum. In general, knowing the power spectrum is insufficient to unambiguously describe the statistical properties of any density field; as such, the power spectrum is often used alongside the correlation function $\xi(x)$. Given observation data, either the power spectrum or the correlation function may be easier to compute, or another may be easier to obtain from models or simulations. In addition, our intuitive understanding of these functions may vary in different situations.

The best measure of anisotropies in the CMB is also the two-point correlation function of the temperature distribution. There is a subtle technical difference between the two power spectra which are used to measure the galaxy distribution and the CMB, though. The difference arises because the CMB temperature is a two-dimensional field, measured everywhere on the sky (i.e., with two angular coordinates). Instead of Fourier transforming the CMB temperature, then, one typically expands it in spherical harmonics, a basis more appropriate for a 2D field on the surface of a sphere. Therefore the two-point correlation function of the CMB is a function of multipole moment ℓ , not wave number k . Figure 5 shows the measurements of dozens of groups since 1992, when COBE first discovered large-angle (low ℓ in the plot) anisotropies.

Q3) Campbell Cosmo Q3

It is assumed that in very early epochs, the matter density field obeyed Gaussian statistics. This is a prediction of a large class of inflationary models which are supposed to generate the primordial density fluctuations of the Universe. An important property of Gaussian statistics, or *Gaussian random fields*, is that their distributions are uniquely determined by their power spectrum $P(k)$. Among the properties which characterize these Gaussian random fields, the probability distribution of the density fluctuations $\delta(\mathbf{x})$ at any point is a Gaussian distribution. Observational evidence for the Gaussian nature of the early density fluctuations comes from observations of anisotropies in the CMB which very strongly constrain any possible deviation from a Gaussian random field in the early Universe.

The early linear stages of structure formation have been successfully and completely worked out within the context of *linear perturbation theory*. For this discussion of the growth of density perturbations, we are concentrating on length scales that are substantially smaller than the Hubble radius. On such scales, structure growth can be described in the framework of the Newtonian theory of gravity; the effects of spacetime curvature and thus General Relativity need only be accounted for when density perturbations are on length scales comparable to, or larger than, the Hubble radius. In addition, we assume for simplicity that the matter in the Universe consists of only pressure-free matter described in the *fluid approximation*. Approximate solutions of the set of equations which describe small deviations from the homogeneous solution have the form

$$\delta(\mathbf{x}, t) = D(t)\tilde{\delta}(\mathbf{x}) \text{ [dimensionless]},$$

i.e., the spatial and temporal dependencies factorize in these solutions. Here, $\tilde{\delta}(\mathbf{x})$ is an arbitrary function of the spatial coordinate, and $D(t)$ satisfies the equation

$$\ddot{D}(t) + \frac{2\dot{a}}{a}\dot{D}(t) - 4\pi G\bar{\rho}(t)D(t) = 0 \text{ [dimensionless]}.$$

The differential equation has two linearly independent solutions. One can show that one of them increases with time, whereas the other decreases. If, at some early time, both functional dependencies were present, the increasing solution will dominate at later times, whereas the solution decreasing with t will become irrelevant. Therefore, we will consider only the increasing solution, which is denoted by $D_+(t)$, and normalize it such that $D_+(t_0) = 1$. Then, the density contrast becomes

$$\delta(\mathbf{x}, t) = D_+(t)\delta_0(\mathbf{x}) \text{ [dimensionless]}.$$

This mathematical consideration allows us to draw immediately a number of conclusions. First, the solution implies that in linear perturbation theory the spatial shape of the density fluctuations is frozen in comoving coordinates, only their amplitude increases. The growth factor $D_+(t)$ of the amplitude follows a simple differential equation that is readily solvable for any cosmological model. In fact, one can show that for arbitrary values of the density parameters in matter and vacuum energy, the growth factor has the form

$$D_+(t) \propto \frac{H(a)}{H_0} \int_0^a \frac{da'}{[\Omega_m/a' + \Omega_\Lambda a'^2 - (\Omega_m + \Omega_\Lambda - 1)]^{3/2}},$$

where the factor of proportionality is determined from the condition $D_+(t_0) = 1$.

Primordial density perturbations on a small scale appear to have a much higher amplitude than those on larger scales. This leads to a hierarchical process of structure formation, with small-scale perturbations being the first one to become nonlinear and develop into cosmic objects.

Returning to the two-point correlation function and Fourier transform, $P(k)$ and $\xi(x)$ both depend on cosmological time or redshift because the density field in the Universe evolves over time. Therefore, the dependence on t is explicitly written as $P(k, t)$ and $\xi(x, t)$. Note that $P(k, t)$ is linearly related to $\xi(x, t)$ and $\xi(x, t)$ in turn depends quadratically on the density contrast δ . If x is the comoving separation vector, we then know the time dependence of the density fluctuations, $\delta(x, t) = D_+(t)\delta_0(\mathbf{x})$. Thus,

$$\xi(x, t) = D_+^2(t)\xi(x, t_0),$$

and accordingly,

$$P(k, t) = D_+^2(t)P(k, t_0) \equiv D_+^2(t)P_0(k) [\mu\text{K}^2],$$

We stress that these relations are valid only in the framework of Newtonian, linear perturbation theory in the matter dominated era of the Universe. This last equation states that the knowledge of $P_0(k)$ is sufficient to obtain the power spectrum $P(k, t)$ at any time, again within the framework of linear perturbation theory.

Q3) Campbell Cosmo Q3

Initially it may seem as if $P_0(k)$ is a function that can be chosen arbitrarily, but one objective of cosmology is to calculate this power spectrum and to compare it to observations. More than 30 years ago, arguments were already developed to specify the functional form of the initial power spectrum.

At early times, the expansion of the Universe follows a power law, $a(t) \propto t^{1/2}$ in the radiation-dominated era. At that time, no natural length-scale existed in the Universe to which one might compare a wavelength. The only mathematical function that depends on a length but does not contain any characteristic scale is a power law; hence for very early times one should expect

$$P(k) \propto k^{n_s} [\mu\text{K}^2].$$

Many years ago, Harrison, Zeldovich, Peebles and others argued that $n_s = 1$, as for this slope, the amplitude of the fluctuations of the gravitational potential are constant (i.e., preferring neither small nor large scales). For this reason, this spectrum with $n_s = 1$ is called a scale-invariant spectrum, or Harrison-Zeldovich spectrum. With such a spectrum, we may choose a time t_i after the inflationary epoch and write

$$P(k, t_i) = D_+^2(t_i) A k^{n_s} [\mu\text{K}^2],$$

where A is a normalization constant that cannot be determined from theory but has to be fixed by observations. However, this is not the complete story: the result needs to be modified to account for the different growth of the amplitude of density fluctuations in the radiation-dominated epoch of the Universe, compared to that in the later cosmic epochs from which our result was derived.

Furthermore, these modifications depend on the nature of the dark matter. One distinguishes between cold dark matter (CDM) and hot dark matter (HDM). These two kinds of dark matter differ in the characteristic velocities of their constituents. Cold dark matter has a velocity dispersion that is negligible compared to astrophysically relevant velocities (e.g., the virial velocities of low-mass dark matter halos). Therefore, their initial velocity dispersion can well be approximated by zero, and all dark matter particles have the bulk velocity of the cosmic ‘fluid’ (before the occurrence of multiple streams). In contrast, the velocity dispersion of hot dark matter is appreciable; neutrinos are the best candidates for HDM, in view of their known abundance, determined from the thermal history of the Universe, and their finite rest mass. The characteristic velocity of neutrinos is fully specified by their rest mass; despite their low temperature of $T_\nu = 1.9\text{ K}$ today, their thermal velocities of

$$v_\nu \sim 150(1+z) \left(\frac{m_\nu}{1\text{ eV}}\right)^{-1} [\text{km s}^{-1}]$$

prevent them from forming matter concentrations at all mass scales except for the most massive ones, as their velocity is larger than the corresponding escape velocities. In other words, the finite velocity dispersion of HDM is equivalent to assigning to it a pressure, which prevents them to fall into shallow gravitational potential wells. We will see below the dramatic differences between these two kinds of dark matter for the formation of structures in the Universe. In particular, this estimate shows that neutrinos cannot account for the dark matter on galaxy scales, and thus cannot explain the flat rotation curves of spiral galaxies.

If density fluctuations become too large on a certain scale, linear perturbation theory breaks down and the linear approximation to the solution of $P(k, t)$ is no longer valid. Then the true current-day power spectrum $P(k, t_0)$ will deviate from $P_0(t)$. Nevertheless, in this case it is still useful to examine $P_0(t)$ – it is then called the linearly extrapolated power spectrum.

Within the framework of linear Newtonian perturbation theory in the ‘cosmic fluid’, $\delta(\mathbf{x}, t) = D_+(t)\delta(\mathbf{x})$ applies. Modifications to this behavior are necessary for several reasons:

- If dark matter consists (partly) of HDM, this may not be gravitationally bound to the potential well of a density concentration. In this case, the particles are able to move freely and to escape from the potential well, which in the end leads to its dissolution if these particles dominate the matter overdensity. From this argument, it follows immediately that for HDM, small-scale density perturbations cannot form. For CDM this effect of free streaming does not occur.
- At redshifts $z \gtrsim z_{\text{eq}}$, radiation dominates the density of the Universe. Since the expansion law $a(t)$ is then distinctly different from that in the matter-dominated phase, the growth rate for density fluctuations will also change.
- A cosmic horizon exists with comoving scale $r_{\text{H,com}}(t)$. Physical interactions can take place only on scales smaller than $r_{\text{H,com}}(t)$. For fluctuations of length-scales $L \sim 2\pi/k \gtrsim r_{\text{H,com}}(t)$, Newtonian perturbation theory will cease to be valid, and one needs to apply linear perturbation theory in the framework of the General Relativity.

Q3) Campbell Cosmo Q3

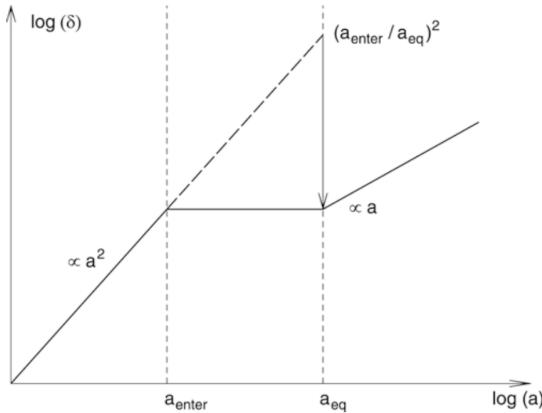


Figure 6: A density perturbation that enters the horizon during the radiation-dominated epoch of the Universe ceases to grow until matter starts to dominate the energy content of the Universe. In comparison to a perturbation that enters the horizon later, during the matter-dominated epoch, the amplitude of the smaller perturbation is suppressed by a factor $(a_{\text{enter}}/a_{\text{eq}})^2$, which explains the qualitative behavior of the transfer function. Adapted from: M. Bartelmann & P. Schneider 2001, Weak Gravitational Lensing, Phys. Rep. 340, 291. Image taken from Schneider (2006).

These effects together will lead to a modification of the shape of the power spectrum, relative to the relation $P(k, t_i) = D_+^2(t_i) A k^{n_s}$; for example, the evolution of perturbations in the radiation-dominated cosmos proceeds differently from that in the matter-dominated era. The power spectrum $P(k)$ is affected by the combination of the above effects, and will be different from the primordial spectral shape, $P \propto k^{n_s}$. The modification of the power spectrum is described in terms of the transfer function $T(k)$ in the form

$$P(k, t) = D_+^2(t) A k^{n_s} T^2(k) [\mu\text{K}^2].$$

The transfer function can be computed for any cosmological model if the matter content of the Universe is specified. In particular, $T(k)$ depends on the nature of dark matter.

The first of the above points immediately implies that a clear difference must exist between HDM and CDM models regarding structure formation and evolution. In HDM models, small-scale fluctuations are washed out by free-streaming of relativistic particles (i.e., the power spectrum is completely suppressed for large k , which is expressed by the transfer function $T(k)$ decreasing exponentially for large k). In the context of such a theory, very large structures will form first, and galaxies can form only later by fragmentation of large structures. However, this formation scenario is in clear contradiction with observations. For example, we observe galaxies and QSOs at $z > 6$ so that small-scale structure is already present at times when the Universe had less than 10% of its current age. In addition, the observed correlation function of galaxies, both in the local Universe (see Figure 6) and at higher redshift, is incompatible with cosmological models in which the dark matter is composed mainly of HDM. Therefore we can exclude HDM as the dominant constituent of dark matter. For this reason, it is now commonly assumed that the dark matter is cold. The achievements of the CDM scenario in the comparison between model predictions and observations fully justify this assumption.

In linear perturbation theory, fluctuations grow at the same rate on all scales, or for all wave numbers, independent of each other. This applies not only in the Newtonian case, but also remains valid in the framework of General Relativity as long as the fluctuation amplitudes are small. Therefore, the behavior on any (comoving) length-scale can be investigated independently of the other scales. At very early times, perturbations with a comoving scale L are larger than the (comoving) horizon, and only for $z < z_{\text{enter}}(L)$ does the horizon become larger than the considered scale L . Here, $z_{\text{enter}}(L)$ is defined as the redshift at which the (comoving) horizon equals the (comoving) length-scale L ,

$$r_{\text{H,com}}(z_{\text{enter}}(L)) = L [\text{Mpc}].$$

It is common to say that at $z_{\text{enter}}(L)$ the perturbation under consideration ‘enters the horizon’, whereas actually the process is the opposite – the horizon outgrows the perturbation. Relativistic perturbation theory shows that density fluctuations of scale L grow as long as $L > r_{\text{H,com}}$, namely $\propto a^2$ if radiation dominates (thus, for $z > z_{\text{rm}}$), or $\propto a$ if matter dominates (i.e., for $z < z_{\text{rm}}$). Free-streaming particles or pressure gradients cannot impede the growth on scales larger than the horizon length because, according to the definition of the horizon, physical interactions (which pressure or free-streaming particles would be) cannot extend to scales larger than the horizon size.

The evolution of density fluctuations of baryons differs from that of DM. The reason for this is essentially the interaction of baryons with photons: although matter dominates the Universe for $z < z_{\text{rm}}$, the energy density of baryons remains smaller than that of the photons for a longer time after recombination begins, as can be seen as follows: the baryon-to-photon density ratio is

Q3) Campbell Cosmo Q3

$$\frac{\rho_b}{\rho_\gamma} = \frac{\Omega_b a^{-3}}{\Omega_\gamma a^{-4}} = a \frac{\Omega_b \Omega_m \Omega_r}{\Omega_m \Omega_r \Omega_\gamma} = 1.68 \frac{a}{a_{\text{rm}}} \frac{\Omega_b}{\Omega_m} \sim 0.28 \frac{a}{a_{\text{rm}}} \text{ [dimensionless].}$$

Hence, if radiation-matter equality happens at $z_{\text{rm}} \sim 3,000$, then the photon density is larger than that of the baryons for $z \gtrsim 800$.

Since photons and baryons interact with each other by photon scattering on free electrons, which again are tightly coupled electromagnetically to protons and helium nuclei, baryons and photons are strongly coupled before recombination, and form a single fluid. Due to the presence of photons, this fluid has a strong pressure, which prevents it from falling into potential wells formed by the dark matter. Thus, the pressure prevents strong inhomogeneities of the baryon-photon fluid.

To discuss the evolution of baryon perturbations in a bit more detail, we consider again a perturbation of comoving scale L . As long as the perturbation is larger than the horizon size, pressure effects can not affect the behavior of the fluid, and thus baryons and photons behave in the same way as the dark matter – the amplitude of their perturbations grow. As soon as the perturbation enters the horizon, the situation changes. Although the baryons are gravitationally pulled into the density maxima of the dark matter, pressure provides a restoring force which acts against a compression of the baryon-photon fluid. As a result, this fluid will develop sound waves.

The maximum distance sound waves can travel up to a given epoch is called the sound horizon. Loosely speaking, it is given by the product of the sound speed and the cosmic time. The sound speed in this photon-dominated fluid is given by $c_s \approx c/\sqrt{3}$. Thus, the sound horizon is about a factor of $\sqrt{3}$ smaller than the event horizon. As soon as a perturbation enters the sound horizon, the amplitude of the baryon-photon fluctuations can not grow anymore; instead, they undergo damped oscillations.

The adiabatic sound speed c_s of a fluid is given in general by

$$c_s = \sqrt{\frac{\partial P}{\partial \rho}} \text{ [m s}^{-1}\text{].}$$

The pressure of the fluid is generated by the photons, $P = c^2 \rho_\gamma / 3 = c^2 \rho_c \Omega_\gamma$, and the density is the sum of that of baryons and photons, $\rho = (\Omega_b a^{-3} + \Omega_\gamma a^{-4}) \rho_c$. Thus, the sound velocity is

$$\begin{aligned} c_s &= \sqrt{\frac{\partial P}{\partial \rho}} \\ &= \sqrt{\frac{dP/da}{d\rho/da}} \\ &= \frac{c}{\sqrt{3}} \sqrt{\frac{4\Omega_\gamma a^{-5}}{3\Omega_b a^{-4} + 4\Omega_\gamma a^{-5}}} \\ &= \frac{c}{\sqrt{3(1+\mathcal{R})}} \text{ [m s}^{-1}\text{]}, \end{aligned}$$

where \mathcal{R} is defined to be

$$\mathcal{R} = \frac{3}{4} \frac{\rho_b}{\rho_\gamma} = \frac{3}{4} \frac{\Omega_b}{\Omega_\gamma} a \text{ [dimensionless].}$$

Note that \mathcal{R} is smaller than unity until recombination, and thus $c_s \approx c/\sqrt{3}$ provides a reasonable first approximation.

