Introduction to Particle Cosmology

Contents

These notes are intended to introduce students to many of the basic ideas in particle cosmology needed to begin research. They were written to be accessible to first-year physics students at Harvey Mudd College, who have taken courses in calculus and calculus-based physics, including special relativity. Even if you don't have this background, however, you can skip over the more technical bits and still get some useful insight into underlying ideas in particle physics. The reader should, however, have looked through my notes introducing foundational concepts in particle physics.

1 Connecting the Large and Small

Particle physics is the study of the fundamental constituent particles and forces that make up all matter that we know and understand today. It is predominantly the study of sub-atomic and sub-nuclear particles like electrons and quarks and the forces that bind them together into more familiar forms of matter (atoms, molecules, and everything on up). **Cosmology** is the study of the origin and structure of the universe as a whole and are therefore concerned with the largest objects in the Universe: galaxies, galaxy clusters, and so forth.

At face value, these two pursuits appear to be entirely unrelated. One is concerned with sub-nuclear distance scales $< 10^{-15}$ m, while the other seeks to understand the universe over distances of billions of light years ($\sim 10^{25}$ m). Furthermore, the dynamics of our universe as a whole is driven by the gravitational interactions between massive objects, while the gravitational force is famously absent from our current Standard Model of particle physics: this is largely due to the fact that the gravitational interactions felt between individual elementary particles are utterly negligible compared to other forces.

However, there is one fact of cosmology that indirectly showcases its link to particle physics: the ongoing expansion of the universe, which was first observed by Edwin Hubble in 1929. We observe that space is expanding, meaning that distant objects in all directions appear to be getting farther away from us. This is not due to actual motion of these objects, but because lengths themselves are getting bigger due to the expansion of spacetime.

The converse of the universe's expansion is that, if we hit the rewind button, the universe was *smaller* in the past. We know from thermodynamics that systems tend to heat up under compression, and at earlier times the energy in the universe was confined to a smaller volume and so the energy density was higher in the early universe. If we go back as far as we can currently understand, which is about 14 billion years into the past down to mere seconds after the Big Bang, we find that the universe used to be compressed into such a small volume and with such a high energy density that elementary particles were continually bashing into each other at energies exceeding those of our most powerful particle colliders. Unlike our colliders, which are highly engineered marvels, these collisions were happening *all the time* in the early universe. Higher energies in particle collisions allow for the production of heavier and more exotic forms of matter, and so the sub-atomic and sub-nuclear realm of elementary particles and their forces is extremely relevant in understanding what was going on in these very early times. This is the essence of the **hot Big Bang theory**[1].

What's more is that our current theory of gravity, general relativity, says that the evolution and expansion of spacetime itself depends on the energy and motions of objects located within

[1]It is a common misconception that the "Big Bang" refers to the *precise beginning* of the universe in an energetic explosion. However, it is more accurately described as the theory hypothesizing that the universe was once in a much hotter, compressed state and subsequently cooled into the universe we see today. This theory has been extremely successful at predicting and explaining our earliest observations of the universe, although both theory and observation become murkier at predicting what happened at ever earlier times. When we refer to times "after the Big Bang", we don't necessarily mean after the start of the universe, but rather the time since the universe entered this very hot, compressed state. We are ultimately agnostic about what came before.

that spacetime. Therefore, the presence and interactions of elementary particles at early times is not merely a by-product of the universe being in a more dense state, but they actually *drove* the expansion of the early universe. Consequently, the particle interactions of the early universe shape and mould the eventual state of the universe today and leave imprints that we observe via astrophysical observations. We can therefore use our astrophysical measurements of the universe to shape our understanding of elementary particle physics (namely, the particles and forces present at various times in the early universe), while using our knowledge of fundamental particles derived from lab-based experiments to make predictions of how the universe evolved at early times.

For the most part, this interplay between cosmology and particle physics has been successful at explaining various observed properties of our universe. As an example, when we take collision rates known from particle and nuclear physics experiments and apply them to the early universe, we can make predictions of the relative abundances of hydrogen, helium, and heavier elements in the universe today. This is the study of **Big Bang Nucleosynthesis**, and the theoretical predictions agree very well with measurements of light elemental abundances in the present day. In a similar vein, there is a flash of light known as the **Cosmic Microwave Background** originating from the epoch when free protons and electrons bound into neutral atoms, and we can study particle interactions in the early universe by making precise observations of this oldest source of light. While we will return to discussing both of these eras in more detail, these successes tell us that we are on the right track in applying the knowledge of particle physics that we have uncovered today to the very early universe.

There are, however, several tantalizing mysteries where our current theory of particle physics does not line up with cosmological and astrophysical observations. Each of these discrepancies suggests that there are actually new particles and forces in nature that are, as yet, undiscovered in laboratory studies of elementary particles but which leave a clear imprint on the structure of our universe. I will discuss a few of them here.

Dark Matter: Since the 1930s, it was noted by astrophysicists by independently determining two quantities: the amount of visible matter in the universe from stars and gas, and the amount of matter inside of objects such as galaxies as inferred by their gravitational pull. When these two quantities are compared, it appears that there is *much* more matter exerting a gravitational pull inside of galaxies than can be explained from the visible light. The necessity of dark matter in explaining astrophysical measurements was put on solid footing in the pioneering work of Vera Rubin and Kent Ford in the 1960s and 1970s, and since then evidence for dark matter has turned up in astrophysical and cosmological measurements from a variety of different epochs in the universe's history. The problem is that *none* of the elementary particles we currently know about can account for this dark matter: it must be something new[2]!

This is not a small problem either: we now know that there is *more than five times more* dark matter than visible matter. The consensus is that this is some new particle beyond the Standard Model, and may come with its own forces and interactions. It is deeply unsatisfying that we don't know what more than 80% of the matter in the universe is, and as particle physicists we want to think of ways that we can learn more about what this dark matter, beyond that we can't see it!

The Matter-Antimatter Asymmetry: There are very few problems in physics (or other fields)

[2]An alternative to the existence of dark matter is that perhaps *gravity* behaves differently than we expect. Such theories fall under the general heading of modified gravity. While this is still a possibility, modified gravity theories have a hard time explaining all of the observations of dark matter (indeed, it is challenging to even understand how modified gravity theories work at very early times). Most of the astrophysics community accepts the existence of dark matter, and we will adopt that view here, although work continues on developing viable and compelling theories of modified gravity as alternatives to dark matter.

that more directly relate to the question of our existence as the matter-antimatter asymmetry. The discovery of the positron in 1932 confirmed the earlier theory of Dirac that elementary particles such as electrons have a corresponding **antiparticle** that has the same mass and spin but with opposite electric charge. Indeed, all of the particles in the Standard Model have antiparticles[3]. But this immediately raises a very important question: why did we have to wait until 1932 to discover antimatter when we've known about regular old matter since the earliest times? Where are all of the antiprotons? Anti-hydrogen? Anti-molecules? The only possible explanation is that there existed, at early times, an excess of matter over antimatter; all the antimatter annihilated away, leaving the remnant matter that makes up everything in the universe today (including us). Indeed, if there had not been an excess of matter over antimatter, all of the matter and antimatter would have annihilated away, leaving. . . nothing.

As we have continued to do studies in particle physics, we have learned that the Standard Model treats matter and antimatter on *almost* the same footing, and whatever differences exist are far too small to explain a universe with as much matter as we see relative to antimatter. It may be that the universe just started in a state with more matter than antimatter, but this is unlikely in light of the next cosmological puzzle discussed below. We are therefore left with the conclusion that there exist new particles and forces in nature that break the symmetry between matter and antimatter in a more significant way than is true in the Standard Model, allowing an accumulation of excess matter over antimatter and ultimately our existence.

Cosmic Inflation: There are some features of cosmological observations that are perhaps more subtle than the first two, but no less surprising. One is that the universe seems extremely isotropic: it looks the same in every direction. This is maybe not so shocking until you realize that parts of the universe that we see in polar opposite directions were *never* in causal contact (meaning that light from one patch has never reached any other during the entire history of the universe). How is it, then, that these two different patches of the universe "know" to be the same? In principle, they should have different temperatures, matter densities, and so on if they don't know about one another. Furthermore, the universe on scales we can see appears to be flat, meaning that it is overall not curved. It is not clear why this is the case.

