Appendix A
The First Millisecond

Last update: 5/1/08

§A.1 The First Millisecond

Well, we have to start sometime. Unless our ambition is so great as to attempt to explain the ultimate origin of the universe, we must start the clock at a little after time zero. Many physicists' ambitions are indeed so great as to have the ultimate explanation as their goal. To disguise their hubris they use another name, such as quantum cosmology or M-theory. Our ambitions are far more modest. The advantage of starting the clock at a few milliseconds is that the universe is at its simplest at this time. It will never be so simple again, and was not so simple before. At one millisecond the universe consisted of just photons, protons and neutrons, plus neutrinos and electrons and their antiparticles. We presume it also included dark matter and dark energy, since these appear to be present at later epochs. If dark energy is a constant vacuum energy, then it would have been of negligible consequence at this early time.

Most importantly, the photons, protons, neutrons, neutrinos, antineutrinos, electrons and positrons were in the form of a homogeneous, extremely hot, gas. The number of electrons and positrons, and each type of neutrino and antineutrino, would have been the same[1], and equal to ¾ of the number of photons at around the first millisecond. The numbers of neutrons and protons would have been almost equal, but far, far less than the numbers of the lighter particles. The reason for not starting a little earlier than one millisecond is that other particles would then also be present. As we consider successively earlier times, firstly muons and their antiparticles would appear in large numbers, and then the pions and their antiparticles. At earlier and earlier times, so particles of ever greater mass appear in large numbers, most of which are strongly interacting via the nuclear force.

The physics becomes far more challenging at such times because of the strong interactions between the particles. Pretty soon we find a whole zoo of strongly interacting particles (hadrons) to be present. In fact, there comes a point at which the hot pottage of hadrons undergoes a phase change and becomes a soup of quarks and gluons instead. But this happens at a few tens of microseconds, and we'll not go that far back. There is little point in studying the physics at these times (for our purposes). It has left no imprint on the later universe. The fact that there was once a similar number of neutral pions as photons is of no relevance now – and indeed of no relevance after a few milliseconds. So we loose nothing by starting our clock at around a millisecond other than some difficult, but irrelevant, physics.

If this seems too good to be true, it is. There are some exceptions. There are a small number of extremely important relics from the very early universe which have a crucial impact on its later development. The first of these are the neutrons and the protons, the major constituents of ordinary matter. Since these are heavier particles than muons and

[1] Well, very nearly

pions, they should have disappeared from the universe well before one millisecond. Now the mechanism which leads to the hadrons disappearing is that they annihilate with their antiparticles. So, for all of them to disappear, there must be the same number of particles and antiparticles. At about, say, a tenth of a microsecond there would have been *almost* the same number of protons as antiprotons, and *almost* the same number of neutrons as antineutrons[2]. But there must have been just a tiny excess of protons and neutrons over their antiparticles, an excess of only about one in a billion. By one millisecond it is only this tiny excess which survives – and the antiprotons and antineutrons have all gone.

This is why, at one millisecond, there are far fewer neutrons and protons than there are photons, neutrinos, electrons and positrons. By rights, there should not be any at all! So we can now see that the reason for there being any protons and neutrons left over is due to the tiny excess that exists at very early times. And this is due to some complex physics at these times which we are going to avoid addressing[3]. The good news, though, is that we can simply regard the numbers of neutrons and protons as another universal constant. More precisely, the number of nucleons per photon is our new universal constant. Of course, we loose something in terms of fundamental understanding by this approach. Physicists will always wish to explain *why* the photon:nucleon ratio takes the value that it does (namely, about 2×10^9). However, as discussed in chapter 2 of the "Cosmic Coincidences", the standard model of particle physics contains many more constants than we shall employ. A valid objective for our purposes is to seek the smallest set of universal constants which suffices to uniquely constrain the later universe. Strangely, it might be that this is smaller than the number of independent constants in a more fundamental theory. Whilst one day there might be a better theory with fewer constants, such as string theory or its variants, we do not currently have any such theory.