At recombination, the free electrons recombined with the hydrogen and helium nuclei, after which there are essentially no more free electrons which couple to the photon field. Hence, after recombination the baryon fluid lacks the pressure support of the photons, and the sound speed drops to zero – the sound waves do no longer propagate, but get frozen in. Now the baryons are free to react to the gravitational field created by the dark matter inhomogeneities, and they can fall into their potential wells. After some time, the spatial distribution of the baryons is essentially the same as that of the dark matter.

Hence, there is a maximum wavelength of the sound waves, namely the (comoving) sound horizon at recombination,

$$r_{\text{H,com}}(z) = \int_0^t \frac{cdt}{a(t)} = \int_0^{(1+z)^{-1}} \frac{cda}{a^2 H(a)} = \int_0^{a_{\text{rec}}} \frac{cda}{\sqrt{3(1+\mathcal{R})} a^2 H(a)} \text{ [Mpc],}$$

where we exchanged the speed of light by the speed of sound.

Figure 7 illustrates the physical significance of this length scale, showing the time evolution of an initial

Q3) Campbell Cosmo Q3

density peak of all four components in the Universe. The length scale r_s is the distance the baryon-photon fluid propagates outwards from the initial density peak before baryons and photons decouple, after which the density perturbation of baryons gets frozen. The x-axis shows the comoving radial coordinate, the y-axis displays the density, multiplied by $(\text{radius})^2$. The different snapshots show the spatial distribution of the various species at later epochs. In particular, because the region is overdense in photons, it is overpressured relative to its surroundings. This overpressure must equilibrate by driving a spherical sound wave out into the baryon-photon plasma which propagates at the speed of sound, $c_s \approx c/\sqrt{3}$. Neutrinos freely stream out of the perturbation at the speed of light. The photon and baryons are strongly coupled before recombination, and thus have the same spatial distribution. At the time of decoupling, the wave stalls as the pressure supplying the photons escape and the sound speed plummets. One ends up with a CDM overdensity at the center and a baryon overdensity in a spherical shell 150 comoving megaparsecs in radius for the concordance cosmology. At $z \ll 10^3$, both of these overdensities attract gas and CDM to them, seeding the usual gravitational instability. However, some of the matter also falls into the density peaks (in the example of this figure, it is an overdense spherical shell) created by baryons, whereas the density profile of neutrinos and photons becomes flat. At late times, the distributions of baryons and dark matter become identical (before the onset of non-linear processes such as halo formation). The central density peak, and the secondary peak have a well-defined separation, given by the distance a sound wave could travel before the baryons decoupled from the photons. Galaxies are more likely to form in these overdensities. The radius of the sphere marks a preferred separation of galaxies, which we quantify as a peak in the correlation function on this scale.

The Universe is of course a superposition of these point-like perturbations, but as the perturbation theory is exquisitely linear at high redshift, we can simply add the solutions. The width of the acoustic peak is set by three factors: silk damping due to photons leaking out of the sound wave, adiabatic broadening of the wave as the sound speed changes because of the increasing inertia of the baryons relative to the photons, and the correlations of the initial perturbations.

There are some other interesting aspects of the physics of this epoch that are worth mentioning. First is that the outgoing wave does not actually stop at $z \sim 10^3$ but instead slows around $z \sim 500$. This is partially due to the fact that decoupling is not coincident with recombination but is also because the coupling to the growing mode is actually dominated by the velocity field, rather than the density field, at $z \sim 10^3$. In other words, the compressing velocity field in front of the wave actually keys the instability at a later time.

Two other aspects that may be surprising at first glance are that the outgoing pulse of neutrino overdensity does not actually remain as a delta function, as one might expect for a population traveling radially outward at the speed of light, and that the CDM perturbation does not remain at the origin, as one would expect for a cold species. Both of these effects are due to a hidden assumption in the initial conditions: although the density field is homogeneous everywhere but the origin, the velocity field cannot be for a growing mode. To keep the bulk of the Universe homogeneous while growing a perturbation at the origin, matter must be accelerating toward the center; this acceleration is supplied by the gravitational force from the central overdensity. However, in the radiation-dominated epoch the outward-going pulse of neutrinos and photons is carrying away most of the energy density of the central perturbation. This outward-going pulse decreases the acceleration, causing the inward flow of the homogeneous bulk to deviate from the divergenceless flow and generating the behavior of the CDM and neutrinos mentioned above. Essentially, the outgoing shells of neutrinos and photons raise a wake in the homogeneous distribution of CDM away from the origin of the perturbation.

The smoothing of the CDM overdensity from a delta function at the origin is the famous small-scale damping of the CDM power spectrum in the radiation-dominated epoch. The overdensity raised decreases as a function of radius because the radiation is decreasing in energy density relative to the inertia of the CDM; in the matter-dominated regime, the outward-going radiation has no further effect. A Universe with more radiation causes a larger effect that extends to larger radii; this corresponds to the shift in the CDM power spectrum with the matter-to-radiation ratio.

Returning to the major conceptual point, that of the shell of overdensity left at the sound horizon, we see immediately that the sound horizon provides a standard ruler. The radius of the shell depends simply on the sound speed and the amount of propagation time. The sound speed is set by the balance of radiation pressure and inertia from the baryons; this is controlled simply by the baryon-to-photon ratio, which is $\Omega_b h^2$. The propagation time depends on the expansion rate in the matter-dominated and radiation-dominated regimes; this in turn depends on the redshift of matter-radiation equality, which depends only on $\Omega_m h^2$ for the standard assumption of the radiation density (i.e., the standard cosmic neutrino and photon backgrounds and nothing else).

The sound waves in the baryon-photon fluid, the baryonic acoustic oscillations (BAOs), are observable today. Since at recombination, the photons interacted with matter for the last time, the CMB radiation provides us with a picture of the density fluctuations at the epoch of recombination. Our observable

Q3) Campbell Cosmo Q3

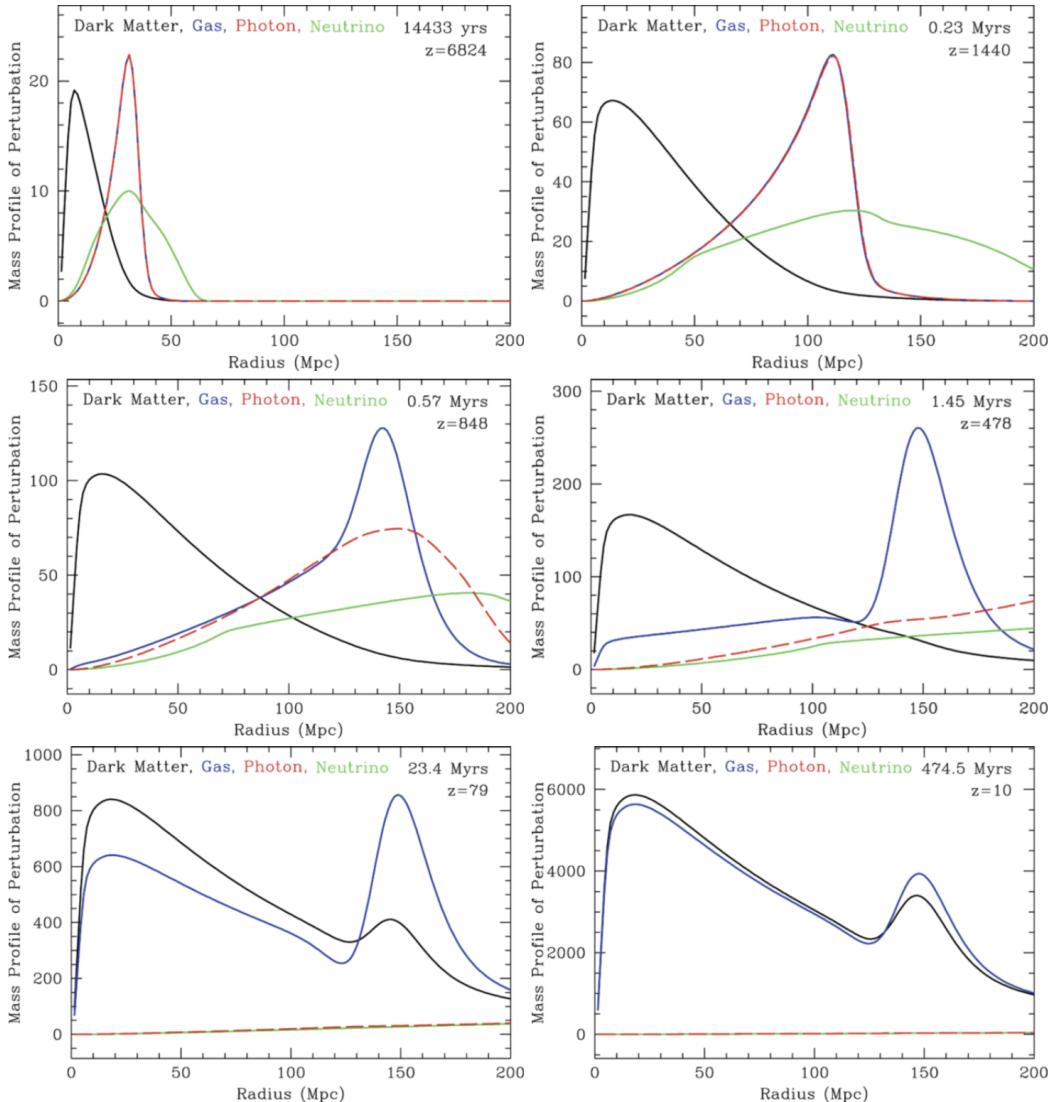


Figure 7: Evolution in time of an initial density peak in all components of the cosmic matter. *Top left:* Near the initial time, the photons and baryons travel outward as a pulse. *Top right:* Approaching recombination, one can see the wake in the cold dark matter raised by the outward-going pulse of baryons and relativistic species. *Middle left:* At recombination, the photons leak away from the baryonic perturbation. *Middle right:* With recombination complete, we are left with a CDM perturbation toward the center and a baryonic perturbation in a shell. *Bottom left:* Gravitational instability now takes over, and new baryons and dark matter are attracted to the overdensities. *Bottom right:* At late times, the baryonic fraction of the perturbation is near the cosmic value, because all of the new material was at the cosmic mean. Source: D.J. Eisenstein et al. 2007, On the Robustness of the Acoustic Scale in the Low-Redshift Clustering of Matter, ApJ 664, 660, p. 662, Fig. 1. Figure taken from Schneider (2006).

cosmic microwave sky essentially is a picture of a two-dimensional cut at fixed time (the time of last scattering) through the density field of the baryons. A cut through an ensemble of sound waves shows an instantaneous picture of these waves. Hence, they are expected to be visible in the temperature distribution of the CMB. This is indeed the case: these BAOs imprint one of the most characteristic features on the CMB anisotropies. Since the sound waves are damped once they are inside the sound horizon, the largest amplitude waves are those whose wavelength equals the sound horizon at recombination. We have argued that the baryons, once they are no longer coupled to radiation and thus become pressureless, fall into the potential wells of the dark matter. This happens because the dark matter fluctuations can grow while the baryonic fluctuations could not due to the photon pressure, and because the mean density of dark matter is substantially larger than that of the baryons. This is almost the full story, but not entirely: baryons make about 15% of the total matter density, and are therefore not negligible. After recombination, the BAOs are frozen, like standing waves, and thus the total matter fluctuations are a superposition of the dark matter inhomogeneities and these standing waves. Whereas the dark matter dominates the density fluctuations, a small fraction of the matter also follows the inhomogeneities created

Q3) Campbell Cosmo Q3

by the standing waves. Since these waves have a characteristic length scale (the sound horizon at recombination) this characteristic length scale should be visible in the properties of the matter distribution even today. The correlation function of galaxies contains a characteristic feature at the length scale r_s . Hence, relics of the sound waves in the pre-recombination era are even visible in the current Universe. The effects of the BAOs are included in the transfer function $T(k)$, which thus shows some low-amplitude oscillations, often called ‘wiggles’.

The distance that acoustic waves can propagate in the first million years of the Universe is measurable not only in the cosmic microwave background (CMB) anisotropies but also in the late-time clustering of galaxies.

1.4.3 Follow-up Questions

- How do we observe the power spectrum?
- What is its relation to BAOs and the C_ℓ power spectrum?
- Do we see super-horizon modes (modes larger than the Universe’s event horizon) in the evolved matter power spectrum?
- What actually causes the damping during the radiation-dominated era?
- Why do modes outside the horizon grow?
- What is Silk damping and what is physically happening?
- Why does Silk damping only affect small modes?
- How do baryon and photon perturbations grow?
- How does this relate to whether structure grows in a top-down or bottom-up approach?

Q3) Zhu Cosmo Q3

12

Note that l is known, but S_κ depends on the co-moving distance between us and the CMB. This requires some knowledge of the subsequent expansion history of the universe, or else there is a degeneracy between Ω_m , Ω_Λ and Ω_κ (Komatsu et al. 2009). An additional constraint, such as a measurement of H_0 , or the series of luminosity distance measurements using high- z SNe, allows us to constrain Ω_κ (Komatsu et al. 2009). See Fig. 5.

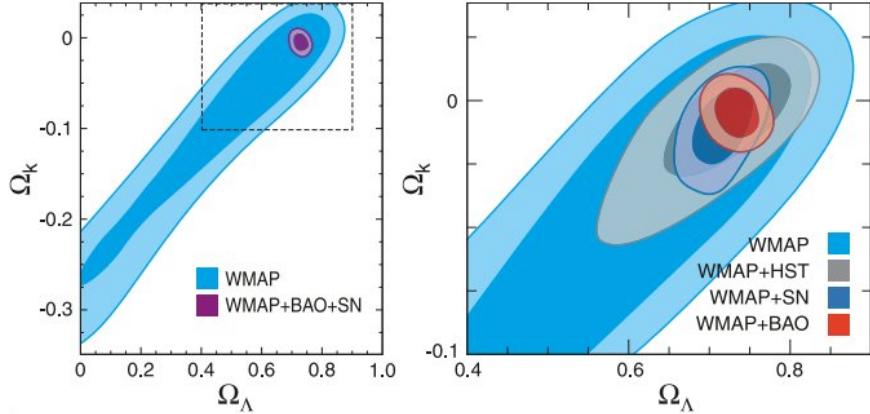


FIG. 5.— Joint two-dimensional marginalized constraint on the dark energy density Ω_Λ , and the spatial curvature parameter, Ω_κ . The contours show the 68% and 95% confidence levels. Additional data is needed to constrain Ω_κ : HST means H_0 from Hubble measurements, SN means luminosity distances from high- z SN, and BAO means baryon acoustic oscillation measurements from galaxy surveys. From Komatsu et al. (2009), their Fig. 6.

1.4. Question 3

QUESTION: Outline the development of the Cold Dark Matter spectrum of density fluctuations from the early universe to the current epoch.

Most of this information is from Schneider (2006), Ch. 7.3 - 7.5.

The growth of a single perturbation (described as one of the follow-up questions) in a matter-dominated universe can be described in the following way. We define the relative density contrast $\delta(\vec{r}, t) = (\rho(\vec{r}, t) - \bar{\rho})/\bar{\rho}$; from this $\delta(\vec{r}, t) \leq -1$. At $z \sim 1000$ $|\delta(\vec{r}, t)| \ll 1$. The mean density of the universe $\bar{\rho}(t) = (1 + z^3)\bar{\rho}_0 = \bar{\rho}_0/a(t)^3$ from Hubble flow. Like in the classic Newtonian stability argument of an infinite static volume of equally space stars, any overdense region will experience runaway collapse (and any underdense region will become more and more underdense). In the linear perturbative regime, the early stages of this collapse simply make it so that the the expansion of the universe is delayed, so $\delta(\vec{r}, t)$ increases. As it turns out, $\delta(\vec{r}, t)$ can be written as $D_+(t)\delta_0(\vec{x})$ in the linear growth regime. $D_+(t)$ is normalized to be unity today, and $\delta_0(\vec{x})$ is the linearly-extrapolated (i.e. no non-linear evolution taken into account) density field today.

The two-point correlation function $\xi(r)$ (Sec. 1.18) describes the over-probability of, given a galaxy at $r = 0$, there will be another galaxy at r (or x , here). It describes the clustering of galaxies, and is key to understanding the large-scale structure of the universe. We define the matter power spectrum (often shortened to just “the power spectrum”) as

$$P(k) = \int_{-\infty}^{-\infty} \xi(r) \exp(-ikr) r^2 dr \quad (20)$$

Instead of describing the spatial distribution of clustering, the power spectrum decomposes clustering into characteristic lengths $L \approx 2\pi/k$, and describes to what degree each characteristic contributes to the total overprobability.

Since the two-point correlation function depends on the square of density, if we switch to co-moving coordinates and stay in the linear regime,

$$\xi(x, t) = D_+^2(t) \xi_0(x, t_0). \quad (21)$$

Likewise,

$$P(k, t) = D_+^2 P(k, t_0) \equiv D_+^2 P_0(k), \quad (22)$$

i.e. everything simply scales with time. This the evolution of the power spectrum is reasonably easily described.

The initial power spectrum $P_0(k)$ was generated by the quantum fluctuations of inflation. It can be argued (pg. 285 of Schneider (2006)) that the primordial power spectrum should be $P(k) = Ak^{n_s}$, where A is a normalization factor

that can only be determined empirically. $P(k)$ when $n_s = 1$ is known as the Harrison-Zel'dovich spectrum, which is most commonly used.