A theory was developed to explain these known as **inflation**. Inflation hypothesizes the existence of a phase of very rapid, exponential expansion of the universe *prior* to the hot and dense state that is the subject of the Hot Big Bang theory. Such a period of rapid expansion can solve both of the above problems: the universe looks the same in all directions because we are looking at one very tiny piece of the original universe which was rapidly blown up to be the distinct patches we see today, and so it looks the same in every direction. Also, the rapid expansion tends to cause local patches of the universe to flatten: consider a balloon when it is mostly deflated and when it is blown up to its greatest extent, and you can see that the latter is less curved.

Inflation also resolves another important issue: why we exist as objects that are clumped into planets, stars, and galaxies. If the universe started in a hot, dense, but largely uniform state, we would expect it to remain in a uniform (but less dense) state as the universe expanded. However, what we see is that there were tiny regions that were more clumped than others, and gravitational attraction caused surrounding matter to fall into the clumpier regions, making them ever more clumpy. The period of inflation is hypothesized to generate tiny wiggles in density due to quantum mechanical fluctuations that seeded the over-dense regions we see today.

[3] This includes neutral particles such as the photon, which are their own antiparticles. Neutrons are neutral but are *not* their own antiparticles. The reason is that the quarks making up neutrons are electrically charged, and the antineutrons are made up of antiquarks that have opposite charge to the quarks in neutrons.

Inflation is also commonly (although not universally) held by cosmologists to be true. If inflation happened, there has to be a new type of particle that was associated with driving this period of rapid expansion. This particle is called the **inflaton**, and just like dark matter there is no inflaton candidate in the Standard Model[4].

In these notes, I will elaborate on the connections between particle physics and cosmology, referring frequently to the problems of dark matter and the matter-antimatter asymmetry. We will examine these problems in more detail and use our astrophysical and cosmological observations to try and pinpoint what *classes* of particle theories could explain these phenomena. Before we can do this, however, we need to learn more about the theory of the expanding universe because this is integral to our discussion of all of the other aspects of particle cosmology.

2 Evolving Spacetime and Our Expanding Universe

2.1 Overview of Gravity and General Relativity

Because of colonialism and globalism, nearly everyone in the world today is raised with a view of the nature of space and time that is inherited largely from the work of Newton (although there are indications of this view as early as classical Greece). When Newton formulated the classical laws of motion, it was a foundational assumption that space and time are distinct quantities and are *absolute*. This means that space and time exist universally for all observers, and are a mere passive stage on which the wonders of physics played out in various ways[5]. We can see this embedded deeply in Newton's second law,

$$\vec{F}(\vec{r}) \;\; = \;\; \frac{d\vec{p}(\vec{r})}{dt}. \tag{1}$$

According to Newtonian mechanics, the interesting physics is in how objects respond to applied forces at various points in space. The coordinate system merely labels *where* and *when* in space time the various applied forces occurred, acting as mere sign posts for the whole spectacle. Newton's law of gravitation states that the gravitational force acting on mass M_1 due to a mass M_2 is

$$\vec{F}(\vec{r}) \;\; = \;\; \frac{GM_1M_2}{r^2}\hat{r}, \tag{2}$$

where $\hat{r}$ is the unit vector pointing from mass 1 to 2, and G is Newton's gravitational constant.

The situation began to change with the theory of special relativity. Special relativity tells us that different observers perceive space and time differently, and that transformations between observers tend to mix up space and time (hence the fact that moving clocks run slow, moving lengths are contracted, and leading clocks lag). This tells us that space and time cannot be kept separate, but must be dealt with together as one **spacetime**. In special relativity, however, spacetime itself is still static even though different observers could perceive it differently depending on their relative motions.

[4]It has been conjectured that the Higgs field could drive inflation, although in this case it must possess unusual gravitational interactions.

[5]It is unfortunately common for students to be taught that absolute, static space and time are"intuitive" and, consequently, relativity is not. This neglects the fact that, both before Newton and in cultures that had not adopted Enlightenment worldviews, people had different conceptions of space and time. Even now, people who experience the world in different ways might have differing opinions on what is intuitive and what is not. I highly recommend the chapter "Spacetime Isn't Straight" in Prof. Chanda Prescod-Weinstein's book *The Disordered Cosmos* (and, indeed, the whole book).

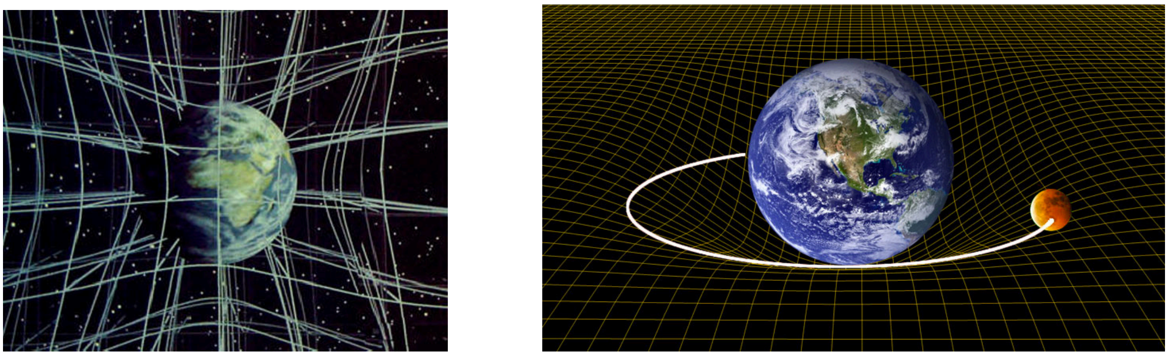

Figure 1: (Left pane) Spacetime is curved in the presence of a massive object such as the Earth. In general relativity, gravity does not exert a force; rather, objects navigate this curved spacetime by travelling in as straight lines as possible. The result is an apparent acceleration towards the Earth. (Right pane) The same thing, but showing only a 2D slice of the curved spacetime and an object orbiting the Earth. (Image credit: Physics Stack Exchange)

Special relativity does, however, raise questions about the validity of Newton's theory of gravity. For starters, special relativity suggests that nothing should travel faster than the speed of light, c. However, let's suppose I suddenly change the mass M_2. According to Eq. (2), mass M_1 should immediately detect a change in the gravitational force due to the change. However, if M_1 and M_2 are very far apart, this suggests that the change has been communicated instantaneously and, in particular, much faster than speed c! A similar situation exists in electrostatics: however, the resolution is found by taking into account the dynamics of the electric and magnetic fields in the presence of moving charges. In particular, the movement of a charge results in the emission of electromagnetic radiation that travels at a speed c, and it is this change in the field that communicates the change in potential to a distance point charge. Perhaps there is an analog to the electromagnetic field that communicates the change to masses in a theory of gravity?

The resolution to these puzzles came with the development of the **general theory of relativity** in 1915, which was a brand new way of thinking about gravity. In general relativity, we dispense entirely with the notion of gravitational force. Instead, in general relativity objects always move in straight lines: however, what we perceive as a gravitational force is actually the distortion of the geometry of spacetime itself in the presence of massive objects. Because spacetime curves in the vicinity of heavy objects (like the Sun), then particles that are trying to move in "straight lines" appear to be bent around the heavy object into orbits, or else appear to be accelerated into the centre of a large object (as we experience in "falling towards the Earth"); see Fig. 1. In other words, when we fall towards the Earth, it is not that there is a physical force acting on us, but that the curvature of spacetime causes us to drift toward the object as we attempt to move in a straight line. This malleable, curved spacetime can also be distorted and create waves and ripples which communicate changes to the mass structure of objects embedded in spacetime.

The development of general relativity is fascinating and there are many deep insights into the nature of gravity and motion that have been perfected over the past century. We do not have the time to review all of them here. For our purposes, we can content ourselves with knowing that this geometric picture of gravity as resulting from a warped and distorted spacetime has been confirmed in a spectacular way in the intervening years since its proposal. For example, our modern GPS systems would not work without the corrections resulting from a general relativistic picture of gravity. Similarly, the detection of gravitational waves was announced in 2016 by the LIGO collaboration after a long hunt: these are the tiny ripples in spacetime that communicate changes in particle masses as motivated by the above analogy with electromagnetism. These are but two of many examples of the empirical successes of general relativity.