The photon:nucleon ratio (or photon:baryon ratio, as it is generally called) is not the only relic from the very early universe. Dark matter and dark energy are also presumably relics from the earliest era. Like the photon:baryon ratio they can be taken into account through additional universal constants. Like the photon:baryon ratio, it may be that one day these dark matter and dark energy constants can be explained in terms of fundamental physics – but not as yet. Finally, there is at least one more constant that has its origins in the very early period. This is the degree of inhomogeneity in the early universe, at a time before galaxies began to form.

The universe is in fact extremely homogeneous on a large enough scale. This is apparent in the distribution of the galaxies, provided one looks on a large enough spatial scale (on smaller scales there are clusters of galaxies, and clusters of clusters of galaxies). The very high degree of isotropy in the cosmic microwave background radiation provides conclusive evidence for the homogeneity of the early universe. Nevertheless, some slight deviations from perfect homogeneity must have been present to seed the gravitational collapse that would lead ultimately to the formation of the galaxies. Great excitement was

---

[2] Not strictly true, of course, because they would be dissociated into their constituent quarks. But if you collected the quarks together and re-assembled hadrons, this is what it would amount to.

[3] In any case it is not fully understood, though the asymmetry between particles and antiparticles implies that the usual time-reversal symmetry which most physical laws display does not apply exactly under these conditions of very high energy.

caused when the COBE satellite produced the first high resolution maps of the microwave background, and quantified the departure from perfect isotropy. The inhomogeneity parameter (Q) measures by how much the background microwave radiation varies from place to place. It is a very small fraction, namely ~2 x 10^{-5}. Once again, the physical origins of this parameter are mysterious, but presumably lie in the extremely early universe. It may arise due to quantum fluctuations (see below). But we shall merely treat Q as another of our universal constants.

In summary, by starting our clock at around a millisecond we avoid some difficult, contentious, and largely unknown physics. Nothing is lost by so doing, since the effect of the physics in the first millisecond on the later universe can be accounted for by a small number of extra universal constants of cosmological origin (at least four of them, possibly a few more – see chapter 2 of the “Cosmic Coincidences”). Physicists would wish to reduce all constants to basic physics, i.e. to derive the cosmological constants from physical constants by starting the clock closer to, perhaps at, time zero. This is prevented currently by the absence of an adequate theory. In any case, judging from the current standard model of particle physics, it is not clear that such an approach would involve fewer constants than we use here. This matter is contentious.

§A.2 The Creation – A Scientific Story

Here is a mythopoeic version of the modern scientific creation story. Ifs and buts and exclusion clauses will follow:-

In the beginning there was nothing. There was neither matter nor energy. There was no space and no time. But there was Law, which is everlasting. One of the Laws is called Quantum Mechanics. This Law permits mass-energy to arise out of nothing – but only very briefly, a mere fluctuation. Quantum Mechanics is a loan bank with stringent bookkeeping. The mass-energy debt must be repaid, and quickly. The terms are harsh: the greater the loan, the shorter the loan period. Under the Law of Quantum Mechanics, many small acts of creation are possible, if only for an instant. Each one is a tiny snap, crackle and pop amounting to nothing of significance. Each one dies as soon as it arises, leaving nothing once again. A single pair of nucleons, the building blocks of matter, created in this way would last less than a millionth of a millionth of a millionth of a millionth of a second (10^{-24} s). The creative loophole allowed by Quantum Mechanics is prodigiously tiny. And the creation of large quantities of matter would last proportionately less time. This is no way to create a universe, you may think.

But there is another Great Law, called Gravity, which conspired with Quantum Mechanics. Under the Law of Gravity, all mass is attracted to other mass. As a consequence, a form of negative energy arises when a mass is close to another mass, called “gravitational potential energy”. Because of this, the Law of Gravity permits spacetime to exist containing ponderable matter such that the net mass-energy content is exactly zero. So Quantum Mechanics and Gravity combined their powers to create a fluctuation in which the positive mass-energy was exactly balanced by the negative gravitational potential energy. The net energy was zero, and thus the repayment period was permitted to be infinitely long. By this ruse, a long lived universe was brought into *being from nothing by these Laws. It has been called the ultimate free lunch (Guth, 1997).*

But there was a price to pay for this success: the content of such a universe adds up to nothing. Nor can time alter this. The apparent content of the universe consists of the separation of the positive and negative parts of nothing. No matter how complex and full of wonderful structure the universe may become in its middle age, recall that it will always amount to nothing. A free lunch, maybe, but the lunch is a mirage.