An additional correction term needs to be inserted is the transfer function to account for evolution in the radiation-dominated universe, where our previous analysis does not apply. We thus introduce the transfer function $T(k)$, such that $P_0(k) = Ak^{n_s}T(k)^2$. $T(k)$ is dependent on whether or not the universe consists mainly of cold or hot ($k_B T \ll mc^2$, where T is the temperature at matter-radiation equality) dark matter. If hot dark matter dominate the universe, they freely stream out of minor overdensities, leading to a suppression of small-scale perturbations. Since our universe is filled with cold dark matter, this need not be taken into account (and indeed gives results inconsistent with observations). $T(k)$ also accounts for the fact that $a(t) \propto t^{1/2}$ rather than $t^{2/3}$ during radiation domination, and that physical interactions can only take place on scales smaller than $r_{H,c}(t)$ (the co-moving particle horizon) - on scales larger than this GR perturbative theory must be applied.

Growth of a perturbation of length scale L is independent of growths at other length scales. The growth of a density qualitatively goes like this:

1. In the early universe, a perturbation of comoving length L has yet to enter the horizon. According to relativistic perturbation theory, the perturbation grows $\propto a^2$ in a radiation-dominated universe, and $\propto a$ in a matter-dominated universe.
2. At redshift z_e , when $r_{H,c}(z_e) = L$, the perturbation length scale becomes smaller than the horizon. If the universe is still radiation-dominated, the Mészáros effect prevents effective perturbation growth, and the overdensity stalls (Mészáros 2005).
3. Once the universe becomes matter dominated ($z < z_{\text{eq}}$), the perturbation continues to grow $\propto a$.

There is therefore a preferred length scale $L_0 = r_{H,c}(z_{\text{eq}}) \approx 12(\Omega_m h^2)^{-1}$ Mpc. The transfer function then has two limiting cases: $T(k) \approx 1$ for $k \ll 1/L_0$, and $T(k) \approx (kL_0)^{-2}$ for $k \gg 1/L_0$. This generates a turnover in $P_0(k)$ where $k = 1/L_0$. Note that due to the dependence on the sound horizon on $\Omega_m h^2$ we often define the shape parameter $\Gamma = \Omega_m h$.

One last modification must be made to this picture: at a certain point growth becomes non-linear, and our analysis must be modified.

Fig. 6 shows the schematic growth of a perturbation, as well as both the primordial Harrison-Zel'dovich spectrum and the modern-day power spectrum for a series of different cosmological parameters.

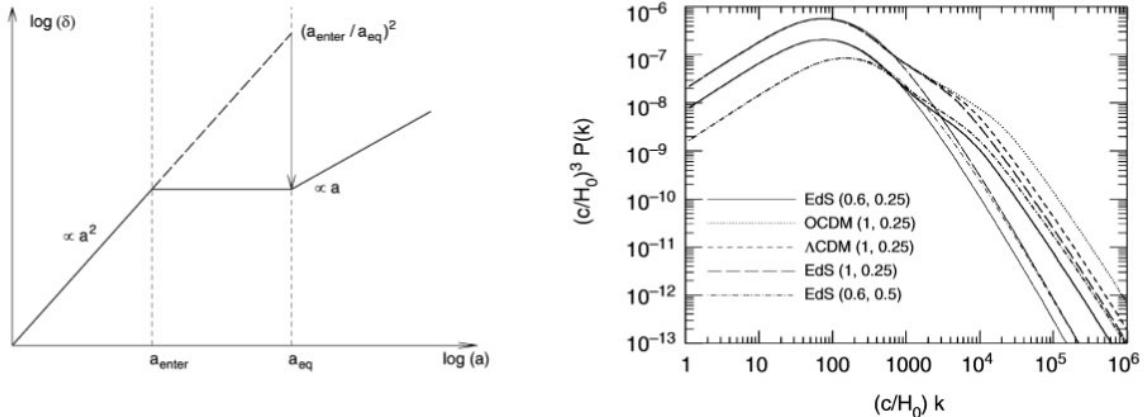


FIG. 6.— Left: evolution of a density perturbation. The $(a_{\text{enter}}/a_{\text{eq}})^2$ line indicates the degree of suppression during radiation domination. Right: the current power spectrum of density fluctuations for CDM models. The various curves have different cosmological models (EdS, $\Omega_m = 1$, $\Omega_\Lambda = 0$, OCDM, $\Omega_m = 0.3$, $\Omega_\Lambda = 0$, Λ CDM, $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$). Values in parentheses specify (σ_8, Γ) . The thin curves correspond to power spectra linearly extrapolated, while the thick curves include non-linear corrections. From Schneider (2006), his Figs. 7.5 and 7.6.

1.4.1. How do baryon and photon density perturbations grow?

This information is from Schneider (2006), pg. 288 - 289.

Baryon and photon density perturbations grew alongside dark matter perturbations until z_e , at which point baryon acoustic oscillations began, stymying any growth until recombination, $z_r < z_{\text{eq}}$. Following this, the photon overdensities escaped while the baryon overdensities began to track the dark matter overdensities.

1.4.2. How does an individual density perturbation grow?

This is described in much greater detail in Schneider (2006), Ch. 7.2.

If we assume a pressure-free ideal fluid, we can write the Euler's and continuity equations in comoving coordinates and linearize them to obtain $\frac{\partial^2 \delta}{\partial t^2} + \frac{2\dot{a}}{a} \frac{\partial \delta}{\partial t} = 4\pi G \bar{\rho} \delta$. This means we can separate $\delta(\vec{x}, t)$ into $D(t)\delta_0(\vec{x})$; i.e. at all comoving points \vec{x} the overdensity rises in the exact same manner over time. Our equation of motion then becomes

$$\ddot{D} + \frac{2\dot{a}}{a} \dot{D} - 4\pi G \bar{\rho}(t) D = 0. \quad (23)$$

There are two solutions to this equation, and we call the increasing one the growth factor, $D_+(a)$. In general $D_+(a)$ is normalized so that $D_+(1) = 1$, so that $\delta_0(\vec{x})$ is the density distribution we would have today if no non-linear effects take hold. We can show that the general increasing solution to this equation, when we switch from time to a , is

$$D_+(a) \propto \frac{H(a)}{H_0} \int_0^a \frac{da'}{(\Omega_m/a' + \Omega_\Lambda a'^2 + \Omega_k)^{3/2}} \quad (24)$$

For an Einstein-de Sitter universe, we can show, through an ansatz that $D \propto t^q$ and Eqn. 23 that $D_+(t) = (t/t_0)^{2/3}$. During matter domination, overdensities grew with the scale length.

Eventually $D(t)\delta(\vec{x})$ approaches 1, and the linear approximation fails. Growth increases dramatically (Fig. 7).

In the case of a uniform homogeneous sphere with density $\rho = \bar{\rho}(1 + \delta)$, where δ is the average density perturbation in the sphere, and we have switched back to proper distances rather than comoving. The total mass within the sphere is $M \approx \frac{4\pi}{3} R_{\text{com}}^3 \rho_0(1 + \delta_i)$, where $R_{\text{com}} = a(t_i)R$ is the initial comoving sphere radius (R is the physical radius of the sphere), and $\rho_0 = \bar{\rho}/a^3$ is the present average density of the universe. We may then model the mass and radius of the sphere as a miniature universe governed by the Friedmann equations. If the initial density of the system is greater than critical, the sphere collapses. Because of the time-reversibility of the Friedmann equations, if we know the time t_{max} where R_{com} is maximum, we know the time t_{coll} when the universe collapses back into a singularity.

This collapse is unphysical. In reality violent relaxation will occur - density perturbations in the infalling cloud will create knots due to local gravitational collapse; these knots then scatter particles, which create more perturbations, creating more knots. This creates an effective translation of gravitational potential energy to kinetic (thermal) energy, within one dynamical time. It turns out a virialized cloud has density

$$\langle \rho \rangle = (1 + \delta_{\text{vir}}) \bar{\rho}(t_{\text{coll}}), \quad (25)$$

where $1 + \delta_{\text{vir}} \approx 178\Omega_m^{-0.6}$.

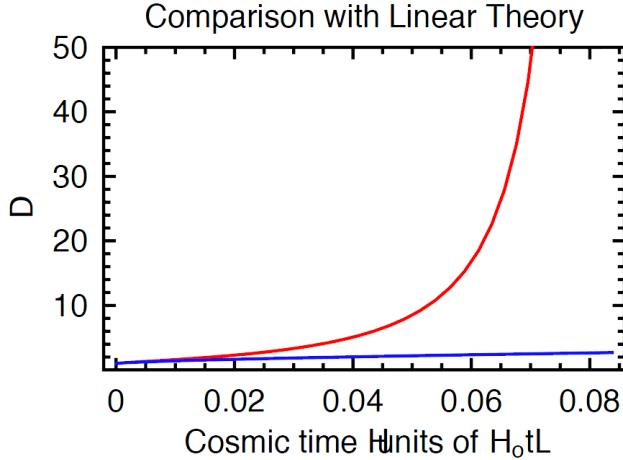


FIG. 7.— Growth of a density fluctuation taking into account non-linear evolution, versus the equivalent linear evolution. The singularity that eventually forms in the non-linear case is not physical, as the dust approximation eventually fails and virialization occurs. In baryonic material dissipative cooling also occurs. From Abraham (2011b).

1.4.3. What is violent relaxation?

This information is from Schneider (2006), pg. 235 and 290.

Violent relaxation is a process that very quickly establishes a virial equilibrium in the course of a gravitational collapse of a mass concentration. The reason for it are the small-scale density inhomogeneities within the collapsing matter distribution which generate, via Poisson's equation, corresponding fluctuations in the gravitational field. These then scatter the infalling particles and, by this, the density inhomogeneities are further amplified.

This virialization occurs on a dynamical time, and once virialization is complete, the average density of the perturbation becomes, as noted earlier, $\langle \rho \rangle \approx 178\Omega_m^{-0.6}\bar{\rho}(t_{\text{collapse}})$

1.4.4. What are top-down and bottom-up growth?

This information is from Schneider (2006), pg. 286.

In a universe dominated by hot dark matter, all small perturbations cease to exist, and therefore the largest structures in the universe must form first, with galaxies fragmenting during the formation of larger structures. This top-down growth is incompatible with the fact that galaxies appear to have already collapsed, while superclusters are still in the linear overdensity regime. In a universe dominated by cold dark matter, small overdensities collapse first, and this bottom-up growth is consistent with observations.

1.4.5. How can the power spectrum be observed?

The matter power spectrum, if one assumes that baryons track dark matter, can be determined observationally recovering the two-point correlation function from galaxy surveys (Sec. 1.18).

1.4.6. How can the power spectrum constrain cosmological parameters?

This information is from Schneider (2006), Ch. 8.1.

The turnover of the power spectrum is determined by the wavenumber corresponding to the sound horizon at matter-radiation equality. This allows us to determine the shape parameter $\Omega_m h$, which can be combined with measurements of H_0 to obtain Ω_m . Detailed modelling of the power spectrum shows that the transfer function depends on Ω_b as well as Ω_m . As a result, this modelling can also derive the baryon to total matter ratio.

One important use of the dark matter power spectrum is to determine the shape and frequency of the baryon acoustic oscillations (Sec. 1.14).

1.4.7. How can we determine the dark matter mass function from perturbation analysis?

This information is from Schneider (2006), pg. 291 - 292.

As noted previously, a spherical region with an average density δ greater than some critical density will collapse. We can therefore back-calculate $\delta(\vec{x}, t)$ from the power spectrum³, smooth it out over some comoving radius R to determine the average density, and determine using the critical density (given the redshift and cosmological model), the normalized number density of relaxed dark matter halos. Since the power spectrum has a normalization factor that must be determined empirically, this normalized number density can then be scaled to the true number density using observations. The result is the Press-Schechter function $n(M, z)$ which describes the number density of halos of mass $> M$ at redshift z . See Fig. 8.

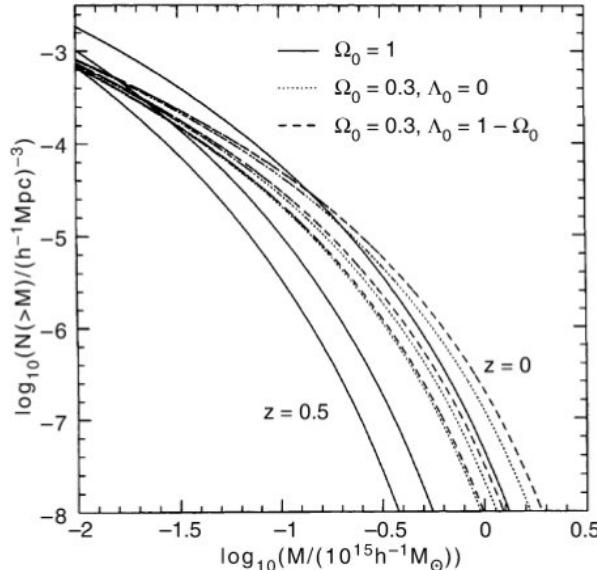


FIG. 8.— Number density of dark matter halos with mass $> M$ (i.e. a reverse cumulative model), computed from the PressSchechter model. The comoving number density is shown for three different redshifts and for three different cosmological models. The normalization of the density fluctuation field has been chosen such that the number density of halos with $M > 10^{14}h^{-1} M_\odot$ at $z = 0$ in all models agrees with the local number density of galaxy clusters. From Schneider (2006), his Fig. 7.7.

³ This only works if $P(k)$ alone is sufficient to describe δ , but because this distribution is Gaussian (for complicated reasons), $P(k)$ completely constrains δ .

Since there is only one normalization, a survey of galaxy cluster total masses at different redshifts (using hot gas and a mass-to-light conversion, gravitational microlensing, etc) can be used to determine cosmological parameters. This is because the minimum overdensity for collapse δ_{\min} is dependent on both the growth rate of overdensities and the expansion of the universe. Increasing Ω_m , for example, decreases $n(M, z)/n(M, 0)$, since massive halo growth is more extreme the higher Ω_m is. A large Ω_Λ damps massive halo growth.

1.5. Question 4

QUESTION: State and explain three key pieces of evidence for the Big Bang theory of the origin of the Universe.

This information is cribbed from Emberson (2012).

The Big Bang theory is the theory that the universe started off in an extremely hot, dense state, which then rapidly expanded, cooled, and became more tenuous over time. The Big Bang theory requires that at some point in the past a). the universe was born, b). the universe was extremely hot and c). objects were much closer together. The three key pieces of evidence are:

1. **Hubble's Law:** galaxies isotropically recede from our position with the relationship

$$\vec{v} = H_0 \vec{r} \quad (26)$$

known as Hubble's Law. As it turns out, moving into the frame of another galaxy ($\vec{r}' = \vec{r} - \vec{k}$, $\vec{v}' = \vec{v} - H_0 \vec{k} = H_0(\vec{r} - \vec{k}) = H_0 \vec{r}'$) does not change any observations. At larger distances, Hubble's Law breaks down (see Sec. 1.7), but the rate of expansion only increases with distance. Because of this isotropic radial motion outward, we can back-calculate a time when all the galaxies ought to be together at one point. This time is $t_0 = r/v = 1/H_0 \approx 14$ Gyr, the Hubble Time. This gives an age to the universe, and indicates that in the distant past everything was closer together.

2. **The Cosmic Microwave Background:** the cosmic microwave background (CMB) is a near perfect isotropic blackbody with a (current) $T_0 \approx 2.73$ K. For a blackbody, $\lambda_{\text{peak}} = 0.0029 \text{ mK}/T$, $U = aT^4$ and $n = \beta T^3$, which gives us $n \approx 400 \text{ cm}^{-3}$, $\epsilon \approx 0.25 \text{ eV cm}^{-3}$, and $\lambda \approx 2 \text{ mm}$. In Big Bang cosmology, this microwave background is the redshifted ($T \propto a^{-1}$) vestige of the surface of last scattering, when $T \approx 3000$ K and the universe became neutral enough for photons to travel unimpeded. This is evidence that the universe used to be hot.
3. **Big Bang Nucleosynthesis:** in the Big Bang theory, the lightest elements were created out of subatomic particles when the temperature dropped enough that the average photon was significantly below the binding energy of light elements. A detailed calculation of nucleosynthetic rates of H, D, He and Li during the first few minutes of the universe is consistent with the current abundances of light elements in the universe. See Sec. 1.8.

Additionally, no object has been found to be older than the currently accepted age of the universe, 13.7 Gyr. As we look back in time, we notice that the average galaxy in the universe looked considerably different - this evolution is consistent with Λ CDM cosmology, which has small, dense cores of dark matter forming due to gravitational instability, and then merging to form larger cores.

1.5.1. What is Olbers's Paradox?

Olbers's paradox is the apparent contradiction one has when an infinitely old, infinitely large universe with a fixed stellar density is assumed. In such a universe every single line of sight would eventually reach a star's photosphere. Since a typical photospheric temperature is ~ 5000 K and surface brightness is independent of distance, we would expect the entire sky to be at ~ 5000 K, roasting the Earth. Setting a finite age to the universe is one solution to the paradox; another would be that stars only formed in the last several billion years, and light from more distant stars have yet to reach us.

1.5.2. Are there Big Bang-less cosmologies?

It is impossible to generate a matter dominated universe for which there is no Big Bang. It is possible, however, for a Λ -dominated universe to be infinitely old, since an exponential (see Sec 1.1.4) never goes to zero. This is consistent with the steady state theory (Sec. 1.6).

1.6. Question 5

QUESTION: Define and describe the "tired light hypothesis" and the "steady state universe" as alternatives to the Big Bang. How have they been disproved observationally?

