Deciphering Modern Cosmology: Energy in the Universe

A. Macleod

macleod.cnoc@gmail.com

Abstract. The problem of energy conservation in the universe is investigated both in the context of conventional cosmology and alternatives.

Table of Contents

Overview

The Societal Energy Problem

We are being made aware almost every day of an approaching energy crisis. There seems to be no such thing as free energy, energy without cost. Continuing consumption of fossil fuels will almost certainly result in damaging climate change, and all the alternatives come with a range of drawbacks. Addressing the problem is becoming critical, and is likely to become the most important societal challenge of the next 50 years.

There is a huge dependence on concentrated energy sources to power our technology. The unrelenting industrialisation of modern society continues to be powered by rich energy sources found within the earth. Energy from the sun that was captured by plant life over millions of years has accumulated in the ground as dead organic matter. Heat and pressure underground have cooked the organic molecules into familiar hydrocarbon mixtures such as oil, coal and gas. These fossil fuels began being removed from the ground in the 19th century to provide the raw energy needed to drive the industrial revolution, and fossil fuels still continue to be extracted from the earth in vast quantities and now work just about everything around us. But what has taken millions of years to accumulate in deep stores within the earth, we seemingly are blithely raiding and will perhaps consume in a mere two hundred years.

Scarcity must eventually become a problem for a world that is heavily dependent on fossil fuels and it seems like this dependence is growing as more and more of the world becomes industrialised. What will we do when there is no fossil fuel left? And are there any alternative sources of energy? Whatever the answer to these questions, change seems inevitable.

Changes in the way energy is produced and consumed will have to take place and will have a huge impact on the individual and society, and there is an increasing belief that making a significant transition now (through choice) from cheap reliable energy to the renewable energy that is available all around will ultimately benefit everyone. Energy will be extracted as required from what is available around about, without a need to draw on the energy of the past (or the future). The 'new' energy will likely be unreliable, intermittent and difficult to capture in sufficient quantity; we will have to take it as it comes – it will not be available 'on tap' just as we want it and it may not be feasible to use energy in the same wasteful and reckless way we currently do.

And it is not solely a matter of supply and demand - there is the additional problem of climate change associated with the burning of carbon. A rising atmospheric carbon dioxide concentration is likely to result in an overall global temperature rise. However, the uncertainties on the modelling process used to anticipate future effects makes it hard to predict what will really happen: predictions range from a runaway greenhouse making the planet uninhabitable, to a 300 year blip which disappears into a highly variable temperature history as the planet self regulates after most of the available fossilised carbon is consumed. It is not even possible to make short term predictions concerning warming effects it would seem. But given the risk, a ‘do nothing’ approach is not to be recommended.

Perhaps surprisingly, the global effort in dealing with carbon emissions has been modest. There is definitely an awareness of the issues, but this has not yet translated into globally significant legislation and behaviour change. This could be because it is known that there is already have a solution available, a massive shift to nuclear power. However, this is most people’s plan B or C, very much a last resort.

What can Science Do?

The energy problem is largely societal in nature and can potentially be controlled through effective economic and political actions, but is there anything science and engineering can do to avoid the inconvenience of having to change the way we consume energy? The current political approach is to turn over the problem to the free markets by incentivising the transition to a large-scale exploitation of renewable sources, but this policy will come under increasing scrutiny as the ‘switch-over’ point is approached. This is when the energy system can be considered primarily renewable backed by conventional energy rather than what we have at the moment - conventional energy with some (disruptive) renewable energy input. This is the point we will start noticing an even more substantial increase in the price of energy.

Can we rely on scientists to get us out of the fix? There are certainly precedents. World starvation was predicted in 1900 as the supply of natural fertilizer used in agriculture, guano, was beginning to run out. But in 1908 Fritz Haber discovered an energy intensive process for capturing nitrogen from air which led to the manufacture of artificial fertilizers. Crisis averted. But there is unlikely to be a quick fix to the current energy problem because we are really at the end of the line (if our current understanding of science is correct). We are running out of energy itself. And are we not entitled to expect a solution, given the huge investment we as a global society make in science?