This is a nice picture, due originally to Tryon (1973). It may be true. But the rigorous scientific position is rather more uncertain. One reason for this is that the status of the gravitational potential energy is problematic. There are different possible definitions for this energy, and it is not clear that they must be equivalent in all cases. Nevertheless, it has been shown that reasonable definitions of the total energy are indeed zero for many cases of cosmological interest. Most often this result is associated with closed universes, e.g., Rosen (1994), Cooperstock (1994), Banerjee and Sen (1997), and Xulu (1999). However, cosmologically relevant cases of open universes have also been argued to have zero net energy, e.g., Faraoni and Cooperstock (2002). Another contentious component of the preceding story is that the physical Laws are immutable, already existing at the instant of creation. One school of thought [Rees (2003), Tegmark (2003), Tegmark, Aguirre, Rees, and Wilczek (2006)] holds that the mathematical form of the laws of physics are themselves subject to some form of emergence and evolution.

The event of creation is called the Big Bang. It is not known whether this occurred at a single point, or whether the universe was created infinite. In either case, the universe must have been created in a state of divergently large density. The reason is that, to create such a very large inventory of matter, the balancing (negative) gravitational potential energy was required to be equally large. But large (negative) gravitational energies occur when a lot of matter is confined to a very small volume, in other words – very high density.

What happens next? In particular, why should all that positive and negative energy, confined together in a small volume, not immediately neutralise? Why should this large quantum fluctuation not just go pop, back to nothing, like all the microscopic fluctuations? The loan bank of Quantum Mechanics may not require repayment, but are there not other Laws, such as the great law Thermodynamics, which would demand immediate annihilation of this presumptuous embryo? The survival of the infant universe depends upon it expanding so quickly that it escapes this fate.

Why should the universe be expanding? The true explanation lies in general relativity. However, a non-relativistic rationalisation of the expansion can be made. For this purpose consider the “universe” to mean the cloud of mass-energy which has just been created, but imagine this to have been placed in an infinite Euclidean space (this is why it is a non-relativistic view). From the point of view of this non-relativistic model, the reason for the expansion of the “universe” lies in the conditions required for a system of masses to be bound together by gravity. This condition is that the total kinetic energy of the masses equals half the magnitude of their gravitational potential energy (this is known as the Virial Theorem). Any system of gravitating masses with less kinetic energy will undergo gravitational collapse. Conversely, any system with more kinetic energy is not bound by gravity. In such a case, the masses will fly apart from one another. But we have seen that, in our embryonic universe, the mass-energy of the matter is equal in magnitude to the gravitational potential energy (but opposite in sign, so they add to zero). Now the energies of the particles of matter are so prodigious at this time that they exceed their rest-mass energies (mc^2) greatly. Consequently their kinetic energy is essentially the same as their total mass-energy, and is therefore equal, on average, to (the magnitude of) their gravitational potential energy. It follows that the "universe" is not gravitationally bound. Moreover, with such huge energies, the particles of matter must be moving at very high speeds, approaching the speed of light. The same is true, of course, of the quanta which comprise the energy fields, such as photons – the particles of light. In short, the whole universe is expanding very rapidly.

Gravity, having conspired with Quantum Mechanics to take the lid off Pandora's box, cannot stop the contents flying out.

This is rather a travesty of the correct picture, which is essentially relativistic. It serves only to make the universal expansion seem intuitively reasonable. In truth, there is no pre-existing Euclidean space in which the Big Bang occurs as some local event. Instead spacetime is created along with the mass-energy content of the universe. The Big Bang fireball exists homogeneously at all points in the space thus created. What is expanding is not only the fireball but also the space in which it exists. Indeed, if space did not expand the fireball would have nowhere to expand into, since it occupies the whole of space at all instants. The details of the expanding spacetime geometry are explained in Chapters 5B and 5C of the Cosmology Tutorial.