There is, however, one *huge* change that comes with imagining spacetime as a dynamical, changing, curving background. In addition to being stretched, spacetime can also **expand** or **contract**. This means that the universe is not some fixed object, but one whose size and other properties are continually changing. Einstein found this so distasteful that he invented a fudge factor into his equations that permitted a universe of static size (essentially by tuning attractive and repulsive forces to keep the universe a fixed size). However, we now know that the universe does indeed expand[6].

Without going into the sophisticated mathematics of differential geometry, we cannot cover the details of how the expansion of spacetime originates from the equations of general relativity. However, we can say that the key insight of general relativity is that a combination of mass, energy, and momentum localized to a region causes spacetime in its vicinity to curve and change. We can therefore tie the rate of the universe's expansion to the amount of stuff in the universe and how it moves. This allows us to link the properties of matter with the evolution of the universe and its expansion. We are specifically interested in the role that elementary particles and their forces play in this process.

2.2 Characterizing the Universe's Expansion

What does it mean when we say that the universe is expanding? What we mean is that **every point in space is getting farther away from every other point in space**. For example, suppose that at time t_0 we are a distance d_0 from a particular distant galaxy. If the universe is expanding, then at a later time t_1 this distance will have changed to a new distance d_1. It is useful to think of the universe as the surface of an inflating balloon: after putting more air into the balloon, every point on the surface becomes farther away from every other point. However, the key distinction is that with a balloon, every point is on the two-dimensional surface is expanding *away* from the centre of the three-dimensional balloon. With the universe's expansion, the *entire* universe is expanding: there is no point from which we are "expanding away"; in other words, for the balloon analogy to work, we have to imagine that the surface of the balloon is itsefl the entirety of the universe (see Fig. 2).

The formulation of the expanding universe is due to Friedmann, Lemaître, Robertson, and Walker. We define a **scale factor**, $a(t)$, that is an increasing function of time[7]. Distances between two objects at fixed positions in spacetime[8] obey the relation

$$\Delta r(t) \quad = \quad a(t)\,\Delta r(t_0), \tag{3}$$

where t_0 is some reference time; the distances get larger due to the expansion of spacetime. In other words, a distance between two points at time t is larger than the distance at time t_0 by a factor of $a(t)$. If you like, we are defining a ruler as specifying the distance between two objects at t_0, and $a(t)$ tells us how that ruler is stretched due to the spacetime expansion[9].

[6] That being said, the "fudge factor", which we call the **cosmological constant**, has actually turned out to have physical significance and is an important part of our current theory of cosmology!

[7] It is important not to confuse the scale factor, $a(t)$, with the acceleration. Unfortunately, as you progress in physics you will find that we run out of letters and some have to do double duty!

[8] By fixed, we mean that the objects do not undergo relative motion in spacetime; they only appear to "move" because distances in spacetime are themselves changing.

[9] If you have studied special relativity, you are hopefully wondering what observer we are using to define the time, t, that characterizes the expansion. This is because special relativity tells us that observers moving with different relative velocities perceive time differently. In this case, the scale factor equation is derived in a special reference frame in

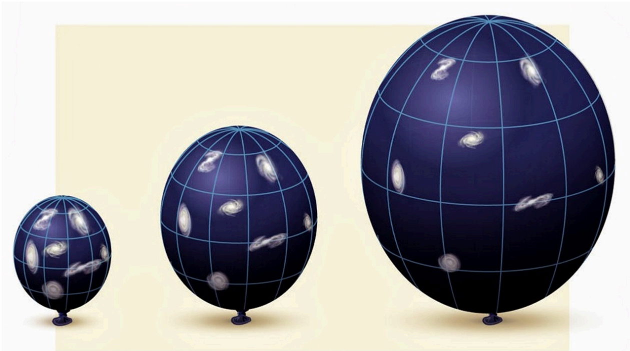

Figure 2: Expansion of the universe imagined as occurring on the surface of a balloon that is being inflated. Everything in the universe is confined to the surface of the balloon, and every point in the universe is getting farther away from every other point due to the expansion. (Image credit: Eugenio Bianchi, Carlo Rovelli, Rocky Kolb)

The scale factor tells us how distances between objects change between different times. We may also be interested in how *fast* the universe is expanding: if the expansion is incredibly slow, we can ignore it and treat spacetime as being essentially fixed, whereas if the expansion is fast then we have to worry about it. The way that we characterize the expansion is through the quantity

$$H(t) = \frac{\dot{a}(t)}{a(t)}, \quad (4)$$

where the dot refers to a derivative with respect to time, $\dot{a} \equiv da/dt$. H is referred to as the **Hubble rate**, named after Edwin Hubble who first provided evidence of the expansion of the universe (indeed, the universe's expansion is often referred to as **Hubble expansion**). Notice that H has units of $(\text{time})^{-1}$: we can understand H^{-1} as the time it takes for the distance between two objects to change by an amount comparable to its distance at time t. The way we can see that is that if we write the derivative as $\dot{a} \approx \Delta a/\Delta t$ where Δt is the time it takes the distance to change by a value Δa, then if $\Delta a \sim a(t)$ we have $H \sim 1/\Delta t$. Right now, the Hubble rate is $\sim 2 \times 10^{-18}\ \text{s}^{-1}$, so we expect distances in the universe to double in approximately 15 billion years.

The scale factor, $a(t)$, and the Hubble rate, $H(t)$ (as well as the differences between them), will come up repeatedly in these readings. It's ok if you don't completely understand them now, but if you find yourself very confused you may want to do some more readings about them online before proceeding.

3 Elementary Particles and the Universe's Expansion

For much of the Universe's history, the expansion of the Universe has been due to the energy present in the matter contained within the Universe. In turn, the expansion of the Universe has profound effects on the matter in the Universe: it dilutes the matter density and can change the rates of reaction occurring between particles. This means that, by studying the expansion of the early Universe, we can learn about the particles present in the Universe at that time. We can also use our knowledge of

which all the matter in the universe looks the same in every direction. We call this the "cosmic rest frame". A different observer will see an expansion that is different in time, but they will also perceive that the universe no longer looks the same in all directions.

the expansion to understand how the expanding early Universe sets the stage of the planets, stars, and galaxies we observe today.

In this section, we survey the qualitative effects of the Universe's expansion on the matter in the Universe, and use it to construct a history of the Universe as we currently understand it.

3.1 Effects of the Hubble Expansion

Looking at Fig. 2, we see that one effect of the expansion is the *dilution* of the number density of particles. Let us for the sake of argument assume that the universe is of finite extent and carries a number N_X of particles of species X. The expansion of the universe does not change the number of particles contained in the universe as a whole. However, the volume of the universe increases proportional to the cube of the scale factor, a^3. Therefore, the number density of particles n_X/unit volume changes due to Hubble expansion according to the relation

$$n_X(t_2) \quad = \quad n_X(t_1) \left(\frac{a(t_2)}{a(t_1)} \right)^{-3} . \tag{5}$$

Now, we don't actually know whether the Universe is finite in extent. However, we can pick an arbitrary volume at time t_1 and track the same patch of space as it blows up according to the Hubble expansion. This patch is called a *comoving volume* (because we are looking at a patch whose size expands at the same rate as the Hubble expansion), and repeat the above arguments: the number of particles N_X inside the comoving patch does not change due to the expansion by definition, but because the patch gets bigger, the number density $n_X \equiv N_X/V$ gets smaller.

Why do we care about number density rather than total number? Let's turn away from particle physics to something more familiar: high-school chemistry. We know that hydrogen gas is extremely explosive; it is also found in trace amounts in the atmosphere, but in this form it is too dilute to actually explode. If, however, we compress the hydrogen gas into a cylinder (or into the Hindenberg airship), then it becomes extremely explosive. In order to sustain the chemical reaction, the hydrogen atoms must be present in a sufficiently high *number density* to have any physical effect.