Can science supply us with a completely new free energy source? On the face of it, there are difficulties in finding a scientific solution to the energy supply problem as we come up against the fundamental laws of energy conservation and the very nature of energy which makes hopes of finding a completely new and plentiful source of energy rather unlikely. It is not impossible though; the accepted model of the cosmos requires the existence of an unknown substance called dark energy to fine-tune the observed expansion rate of the universe, and invisible dark matter to explain anomalous behaviour on the scale of galaxies and above[1]. At the other end of the scale, quantum physics describes the vacuum of

---

[1] When we do the sums, we find that only 5% of the mass of the universe is normal matter. The rest is dark matter and dark energy. If this dark material is distributed throughout space, could it not be harvested by collectors here on earth as the planet weaves a path through space? This may seem like jumping the gun a little when we know nothing about the material, but the point is that even if it could be gathered,

space as seething with transient particles which appear and disappear as energy is borrowed and returned. This is by no means fantasy as precise calculations have predicted that these particles can exert a measurable force, a force that has been observed as the Casimir Effect – the unexpected attraction of two metal plates in very close proximity. Another potential source of energy that will undoubtedly add to the list of 'free energy' proposals in future years is the Higgs field. This is the mechanism by which it is suggested particles acquire mass. The validity of the theory is well supported by the discovery in 2012 of the Higgs boson, a particle essential for the mechanism to operate. One other force you should be aware of is the so-called exclusion force. Once an atomic level is filled with the allowed number of electrons, it is not possible for any other electrons to occupy that level. This is presumably because an additional electron would be indistinguishable from the ones already there. No one know how to extract energy from this incredibly strong force.

Gravity could be an unexpected source of energy. Very curious behaviour observed in galaxies suggests the gravitational force is not as behaving as it should. The tangential velocity of stars orbiting spiral galaxies can be worked out from the Doppler shift associated with the motion. The results when viewing 'side-on' spiral galaxies are surprising. Instead of velocity falling off with distance (as happens with the planets orbiting the sun) the speed tends to level off. Though this can be the result of very specific distributions of dark matter, it could also be sign of gravity misbehaving on this scale, as suggested by a theory called MOND proposed by Mordehai Milgrom. This is an elementary modification to the gravitational force at low acceleration. Whilst the simplicity of the idea and the effectiveness is impressive, much more so than dark matter, the theory has very few supporters in the establishment. The problem is that it undermines 400 years of physics.

These are all tenuous possibilities, and, on balance, it is considered unlikely that a new source of energy will be found.

Energy and Science

Rather than speculate on possible new energy sources, a more solid approach might be to develop a complete understanding of physics on the large scale (particle physics) and the small scale (cosmology), then investigate the real energy possibilities that emerge. That end point is labelled the Theory of Everything (ToE), a theory which explains all physics. There has been much progress in understanding the universe on the large and small scale and over the last century

---

there is just too little of it to satisfy our energy requirements. The mean density of normal matter in space is only about 2 x 10^{-27} kg per m^3. The dark component density will therefore be 4 x 10^{-26} kg per m^3. A fixed collecting device with a 1 m^2 aperture will move at about 250 km s^{-1} relative to the centre of the galaxy would therefore gather 10^{-20} kg of material each second. Even if this material could all be converted to energy, this equates to only 0.001 J of energy , a power output of 1 mW!

and the imminent development of a ToE is considered a real possibility by many scientists.

But we have been here before, close to completing physics. All appeared well with science towards the end of the 19th century. An effective description of the physical world had been developed and stimulated the technologies that started the rapid industrialisation of western society: machines were invented that obediently followed the mechanical laws of Newton and the thermodynamic rules of Carnot, while electricity could be controlled through the application of Maxwell's equations.

There were niggling problems though; in particular, regarding the nature of light. Experiments appeared to show the measured speed of light was constant, irrespective of the movement relative to the light source of the equipment used to make the measurement. And if light was really a form of wave motion, then the theories predicted that a heated black body in equilibrium with its surroundings should emit radiation with infinite power. This clearly did not happen.

But few scientists were overly concerned with issues that were rather distant from real life applications of science and in all likelihood would eventually be explained in a way that might perhaps be obscure, but would still be consistent with the accepted world view.