The very high speeds of the particles of matter at this time imply very high temperatures[4]. As the universe expands, so its gravitational potential energy reduces in magnitude as the particles move further apart. The mean kinetic energy, and hence the temperature, therefore also reduce. The physics of the first microsecond is fascinating, difficult and highly uncertain in most parts. A brief summary is provided below.

§A.3 The First Microsecond – Another Scientific Story

The earliest time which makes any sense at all is the so-called Planck era at $\sim 10^{-43}$ seconds. At a time scale of this order, spacetime itself ceases to behave as a continuum. Hence $\sim 10^{-43}$ of a second can be regarded as a quantum of time. The corresponding quantum of space, the Planck length, is $\sim 10^{-35}$ m. These quantities are $\sqrt{\hbar G / c^5}$ and $\sqrt{\hbar G / c^3}$ respectively. Up until $\sim 10^{-38}$ seconds, when prevailing thermal energies exceed $\sim 10^{16}$ GeV, the strong, weak and electromagnetic forces behave as a single unified force. After this time the three forces will start to become distinct. At perhaps $\sim 10^{-36}$ to $\sim 10^{-33}$ seconds, a most important event called "inflation" is believed to occur [see Guth (1998) for an entertaining first hand account of how the theory arose].

---

[4] Strictly, the concept of temperature only applies where there is thermodynamic equilibrium. Despite its rapid expansion, the universe does maintain thermodynamic equilibrium at this time – and in fact continues to do so for ~10 million years, Peebles (1993), Equ.(6.138).

Appendix A - The First Millisecond

The Big Bang model as described so far has several shortcomings. It fails to explain four things: the flatness problem, the horizon problem, the absence of magnetic monopoles and other exotica, and the origin of density fluctuations which will later give rise to structure in the universe. Inflation theory provides explanations of all four of these issues.

The flatness problem is dealt with in detail in Chapter 2B of the Cosmology Tutorial. In brief, the flatness problem refers to the fact that the mean density of the universe is of the same order as the critical density that marks the boundary between a closed (finite) universe and an open (infinite) universe. This implies that the universe started off extremely close to being exactly flat, since otherwise the mean density would have diverged from the critical density very rapidly. The question is, why should the universe be so close to perfectly flat? Inflation theory provides an answer. As the name implies, this theory envisages the universe as inflating extremely rapidly at around $\sim 10^{-36}$ to $\sim 10^{-33}$ seconds. Although the universe is expanding anyway, during inflation it expands dramatically faster. The idea is that a repulsive gravitational field created by a false vacuum causes exponential expansion. This rapid expansion caused the universe to increase in size by at least a factor of $\sim 10^{30}$ in a mere $\sim 10^{-33}$ seconds. Just before inflation, the current observable universe, of order 10^{26} m in size, would have been comparable with the Planck size scale. Without inflation it would have been about a metre in size at 10^{-33} seconds. The period of rapid inflation, brief though it is, provides a mechanism for forcing the universe to become very flat indeed. The result of inflation is to force the density of the universe to equal the critical density to an extremely high degree of precision.

The horizon problem is discussed in Chapter 2 of the Cosmology Tutorial. It relates to the fact that, in the standard Big Bang model, not all of the universe which is currently observable would have been in mutual causal contact in earlier epochs. This would mean that two widely separated regions of the sky would never have been in contact with each other, e.g. via exchange of light or other radiation. The size of the region containing the matter which is now observable varies as we look back in time as $r \approx r_0 (t/t_0)^{2/3}$, where r_0 and t_0 are the size and age of the observable universe now. So, although this region shrinks as we look back to earlier times, it does not shrink as quickly as the maximum distance that can be covered by a beam of light, which is of order ct. (Actually closer to 3ct due to the expansion of space, but this does not affect the argument). Hence, the fraction of the currently observable universe which was causally connected at time t is $(t/t_0)^{1/3}$, and this becomes arbitrarily small as we consider earlier times. For example, at the time from which the cosmic microwave background appears to originate (~360,000 years), this gives a fraction ~0.03 (using 13.7 Byrs as the age of the universe). This corresponds to 1.7 angular degrees in the sky.