2. Galaxies

General Qualifying Exam Solutions: Galaxies

Starkman, Nathaniel

Lokken, Martine

Ludwig, Bethany

Sheere, Connor

Winch, Harrison

INTRODUCTION

Q1) MASS OF THE MILKY WAY

What is the total mass (in both dark matter and in stars) of the Milky Way galaxy? How does this compare to M31 and to the LMC? How is this mass determined?

Short Answer

Table 2.1: Galaxy Masses

Object	Total [M_{\odot}]	Stars [M_{\odot}]
Milky Way	$\approx 10^{12}$	$\approx 10^{10}$
Andromeda (M31)		$\approx 2 \times M_{MW}$
LMC	$\approx 10^{10}$	$\approx 10^{19}$

There are variety of methods to determine the total masses of these objects. For more details, besides the included notes, see Jo Bovy's detailed Galaxy Dynamics notes.

Methods:

1. Rotation curves: $M(< r) = \frac{rv_c^2}{G} \propto r$
2. Virial theorem: $M = \frac{1}{G} \frac{\sum w_i |v_i|^2}{\sum w_i / r_i}$
3. Velocity distribution cutoff of stars
4. Via Globular Clusters:
 - Escape velocity: $v_{esc} = \sqrt{\frac{GM(<r)}{r}}$
 - Photometrically for baryonic mass + scaling relations for the DM
5. Local group timing argument / Spherical Jeans equation $M(< r) = -\frac{r\sigma_r^2}{G} \left(\frac{d\ln(\nu\sigma_r^2)}{d\ln r} + 2\beta \right)$
6. Wolf Mass Estimate: $M(< r_{1/2}) = 3G^{-1}\sigma_{los}^2 r_{1/2}$

The easiest method for the LMC is Spherical Jeans.

The Spherical Jeans is hard for the Milky Way because we do not (yet!) have large data samples with three-dimensional velocities at large distance. Determining $\sigma_r(r)$ and especially $\beta(r)$ is difficult. At $r \gg R_0$, the solar radius, $\sigma_r(r)$ is approximately equal to $\sigma_{los}(r)$, because to a good approximation the Sun is sitting at the center of the Galaxy. In this case $v_{los} \approx v_r$. For the same reason, $\beta(r)$ can only be measured using tangential velocities (i.e., proper motions).

Q1) Herman ExtraGal Q2

EG 2

- List Masses
- invoke virial theorem to relate $M \propto v$.
- get v from 21 cm line width (Doppler broadened),
Tully-Fisher (max rot. v $\propto L$ profile of spiral)

- Mass determination of MW Galaxy

Galaxy	Mass w/ DM	Mass w/out DM
MW	$\sim 10^{12} M_{\odot}$	$\sim 10^{10} M_{\odot}$
M31	$\sim 10^{12} M_{\odot}$	$\sim 10^{11} M_{\odot}$
LMC	$\sim 10^{10} M_{\odot}$	$\sim 10^9 M_{\odot}$

- Assume galaxy is virialized (dynamically relaxed, in equilibrium) and determine rotation curve

$$U = -\frac{1}{2}K$$

$$-\frac{GMm}{r} = -mv^2$$

$$v = \sqrt{\frac{GM}{r}} \rightarrow M = \frac{v^2 r}{G}$$

- measure velocities (or velocity dispersion) of galaxy components as function of radius
- measure width of Doppler broadened 21 cm line from HI to get velocity dispersion
- could use Tully-Fisher relation to get max. rotational velocity based on luminosity profile of spiral; and Faber-Jackson relation to get velocity dispersion based on luminosity of ellipticals.

R - Max rotational velocity v. L

D - velocity dispersion v. L

S - spirals

E - ellipticals

T - Tully-Fisher

F - Faber-Jackson

- The mass of a galaxy is often expressed in terms of its M/L ratio.

Q1) Emberson ExtraGal Q2

87

QUESTION 2

What is the total mass (in both dark matter and in stars) of the Milky Way galaxy? How does this compare to M31 and to the LMC? How is this mass determined?

The masses of the MW and Andromeda galaxy (M31) have been measured by a variety of methods, but often with conflicting results that have led to a debate around which galaxy is more massive. Mass judging criteria based on observations of the surface brightness of the stellar halo, the number of globular clusters (which correlates with total mass albeit with scatter), and the amplitude of the inner gas rotation curve suggest that M31 is more massive. On the other hand, if the mass estimate is based on criteria such as the velocities of satellite galaxies, distant globular clusters, or tidal radii of nearby dwarf spheroidals, then the MW appears more massive. The current consensus, however, is that the two galaxies are roughly of the same mass ($\sim 10^{12} M_{\odot}$), with M31 probably the slightly more massive of the two, though this is based on the rather indirect mass estimates described above for M31 (Watkins et al. 2010).

On the other hand, the masses of the two galaxies are reasonably well constrained within the first few kpc from knowledge of their gas rotation curves via 21-cm radio observations (Carroll & Ostlie 2007, pg. 914). Of course, this only samples the inner regions of the galaxies, and in order to probe further out into the vast dark matter halos it is necessary to resort to satellite kinematics. Unfortunately, the uncertainties in such techniques are plagued by low sample sizes as well as the fact that there is seldom knowledge of the proper motion of the satellites to complement their observed radial velocity and distance data. Because of the latter, assumptions must be made on the eccentricities of the satellite orbits thereby affecting the mass determination; see equation (170) below.

In order to convert satellite kinematics into mass estimates we begin by analyzing the virial theorem for a spherically symmetric collection of N test particles (e.g. planetary nebulae, stars, globular clusters, satellite galaxies) orbiting a point mass M . For this situation the virial theorem dictates that

$$GM = \frac{\langle v^2 \rangle}{\langle 1/r \rangle}, \quad (170)$$

where angular brackets denote average values. If the distribution of test particles is spherically symmetric then $\langle v^2 \rangle = 3\langle v_r^2 \rangle$ and $\langle 1/r \rangle = 2/\pi\langle 1/R \rangle$, where v_r is the observed radial velocity and R the projected separation. Substituting these identities into equation (170) for a collection of N test particles yields a mass estimate of the form

$$M = \frac{3\pi}{2G} \frac{\sum_i v_r^2}{\sum_i 1/R_i}. \quad (171)$$

Despite its easy appearance, the virial theorem does not provide accurate mass estimates. There are many problems associated with its use including its failure to converge as $N \rightarrow \infty$ (Bahcall & Tremaine 1981).

Instead of using the virial theorem, a more reliable mass estimate is based on the projected mass $q \equiv v_r^2 R / G$. The variable q has dimensions of mass and with a suitable multiplicative factor can be used as an estimator to the mass M . For a general distribution of test particles it turns out that the expectation value of q is

$$\langle q \rangle = \frac{\pi M}{32} (3 - 2\langle e^2 \rangle), \quad (172)$$

where $\langle e^2 \rangle$ is the expectation value of the square of the eccentricities of the particles orbits. Using this relation and the definition of q it is straightforward to arrive at a mass estimate of the form

$$M = \frac{C}{G N} \frac{1}{i} \sum_i v_r^2 R_i, \quad (173)$$

where C is a constant of order unity depending on the test particles' eccentricities. Unlike the virial theorem method, the central value theorem applies to this method and guarantees that the sum will converge to the true mass M with an error proportional to $1/\sqrt{N}$ (Bahcall & Tremaine 1981).

Applying a modified form of equation (173) to 26 satellite galaxies of the Milky Way and 23 satellite galaxies of M31, Watkins et al. (2010) determine the masses of the two galaxies within 300 kpc from their centres to be $M_{\text{MW}} \sim 3 \times 10^{12} M_{\odot}$ and $M_{\text{M31}} \sim 1 \times 10^{12} M_{\odot}$. These values are rather volatile inasmuch as the exclusion of the satellite galaxies with ambiguous velocity and distance measures changes the mass estimates by a factor of roughly 2.

The value for M_{MW} is in good agreement with the study by Xue et al. (2008) in which 2401 blue horizontal branch (BHB) stars (which have high luminosities and nearly constant absolute magnitudes within a restricted colour range) from the SDSS are used to constrain the MW's circular velocity curve up to 60 kpc. From this the total mass within 60 kpc is determined and subsequently used to estimate the mass of the entire halo to be $M_{\text{MW}} \sim 1 \times 10^{12} M_{\odot}$. Of course, Newton's theorem asserts that any mass outside of the limiting radius of 60 kpc will have no observational effect in a spherical or elliptical system and so estimating the halo mass in this way requires an initial assumption on the structure of the dark matter halo. Indeed, the estimate by Xue et al. (2008) is based on the assumption of an NFW halo profile.

Schommer et al. (1992) measure the velocities of individual stars in 83 star clusters in the LMC to arrive at a mass estimate for the galaxy. Using equation (173) they find the mass of the LMC to be $M_{\text{LMC}} \sim 2 \times 10^{10} M_{\odot}$, roughly 1/100 that of the Milky Way. They compare this to an estimate based on a rotation curve constructed from their cluster rotation data in addition to earlier

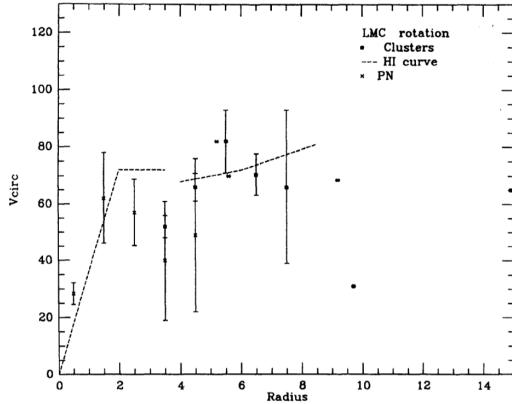


FIG. 29.— The rotation curve of the LMC built from rotational data of star clusters, H I, and PN. The dashed line shows the H I rotation curve whereas the points represent data from the clusters and PN. The error bars are the standard deviations of the mean of the velocities in the given bin while points with no error bars denote single objects. The bottom axis is in units of degrees and can be converted to physical lengths by noting that the LMC is 48 kpc from the MW. Image taken from Schommer et al. (1992).

data on the circular velocities of H I and planetary nebulae (PN) in the LMC. This is shown in Figure 29. To estimate the mass of the LMC from the rotation curve, one needs to compare the the circular velocity of the MW and the LMC at a distance of 8.5 kpc, the distance from the Sun to the centre of the MW. For the MW this value is roughly 200 km s^{-1} while for the LMC it is roughly 20 km s^{-1} since the circular velocity is roughly 20% larger than the rotation velocity (Weinberg 2000). Then since $M \propto Rv^2$ we have that $M_{\text{LMC}} \sim 1 \times 10^{10} M_{\odot}$, in good agreement with the value above. This estimate would obviously be improved with more measurements of mass tracers at large radii. In principle, this can be achieved by H I though at such large radii its kinematics may be disturbed by hydrodynamical processes from tidal interactions (i.e. the Magellanic Stream).

Another quick check on the mass of the LMC is to investigate its tidal interactions with the MW. From the point of view of the MW, the LMC is an oversize globular cluster. Its tidal radius is measurable and depends both on the MW rotation curve and the LMC mass (and also weakly on the LMC mass profile). The tidal radius of the LMC can be estimated by observing the extent of its stellar halo. From this its mass is estimated via

$$M_{\text{LMC}} = 2 \left(\frac{r_t}{R_{\text{LMC}}} \right)^3 M_{\text{MW}}, \quad (174)$$

where r_t is the tidal radius and R_{LMC} is the distance to the LMC. This analysis is considered by Weinberg (2000) in which a tidal radius of 10.8 kpc is used to constrain the mass of the LMC at $M_{\text{LMC}} \sim 2 \times 10^{10} M_{\odot}$.

Q1) Campbell ExtraGal Q2

1.2 Question 2

What is the total mass (in both dark matter and in stars) of the Milky Way galaxy? How does this compare to M31 and to the LMC? How is this mass determined?

1.2.1 Short answer

The MWG has a total mass of about $10^{12} M_{\odot}$. This is about the same as M31 (Andromeda), and larger than the LMC which is roughly $10^{10} M_{\odot}$. Assuming virial equilibrium, the mass can be determined from the flat part of the rotation curve via

$$V_0 = \sqrt{\frac{GM(< R)}{R_0}} [\text{m s}^{-1}]$$

If spiral galaxies are being observed, the Tully-Fisher relation $L \propto v_{\text{max}}^4$ can be used to determine the maximum orbital velocity in replacement of the rotation curve. Similarly, if elliptical galaxies are being observed, the Faber-Jackson relation $L \propto \sigma_v^4$ can be used instead.

1.2.2 Additional context

The components of the Milky Way Galaxy (MWG) have total masses as follows: a disk mass of $4.5 \times 10^{10} M_{\odot}$, bulge mass of $4.5 \times 10^9 M_{\odot}$, dark halo mass of $2 \times 10^{12} M_{\odot}$, and BH mass of $4 \times 10^6 M_{\odot}$. The Galactic disk rotates, with rotational velocity $V(R)$ depending on the distance R from the center. We can estimate the mass of the Galaxy from the distribution of the stellar light and the mean mass-to-light ratio of the stellar population, since gas and dust represent less than $\sim 10\%$ of the mass of the stars. From this mass estimate we can predict the rotational velocity as a function of radius simply from Newtonian mechanics. However, the observed rotational velocity of the Sun around the Galactic center is significantly higher than would be expected from the observed mass distribution. If $M(< R_0)$ is the mass inside a sphere around the Galactic center with radius $R_0 = 8 \text{ kpc}$, then the rotational velocity from Newtonian mechanics is

$$V_0 = \sqrt{\frac{GM(< R)}{R_0}} [\text{m s}^{-1}]$$

From the visible matter in stars we would expect a rotational velocity of 160 km s , but we observe $V_0 = 220 \text{ km s}$ (see Figure 3). This discrepancy, and the shape of the rotation curve $V(R)$ for larger distances R from the Galactic center, indicates that our Galaxy contains significantly more mass than is visible in the form of stars. This additional mass is called dark matter. Its physical nature is still unknown. The main candidates are weakly interacting elementary particles like those postulated by some elementary particle theories, but they have yet not been detected in the laboratory. Macroscopic objects (i.e., celestial bodies) are also in principle viable candidates if they emit very little light.

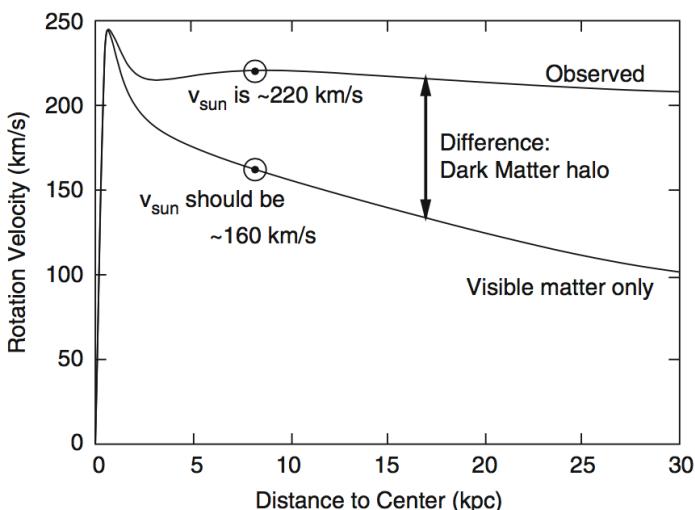


Figure 3: The upper curve is the observed rotation curve $V(R)$ of our Galaxy, i.e., the rotational velocity of stars and gas around the Galactic center as a function of their galactocentric distance. The lower curve is the rotation curve that we would predict based solely on the observed stellar mass of the Galaxy. The difference between these two curves is ascribed to the presence of dark matter, in which the Milky Way disk is embedded. This image is adapted from Nick Strobel's webpage at www.astronomynotes.com. Image taken from Schneider (2006).

Q1) Campbell ExtraGal Q2

1.2.3 Follow-up Questions

- If rotation curve/lensing measurements were instead due to modified gravity, how could we tell?
- What are Ω_m and Ω_b estimated to be? Why does the ratio of the two differ from the star/total mass ratio you have here? Where is all the extra baryonic mass?
- When you say we estimate stellar mass by “counting stars”, what does that mean?

Q2) NUCLEAR BLACK HOLES

What evidence is there that most galaxies contain nuclear black holes? How do those black holes interact with their host galaxies?

Q3) AGN

What are AGN? Describe different observational classes of them and how they may relate to each other.

Q4) QUASARS

Draw a spectrum of a high-redshift quasar. What do quasar emission lines typically look like? Explain what we see in the spectrum at rest wavelengths bluer than 1216 Angstroms.

Q5) METHODS OF GALAXY MASS DETERMINATION

Describe three different methods used in the determination of the mass of a galaxy cluster.

Q6) SEDs OF SINGLE-BURST GALAXIES

Draw the spectral energy distribution (SED) of a galaxy formed by a single burst of star formation at the ages of 10 Myrs, 2 Gyrs, and 10 Gyr. Please highlight the change over time in the 4000 Angstrom break.

Q7) GALACTIC SPIRAL STRUCTURE

What is galactic spiral structure and why is it thought to occur?

Q8) STELLAR INITIAL MASS FUNCITON (IMF)

What is a stellar Initial Mass Function (IMF)? Explain how it is determined and how it is used.

Q9) STELLAR POPULATIONS IN THE GALAXY

Characterize the stellar populations in the following regions: i) the Galactic bulge ii) the Galactic disk, outside of star clusters iii) open star clusters iv) globular clusters v) the Galactic halo

Q10) G-DWARF PROBLEM IN THE SOLAR NEIGHBOURHOOD

What is the G-dwarf problem in the solar neighbourhood?