In fact, the resolution of these seemingly trivial problems led to a scientific revolution. Space and time were discovered not to be absolute and separate entities as was previously thought, but were connected in a complex way that could result in geometrical distortions which seemed to explain the force of gravity. And light proved to be not just a wave, but both a wave and a particle. The newly developed quantum theory implied that electromagnetic interaction processes at a fundamental level are based on randomness and are completely unpredictable. The new physics undermined the notion of a universe operating along simple rules and has endured to this day[2], but this knowledge has enriched the modern world with many new devices based on quantum effects, not least the transistor, the building block of the modern computer era.

In the century since the start of this revolution, incredible progress has been made in making sense of this strange new world and the models have been refined into elegant structures which make predictions that almost completely

---

[2] And not just in science - around the beginning of the 20th century, classically accepted views where being replaced by a more creative and less restrictive interpretation of the subjects. Classical art developed in the renaissance and after into highly refined and lifelike representations, but this very limited mode of depiction collapsed with a number of movements around the beginning of the 20th century showing that art could do a lot more. The dramatic change in painting style is demonstrated by comparing 'The Ambassadors' from 1533 by Holbein the Younger with 'The Persistence of Memory' painted in 1931 by Salvador Dali. Holbein's painting shows carefully shelved objects representing the cosmic and religious domain, and items from the earthly domain. There is a clear emphasis on the importance of knowledge. Dali's image is a surreal landscape containing much more complex metaphors and the melting clocks can be associated with the discovery that time is not as rigid as was once thought.

match reality. Many scientists consider these standard models to be close to completion with just a few loose ends to tidy up. But are we making the same mistake as the physicists of a century and a half ago?

Energy in the Universe

But in spite of all scientific progress to date, it is perhaps surprising that in fundamental science there is a poor understanding of what energy really is. Energy appears to be conserved, hence it is technically possible to follow each joule of energy we consume in our homes step-by-step backwards, possibly all the way to the big bang. For this reason, it would seem the place to go to investigate the origin of energy is cosmology. The elaborate cosmological model that has been developed describes the flow of energy, and a close study of this model can potentially help us gain a better understanding of energy. The model is extremely effective as a description of the universe but it is hard to understand what it is telling us about the workings and origin of the universe. It is evident that current theories represent the efficient reduction of the data to the point of revealing patterns but there is a need to interpret (or decipher) the model in some way to extract meaning.

The model is a curious marriage of analogy[3] and mathematics and given that it is rather contrived in places and not entirely consistent, it is worth reviewing the basic assumptions in detail. In the chapters that follow, each a standalone piece of work, we look closely at the accepted cosmological model and check some of the foundational ideas[4].

This is far from straightforward. We are as a society dependent on energy and preparing for an energy crisis, but in truth, at a fundamental level, we still have no idea what energy is. But the situation is even worse than that - we do not understand space and time either. Given the surprising lack of progress answering fundamental questions, where should we go from here?

---

[3] A common device in cosmology is the use of metaphors and analogies to deal with abstract ideas, but this can be as confusing as it is enlightening because it is sometimes unclear whether or not the metaphors (models) are features of the real system. Care must be taken with any analogy. For example, David Mitchell writing about telomeres parrots the sentence: 'Telomeres are just the bits of DNA that protect our genetic code', whilst admitted not understanding what that means. He quotes the BBC explaining that telomeres are comparable to the plastic bit at the end of shoelaces, *flugelbinders*, that keeps the laces from unravelling. He comments that this analogy merely creates the illusion of understanding complex and still uncertain biology on the basis of an understanding of how shoelaces work. Whilst this may be helpful, it is possible for analogies to promote misunderstanding and it is very hard later to break a faulty association.

[4] On the cosmological scale only; the standard particle model will be critically evaluated elsewhere.

An Evaluation of Consensus Cosmology

The conventional cosmological model is very impressive in the way it matches up with observational data. It should be noted that the model is rarely reviewed at a foundational level, presumably on the belief that a theory this effective cannot possibly be wrong on a fundamental level. However, an objective look at the underlying model reveals the basic concepts are simplistic and thoroughly unconvincing. Because the original idea is safely buried under a mountain of mathematics the practitioner, almost certainly with a vested interest in the correctness of the model, will rarely if ever think about the basic model. But for an impartial external observer, the underlying concepts are extremely weak[5].