Similarly, the number densities of particles are very important for chemical, nuclear, and particle-physics reactions. As the Universe expands, the densities of every species of particle gets lower, and may drop below the critical threshold that required by certain reactions to proceed. The process by which reactions cease to occur due to a drop in density from the Hubble expansion is called **thermal freeze-out**, and we will look into how this happens in more detail later.

The second effect of the Hubble expansion on particle properties in the early universe is through the process of **redshift**, which is a phenomenon that affects the propagation of waves through an expanding medium. Consider some type of wave (for example, a light wave): one of the defining characteristics of a wave is its *wavelength*, or the distance in physical space between two adjacent peaks in the waveform. If we examine the picture in Fig. 3, we see in the top pane a wave with 5 crests and 5 troughs. When the background space it is in expands, we still have the same number of peaks and troughs, but distributed over a longer distance. This results in a larger wavelength $\lambda > \lambda_0$ in the expanded space. We therefore find that

$$\lambda(t_2) \quad = \quad \lambda(t_1) \left(\frac{a(t_2)}{a(t_1)} \right) , \tag{6}$$

and that wavelengths get *longer* as spacetime expands. If we consider a visible light wave, larger wavelengths correspond to redder colours, and so this phenomenon is called redshift. The opposite phenomenon, coming from the contraction of wavelengths in a shrinking space, is called **blueshift**.

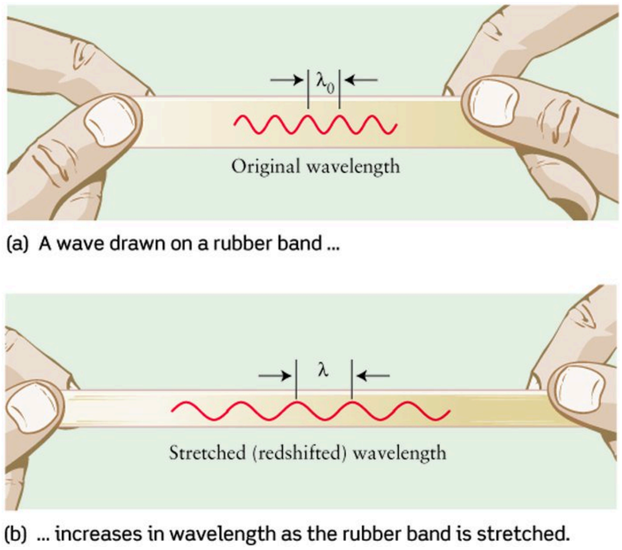

Figure 3: A wave of wavelength λ_0 gets stretched if the background spacetime surface is stretched. This results in an increase in the wavelength to λ. Since larger wavelengths are associated with redder colours of light, this is called *redshifting*. (Image credit: University of Alberta)

What is the effect of redshift on particle properties? Quantum mechanics tells us that, for a massless (or relativistic) particle, the energy per particle is proportional to the frequency, ν:

$$E = 2\pi\hbar\nu, \tag{7}$$

where $\hbar = h/2\pi$ is the reduced Planck's constant. This may be familiar to you as the photon energy as a function of the frequency of light in explaining the photoelectric effect. The frequency and wavelength are, in turn, related via the wave equation,

$$\lambda\nu = c, \tag{8}$$

where c is the speed of the wave (for a relativistic particle, c is the speed of light). We therefore find

$$E = \frac{2\pi\hbar c}{\lambda}. \tag{9}$$

Since λ increases with the expansion of the universe due to redshift, *the energy of each relativistic particle decreases with the expansion of the universe.* In other words, relativistic particles like photons lose energy as the Universe grows older. For example, the light from the cosmic microwave background is about 10^{-4} times less energetic now than when it was produced due to the effects of redshift.

We can now combine the effects of dilution and redshift to study the effects of spacetime expansion on **energy density**, defined as the total energy per unit volume $\rho \equiv E/V$. To find the energy density, we must add the energies of every type of particle present in the universe at that time. It is important because the gravitational force acts proportionally to the total energy-momentum of particles, and so the expansion of the universe (which results from gravity) depends on the energy density.

Let us consider a collection of particles of species X, each of which has energy E_X. If the particles are non-relativistic, most of their energy is in their mass, and so $E_X \approx M_X c^2$, which does not ever change; the mass of a particle is a fundamental invariant property of the particle. We have the energy density in species X,

$$\rho_X = \frac{N_X \cdot E_X}{\text{volume}} \quad (10)$$

$$\approx \frac{N_X \cdot M_X c^2}{\text{volume}}. \quad (11)$$

Since the volume scales like a^3, and N_X and M_X are fixed, we have

$$\rho_X(t_2) = \rho_X(t_1) \left(\frac{a(t_2)}{a(t_1)} \right)^{-3} \qquad (\text{non}-\text{relativistic}). \quad (12)$$

We see that the energy density of a *non-relativistic species* decreases cubically with the expansion of the universe, and this is due to the dilution of the particle species X: as the universe expands, there are fewer particles per unit volume, and therefore less energy.

The effects of the expansion for a *relativistic particle* are somewhat different. The reason is that, for a relativistic particle, most of its energy is in kinetic energy rather than in the mass. As a result, there is an *additional* suppression of the energy due to the redshift, $E_X \propto a^{-1}$. Combined with the a^3 scaling of the volume, we get:

$$\rho_X(t_2) = \rho_X(t_1) \left(\frac{a(t_2)}{a(t_1)} \right)^{-4} \qquad (\text{relativistic}). \quad (13)$$

Because the energy density of relativistic particles (also known as the **radiation density**) is reduced more severely by the Hubble expansion than the energy density of non-relativistic particles (also known as the **matter density**), we find that as more time elapses and the Universe expands to ever larger sizes, the matter density comes to dominate the total energy in the Universe. This is illustrated in Fig. 4.

We now summarize the results of this subsection:

1. The expansion of the universe reduces the *number density* of particles by a factor of a^{-3} as the scale factor grows. This is due to the dilution of the number density of particles because of the increasing volume of spacetime.

2. The expansion of the universe also causes relativistic particles to undergo *redshift*, reducing their energy by a factor of a^{-1}. Non-relativistic particles are unaffected because most of their energy is in mass rather than in the wavelength.

3. The combination of these effects means that the energy density of non-relativistic species scales like a^{-3} with the Hubble expansion, while the energy density of relativistic species scales like a^{-4}.

3.2 The Expanding Universe on Rewind

Today, the Universe consists of clumps of matter (us!), as well as a bath of electromagnetic radiation called the **cosmic microwave background (CMB)**. The current temperature of this background is approximately 2.7 K, which means that the average photon has an energy of about 0.235 meV (milli-electronVolts). Given that the energy required to ionize hydrogen is about 13 eV and most of the matter in the universe is now made of electrically neutral atoms, there is not really a whole lot of exciting stuff that the CMB does nowadays.

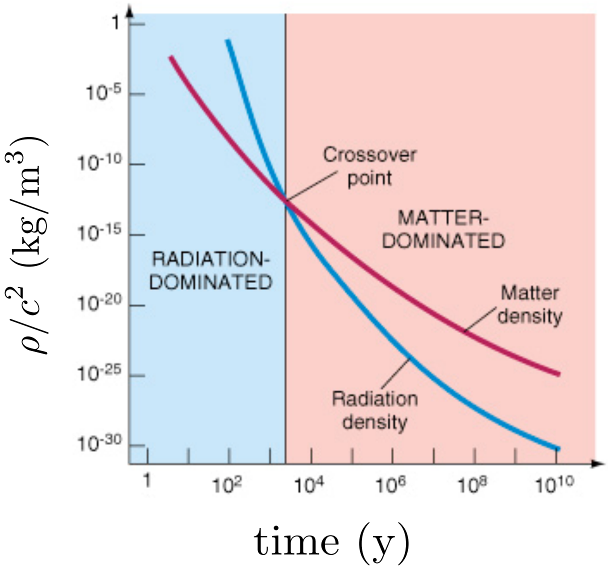

Figure 4: The expansion of the Universe reduces the energy density of relativistic particles (also known as the **radiation density**) faster than the energy density of non-relativistic particles (also know as the **matter density**). As a result, the effects of radiation dominated the Universe in the earliest times after the Big Bang, whereas the Hubble expansion reduces its relative importance over time until the non-relativistic matter comes to dominate (Image credit: Nanjing University).