Q11) STELLAR ORBITS IN POTENTIAL

Describe the orbits of stars in a galactic disk and in galactic spheroid.

Q12) DYNAMICAL RELAXATION

What is dynamical relaxation? Explain why this operates in star clusters but not in an elliptical galaxy.

Q13) DYNAMICAL FRICTION

What is dynamical friction? Explain how this operates in the merger of a small galaxy into a large one.

3. Stars and Planets

General Qualifying Exam Solutions: Physics and Fundamentals

Starkman, Nathaniel

Lokken, Martine

Ludwig, Bethany

Winch, Harrison

INTRODUCTION

4. Physics and Fundamentals

General Qualifying Exam Solutions: Physics and Fundamentals
Starkman, Nathaniel
Lokken, Martine
Ludwig, Bethany
Winch, Harrison

INTRODUCTION

Q1) INTERFEROMETRY

A two-element interferometer consists of two telescopes whose light is combined and interfered. Explain how this might be accomplished in practice, and sketch the response of such an interferometer to a nearby red giant star, as a function of the (projected) separation between the two telescopes.

Campbell Physics Q2

1.2 Question 2

A two-element interferometer consists of two telescopes whose light is combined and interfered. Sketch the response of such an interferometer to a nearby red giant star, as a function of the (projected) separation between the two telescopes. The red giant subtends one-fiftieth of an arc second on the sky, and the telescope operates at a wavelength of 2 microns.

1.2.1 Short answer

In a similar way that a single slit experiment results in an interference pattern depending on the slit size, an analogous interference pattern will emerge from a two-element interferometer and will depend on the projected baseline. Since the red giant star is resolved to 1/50 arcseconds, let's determine the projected baseline this corresponds to. Before we can do this, we must first convert the angular size from arcseconds into radians:

$$\theta = \left(\frac{1}{50} \text{ arcsec} \right) \left(\frac{1 \text{ deg}}{3600 \text{ arcsec}} \right) \left(\frac{1 \text{ rad}}{57 \text{ deg}} \right) = 10^{-7} \text{ rad.}$$

Now to determine the projected baseline:

$$\begin{aligned} \theta &\sim 2.2 \left(\frac{\lambda}{D} \right) [\text{rad}] \\ D &\sim 2.2 \left(\frac{\lambda}{\theta} \right) [\text{m}] \\ &= 2.2 \left(\frac{2 \times 10^{-6} \text{ m}}{10^{-7} \text{ rad}} \right) = 44 \text{ [m].} \end{aligned}$$

Thus, the resulting response function will be a Bessel function whose minima are at integer multiples of the 44 m projected baseline:

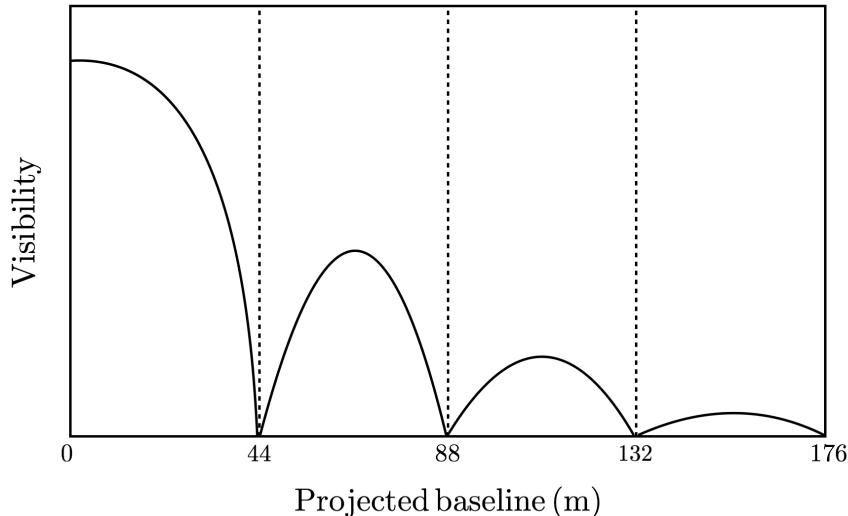


Figure 5: Response function of a two-element interferometer as a function of the projected baseline observing a resolved red giant star with an angular resolution of 1/50 arcseconds.

1.2.2 Additional context

The two-element quasi-monochromatic interferometer: The simplest radio interferometer is a pair of radio telescopes whose voltage outputs are correlated (multiplied and averaged), and even the most elaborate interferometers with $N \gg 2$ antennas, often called **elements**, can be treated as $N(N - 1)/2$ independent two-element interferometers. Figure 6 shows two identical dishes separated by the **baseline vector** \vec{b} of length b that points from antenna 1 to antenna 2. Both dishes point in the same direction specified by the unit vector \hat{s} , and θ is the angle between \vec{b} and \hat{s} . Plane waves from a distant point source in this direction must travel an extra distance $\vec{b} \cdot \hat{s} = b \cos \theta$ to reach antenna 1, so the output of antenna 1 is the same as that of antenna 2, but it lags in time by the geometric delay

Campbell Physics Q2

$$\tau_g = \frac{\vec{b} \cdot \hat{s}}{c} [\text{s}].$$

For simplicity, we first consider a quasi-monochromatic interferometer, one that responds only to radiation in a very narrow band $\Delta\nu \ll 2\pi/\tau_g$ centered on frequency $\nu = \omega/(2\pi)$. Then the output voltages of antennas 1 and 2 at time t can be written as

$$V_1 = V \cos[\omega(t - \tau_g)] [\text{V}]$$

$$V_2 = V \cos(\omega t) [\text{V}].$$

These output voltages are amplified versions of the antenna input voltages; they have not passed through square-law detectors. Instead, a **correlator** multiplies these two voltages to yield the product

$$V_1 V_2 = V^2 \cos[\omega(t - \tau_g)] \cos(\omega t) = \left(\frac{V^2}{2} \right) [\cos(2\omega t - \omega\tau_g) + \cos(\omega\tau_g)] [\text{V}]$$

which follows directly from the trigonometric identity $\cos x \cos y = [\cos(x+y) + \cos(x-y)]/2$. The correlator also takes a time average long enough ($\Delta t \gg (2\omega)^{-1}$) to remove the high-frequency term $\cos(2\omega t - \omega\tau_g)$ from the **correlator response** (output voltage) R and keep only the slowly varying term

$$R = \langle V_1 V_2 \rangle = \left(\frac{V^2}{2} \right) \cos(\omega\tau_g) [\text{V}^2].$$

The voltages V_1 and V_2 are proportional to the electric field produced by the source multiplied by the voltage gains of the two antennas and receivers. Thus the correlator output amplitude $V^2/2$ is proportional to the flux density S of the point source multiplied by $\sqrt{A_1 A_2}$, where A_1 and A_2 are the effective collecting areas of the two antennas.

Notice that the time-averaged response R of a multiplying interferometer is zero. There is no DC output, so fluctuations in receiver gain do not act on the whole system temperature T_s as for a total-power observation with a single dish. Uncorrelated noise power from very extended radio sources such as the CMB and the atmosphere over the telescopes, also averages to zero in the correlator response. Short interference pulses with duration $t \ll |b|/c$ are also suppressed because each pulse does not reach both telescopes simultaneously. Likewise, a multiplying radio interferometer differs from a classical **adding interferometer**, such as the optical Michelson interferometer, that adds the uncorrelated noise power contributions.

The correlator output voltage $R = (V^2/2) \cos(\omega\tau_g)$ varies sinusoidally as the Earth's rotation changes the source direction relative to the baseline vector. These sinusoids are called **fringes**, and the **fringe phase**

$$\phi = \omega\tau_g = \frac{\omega}{c} b \cos \theta [\text{rad}]$$

depends on θ as follows:

$$\frac{d\phi}{d\theta} = \frac{\omega}{c} b \sin \theta = 2\pi \left(\frac{b \sin \theta}{\lambda} \right) [\text{dimensionless}].$$

The **fringe period** $\phi = 2\pi$ corresponds to an angular shift $\theta = \lambda/(b \sin \theta)$. The fringe phase is an exquisitely sensitive measure of source position if the **projected baseline** $b \sin \theta$ is many wavelengths long. Note that fringe phase and hence measured source position is not affected by small tracking errors of the individual telescopes. It depends on time, and times can be measured by clocks with much higher accuracy than angles (ratios of lengths of moving telescope parts) can be measured by rulers. Also, an interferometer whose baseline is horizontal is not affected by the plane-parallel component of atmospheric refraction, which delays the signals reaching both telescopes equally. Consequently, interferometers can determine the positions of compact radio sources with unmatched accuracy. Absolute positions with errors as small as $\sigma_\theta \approx 10^{-3}$ arcsec and differential positions with errors down to $\sigma_\theta \approx 10^{-5}$ arcsec $< 10^{-10}$ rad have frequently been measured.

If the individual antennas comprising an interferometer were isotropic, the interferometer point-source response would be a sinusoid spanning the sky. Such an interferometer is sensitive to only one Fourier component of the sky brightness distribution: the component with angular period $\lambda/(b \sin \theta)$. The response R of a two-element interferometer with directive antennas is that sinusoid multiplied by the product of the voltage patterns of the individual antennas. Normally the two antennas are identical, so this product is the power pattern of the individual antennas and is called the **primary beam** of the interferometer. The primary beam is usually a Gaussian much wider than a fringe period, as indicated in Figure 6. The convolution theorem states that the Fourier transform of the product of two functions is the convolution of their Fourier transforms, so the interferometer with directive antennas responds to

Campbell Physics Q2

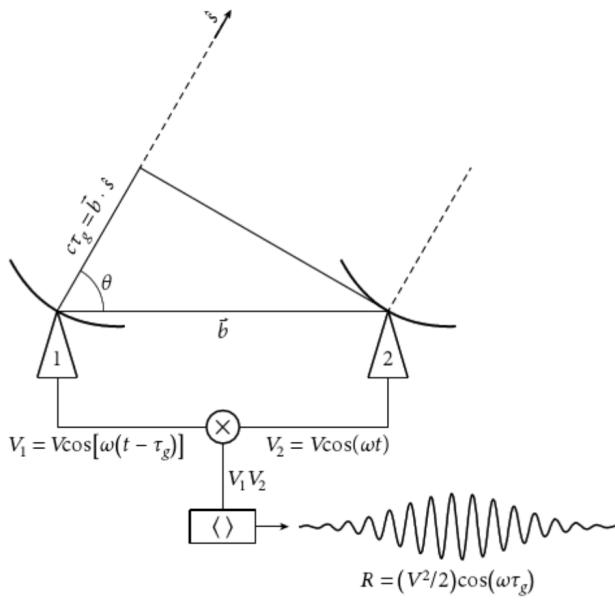


Figure 6: This block diagram shows the components of a two-element quasi-monochromatic **multiplying interferometer** observing in a very narrow radio frequency range centered on $\nu = \omega/(2\pi)$. \hat{s} is the unit vector in the direction of a distant point source and \vec{b} is the baseline vector pointing from antenna 1 to antenna 2. The output voltage V_1 of antenna 1 is the same as the output voltage V_2 of antenna 2, but it is retarded by the geometric delay $\tau_g = \vec{b} \cdot \hat{s}/c$ representing the additional light-travel delay to antenna 1 for a plane wavefront from a source at angle θ from the **baseline vector**. These voltages are amplified, multiplied (\times), and time averaged ($\langle \rangle$) by the **correlator** to yield an output response whose amplitude R is proportional to the flux density of the point source and whose phase $(\omega \tau_g)$ depends on the delay and the frequency. The quasi-sinusoidal output **fringe** shown occurs if the source direction in the interferometer frame is changing at a constant rate $d\theta/dt$. The broad Gaussian envelope of the fringe shows the primary-beam attenuation as the source passes through the beam of the dishes. Figure taken from Condon & Random (2016).

a finite range of angular frequencies centered on $b \sin \theta / \lambda$. Because the antenna diameters D must be smaller than the baseline b (else the antennas would overlap), the angular frequency response cannot extend to zero and the interferometer cannot detect an isotropic source – the bulk of the 3 K CMB for example. The missing short spacings ($b < D$) can be provided by a single-dish telescope with diameter $D > b$. Thus the $D = 100$ m GBT can fill in the missing baselines $b < 25$ m that the $D = 25$ m VLA dishes cannot obtain.

Improving the instantaneous point-source response pattern of an interferometer requires more Fourier components; that is, more baselines. An interferometer with N antennas contains $N(N - 1)/2$ pairs of antennas, each of which is a two-element interferometer, so the instantaneous **synthesized beam** (the point-source response obtained by averaging the outputs of all of the two-element interferometers) rapidly approaches a Gaussian as N increases. The instantaneous point-source responses of a two-element interferometer with projected baseline length b , a three-element interferometer with three baselines (projected lengths $b/3, 2b/3$, and b), and a four-element interferometer with six baselines (projected lengths $b/6, 2b/6, 3b/6, 4b/6, 5b/6$, and b) are shown in Figure 7.

Most radio sources are stationary; that is, their brightness distributions do not change significantly on the timescales of astronomical observations. For stationary sources, a two-element interferometer with movable antennas could make $N(N - 1)/2$ observations to duplicate one observation with an N -element interferometer.

Slightly extended sources and the complex correlator: The response $R = (V^2/2) \cos(\omega \tau_g)$ of the quasi-monochromatic two-element interferometer with a “cosine” correlator (Figure 6) to a spatially incoherent slightly extended (much smaller than the primary beamwidth) source with sky brightness distribution $I_\nu(\hat{s})$ near frequency $\nu = \omega/(2\pi)$ is obtained by treating the extended source as the sum of independent point sources:

$$R_c = \int I(\hat{s}) \cos(2\pi\nu \vec{b} \cdot \hat{s}/c) d\Omega = \int I(\hat{s}) \cos(2\pi \vec{b} \cdot \hat{s}/\lambda) d\Omega [V^2].$$

Notice that the even cosine function in this response is sensitive only to the even (inversion-symmetric) part I_E of an arbitrary source brightness distribution, which can be written as the sum of even and odd (anti-symmetric) parts: $I = I_E + I_O$. To detect the odd part I_O we need a “sine” correlator whose output is odd, $R = (V^2/2) \sin(\omega \tau_g)$. This can be implemented by a second correlator that follows a $\pi/2$ rad = 90° phase delay inserted into the output of one antenna because $\sin(\omega \tau_g) = \cos(\omega \tau_g - \pi/2)$. Then

$$R_s = \int I(\hat{s}) \sin(2\pi \vec{b} \cdot \hat{s}/\lambda) d\Omega [V^2]$$

The combination of cosine and sine correlators is called a **complex correlator** because it is mathematically convenient to treat the cosines and sines as complex exponentials using Euler’s formula

$$e^{i\phi} = \cos \phi + i \sin \phi [\text{rad}].$$

Campbell Physics Q2

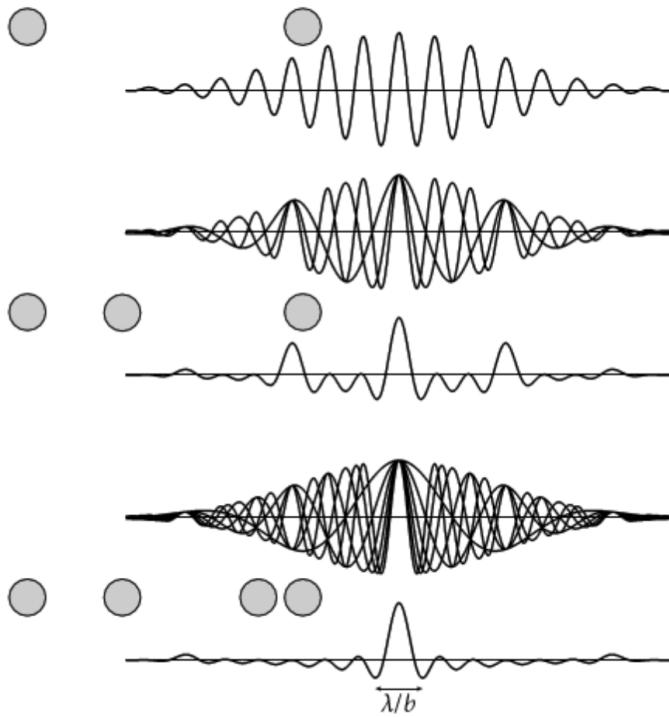


Figure 7: The instantaneous point-source responses of interferometers with overall projected length b and two, three, or four antennas distributed as shown are indicated by the thick curves. The synthesized main beam of the four-element interferometer is nearly Gaussian with angular resolution $\theta \approx \lambda/b$, but the sidelobes are still significant and there is a broad negative ‘bowl’ caused by the lack of spacings shorter than the diameter of an individual antenna. Thus the **synthesized beam** is sometimes called the **dirty beam**. The instantaneous dirty beam of the multi-element interferometer is the arithmetic mean of the individual responses of its component two-element interferometers. The individual responses of the three two-element interferometers comprising the three-element interferometer and of the six two-element interferometers comprising the four-element interferometer are plotted as thin curves. Figure taken from Condon & Random (2016).

The **complex visibility** is defined by

$$\mathcal{V} \equiv R_c - iR_s \text{ [V}^2\text{]}$$

which can be written in the form

$$\mathcal{V} = Ae^{-i\phi} \text{ [rad]},$$

where

$$A = \sqrt{R_c^2 + R_s^2} \text{ [dimensionless]}$$

is the **visibility amplitude** and

$$\phi = \arctan\left(\frac{R_s}{c}\right) \text{ [rad]}$$

is the **visibility phase**. The response to an extended source with brightness distribution $I(\hat{s})$ of the two-element quasi-monochromatic interferometer with a complex correlator is the complex visibility

$$\mathcal{V} = \int I(\hat{s}) \exp(-i2\pi\vec{b} \cdot \hat{s}/\lambda) d\Omega \text{ [V}^2\text{].}$$

Fringe patterns and Fourier transforms: Interferometry begins with the Youngs slits fringe pattern (Fig. 1). With a single point source emitting coherent radiation, interference fringes are observed, with constructive and destructive interference observed as the relative delay of the two interfering rays changes; the separation of the fringes is $/d$, the wavelength of the light divided by the slit separation.