What are the problems? The universe is expanding into nothing following the rather childish analogy of an inflating balloon. Because of the expansion, mass that is moving with the expansion is slowed down as work is done against gravity. Whilst this would certainly be true of mass moving *through* space, where is the evidence that the expansion process, the growth of space, is coupled with matter in this unusual way? Frequently questions of this type which eventually converge on energy and energy conservation come up against the statement that the model is general relativistic in origin and that energy conservation is a grey area in general relativity. But why then the huge effort being made to ensure energy is conserved, even to the extent of introducing additional material with unusual properties in order to hold everything together?

To the external observer, the theory is a hotchpotch of fixes and the model as a whole is completely incoherent. The onus is on those supporting the theory to show that it is correct at a basic level. The key question is this: If the expansion is slowed down by the work done against gravity, there should be local variations seen in the expansion depending on the local mass distribution (after proper motion has been accounted for) - where is the evidence for this? And if the 'energy corrections' are global rather than local, what is the mechanism for maintaining this complex energy balance?

And these questions must be asked - now in 2020 there are problems emerging with the model that indicates another new patch is needed. What if the entire concept really is wrong? Should we be considering alternatives?

However, there are no alternatives - the consensus model has eliminated the competition! Evidence for the expansion is conclusive and it is hard to argue for any steady-state model. The only real possibility is to resurrect the discredited Milne cosmology from the 1930s. Is it worth revisiting?

But let us go back to the 1970s. Were supernovae data available then, the early version of the big bang model could be compared with the Milne model based on special relativity. The Milne model has not changed since - it provides a reasonable match for supernovae now and would have done so then as well. In comparison, the immature big bang cosmological model would have performed

---

[5] Care is needed when pointing this out as those active in this field are extremely sensitive to criticism and intolerant of criticism (though many must harbour private suspicions that all is not right).

very badly. The Milne model would at that time only have needed a small modification to perfectly match the data.

But the Milne model makes a huge assumption that seemingly cripples the model and requires serious investigation. The claim is that the expansion is not affected by gravity. In other words the expanding masses do no work as they are pulled apart by the expansion. We would need to investigate if this is really a problem if we were to push Milne cosmology as a modern alternative. We will do this in a serious of short standalone limited investigations. This is not an attempt to create a new theory but to show that other possibilities do exist and are worthy of exploration[6].

---

[6] Using plausibility arguments rather than trying to demonstrate a proof of concept. No attempt has been made to make the analysis rigorous.

1 An Interpretation of Modified Newtonian Dynamics

Summary

Modified Newtonian Dynamics (MOND) has been extremely effective as an explanation for the anomalous rotation curves of spiral galaxies. Although more elegant than the competing dark matter explanation, it is much less credible because the theory is in severe conflict with well established physical principles. A plausibility argument is presented here suggesting a direction by which the MOND effect might be brought into mainstream physics.

1.1 Introduction

Although the gravitational theories of Newton and Einstein are locally well established, it has been known for almost a century that gravity on the cosmic scale does not operate as expected. The anomalies begin at galaxy level where the dynamics of disk galaxies deviate significantly from expectation. Specifically, the rotation velocity of the outer disk material suggests the gravitational influence of much more mass than can be accounted with by tallying the visible stars and interstellar gas. The obvious explanation is to postulate invisible mass, but a huge amount of hidden material is then required, and it cannot interact like ordinary matter as to do so would directly reveal its presence. A dark matter solution is consistent with established physics and it is possible to account for all gravitational anomalies with appropriate dark matter aggregations. However, it is not an ideal solution because the nature of dark matter is unknown and has up to now eluded all searches.