However, we can take the existence of the CMB and the fact that the Universe is expanding to draw some very important conclusions about the history of the Universe. Indeed, all we have to do is hit rewind and go *backwards* to imagine what happened at earlier times. If we do this, we find that the Universe contracts as go further into the past. This contraction is precisely the opposite of the Hubble expansion, and again has two important effects:

- As spacetime contracts, particles are compressed into a smaller volume of space; this makes them more likely to interact with one another.
- As spacetime contracts, any radiation undergoes *blueshift*, meaning that the energies are increased by the fact that wavelengths are condensed with the contracting spacetime.

Therefore, at earlier times, the Universe was a hotter, denser place than it is today. This means that many particle reactions that don't really occur in nature today due to lack of energy were happening abundantly at early times.

3.2.1 The Ionization of Hydrogen

As I mentioned earlier, the energy required to kick an electron out of a hydrogen atom is approximately 13 eV. This means that, as we rewind in our history of the Universe, the photons in the CMB are blueshifted until their energies are about 10 eV. At this point, each photon in the CMB can kick out an electron from an atom, and so instead of finding neutral atoms in the Universe, we find freely floating positive nuclei, electrons, and photons. This means the world was a *very* different place when the temperature of the universe had photons with energy $E \sim 10$ eV: you can't form atoms and molecules if the electrons and nuclei don't stay stuck together!

Indeed, the origins of the CMB are from the very period where the Universe transitioned from a plasma of free electrons and nuclei into a collection of neutral hydrogen atoms. In the plasma phase, any photon floating through the Universe will quickly get absorbed by a passing nucleus or electron, while other photons will be emitted. It is a constant dance of emission and absorption of particles. Once the protons and neutrons bind up into neutral atoms, however, the transient photons no longer "see" any charged particles to scatter off of or be absorbed by: this results in the release of a burst of photons that we now see as the CMB. Given our current theory of how nuclei and electrons interact with photons, we can make precise predictions of what the CMB should look like and compare with data.

3.2.2 Big Bang Nucleosynthesis

Let's rewind further. The protons and nuclei, which are non-relativistic during this time, simply become more densely packed as the Universe contracts, whereas the photons continue to become more energetic. At a certain point, the photon energy increases so that the characteristic energy is $\sim$ MeV. This happens to be the energy required to eject a proton or neutron from a nucleus. Therefore, as we proceed to earlier times, the collisions of photons with atomic nuclei splits them apart into their constituent protons and neutrons. We therefore end up with a hot bath of unbound protons, neutrons, electrons, and photons. Note that this is basically the same thing as the CMB time period, except we are now talking about nuclei splitting apart or re-assembling (vs. atoms).

The process by which free protons and neutrons fused into nuclei in the early Universe is called **Big Bang Nucleosynthesis** (BBN). As with the CMB, we can take our theories and make predictions about the outcome of BBN: for instance, the ratio of hydrogen to helium in the Universe. If there are new particles or forces beyond the ones we know about, this can alter the process of BBN and give a very different looking Universe. Indeed, our measurements of BBN (and the CMB) provide powerful evidence for the existence of new particles beyond the ones we know about (dark matter), and place stringent constraints on new forces that can interact with us. This gives a rough illustration of how we can use our "rewound" knowledge of the Universe to learn more about the fundamental particles in nature.

3.2.3 The Production of Antimatter

Around the same time as BBN (or perhaps a bit earlier), the bath of photons becomes sufficiently energetic to allow a new type of process: the production of antiparticles. Here, I'm talking specifically about the production of electron-positron pairs in photon collisions. The electron mass is about $m_e = 511$ keV/c^2, and so in a collision of $\gamma\gamma \to e^+e^-$, then (in the centre-of-mass frame) each photon must have above 511 keV in order to produce the e^+e^- pair. This means that electrons and positrons are copiously produced once the Universe contracts to a sufficiently small size (and hence the photons have a large enough energy).

Exercise: Draw all Feynman diagrams for the process $\gamma\gamma \to e^+e^-$. Why is the process $\gamma \to e^+e^-$ not allowed?

As we go back further in time, the plasma heats up even more. When the characteristic energy of the particles in the Early Universe becomes much larger than $m_e c^2$, the electrons *also become relativistic.* Recall that the relativistic boost factor for electrons is

$$\gamma \equiv \frac{E}{m_e c^2}, \tag{14}$$

and so $\gamma \gg 1$ for $E \gg m_e c^2$. So, at early times, even the "matter" particles like electrons (and, even earlier protons) are part of the radiation of the universe. As a result, the changing size of the Universe also *blueshifts* the electron energy under contract once they are relativistic (or, we play the Universe's expansion the right way around, *redshifts* them). At even earlier times, the protons are sufficiently energetic enough that they blast each other apart into their constituent quarks.

Eventually in our time travel back through the history of the Universe, we find that the Universe was made up of a hot, dense plasma consisting of *every particle in the Standard Model.* The heaviest particle in the Standard Model is the top quark, $m_t = 173$ GeV$/c^2$, and so once the typical energy of photons, electrons, and other particles is larger than 175 GeV, the Universe is filled with every type of particle of which we are currently aware, and perhaps some that we currently know nothing about! We are curious about how these new types of particles can affect the evolution of the Universe.

4 Thermodynamics in the Early Universe

In the last section, we saw that the presence of interactions such as $\gamma\gamma \to e^+e^-$ means that a hot, dense collection of photons will eventually lead to the production of electrons and positrons, as well as every other type of particle in the Standard Model (or, at least for any species such that $E_\gamma > Mc^2$). In this section, we explore some of the implications of these processes, and will find that these huge, complicated collections of particles can in fact be characterized by just a few properties, such as the temperature. This is similar to how a room full of air contains $> 10^{24}$ molecules, and yet can be simply described in terms of number density, volume, temperature, and pressure. The reason why this simplification is possible is because the individual behaviours of particles average out over the entire collection, allowing the system as a whole to be characterized by just a few numbers. At the beginning of our study, we neglect the background effects of the expansion of the Universe. Once we have seen the general idea of the thermodynamics of the Early Universe on its own, we will restore the expansion to see how it affects this gas or plasma of particles.

4.1 A Non-Expanding Universe

4.1.1 Chemical and Kinetic Equilibrium

Let's conduct a thought experiment and suppose that we start off in a non-expanding Universe consisting entirely of photons with typical energy $E > M_e c^2$. This means that, when the photons collide, they can produce e^+e^- pairs. Each collision depletes two of the photons and, in their place, produces an electron-positron pair. If we started with N_γ photons and 0 electrons or positrons, then after a single collision we have $N_\gamma - 2$ photons, 1 electron, and 1 positron. We don't really have to worry about the reverse process $e^+e^- \to \gamma\gamma$ provided $N_\gamma \gg 1$, because there are many photons that can convert into electron-positron pairs but almost no electron-positron pairs to go back into photons. This process continues so that, after n collisions, we have $N_\gamma - 2n$ photons, n electrons, and n positrons.

It gets a bit tricker once n gets to be comparable to $N_\gamma/2$. When this occurs, we end up with a situation where the $e^-e^+ \to \gamma\gamma$ process becomes important. In general, *all processes that can happen in nature can also happen in reverse.* In other words, if electron-positron pairs can be created, they can also be destroyed.

Does the conversion of photons into e^+e^- pairs ever stop? The simple answer is no; any time a photon happens to bump into another photon, there is a chance that they will create an e^+e^- pair. Similarly, any time an e^+e^- pair collide, there is a chance they will create a pair of photons. However, at a certain point the rate of forward reaction will *equal* the rate of reverse reaction; this configuration is known as **chemical equilibrium**. Short of any disruption to the system, the *average* numbers of photons and electrons will not change from this point forward because the rate of production of e^+e^- pairs via photon collisions is exactly equal to the rate of production of $\gamma\gamma$ pairs in e^+e^- collisions. If we track individual photons, they may be destroyed, but if we only look at the total numbers, they remain unchanging. Thus, we say that chemical equilibrium is characterized by

$$\frac{dN_{e^-}}{dt} = 0 \quad (15)$$

$$\frac{dN_{e^+}}{dt} = 0 \quad (16)$$

$$\frac{dN_{\gamma}}{dt} = 0. \quad (17)$$

In general, chemical equilibrium implies that systems *on the whole* are not changing, and so time derivatives of total or average quantities are zero.