If the source is made wider (Figure 8b), we can think of it as a sequence of point sources each of which emit radiation which is uncorrelated with the emission from the others. It follows that the total interference intensity pattern is the sum of the individual patterns. Since an angular displacement in the source produces an equal angular displacement in the fringe pattern, as the source size approaches λ/d the fringe patterns will add to give a constant illumination (Figure 8c). In this case, the fringe visibility (defined as the difference between maximum and minimum intensity, normalized by the sum of maximum and minimum intensity) drops to zero. Conversely, when the angular size of the source is λ/d , the fringe visibility is 1; this corresponds to a situation in which the source size is smaller than the angular resolution of the interferometer, and only an upper limit of order λ/d can be obtained on it.

Now suppose that the slit spacing d is decreased. For the same size of source, this produces less ‘washing-out’ of the fringes, because the same displacement of the source now produces much less displacement of the fringe patterns as a fraction of the fringe separation λ/d (Figure 8d). The smaller the slit separation,

Campbell Physics Q2

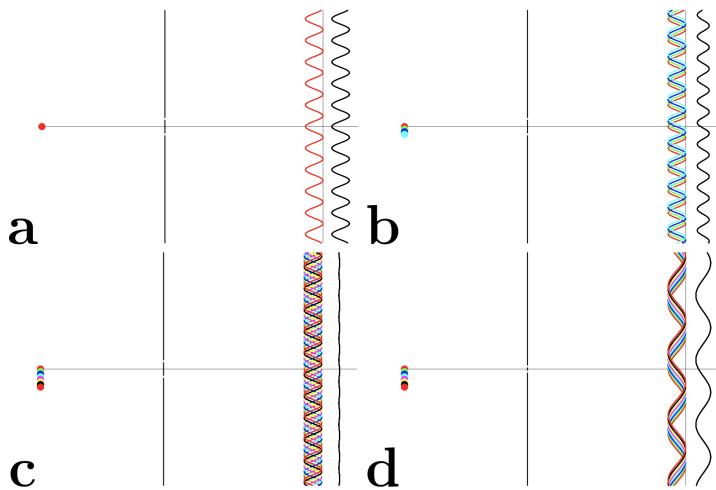
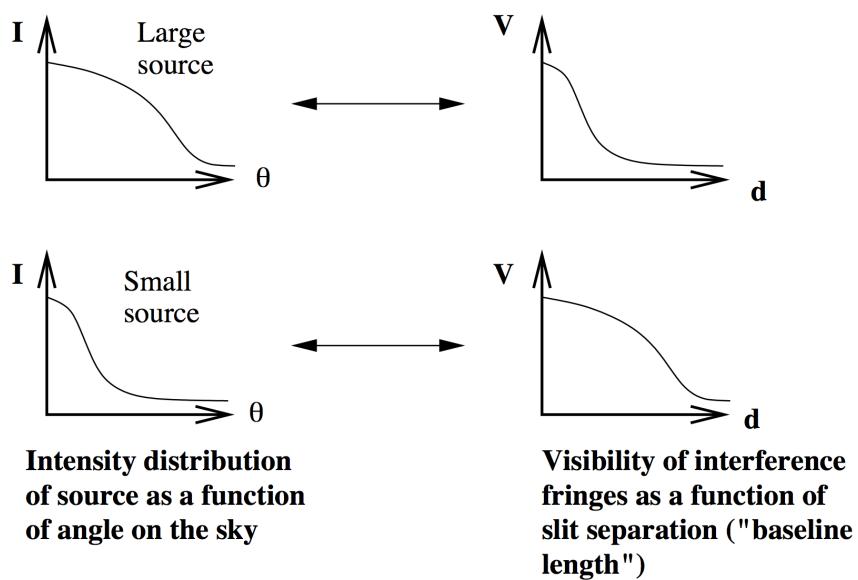


Figure 8: Young's slits in various situations. In each panel the source is shown on the left, and on the right of the slit are shown the fringe patterns separately for each part of the source and then the added fringe pattern. (a): The basic two-slit pattern, showing fringes an angular distance λ/d apart. (b): The effect of increasing the source size. An angular shift of the source position by θ shifts the fringe patterns by θ the other way. Since the patterns come from mutually incoherent sources, the intensity patterns add to give a pattern of reduced visibility. (c): When the size of the source reaches λ/d , the fringes add to give zero visibility. (d): If the slit spacing is then reduced, the fringe spacing increases, and the same size of source is still able to give visible fringes: the source would need to be increased in size to λ/d_{new} in order to wash out the fringes. Figure taken from Jackson.

Figure 9:
Relation between source brightness as a function of angular distance and visibility of interference fringes as a function of slit separation (baseline length). Figure taken from Jackson.



the larger the source size that can be probed using interferometry.

The situation is summarized in Figure 9. If we plot, for a given source distribution, the way in which visibility varies with slit separation, it can be seen that for small sources the visibility remains high out to large slit separation (in the limit of point sources, to infinite slit separation), whereas large sources produce visibility patterns which fall off quickly as the slit separation increases.

The relation between $I(\theta)$ and $V(d)$ represented here is one which maps a large Gaussian into a small Gaussian, and vice versa, and it is fairly obvious that it is a Fourier transform.¹

1.2.3 Follow-up Questions

- What do the minima in the response function tell you?

¹This is known as the Van Cittert-Zernicke theorem.

Campbell Physics Q2

1.3 Question 3

What's the minimum mass of a black hole you could survive a fall through the event horizon without being ripped to shreds? Why would you be ripped to shreds for smaller black holes? How does this relate to the BH mass range for which we expect tidal disruption flares caused by shredding main-sequence stars?

1.3.1 Short answer

We can find the minimum mass using the equation for the tidal force one would experience:

$$F_T = \left(\frac{dg}{dr} \right) \ell = \frac{d}{dr} \left(-\frac{GM}{r^2} \right) \ell = -GM\ell \frac{d}{dr} r^{-2} = -GM\ell(-2)r^{-3} = \frac{2GM\ell}{r^3} [\text{N}].$$

Setting the radius equal to the Schwarzschild radius $r_s \equiv 2GM/c^2$,

$$F_T = \frac{2GM\ell}{r_S^3} = 2GM\ell \left(\frac{c^2}{2GM} \right)^3 = \frac{\ell c^6}{4G^2 M^2} [\text{N}],$$

we can now solve for the mass of the black hole M :

$$M = \sqrt{\frac{\ell c^6}{4G^2 F_T}} [\text{kg}].$$

Let's set $\ell \sim 1 \text{ m}$ for an order-of-magnitude length of the human body, and $F_T \sim 100 \text{ g} = 10^4 \text{ N}$ for the maximum force that the body can withstand. We can now use this to find the minimum mass of a black hole that you could survive a fall through the event horizon without being ripped to shreds:

$$M = \sqrt{\frac{(1 \text{ m})(2.998 \times 10^8 \text{ m s}^{-1})^6}{4(6.67 \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2})^2 (10^4 \text{ N})}} = 10^{34} \text{ kg} \sim 10^3 \text{ M}_\odot.$$

Since we can see that the tidal force $F_T \sim M^{-2}$, you would be ripped to shreds for smaller black holes because the resulting tidal force on the body would be greater the smaller the black hole mass is.

1.3.2 Follow-up Questions

- How would you estimate the maximum tidal acceleration a star can withstand?
- Why is it enough to know if the surface of the star will be disrupted?

Q2) DETECTOR COOLING

You don't usually need to cool down the detectors for short wavelength (e.g., X-ray) observations, but it's critical to cool down the detectors in long wavelength (e.g., far-IR) observations. Why is this, and why is it usually less essential or unnecessary for radio observations?

Campbell Physics Q11

1.11 Question 11

You don't usually need to cool down the detectors for short wavelength (e.g., X-ray) observations, but it's critical to cool down the detectors in long wavelength (e.g., far-IR) observations. Why is this, and why is it not necessary for radio observations?

1.11.1 Short answer

IR detectors use CCDs which are made up of semi-conducting silicon material; these detectors, when hit with a photon of the 'right' energy, will release a photoelectron which is later read by the CCD. The required energy for an incoming photon to release a photoelectron must be greater than silicon's band gap energy ($> 1.14 \text{ eV}$). Silicon has a useful photoelectric effect range of $1.1 - 10 \text{ eV}$, covering the range of near-IR to soft x-ray. While CCDs are generally more sensitive at higher temperatures, this results in a dark current of photoelectrons which cannot be distinguished from astrophysical sources. Thus, cooling detectors for short wavelength observations is necessary to reduce these dark currents specifically because they are sensitive to the IR whereas short wavelength observations are not. This is not necessary for radio observations because these observations use bolometers (instead of CCDs) which detect electromagnetic fields (instead of photons).

1.11.2 Additional context

Why use CCDs?: Most astronomical detectors in use today at professional observatories, as well as with many amateur telescopes, are CCDs. This fact alone gives an impression that there must be something very special or useful about CCDs; otherwise why all the fuss? CCDs have revolutionized modern astronomy. They will take their place in astronomical history along with other important discoveries such as the telescope, photographic plates, prisms, and spectroscopy. The contribution to our knowledge of the heavens brought about by CCDs is astounding, even more so when one considers that they have been in use for only about thirty years.

First introduced as electronic analogs to magnetic bubble memory at Bell labs, CCDs provided their first astronomical image in 1975 when scientists from the Jet Propulsion Laboratory imaged the planet Uranus at a wavelength of 8900\AA . This observation used the University of Arizona 61" telescope atop Mt. Lemmon and a prototype version of a CCD made by Texas Instruments Corporation as part of a development project for NASA spacecraft missions.

During the past ten years, tremendous progress has been made in the manufacturing process and therefore in the properties of the CCD itself. These improvements have allowed much lower noise properties for CCDs, thereby increasing their overall efficiency in astronomy. In addition, larger format devices have been produced and the readout times are much shorter, approaching $1 - 2$ seconds even for arrays as large as 1024 pixels square. This latter advance is mainly due to the availability of high-speed, low-power and low-noise CCD controllers. The driving technology for CCD manufacturing is for items such as copy machines, TV cameras, and digital cameras, but the requirements for low noise, excellent pixel cosmetics, and nearly perfect performance is still firmly rooted in astronomy. We outline below two of the important reasons why CCDs are considered as essentially the perfect imaging device.

Noise properties: The usefulness of a detector is very often determined by the amount of inherent noise within the device itself. Suffice it to say that modern astronomical CCDs are almost noise free. The original line of photosensitive electronic array detectors, such as television-type imaging detectors, vidicons, silicon intensified targets, and image-dissector scanners, all had very high noise properties. For comparison, silicon intensified target imagers (SITs) had a noise level upon readout of 800 electrons per picture element. Some very good systems of this type could be produced with read noise values of only 200 electrons. The first CCDs had readout noise levels similar to this latter value, while modern CCDs have noise values of ten down to two electrons per pixel per readout. The large noise levels present in early array detectors not only limited the SNR obtainable for a given measurement, they also severely limited the total dynamic range available to the camera. Another "feature" of older, higher noise CCDs was the decision an astronomer had to make about co-addition of frames. Since the read noise adds as its square to the total noise budget, adding two frames resulted in a much higher read noise contribution. Today, with typical read noise values of $2 - 5$ electrons, co-addition is essentially equal to a single exposure of longer integration time.

CCD operation: The simplest and very understandable analogy for the operation of a CCD is also one that has been used numerous times for this purpose. This is the "water bucket" idea in which buckets represent pixels on the CCD array, and a rainstorm provides the incoming photons (rain drops). Imagine a field covered with buckets aligned neatly in rows and columns throughout the entirety of the area (Figure 18). After the rainstorm (CCD integration), each bucket is transferred in turn and metered to

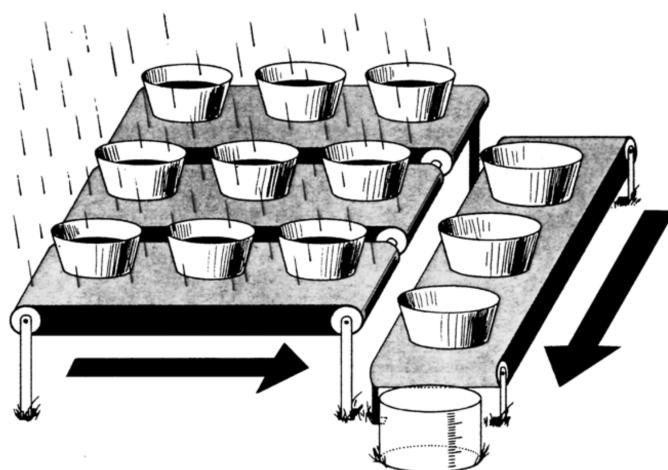


Figure 18:
CCDs can be likened to an array of buckets that are placed in a field and collect water during a rainstorm. After the storm, each bucket is moved along conveyor belts until it reaches a metering station. The water collected in each field bucket is then emptied into the metering bucket within which it can be measured. From Janesick & Blouke (1987). Figure taken from Howell (2006).

determine the amount of water collected. A written record (final CCD image) of the amount of water in each bucket will thus provide a two-dimensional record of the rainfall within the field.

Referring to the actual mechanisms at work within a CCD, we start with the method of charge generation within a pixel: the photoelectric effect.³ Incoming photons strike the silicon within a pixel and are easily absorbed if they possess the correct wavelength (energy). Silicon has a band gap energy of 1.14 eV, and so it easily absorbs light of energy 1.1 to 4 eV (3000 to 11000 Å).⁴ Photon absorption causes the silicon to give up a valence electron and move it into the conduction band. Photons of energy 1.1 eV to near 4 or so eV generate single **electron-hole pairs**, whereas those of higher energy produce multiple pairs. Left to themselves, these conduction band electrons would recombine back into the valence level within approximately 100 μ s. Silicon has a useful photoelectric effect range of 1.1 to about 10 eV, which covers the near-IR to soft X-ray region. Above and below these limits, the CCD material appears transparent to the incoming photons.

Once electrons have been freed to the conduction band of the silicon, they must be collected and held in place until readout occurs. The details of the actual construction of each pixel within a CCD (that is, the formation of the MIS capacitor with its doped silicon, layers of silicon dioxide, etc.) are beyond the scope of this discussion, but suffice it to say that each pixel has a structure allowing applied voltages to be placed on sub-pixel sized electrodes called **gates**. These gate structures provide each pixel with the ability to collect the freed electrons and hold them in a potential well until the end of the exposure. In a typical arrangement, each pixel has associated with it three gates, each of which can be set to a different voltage potential. The voltages are controlled by clock circuits with every third gate connected to the same clock. Figure 19 illustrates this clocking scheme for a typical three-phase device.

We note in Figure 19 that, when an exposure ends, the clock voltages are manipulated such that the electrons that have been collected and held in each pixel's +10 volt potential well by clock voltage V3 can now be shifted within the device. Note that electrons created anywhere within the pixel during the exposure (where each pixel has a surface area equal to the total area under all three gates) will be forced to migrate toward the deepest potential well. When the exposure is terminated and CCD readout begins, the voltages applied to each gate are cycled (this process is called **clocking the device**) such that the charge stored within each pixel during the integration is electronically shifted. A simple change in the voltage potentials (V3 goes to +5 volts, while V1 becomes +10 volts and so on) allows the charge to be shifted in a serial fashion along columns from one CCD pixel to another throughout the array. The transfer of the total charge from location to location within the array is not without losses. Each charge transfer (one of which occurs for each voltage change or clock cycle) has an associated efficiency. This efficiency value is the percent of charge transferred compared with that which was actually collected. Modern values for the charge transfer efficiency (CTE) are approaching 0.999999 (i.e., 99.9999% efficient) for each transfer.

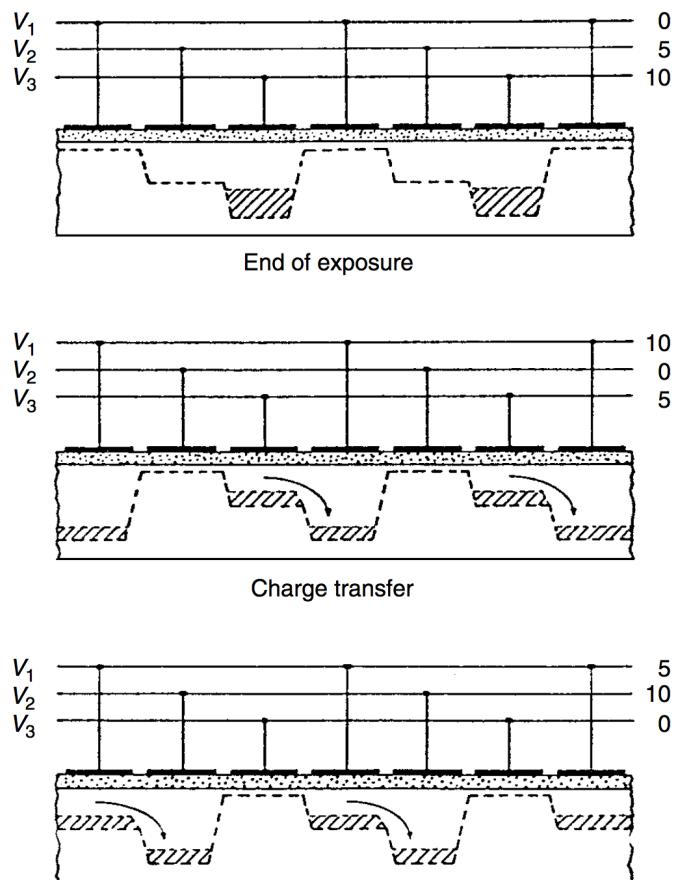
Each column in the array is connected in parallel and thus each pixel shift is mimicked throughout the entire array simultaneously. One clock cycle moves each row of pixels up one column, with the top row being shifted off the array into what is called the **output shift register** or **horizontal shift register**. This register is simply another row of pixels hidden from view (i.e., not exposed to incident light) and serves as the transition between active rows on the array and the output of the device. Once an entire

³Albert Einstein received his Nobel Prize mainly for his work on the photoelectric effect, not, as many think, for relativity.