There are some alternatives, but all require the gravitational force to be changed in some way. Modified Newtonian dynamics, MOND, is one such proposal and it has been particularly successful at explaining the anomalous effects in spiral galaxies by referring only to the visible matter [1], [2]. MOND relies on an interpolation function to manage a transition between 'normal' gravity at high accelerations and stronger than expected gravity at low accelerations (with the transition acceleration a_o approximately equal to 1.2 x 10^{-10} m s^{-2}). Several functions have been proposed on a trial-and-error basis without any compelling theoretical justification. In spite of the successes, which include numerous predictions that have proved correct, it is not widely believed that MOND correctly describes the operation of gravity on any scale.

The reason for the skepticism is evident from a cursory analysis. Consider two masses m_1 and m_2, a distance r apart with the gravitational force weak enough for a classical approach to be adequate. Adopting the most basic MOND interpolation function,

$$a_N = \mu\left(\frac{a}{a_o}\right)a, \tag{1.1}$$

with

$$\mu\left(\frac{a}{a_o}\right) = \frac{1}{1+\frac{a_o}{a}}, \tag{1.2}$$

where a is the actual acceleration and a_N is the expected Newtonian acceleration, the force on each mass because of the influence of the other is:

$$F_{1\leftarrow 2} = m_1 \frac{Gm_2}{2r^2}\left(1+\sqrt{1+\frac{4a_o r^2}{Gm_2}}\right), \tag{1.3}$$

$$F_{2\leftarrow 1} = m_2 \frac{Gm_1}{2r^2}\left(1+\sqrt{1+\frac{4a_o r^2}{Gm_1}}\right). \tag{1.4}$$

The forces are clearly not the same if $m_1 \neq m_2$. Action is not equal to reaction, a violation of Newton's third law that brings into question the validity of the MOND concept[1].

However, it is still possible that this apparent violation is an artefact arising from underlying hidden effects that are consistent with standard physics; a possibility that is investigated here.

1.2 Varying Gravity

One promising avenue of exploration is a varying gravitational constant. This could potentially reproduce MOND effects, but it seems a non-starter: it is straightforward to measure the strength of gravity within the solar system to high precision and current data do not reveal any temporal variation [3].

Nevertheless, a varying gravitational constant is still possible. It may be postulated that G is constant when the gravitational acceleration is above the a_o value that features in MOND, but once the acceleration drops below this level, G will vary, increasing with time. a_o therefore marks a transition acceleration that determines the stability of G: the value of G in the Solar System and within stars is on one side of the transition point and is constant; the outer edges of spiral galaxies are on the other side and are subject to a varying G.

The transition effect may even reflect a physical process. The MOND constant a_o can define a transition acceleration level below which an entity joins the Hubble flow. On joining the expansion, G no longer remains constant but changes over time, increasing with the expansion from its previously constant value.

Once an entity joins the Hubble flow the expansion velocity is related to the distance from a central mass through the usual relation:

$$v_h(T) = H(T)r(T) \approx \frac{r(T)}{T}, \tag{1.5}$$

where the approximation $H(T).T \approx 1$ has been used. T is the current age of the universe.

It is clear that the expansion velocity defined in this way will remain constant in the absence of forces: after interval ΔT in an external observer frame,

---

[1] There are relativistic versions of MOND which address these problems, but the fact the theory in the non-relativistic limit is so badly behaved is a serious concern.

$$
\begin{aligned}
v_h(T+\Delta T) &= \frac{r(T+\Delta T)}{T+\Delta T} \\
&= \frac{r(T)+\frac{r(T)}{T}\Delta T}{T+\Delta T} \\
&= v_h(T).
\end{aligned}
\tag{1.6}
$$

Now consider a central point mass with a test mass at distance r maintaining its radial position because of a balancing centripetal acceleration. The gravitational acceleration of the test mass is GM/r^2. If the acceleration falls to a_o the test mass joins the flow. (Note that the bound status of an object joining the flow is unaffected by the transition.) Assume it then continues moving outwards subject to a constant Hubble velocity. As it does so, the value of G will increase in a way that conserves energy (the work done against the field is balanced by increased gravitational binding)[2]:

$$\frac{dG}{dr} = \frac{G}{r}. \tag{1.7}$$

Because G/r is constant as the test mass moves outwards with the expansion, the proper acceleration will vary with r^{-1} from that point on instead of the normal r^{-2} dependence up to the transition point (assuming no peculiar radial velocity, i.e. perfectly circular orbits). The centripetal acceleration will continue to balance the inward acceleration as the test mass moves outwards because the centripetal acceleration also has a r^{-1} dependence[3].