We can draw a very simple mechanical analogy. Suppose we have an empty bucket with a hole in the bottom. We now start pouring water at a rapid rate into the bucket. Because there isn't much water, the water pressure is low, and the tiny trickle of water out of the bottom of the bucket cannot compensate for the large amount of water being poured in. As the bucket fills, however, the water pressure starts rising, and at a certain point the outflow exactly equals the inflow. If we track individual molecules of water, of course we will see a change: some of them drop into the bucket, swirl around, or plummet out. But if we look at the surface of the bucket, it stays exactly level: the total amount of water in the bucket stays the same. Nothing changes until we change the configuration of the system (for example, by increasing the inflow, plugging up the hole, or changing the type of fluid being poured into the bucket). There are many other examples we could think about: for example, connecting an evacuated room with a room containing normal air and observing the flows of gas molecules between them.

Returning to particle physics, the motions of particles do not depend *only* on the number of particles. We also care about the momenta and positions of each particle. This seems troubling: if we have 10^{10} particles in a particular volume that we are tracking, this (in principle) means that we need to keep track of 6×10^{10} quantities, namely the three components of the momentum vector and the three components of the position vector. This is an impossible task! Fortunately, we will soon see that there is another form of equilibrium that allows us to simplify our description of particle energies. We also assume that the universe is homogeneous and isotropic (the same at every point in space and in every direction), which is pretty consistent with our observations of the universe on large scales and allows us to forget completely about the positions of particles. F

In addition to matter-antimatter creation/annihilation processes like $\gamma\gamma \leftrightarrow e^+e^-$, there also exist **elastic collisions** like $e^-\gamma \to e^-\gamma$ or $e^-e^- \to e^-e^-$, which preserve the nature of each particle in the collision but transfer momenta between particles. Imagine, for example, that we have two electrons with momenta $\vec{p}_1$, $\vec{p}_2$ such that $|\vec{p}_1| \gg |\vec{p}_2|$. In this case, it is likely that they will collide in such a way as to equalize the momenta between them. Thus, rapidly colliding particles tend to exchange momenta in such a way as to give each one comparable momentum; it would be very unusual for a large number of particles to collide and concentrate their collective momenta into a single particle!

If we wait long enough, once again the particles reach equilibrium. This time, it is called **kinetic equilibrium** because we are referring to the distribution of kinetic energies in the particles. When in kinetic equilibrium, an individual particle can change its momentum by bumping into other particles, but the overall distribution of momenta does not change. We characterize the momentum spread among the particles by a probability distribution that a single particle drawn at random has a given momentum $\vec{p}$. This probability distribution is referred to as the **momentum distribution**, $F(|\vec{p}|)$, and is a function of the momentum $\vec{p}$:

$$F(\vec{p})\,dp_x\,dp_y\,dp_z \quad \propto \quad \#\text{ particles with momentum } [\vec{p},\vec{p}+d\vec{p}), \tag{18}$$

where the above relation holds up to a constant of proportionality that is defined by convention to be $(2\pi)^{-3}$. The total number of particles of the species can be found by integrating over all possible momenta,

$$N \;=\; \int_0^\infty \frac{dp_x\,dp_y\,dp_z}{(2\pi)^3} F(\vec{p}). \tag{19}$$

Because we cannot count the total number of particles in the Universe, we are usually instead interested in the number of particles *per unit volume*. This is denoted by $n \equiv N/V$. Similarly, we define the momentum density distribution $f(\vec{p}) \equiv F(\vec{p})/V$ and have

$$n \;=\; \int_0^\infty \frac{dp_x\,dp_y\,dp_z}{(2\pi)^3} f(\vec{p}). \tag{20}$$

Typically, $f(\vec{p})$ does not depend on the direction of the momentum; the reason is that the Universe is isotropic (looks the same in all directions), and so the system has equal probability of finding a vector of magnitude $|\vec{p}| \equiv p$ pointing in any direction. We therefore write instead $f(\vec{p})$ as a function of the magnitude only, $f(p)$. Switching to spherical polar coordinates and integrating over the angular variables, we find

$$n \;=\; \int_0^\infty \frac{dp}{2\pi^2}\, p^2\, f(p). \tag{21}$$

We see that $p^2\, f(p)$ is (up to a normalization factor) the probability of finding a particle of momentum magnitude p.

For a classical system with many rapid collisions, it turns out that $f(p)$ takes on a universal form regardless of the nature of the particles that are scattering. This function is called the **Maxwell-Boltzmann distribution** and is characterized by a single number, T. T is the **temperature** and characterizes the typical momentum (and hence kinetic energy) of particles in the system. We will use units where T has dimensions of energy[10], in which case $f(p) \propto e^{-E(p)/T}$ and $E(p) = \sqrt{p^2c^2 + M^2c^4}$ is the relativistic energy-momentum relation. In particle physics, we often care about the collisions for highly relativistic particles, $p \gg Mc$. Then, we have

$$p^2\, f(p) \;=\; \mathcal{C}\, p^2\, e^{-pc/T}. \tag{22}$$

where $\mathcal{C}$ is a constant of proportionality that depends on various factors that are unimportant for the time being. The shape of the Maxwell-Boltzmann distribution is shown in Fig. 5. The distribution has a maximum at momenta comparable to T, and then there is a sharp exponential fall-off. This is the result of the fact that thermodynamics likes particles to have all comparable energies and momenta, and you pay an exponential penalty for trying to find a particle with energy or momentum $\gg T$.

[10] One can pass between units of energy and the usual units of Kelvins by dividing by a fundamental constant called the Boltzmann constant.

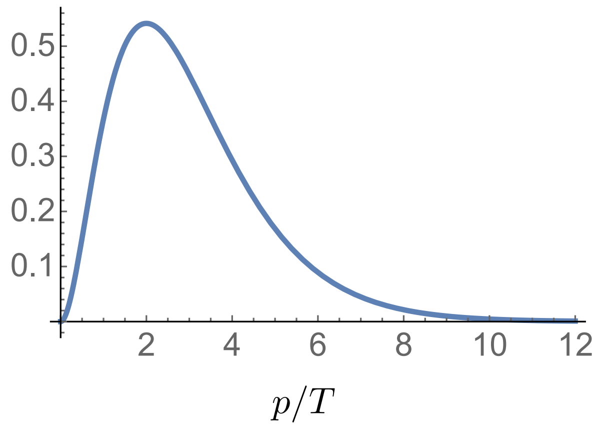

Figure 5: The shape of the Maxwell-Boltzmann distribution for a highly relativistic particle. It is evident that the distribution of particle momenta is peaked around $p/T \sim 2$. The scale of the y-axis is arbitrarily normalized since we are only interested in looking at the shape for now. For the true Maxwell-Boltzmann distribution, the y-axis tells us something about the number density of the particles.

4.1.2 Detour: Natural Units

If you've gotten this far and haven't found any factors of c missing, then I should count myself very lucky. The reality is that many of these factors are annoying to carry around: for instance, we have to remember to divide E by c in the momentum 4-vector, and to multiply t by c in the position 4-vector. It would be simpler if we didn't have to deal with them at all.

For this to work, we need to define a set of units where $c = 1$. It is perfectly reasonable to define such a set of units, since c is the only fundamental constant that relates things like position and time, or energy, momentum, and mass. If I say "The proton has a mass of 1.6×10^{-10} J", there is no ambiguity as to my meaning, which more precisely is "The proton has a mass whose rest energy is equal to 1.6×10^{-10} J". There is no ambiguity because mass and rest energy map onto one another in a simple, one-to-one manner.