The horizon problem refers to the fact that the universe is actually very homogeneous on large scales. Beyond the size scale of the largest superclusters of galaxies (a few tens of Mpc) the universe has a uniform distribution of matter. Even more spectacularly, the cosmic microwave background (CBM) is isotropic to an accuracy of about one part in 100,000. The theoretical problem that this poses is, how does this homogeneity come about in view of the fact that regions of the sky more than a degree or two apart have never been in contact with each other? The isotropy of the CMB implies thermal equilibrium. But how can this ever have been established? Inflation theory also solves this problem by noting that, just before the inflationary period, causal connection would apply throughout the region destined to become the currently observable universe. For example, at ~10^{-33} seconds, causality will stretch over about ct ~ 10^{-25} m, whereas the pre-inflation universe is a mere 10^{-35} m.. Almost no time later (well, ~10^{-33} seconds later), the post-inflation universe will be about a metre in size, and hence causal connection will have ceased and will not recur until the present time. It is the period of exponentially rapid inflation which allows the very small, causally connected universe at ~10^{-33} s, to produce the larger, causally disconnected, universe which is necessary for the subsequent evolution to be consistent with standard Big Bang evolution at later times. From this perspective, the thermal equilibrium, of which we still see the evidence in the homogeneity of the universe, was established prior to the first 10^{-33} of a second.

The third problem which inflation theory explains is the absence of various exotic particles which particle theory suggests should be created in the very early universe. The usual example is the magnetic monopole. Magnetic monopoles are believed to have a stupendously large mass (in elementary particle terms, that is) of ~10^{17} GeV (about 0.1 micrograms). They could therefore be created when ambient thermal energies were of this order or greater, before about 10^{-40} seconds. Inflation theory does nothing to prevent this, but nevertheless gets rid of the unwanted monopoles simply by diluting them out of existence. In fact, all particle species which are present before inflation are diluted away by the exponential expansion. All except the inflaton, the hypothetical particle which is postulated to be responsible for causing the inflation. The inflaton is generally supposed to be a self-interacting scalar field with a potential which dominates its kinetic energy. This leads to a negative pressure which results in a repulsive gravitational effect via essentially conventional general relativistic equations of motion. Because inflation dilutes away the existing inventory of matter, the inflaton field itself carries the whole energy content of the universe during the inflationary epoch. It comes to an end when the unstable inflaton vacuum decays, giving rise to a rebirth of other particles, a phenomenon known as ‘reheating’. But by this time (~10^{-33} s) the temperature achievable will be insufficient to produce particles as massive as monopoles. Consequently, whilst other particles, with masses up to perhaps ~10^{15} GeV, will be formed, the monopole menace has been dealt with.

The final triumph of inflation theory *might be* that it will manage to pull off the trick of explaining why the universe is so homogeneous whilst also providing a mechanism for those small deviations from perfect homogeneity which are essential to structure formation. The mechanism is postulated to be, once again, quantum fluctuations. Quantum fluctuations can occur before, and during the early stages of, inflation. These become ‘frozen-in’ when inflation takes the universe out of causal connection. These are the seeds of the inhomogeneities from which structure develops later in the life of the universe. Observations are consistent with the primordial density perturbations being entirely adiabatic, providing key support for inflation theory.