⁴The energy of a photon of a given wavelength (in electron volts) is given by $E_{\text{eV}} = 12407/\lambda(\text{\AA})$.

Campbell Physics Q11

Figure 19:
Schematic voltage operation of a typical three-phase CCD. The clock voltages are shown at three times during the readout process, indicating their clock cycle of 0, 10, and 5 volts. One clock cycle causes the stored charge within a pixel to be transferred to its neighboring pixel. CCD readout continues until all the pixels have had their charge transferred completely out of the array and through the A/D converter. From Walker (1987). Figure taken from Howell (2006).



row is shifted into the output register, and before any further row shifts on the active area occur, each pixel in the output register is shifted out one at a time (in a similar manner as before) into the output electronics. Here, the charge collected within each pixel is measured as a voltage and converted into an output digital number. Each pixel's collected charge is sensed and amplified by an **output amplifier**. CCD output amplifiers are designed to have low noise and are built directly into the silicon circuitry; thus they are often referred to as **on-chip amplifiers**. These amplifiers must work with extremely small voltages and are rated, as to their sensitivity, in volts per electron. Typical values are in the range of 0.5 to 4 microvolts per electron.

The output voltage from a given pixel is converted to a **digital number** (DN) and is typically discussed from then on as either **counts** or **analog-to-digital units** (ADUs). The amount of voltage needed (i.e., the number of collected electrons or received photons) to produce 1 ADU is termed the **gain** of the device. A typical CCD gain might be 10 electrons/ADU, which means that for every 10 electrons collected within a pixel, the output from that pixel will produce, on average, a count or DN value of 1. For example, with this gain value if a pixel collects 1000 electrons (photons), the output pixel value stored in the computer would be 100 ADUs. For 1500 electrons 150 ADUs would be produced and for 17234 electrons, the output pixel value would be 1723 ADUs (note, not 1723.4). Digital output values can only be integer numbers and it is clear already that the discrimination between different pixel values can only be as good as the resolution of the gain and digital conversion of the device.

Conversion of the output voltage signal into a DN is performed within a device called an **analog-to-digital converter** (A/D or ADC). There is an intimate connection between the number of digital bits available in the A/D and the value of the gain that can or should be used for the CCD. The output DNs are usually stored initially in computer memory and then moved to disk for storage and later manipulation. The process of shifting each entire CCD row into the output register, shifting each pixel along within this register, and finally performing the voltage conversion of each pixel's stored charge by the A/D to produce a DN value is continued until the entire array of pixels has been read out. For large-format CCD arrays, this process can take upwards of a few minutes to complete a single read out of the entire device. Note that for a 2048×2048 CCD, the charge collected in the last pixel to be read out has to be transferred over four thousand times. However, most modern large-format CCDs or mosaic cameras containing many large CCDs use a few tricks to readout faster. Single monolithic CCDs usually have 2

Campbell Physics Q11

or 4 output amplifiers available (one in each corner) and given the proper electronic setup, these large chips are often read out from 2 or 4 corners simultaneously, thus decreasing the total readout time by 2 – 4. For a mosaic of CCDs, this same process can read the entire array (using multiple amplifiers on each CCD) much faster than even one single large CCD.

The array size of a single CCD, as well as the size of a given pixel on a device, is controlled by the current limitations of manufacturing. How large one can make a good quality, large-scale integrated circuit and how small one can make an MIS capacitor, both of which have demanding requirements for near perfect operation, set the scale of CCD and pixel sizes that are available. CCDs as large as 5040×10080 and 7168×9216 pixels and pixels as small as $2 - 10 \mu\text{m}$ have been successfully produced.

Modern CCDs have much higher processing standards than even five years ago. Items such as multi-layer registration on the silicon wafer on the photomasks used in the production of the CCD integrated circuit and the ability to make smaller electrical component parts on the wafers (such as output amplifiers) lead to much lower noise characteristics, better pixel charge manipulation, and the ability for faster readout speeds with lower noise. For example, better alignment of the CCD layers in each pixel allow lower clock voltages to be used (as low as 2 volts has been demonstrated) leading to lower overall power consumption. This fact, in turn, allows for items such as plastic packaging instead of ceramic, reducing overall packaging costs, a cost that often rivals that of the CCD itself.

As you might imagine, astronomy is not the driving force for CCD manufacturing. Video devices, cell phones, security cameras, Xerox machines, etc. are the global markets boosting the economy of CCD makers. The trend today is to produce CCDs with small pixels ($10 - 12 \mu\text{m}$ for astronomy down to $\sim 2 \mu\text{m}$ for other applications) in order to increase image resolution. Small pixels (and small CCDs) have lower cost and higher yield but the small pixels have shallow well depths. This is somewhat compensated for using fast readout techniques and/or microlens arrays, which focus light from an incoming source onto each small CCD pixel. Not all CCD pixels are desired to have shallow wells. The CCDs produced by E2V for the NASA Kepler Discovery mission have $27 \mu\text{m}$ pixels with well depths of nearly 1 million electrons each and a capacity of > 50000 electrons per pixel is quite common in astronomy. Even CCDs with built-in electronic shutters are being experimented with. Each pixel contains a $p^+ - n - p^-$ vertical overflow drain (VOD) photodiode structure on its top through which the incoming light passes. The absorption of incoming light when the “shutter” is open is minimal and, within a few hundred nanoseconds, the electronic shutter can be biased and become opaque.

Quantum efficiency: The composition of a CCD is essentially pure silicon. This element is thus ultimately responsible for the response of the detector to various wavelengths of light. The wavelength dependence of silicon can be understood in an instant by glancing at Figure 20. Shown here is the length of silicon needed for a photon of a specific wavelength to be absorbed. Absorption length is defined as the distance for which 63% ($1/e$) of the incoming photons will be absorbed. Figure 20 clearly shows that, for light outside the range of about 3500 to over 8000 Å, the photons (1) pass right through the silicon, (2) get absorbed within the thin surface layers or gate structures, or (3) simply reflect off the CCD surface. At short wavelengths, 70% or more of the photons are reflected, and for very short wavelengths (as for long wavelengths) the CCD becomes completely transparent. Thus the quantum efficiency of a typical CCD device will approximately mirror the photon absorption curve for silicon. Shortward of $\sim 2500 \text{ \AA}$ (for thinned devices) or about $\sim 25 \text{ \AA}$ (for thick devices) the detection probability for photons increases again. However, owing to their much higher energy, these photons lead to the production of multiple electron-hole pairs within the silicon and may also produce damage to the CCD itself.

CCD quantum efficiencies are therefore very dependent on the thickness of the silicon that intercepts the incoming photons. This relation between absorption probability and CCD thickness is why front-side illuminated (thick) devices are more red sensitive (the photons have a higher chance of absorption) and why they have lower overall (blue) QE (since the gate structures can be close to or even exceed the necessary absorption depths of as small as only a few atomic layers). A few front-side CCDs have been produced with special gate structures that are transparent to incoming blue and UV photons. In thinned devices, the longest wavelength photons are likely to pass right through the CCD without being absorbed at all.

Figure 21 (left) shows the quantum efficiencies for various imaging devices. Note that the y scale is logarithmic and the much superior QE provided by CCDs over previous detectors. Figure 21 (right) shows a selection of modern CCD QE. The large difference in QE that used to exist between thinned and thick CCDs is now mostly eliminated due to manufacturing processes and coatings although other differences (such as location of peak QE, cosmic ray detection, etc.) remain. Quantum efficiency or QE curves allow one quickly to evaluate the relative collecting power of the device as a function of wavelength. Measured QE curves, such as in Figure 21 (right) and those shown in the literature, are generally assumed to be representative of each and every pixel on the device, that is, all pixels of a given device are assumed to work identically and have the same wavelength response. This is almost true, but it is the “almost” that makes flat fielding of a CCD necessary. In addition, the QE curves shown or delivered with a

Campbell Physics Q11

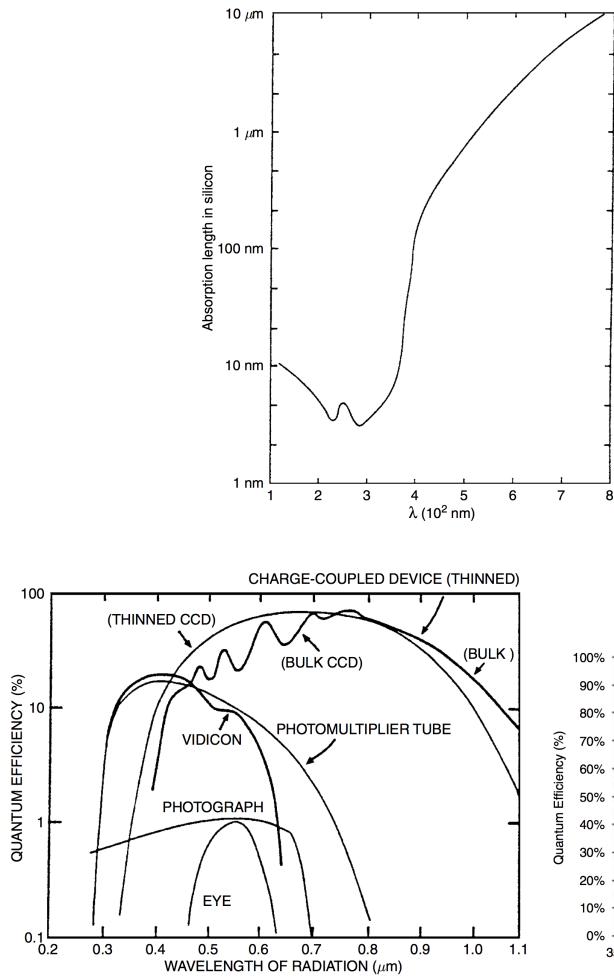


Figure 20:
The photon absorption length in silicon is shown as a function of wavelength in nanometers. From Reicke (1994). Figure taken from Howell (2006).

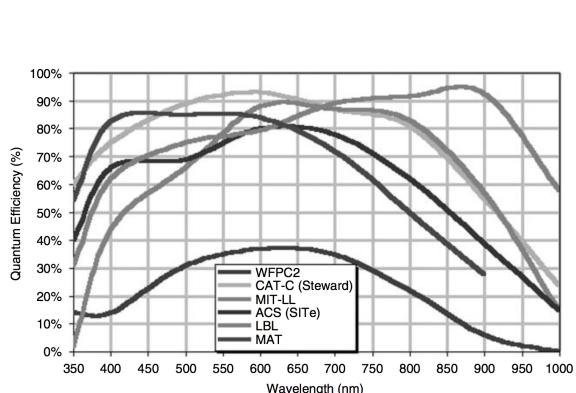


Figure 21: (left): QE curves for various devices, indicating why CCDs are a quantum leap above all previous imaging devices. The failure of CCDs at optical wavelengths shorter than about 3500Å has been essentially eliminated via thinning or coating of the devices. (right): QE curves for a variety of CCDs. WFPC2 is the second generation wide-field/planetary camera aboard HST, CAT-C is a new generation SITe CCD used in a mosaic imager at the University of Arizona's 90'' telescope on Kitt Peak, MIT-LL is a CCD produced at the MIT Lincoln Laboratories and optimized for red observations, ACS is the Hubble Space Telescope Advanced Camera for Surveys SITe CCD, LBL is a Lawrence Berkeley Lab high resistivity, "deep depletion" CCD with high red QE, and MAT is a front-side, processed CCD showing high blue QE. Figures taken from Howell (2006).

particular device may only be representative of a “typical” device of the same kind, but they may not be 100% correct for the exact device of interest.

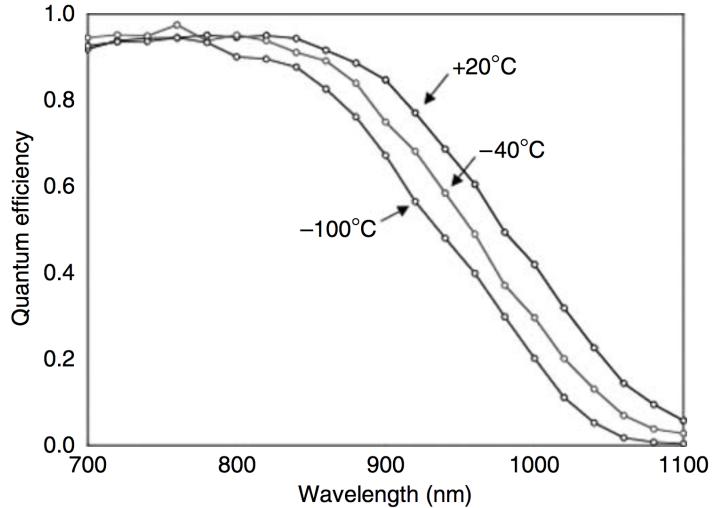
The quantum efficiency of a CCD is temperature sensitive especially in the red wavelength region. It has long been known that measurement of the QE at room temperature is a poor approximation to that which it will have when operated cold. Thus QE curves should be measured at or near the operating temperature at which the CCD will be used. As an example of the temperature sensitivity of the efficiency of a CCD, Figure 22 shows three QE measurements of the same CCD for temperatures of +20° C (~ room temperature), -40° C, and -100° C (~ operating temperature). Note that the variations are small at 8000Å but increase to 20% at 9000 – 10000Å. The curves in this plot would lead one to the conclusion that operating a CCD at room temperature is the best thing to do. However, this is *not* the case and a compromise between operating temperature (i.e., dark current) and delivered QE must be used.

Charge diffusion: Once an electron is captured in a CCD pixel, the voltages applied during integration attempt to hold it in place. However, situations arise within a CCD pixel that provide a finite possibility for any given electron to wander out of its collection pixel and into a neighboring pixel. This process is called **charge diffusion** and until recently it was noted but of low significance compared with other noise and readout issues. Today CCDs are of excellent quality and have very low readout noise, good pixel registration on the array, and reside in high quality optical systems. These facts mean that CCD imaging now has the ability to show great detail of any optical aberrations and geometric distortions. Even items such as better mechanical tolerances in instrumentation can reveal noticeable focus variations

Campbell Physics Q11

Figure 22:

Sensitivity of the quantum efficiency of a MIT/LL CCD for three operating temperatures. The blue sensitivity is little affected by a change in operating temperature but the red QE can change by a factor of two. The use of such devices requires a balance of higher operating temperature and keeping the dark current under control. Figure taken from Howell (2006).



as materials breathe with thermal changes. Given CCDs with deep pixel wells, large format front-side illuminated thinned devices, and the related improvements to modern astronomical instrumentation, the effects of charge diffusion on the point-spread function are noticeable.

A few ways in which charge diffusion can occur may be useful to discuss. Imagine a deep (geometrically long) pixel modeled after that which is shown in Figure 23. Electrons produced by long wavelength photons are captured in the front-side illuminated pixel near the bottom, far from the applied voltages in the front gates. Thus the potential well for these electrons is more like a potential dip. Given the right circumstances, an electron can move into a neighboring pixel. Another example would be impurities in the silicon material the CCD was manufactured from. These variations in the Si lattice can slightly bend or slightly redirect the potential within a pixel and provide weak spots from which electron escape is possible. Ways to mitigate electron loss are the use of higher potential voltages (although this can lead to other issues such as logic glow or shorting on the array), higher resistivity Si to more tightly hold the electrons (the Si lattice) in place, or to use small pixels (but these have lower red QE and small well depths). Again, we see that compromise and specific application come into play and can be tuned into the CCD as a part of its production process.

Charge diffusion often varies significantly across a CCD, especially thinned devices as they are not equal thickness everywhere. For example, the HST ACS wide field camera thinned CCD shows a variation in the core of the PSF, caused by charge diffusion, across the field of view. The variation is 30 – 40% at 5000Å with larger variations at shorter wavelengths. The effects of charge diffusion are not to be taken lightly. The ACS/WFC suffers a loss of about 0.5 magnitudes in its limiting magnitude at short wavelengths and near 0.2 magnitudes elsewhere. Charge diffusion is especially important in thinned devices that under-sample the PSF.