We call this **natural units**. In natural units, we set[11]

$$c = 1, \tag{23}$$

$$\hbar = 1, \tag{24}$$

where $\hbar$ is the reduced Planck's constant, $\hbar = h/2\pi$. Recall that Planck's constant relates energy and time,

$$E = h\nu, \tag{25}$$

where ν is the frequency of light, and so using combinations of $\hbar$ and c it is possible to relate pretty much all physical quantities to a single unit.

By convention, we choose the physical unit to be a unit of energy, and in particle physics we use the energy unit of electron-volts (1 eV $= 1.6 \times 10^{-19}$ J). Using Eq. (25) and setting $\hbar = 1$, we see that time has units of inverse energy. Similarly, since distance is related to time via $x^0 = ct$, we see that distance also has a unit of inverse energy. Physical dimensions of length, time, etc. can be found

[11] For those who have studied statistical mechanics, we also set the Boltzmann constant $k_{\rm B} = 1$ in natural units.

by remembering that $\hbar = c = 1$ in natural units, so you can multiply or divide by $\hbar$ and c wherever you like without changing the answer. Then, sub in the physical values $\hbar = 6.58 \times 10^{-25}$ GeV · s and $c = 3 \times 10^8$ m to convert from natural units back into more familiar units.

Exercise: (a) A particle physicist tells you that the typical lifetime of a particle in its rest frame is $\tau = 10^{14}\ \mathrm{GeV}^{-1}$*. What is this time in seconds? (b) In the lab frame, the particle has a speed* v*. What is the typical decay length in the lab frame?*

Exercise: (a) In natural units, we say "The wavelength of the photon is $1\ \mathrm{MeV}^{-1}$*". What do we mean by this (in other words, translate the energy to a property of the wavelength). (b) In natural units, we say "The momentum of particle A is* 10 GeV*". What do we mean by this (in other words, translate the energy to a property of the momentum)*

4.1.3 Distributions in Thermal Equilibrium

When a species of particle is in both kinetic and chemical equilibrium, we say that it has reached **thermal equilibrium**. Its properties are determined completely by the temperature of the system and the mass of the species. While we cannot derive these results here, they may be familiar to you if you have taken an introductory course in quantum mechanics. Because particles are quantum phenomena, we must take into account their quantum behaviour. What is most relevant is whether the particles are **bosons**, which means they have intrinsic angular momentum that is an integer multiple of $\hbar$ (0, $\hbar$, $2\hbar$, etc.), or **fermions**, which have intrinsic angular momenta that are half-integer multiples of $\hbar$ ($\hbar/2$, $3\hbar/2$, etc.). Fermions obey the **Pauli Exclusion Principle**, which means that two fermions cannot occupy the same quantum state. Electrons are fermions, and for this reason they stack up in atomic orbital energy levels rather than all descending to the ground state. By contrast, bosons *like* to join up in the same quantum state and tend to exhibit clumping effects.

The probability distribution functions, f, which tell us how likely a particle is to have a given energy or momentum, depend on whether the particle species is a boson or a fermion. In general, for a species in thermal equilibrium we have

$$f(E,T) = \frac{1}{e^{E/T} \pm 1}, \tag{26}$$

where E is the particle energy, T is the temperature, the equation with the plus sign corresponds to fermions and the equation with the minus sign corresponds to bosons. For a relativistic species, $E = p$, and for $p \gtrsim T$, we see that the quantum version of f reduces to the Maxwell-Boltzmann distribution introduced in the previous section. What we find different here is the behaviour as $E \to 0$. For small energies, the distribution for bosons blows up, indicating the preference for bosons to all exist in the same (ground) state. For fermions, we instead find the distribution is *smaller* than the Maxwell-Boltzmann distribution for $E \to 0$, indicating the dislike for fermions to occupy the same state.

Note that Eq. (26) gives the distribution in terms of energy rather than momentum. These two are in one-to-one correspondence, since we know that $E = \sqrt{p^2 + M^2}$ for a species of mass M. This allows us to consider also the non-relativistic limit of particles, namely when the temperature T is smaller than the mass energy of the particle. In this case, we have $E \approx M + p^2/2M$ and we see that

$$f(E,T) \approx e^{-M/T}. \tag{27}$$

The probability of finding the particle becomes exponentially suppressed as $T \ll M$. Why? Recall that particles can be created from energy and can also annihilate into energy. When the mass energy of the species exceeds the typical energy of a particle in the system, the particles can annihilate away but there is insufficient energy for them to be produced again by collisions of other particles. This means that the number density of the heavy particle is steadily depleted by annihilations, with no inverse processes to produce them. The result is an exponential depletion of the number density of the heavy particles.

We can perform the integral[12]

$$n = \int \frac{dp_x\, dp_y\, dp_z}{(2\pi)^3} f(E,T) \tag{28}$$

analytically in the relativistic and non-relativistic limits. The results for **relativistic bosons and fermions** are

$$n_B(T) = \frac{\zeta(3)}{\pi^2} T^3, \tag{29}$$

$$n_F(T) = \frac{3\zeta(3)}{4\pi^2} T^3, \tag{30}$$

where $\zeta(x)$ is the Riemann zeta function. The results for **non-relativistic bosons and fermions** are the same:

$$n(T) = \left(\frac{MT}{2\pi}\right)^{3/2} e^{-M/T}. \tag{31}$$

Thus, we see that *the temperature of a system in thermal equilibrium fixes both the number density of particles, as well as their typical momentum distributions.*

4.1.4 The Approach to Equilibrium: Boltzmann Equations

If we wait long enough, we expect the Universe to settle into a state of thermal equilibrium. The reason is that collision processes that produce and destroy particles, and re-distribute their momenta, drive the particle distributions to their equilibrium values. Once equilibrium is reached, we expect that the Universe is in a steady state: while particles are continually being created, destroyed, and scattered, the overall distributions no longer change in time. This is a simple state of the Universe to characterize.

However, if the Universe *starts out* in a non-equilibrium state, we can ask about the dynamics that cause it to come into equilibrium. These might be important for determining how long it takes to arrive at equilibrium, for example. In our thought experiment, we started with a Universe filled only with photons. How long does it take to start producing appreciable numbers of electrons, positrons, and other particles? Or, conversely, if we started with a Universe populated entirely of electrons and positrons, how long would it take to fill with photons, quarks, etc?

In this section, we will explore a toy scenario where there exists a particle called X that can decay into electron-positron pairs, $X \to e^+e^-$. We will use this scenario to argue for the general form of the rate equations that describe the approach to equilibrium. These equations are collectively referred to **Boltzmann Equations**, and they often form the central objects of study in tracking the abundances and properties of particles in the early Universe.

We define the rate of X production in the process $e^+e^- \to X$ in terms of a quantity γ_X; this is the number of X particles produced per unit volume per unit time assuming the electrons and positrons

[12]If you have seen statistical mechanics before, you may know that we actually need to multiply the number density by the number of degrees of freedom by the species. We'll ignore this subtlety for now.

are in thermal equilibrium. We can write the change of the X number density due to production as

$$\frac{dn_X^{\rm prod}}{dt} = +\gamma_X. \tag{32}$$

At the same time, the particle X decays to e^+e^- with lifetime τ_X. The lifetime is typically defined such that, in time τ_X, the number of X particles decreases by a factor of $1/e$. In this case, we have

$$\frac{dn_X^{\rm decay}}{dt} = -\frac{n_X}{\tau_X}; \tag{33}$$

in the absence of production of X, the solution to this equation would be $n_X(t) = n_X(0)e^{-t/\tau}$. In practice, we have to take into account effects of relativistic time dilation but for now we will assume that τ_X gives the relativistically corrected lifetime.