Immediately following the inflationary period, perhaps at 10^{-33} to10^{-30} seconds or so, but possibly extending later than this, comes baryogenesis. Could it be *much* later, like at the electroweak scale, ~10^{-11} sec? The observable universe appears to consist of matter but not anti-matter (Reference?). Yet the established laws of particle physics are symmetrical as regards matter and anti-matter (so-called charge conjugation symmetry, denoted C). Why does matter predominate? Whatever the mechanisms leading to the dominance of matter over anti-matter, they are collectively known as baryogenesis. Clearly, some violation of charge conjugation symmetry must occur. It is generally accepted that a departure from thermal equilibrium is also necessary during baryogenesis. But the physics of baryogenesis remains highly speculative. It is an active, but difficult, research area. See Dologov (1997, 2006), Riotto (1998) and Cline (2006) for reviews. We note in passing that, whilst it is unlikely that there are significant amounts of anti-matter in the observable universe, it could be that anti-matter predominates in unobservable parts of the universe/multiverse.

Leaping forward to $\sim 10^{-11}$ seconds brings us to the electroweak transition, when all particles other than the Higgs boson are supposed to acquire their mass through spontaneous gauge-symmetry breaking. Note that at this time the strongly interacting part of the plasma consists of a quark-gluon soup. At $\sim 10^{-11}$ seconds, all six quark flavours would be present in the quark-gluon soup. The other components of the plasma are the leptons, the W and Z bosons and presumably the Higgs boson (if it exists). There might also be various supersymmetric particles (if these exist).

By $\sim 10^{-8}$ seconds the most popular candidates for the components of cold dark matter, generically called WIMPs, will probably have had their abundance fixed for all time, any reactions leading to their annihilation or creation having frozen-out, Schwartz (2003). By about a millisecond they will have decoupled completely from the rest of the universe, except through gravity. After 10^{-8} seconds we enter the regime of relatively well established physics, at least as regards knowing what particles exist at the corresponding mass.

As one microsecond approaches, the heavier particles annihilate. Thus the W, Z and Higgs bosons would no longer be present, nor would the top quark and few bottom quarks. At all times the anti-quarks are present in almost, but not quite, identical numbers to the quarks. The mechanisms of baryogenesis, whatever they might be, have ensured a very slight preponderance of quarks over anti-quarks. At one microsecond the universe still consists of a quark-gluon soup, albeit with just four of the six quark flavours contributing, together with the leptons, and, of course, high energy photons. Of the three massive leptons, the tau would be rapidly disappearing.

At $\sim 10^{-5}$ seconds the so-called QCD[5] transition occurs. After this QCD phase transition, the hadrons condense out of the quark-gluon soup. This is when the nucleons (protons and neutrons) form. They are accompanied by lower mass hadrons such as pions. Note that the typical thermal energies available at this time are such that particles much more massive than nucleons (i.e. more massive than ~1 GeV) are unlikely to form even transiently. The amusing corollary of this is that the entire zoo of hadrons with masses much greater than 1 GeV had probably never existed anywhere in the universe before being created artificially in accelerators on earth. When thermal energies were large enough for their production, before the first microsecond, the QCD phase transition had not occurred, and hadrons as such did not exist – only a quark-gluon plasma. **Is this true?**

[5] Quantum Chromodynamics, the SU(3) gauge theory of the strong nuclear force.

Appendix A - The First Millisecond

At ~100 microseconds the ambient thermal energy is typically ~100 MeV. This is less than the nucleon mass (~938 MeV) and so there will be negligible numbers of anti-nucleons. The nucleons now present will persist for the life of the universe (though many of the neutrons will decay into protons and electrons). At 100 microseconds the thermal energies are too low to create tau leptons, which have annihilated. But muons and pions are still present, though the latter are disappearing rapidly.

By one millisecond even the muons have disappeared. The universe now consists of photons and the light leptons and their antiparticles, all these in comparable numbers. By the light leptons we mean the electron and all three flavours of neutrino. If cold dark matter exists, then these particles will also be present (be they WIMPs, axions, neutralinos, or whatever). Dark energy, if it exists and if it is a constant vacuum energy density, will be negligible at this period.

In addition there are the remnant nucleons, but there is merely one nucleon for every ~10^9 of the other particles. This thin scum of heavier particles, which has escaped annihilation only through the unknown CP-symmetry violating mechanism, is what will ultimately form the whole of the material world which we know.