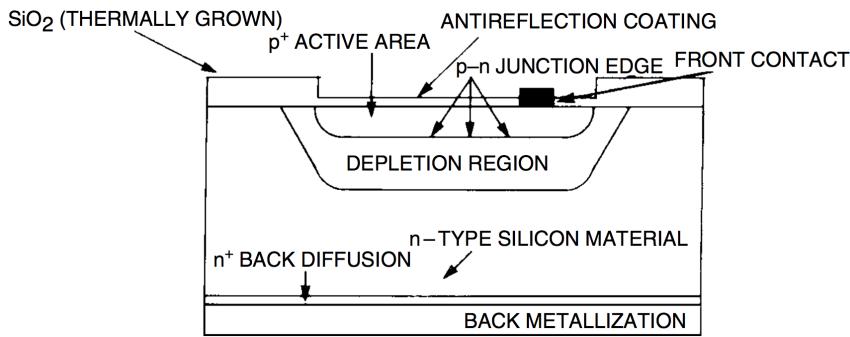


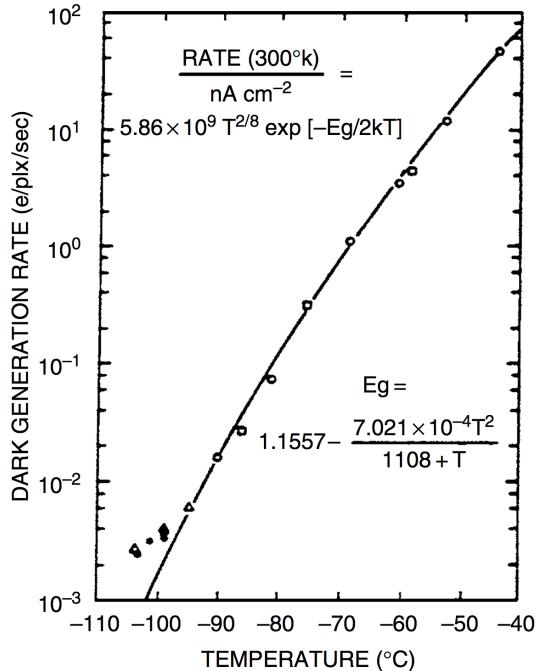
Figure 23:

Schematic view of a single front-side illuminated CCD pixel. The square labeled “front contact” is a representation of part of the overall gate structure. The letters “p” and “n” refer to regions within the pixel consisting of silicon doped with phosphorus and boron respectively. Figure taken from Howell (2006).

Charge transfer efficiency: Charge transfer efficiency or CTE is a measure of the fraction of the charge that is successfully transferred for each pixel transfer. CTE values of 0.999995 or more are typical in good modern CCDs. For a CCD with 1024×1024 pixels, the charge collected in the last pixel readout has shifted 2048 times thus the CTE must be nearly 100% in order to preserve the charge in each pixel during readout. CTI (charge transfer inefficiency) is 1-CTE or numerically near 10^{-5} or 10^{-6} in value. CTI can be and usually is different in the vertical and horizontal directions. The loss in charge from a CCD

Campbell Physics Q11

Figure 24:
 Experimental (symbols) and theoretical (line) results for the dark current generated in a typical three-phase CCD. The rate of dark current, in electrons generated within each pixel every second, is shown as a function of the CCD operating temperature. E_g is the band gap energy for silicon. From Robinson (1988a). Figure taken from Howell (2006).



pixel containing N electrons that is shifted 1024 times vertically and 1024 times horizontally is given by $L(e) = N(1024 \times \text{CTI}(H) + 1024 \times \text{CTI}(V))$, or, if a single CTI value is given, $L(e) = 2048 \times N \times \text{CTI}$. CCDs with poor CTE generally show charge tails in the direction opposite readout for bright stars. These tails are the charge left behind as the image is shifted out.

The standard method for measuring CTE is to use X-ray stimulation of a CCD with a Fe^{55} source. CCDs are good X-ray detectors and for a specific X-ray source such as Fe^{55} , each X-ray photon collected produces ~ 1620 electrons.⁵ A CCD is exposed to X-ray illumination and the resulting image readout. An X-ray transfer plot is made of signal in DN (y -axis) vs. running pixel number on the x -axis. Often hundreds of rows are summed together to increase the signal generated by the weak X-ray source. If the CCD has good CTE, a horizontal line will be seen at 1620 electrons (assuming a gain of 1.0). If the CTE is poor, this line starts at 1620 electrons (for rows close to the output amplifier) but tilts toward lower signal values for pixels further away from the output amplifier. This behavior indicates a loss of charge being transferred due to poor CTE. The CTE of a given CCD will generally degrade rapidly with decreasing operating temperature and is also a function of the clocking pulse shape and speed.

Readout noise: CCDs can be thought of as having three noise regimes: **read noise**, **shot noise**, and **fixed pattern noise**. In astronomy, we speak of these as **read noise limited**, **photon noise limited**, and **flat field uncertainties**. A plot of the log of the standard deviation of the signal (y -axis) vs. the log of the signal itself (x -axis) for a CCD is called the **photon transfer curve**. Read noise (or any noise independent of signal level) sets a noise floor for a device. Upon illumination, photon or Poisson noise raises the sigma measured following a \sqrt{N} slope. Finally, for large signal values, pixel to pixel variations due to processing errors and photomask misalignment begin to dominate. This latter noise is proportional to the signal and rises with a slope of 1.0. Full well sets in at some very high illumination and the slope of the photon transfer curve turns over or breaks.

Readout noise, or just read noise, is usually quoted for a CCD in terms of the number of electrons introduced per pixel into your final signal upon readout of the device. Read noise consists of two inseparable components. First is the conversion from an analog signal to a digital number, which is not perfectly repeatable. Each on-chip amplifier and A/D circuit will produce a statistical distribution of possible answers centered on a mean value.⁶ Thus, even for the hypothetical case of reading out the same pixel twice, each time with identical charge, a slightly different answer may be produced. Second, the electronics themselves will introduce spurious electrons into the entire process, yielding unwanted random fluctuations in the output. These two effects combine to produce an additive uncertainty in the final output value for each pixel. The average (1σ) level of this uncertainty is the read noise and is limited by the electronic properties of the on-chip output amplifier and the output electronics.

⁵Remember that for optical photon detection, one photon collected produces one photoelectron, regardless of its wavelength. For the much higher energy X-rays, a single photon collected produces multiple electrons in proportion to the photon's energy.

⁶The distribution of these values is not necessarily Gaussian

Campbell Physics Q11

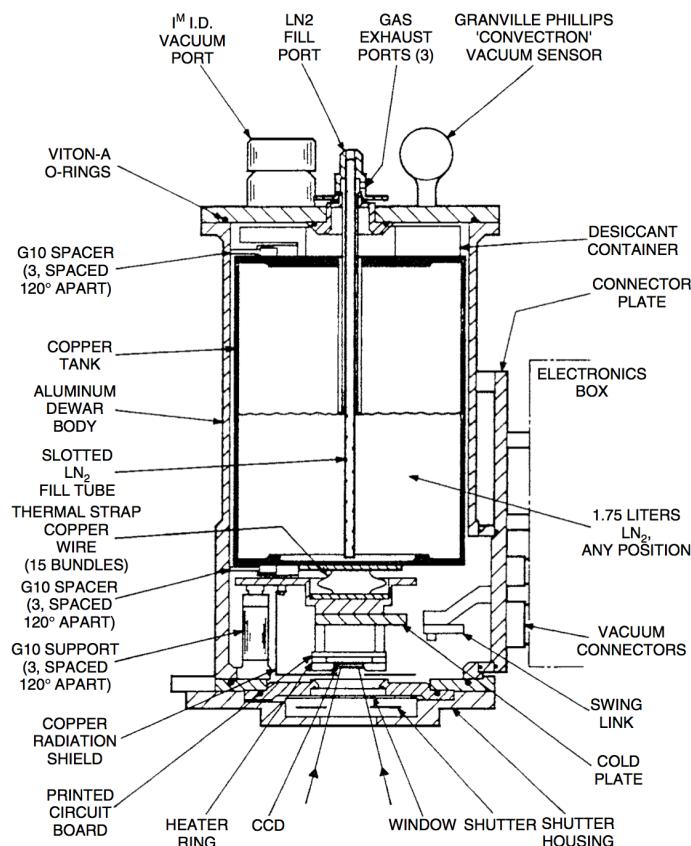


Figure 25:
A typical CCD dewar. This is the Mark-II Universal dewar originally produced in 1984 at Kitt Peak National Observatory. The dewar held 1.75 liters of liquid nitrogen providing a CCD operating time of approximately 12 hours between fillings. This dewar could be used in up-looking, down-looking, and side-looking orientations. From Brar (1984). Figure taken from Howell (2006).

In the output CCD image, read noise is added into every pixel every time the array is readout. This means that a CCD with a read noise of 20 electrons will, on average, contain 20 extra electrons of charge in each pixel upon readout. High read noise CCDs are thus not very good to use if co-addition of two or more images is necessary. The final resultant image will not be quite as good as one long integration of the same total time, as each co-added image will add in one times the read noise to every pixel in the sum. However, for modern CCDs, read noise values are very low and are hardly ever the dominant noise with which one must be concerned. Good read noise values in today's CCDs are in the range of 10 electrons per pixel per read or less. These values are far below read noise levels of ten years ago, which were as high as 50 – 100 electrons, and even those are well down from values of 300 – 500 or more electrons/pixel/read present in the first astronomical CCDs.

Dark current: Every material at a temperature much above absolute zero will be subject to thermal noise within. For silicon in a CCD, this means that when the thermal agitation is high enough, electrons will be freed from the valence band and become collected within the potential well of a pixel. When the device is read out, these dark current electrons become part of the signal, indistinguishable from astronomical photons. Thermal generation of electrons in silicon is a strong function of the temperature of the CCD, which is why astronomical use generally demands some form of cooling. Figure 24 shows a typical CCD dark current curve, which relates the amount of thermal dark current to the CCD operating temperature. Within the figure the theoretical relation for the rate of thermal electron production is given.

Dark current for a CCD is usually specified as the number of thermal electrons generated per second per pixel or as the actual current generated per area of the device (i.e., picoamps cm^{-2}). At room temperature, the dark current of a typical CCD is near 2.5×10^4 electrons/pixel/second. Typical values for properly cooled devices range from 2 electrons per second per pixel down to very low levels of approximately 0.04 electrons per second for each pixel. Although 2 electrons of thermal noise generated within a pixel every second sounds very low, a typical 15 minute exposure of a faint astronomical source would include 1800 additional (thermal) electrons within each CCD pixel upon readout. These additional charges cannot, of course, be uniquely separated from the photons of interest after readout. The dark current produced in a CCD provides an inherent limitation on the noise floor of a CCD. Because dark noise has a Poisson distribution, the noise actually introduced by thermal electrons into the signal is proportional to the square root of the dark current.

Campbell Physics Q11

Cooling of CCDs is generally accomplished by one of two methods. The first, and usually the one used for scientific CCDs at major observatories, is via the use of liquid nitrogen (or in some cases liquid air). The CCD and associated electronics (the ones on or very near the actual CCD itself, called the head electronics) are encased in a metal dewar under vacuum. Figure 25 shows a typical astronomical CCD dewar. The liquid nitrogen (LN2) is placed in the dewar and, although not in direct physical contact with the CCD, cools the device to temperatures of near -100°C . Since LN2 itself is much colder than this, CCDs are generally kept at a constant temperature $\pm 1^{\circ}\text{C}$ with an on-board heater. In fact, the consistency of the CCD temperature is very important as the dark current is a strong function of temperature (Figure 24) and will vary considerably owing to even modest changes in the CCD temperature.

A less expensive and much less complicated cooling technique makes use of **thermoelectric cooling** methods. These methods are employed in essentially all “off-the-shelf” CCD systems and allow operation at temperatures of -20°C to -50°C or so, simply by plugging the cooler into an electrical outlet. Peltier coolers are the best known form of thermoelectric cooling devices. CCD operation and scientific quality imaging at temperatures near -30°C is possible, even at low light levels, due to advances in CCD design and manufacturing techniques and the use of multipinned phase operation.

The amount of dark current a CCD produces depends primarily on its operating temperature, but there is a secondary dependence upon the bulk properties of the silicon used in the manufacture. Even CCDs produced on the same silicon wafer can have slightly different dark current properties. Today’s CCDs are made from high purity epi wafers produced with low occurrences of integrated circuit error. These factors have greatly reduced many of the sources of dark current even at warmer temperatures. As with most of the noise properties of a given CCD, custom tailoring the CCD electronics (such as the bias level and the readout rate) can produce much better or much worse overall dark current and noise performance.

Q3) CAMERA FIELD OF VIEW

What's the field of view of a 2K x 2K CCD camera on a 5-m telescope with f/16 focal ratio? The pixel size of the CCD is 20 micron. How does the field of view change if we bring it to a 10-m telescope?

increasing the diameter decreases FOV b/c resolution is $\theta = \lambda/D$ and have a tradeoff between resolution and FOV

Campbell Physics Q14

1.14 Question 14

What's the field of view of a $2\text{K} \times 2\text{K}$ CCD camera on a 5 m telescope with $f/16$ focal ratio? The pixel size of the CCD is $20\text{ }\mu\text{m}$. Now, lets bring this to a 10 m telescope with the same focal ratio. Explain how the field of view changes on the 10 m telescope (compared to that of the 5 m telescope) based on the Etendue conservation rule.

1.14.1 Short answer

The FOV of a telescope operating with a CCD is given as follows:

$$\text{FOV} = \frac{1 \text{ rad}}{\text{focal length}} \cdot (\# \text{ pixels}) \cdot (\text{pixel size}) = \frac{206265''}{f^\# D} N \Delta x,$$

where we have used the fact that $1 \text{ rad} = 206265''$ and that the focal length is given by $f = f^\# D$, where D is the telescope diameter.

Using this to find the FOV for a $2\text{K} \times 2\text{K}$ CCD camera on a 5 m telescope with $f/16$ focal ratio:

$$\text{FOV} = \frac{206265''}{(16)(5 \text{ m})} (2000)(20 \times 10^{-6} \text{ m}) \sim 100''.$$

We can now use the Etendue conservation rule to find the FOV at a 10 m telescope. Etendue is a property of light which characterizes how “spread out” the light is in an area and solid angle, and the Etendue conservation rule says that this property is conserved following the equation

$$\epsilon = \text{FOV}^2 \cdot D^2 = \text{const} \text{ [dimensionless].}$$

This allows us to equate the etendue of the first telescope setup to that of the second in order to solve for the new FOV:

$$\begin{aligned}\epsilon_{5 \text{ m}} &= \epsilon_{10 \text{ m}} \\ \text{FOV}_{5 \text{ m}}^2 \cdot D_{5 \text{ m}}^2 &= \text{FOV}_{10 \text{ m}}^2 \cdot D_{10 \text{ m}}^2 \\ \text{FOV}_{10 \text{ m}} &= \left(\frac{D_{5 \text{ m}}}{D_{10 \text{ m}}} \right) \text{FOV}_{5 \text{ m}} \\ \text{FOV}_{10 \text{ m}} &= \left(\frac{5 \text{ m}}{10 \text{ m}} \right) 100'' \\ \text{FOV}_{10 \text{ m}} &= \frac{1}{2} 100'' \\ \text{FOV}_{10 \text{ m}} &= 50''.\end{aligned}$$

Thus, using Etendue conservation, we can see that the FOV is reduced by a factor of 2 when brought to a telescope twice the diameter.

1.14.2 Follow-up Questions

- If you wanted a smaller FOV with the same size telescope, what would you change?

A4) SIDELOBES IN DIFFRACTION-LIMITED DETECTORS

Explain why diffraction-limited detectors tend to have sidelobes, and how sidelobes can be suppressed in optical and radio observations.

Herman Physics Q21

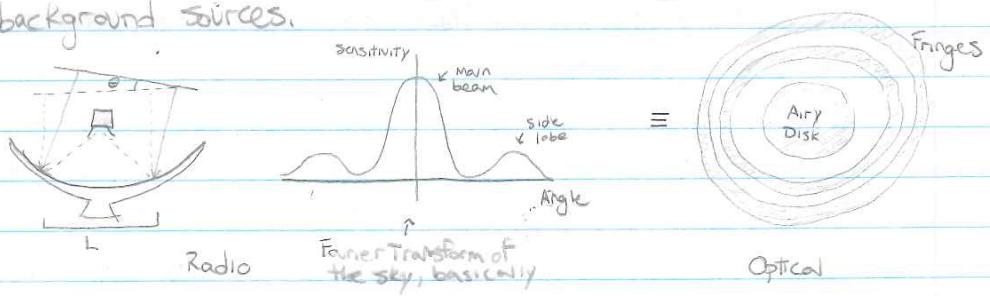
P21

- Side lobes are due to constructive interference. When your source is slightly off axis, there is a phase diff b/w rays of light, which interfere.
- In radio you can add a feed horn to taper the antenna, and in optical you can apodize the mirror, smoothing out the sensitivity on the edges.

• Side lobes and Diffraction - Limited Detectors

Diffraction limited telescopes are ones whose angular resolution only depends on the telescope's optics and the observed wavelength. The refraction of light by the atmosphere is negligible if you're diffraction limited, ie the atmosphere doesn't reduce your resolution, which is really only true for radio observations, I think.

The reason you get sidelobes is due to ^{partial!} constructive interference. If you have a source that's slightly offset from your optical axis, there'll be a phase difference b/w the 2 incoming rays from your plane wave of light. This means the rays can add constructively, so the telescope has a small increase in sensitivity at larger off-axis angles, so you'll pick up background sources.



To suppress radio side lobes, you can taper the antenna by adding a feed horn. These are made specific sizes so they're only sensitive to radiation reflecting off a small part of the dish, which reduces the effects of side lobes (since they're at larger angles, but also reduces the sensitivity of your main beam).

You can also build the telescope with an offset paraboloid dish (instead of a circular one) like the GBT, which also reduces side lobes.

I have no idea about reducing sidelobes/fringe patterns in optical observations that are diffraction limited. If the seeing is bad then adaptive optics would improve your angular resolution, but that doesn't seem to apply if you're already diffraction limited.

- JK apodizing the mirror helps, so you reduce sensitivity of the primary mirror near its outer edges. This reduces airy rings but also reduces collecting area.

No more sticky notes! :)