We can combine the production and decay into a single equation,

$$\frac{dn_X}{dt} = \gamma_X - \frac{n_X}{\tau_X}. \tag{34}$$

We can factor this equation into a suggestive form:

$$\frac{dn_X}{dt} = -\frac{1}{\tau_X}\left(n_X - \gamma_X\tau_X\right). \tag{35}$$

Notice that when $n_X = \gamma_X(t)\tau_X$, then $dn_X/dt = 0$: the abundance does not change! This is, by definition, the equilibrium abundance, since the system has reached a steady state. We can therefore *define* $\gamma_X\tau_X \equiv n_X^{\rm eq}$, and we have

$$\frac{dn_X}{dt} = -\frac{1}{\tau_X}\left(n_X - n_X^{\rm eq}\right). \tag{36}$$

We already have an expression for the equilibrium abundance in terms of the system temperature in Eq. (28). Thus, we can compute the X production rate in terms of the decay rate and equilibrium abundance:

$$\gamma_X = \frac{n_X^{\rm eq}}{\tau_X}. \tag{37}$$

This seems like magic: by computing the X decay rate and knowing the general form of the equilibrium number density, we can automatically find the X production rate without needing to calculate it from first principles *even if the system is not in equilibrium.* What we are using is the fact that, once we know that a state of equilibrium can exist and the rate of the destruction of X, we automatically must know the rate of its creation as well. This principle is known as **the Principle of Detailed Balance** and is generally very useful.

There is one final notation that we introduce. The quantity τ_X is, roughly speaking, how long it takes for a typical X to decay. We can turn this instead into a rate of decays per unit time, Γ_X, by taking the reciprocal:

$$\Gamma_X \equiv \frac{1}{\tau_X}, \tag{38}$$

and so the final form of our Boltzmann equation is

$$\frac{dn_X}{dt} = -\Gamma_X\left(n_X - n_X^{\rm eq}\right). \tag{39}$$

We can interpret it as follows: the destruction rate of X is Γ_X, but this is balanced by the inverse process. If $n_X > n_X^{\rm eq}$, X will decay away until it is balanced by the inverse production processes. Similarly, if we start too few X particles, they will be produced by $e^+e^- \to X$ until the production rate is balanced by the decay rate.

We can now generalize this to any number of processes:

$$\frac{dn_X}{dt} = -(\Gamma_{X\to A_1 B_1} + \Gamma_{X\to A_2 B_2} + \ldots)\left(n_X - n_X^{\rm eq}\right). \tag{40}$$

This is valid as long as all the particles X can decay into are already in thermal equilibrium with one another[13]. A limitation of the Boltzmann equations in this form is that they really only apply when we have one species that is out of equilibrium, along with a large number of particles that *are* in equilibrium with one another. Luckily, this is precisely the scenario in the early universe when all SM particles formed a single equilibrium configuration with temperature T, and then we can imagine one or two out-of-equilibrium species that make up, say, dark matter.

4.2 Particle Abundances in an Expanding Universe

In the last section, we had an in-depth discussion of the thermodynamics of particles in a static Universe. Now, we return to the actual case of an expanding Universe. We immediately run into a difficulty: equilibrium is defined by quantities that do not change with time, whereas we know that the scale factor changes with time, $\dot{a}(t) \neq 0$. Thus, the expanding Universe is by definition *not* in equilibrium!

Do we have to abandon all of the notions of equilibrium we just derived? Fortunately, we do not. The reason is that SM processes are happening very quickly compared to the expansion of the Universe; in other words, the rates of SM particle interactions exceeds the Hubble rate H. One way to understand this is that the Hubble expansion occurs due to gravity, which is much weaker than other SM forces. When the Universe expands a tiny bit, the "gas" of SM particles quickly adjusts its temperature to account for the expansion and reaches a new steady state. We call this a **quasi-equilibrium approximation**, because the Universe is never in true equilibrium, but it reaches an approximate equilibrium at each moment in time.

One of the effects of the expansion is that the temperature of the system becomes *time dependent.* The reason is that SM particles are typically relativistic at high temperature; as the Universe expands, the particles are redshifted and lose energy. In fact, we find that during a typical epoch of the Universe's history,

$$T(t) \propto \frac{1}{a(t)}. \tag{41}$$

In an era where most SM particles are relativistic, we also can calculate the Hubble rate explicitly,

$$H = \frac{1}{2t}. \tag{42}$$

The explicit relationship between temperature and time is

$$\frac{1}{2t} = H(t) = \frac{T^2}{M_0}, \tag{43}$$

where $M_0 = 7 \times 10^{17}$ GeV is a constant related to the strength of the gravitational force.

[13] If this were not true, then there would be no unique temperature and so no single value of $n_X^{\rm eq}$. As you can imagine, this makes life considerably more difficult.

So, one effect of the Universe's expansion is a reduction in the temperature of the particles present in the early Universe. This reduction in temperature in turn affects all the particle destruction rates Γ, so these also become functions of temperature. Are there any other effects of the expansion?

In Section 3.1, we argued that the expansion also *dilutes* the number density of particles. Indeed, we argued that

$$n(t) \propto \frac{1}{a(t)^3} \propto T^3. \tag{44}$$

This means that the number density of our particle species not only changes due to scattering and decays (which we modelled in Section 4.1.4), but also due to the expansion. We conjecture that the change in the number density of a species X due to the Hubble expansion is

$$\left(\frac{dn_X}{dt}\right)_{\text{expansion}} = -3Hn_X. \tag{45}$$

I encourage you to substitute $H = \dot{a}/a$, solve this differential equation assuming no other interactions, and show that n_X satisfies the relation Eq. (44).

We can now write the full Boltzmann equation as

$$\frac{dn_X}{dt} = -3Hn_X - \left(\Gamma_{X\to A_1B_1} + \Gamma_{X\to A_2B_2} + \dots\right)\left(n_X - n_X^{\text{eq}}\right). \tag{46}$$

This is interesting, because we now see there are *competing* rates. If all of the scattering rates Γ_X are $< H$, then we can get an approximate solution by throwing away all of the terms proportional to Γ. The resulting equation is simply $dn_X/dt = -3Hn$: any amount of X that already exists gets diluted by the expansion, but nothing else happens. The reason is that the universe is expanding too quickly for any other particle interactions to have any effect.

By contrast, in the limit $\Gamma_X > H$, we can to a good approximation neglect the term proportional to H and recover the equations for a non-expanding spacetime. The result is that n_X will track its equilibrium distribution. The only effect of the expansion of spacetime is indirect: n_X^{eq} depends on T, which changes with the expansion.

The interesting physics happens when $\Gamma_X \sim H$, which is when the equilibrium condition is marginally satisfied; we find that particles come into or go out of equilibrium at stages of the Universe's history when its expansion rate equals particle production and destruction rates!

Eq. (46) is appealing because it explicitly lays out all the physical processes happening in the expanding early Universe and allows us to see the competition between them. In practice, it is unnecessarily complicated in terms of trying to solve the equation. The reason is that we already know what the effect of the expansion is: it simply reduces the abundance by a factor of $1/a^3$! So we don't need to re-solve this part of the equation every time we want to solve the Boltzmann equation.

To address this, we exchange the number density for a dimensionless quantity that more simply captures the dilution effect of the expansion. We define something known as the **yield**:

$$Y_X(t) \equiv \mathcal{C}\, n_X(t)\, a(t)^3, \tag{47}$$

where C is a normalization constant[14]. Upon taking a time derivative, we get

$$\dot{Y}_X = \mathcal{C}\left(\dot{n}_X\, a^3 + n_X\, 3a^2\dot{a}\right) \tag{48}$$

$$= \mathcal{C}\left(\dot{n}_X + 3Hn_X\right) a^3. \tag{49}$$

[14]For those who have taken statistical mechanics, it is conventional to choose C such that the yield is the number density divided by the entropy density of the system. This is a convenient normalization because entropy is typically conserved under the effects of the expansion of the universe, and consequently the entropy density scales like a^{-3}.

Substituting this into Eq. (46), we find a simplification of the Boltzmann equation with this new choice of function:

$$\frac{dY_X}{dt} = -(\Gamma_{X\to A_1B_1} + \Gamma_{X\to A_2B_2} + \ldots)\left(Y_X - Y_X^{\text{eq}}\right). \tag{50}$$

This has a simpler form because we have already included the effect of the Universe's expansion on the X number density. Indeed, we see that if Γ_X is negligible, then Y_X is approximately constant. The reason that this works is that the decrease in $n_X \propto a^{-3}$ from Hubble expansion is cancelled by the factor of a^3 in the definitino of yield. An additional advantage of using Y_X is that, because the effects of the expansion are already removed from its evolution, we can directly compare values of Y from different periods in the universe's history without needing to worry about how distances in the universe have changed over time.