The observed gravitational acceleration expected from the analysis above is the solid black line in Fig. 1 and shows a sharp gradient change at the transition acceleration. However, this is not consistent with observations around the transition region, which demonstrate a clear pattern that has been referred to as the radial acceleration relation (RAR) [4]. The data is overlaid in Fig. 1 and does not show the expected sharp transition at the critical acceleration. In addition, the concept of an entity joining the flow at a transition acceleration suffers from similar problems to those that plague MOND. Again, consider two masses m_1 and m_2; if $m_1 \neq m_2$ then one mass can be subject to a local acceleration that is below threshold, while the other one is at the same time subject to an acceleration that is above threshold. What happens then?

The concept of an entity joining the flow based on either an absolute or relative acceleration cannot be correct, and a more sophisticated approach is required. To persist with the idea of a transition event whilst maintaining consistency with other physical processes, an alternative approach is needed: is there is an energy condition that will moderate the transition?

---

[2] The validity of defining gravitational energy globally as before when G is changing is explored in detail elsewhere.

[3] The angular momentum increases as the masses move outwards, implying there is a torque acting on the masses. The source of this torque is discussed in detail elsewhere.

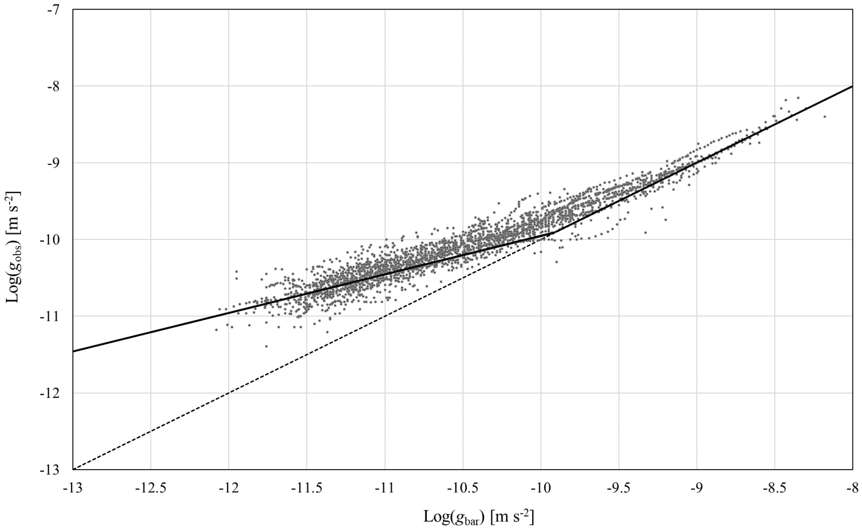


Fig. 1. The actual acceleration inferred from galaxy dynamics is plotted against the expected Newtonian acceleration from the visible mass. The dotted line represents no deviation from the standard Newtonian force. The black line shows a clean transition at a_o but does not match the data around the transition point. In addition, there is an anomalous increase in the observed gravitational acceleration even before the transition point is reached. Note that this is the gravitational acceleration with reference to a single entity over time rather than the actual variation over space at a particular time - this distinction must be made throughout.

1.3 Joining the Flow

Mass m_2 is in a perfectly circular orbit around larger mass concentration m_1 a distance r away. It joins the flow at cosmological time T and immediately gains velocity v_h (as defined by Equation 5). The increase in kinetic energy is

$$\Delta E_k = \frac{1}{2} m_2 \left(\frac{r}{T}\right)^2 . \tag{1.8}$$

Applying a Lorentz transform to reflect the newly acquired radial velocity, the retarded position of m_1 is altered, hence the gravitational energy (assuming no instantaneous change in G) is changed by

$$\Delta E_p = -\frac{G m_1 m_2}{r} \frac{v_h}{c}. \tag{1.9}$$

These balance if[4]

$$\frac{G m_1}{r^2} = \frac{c}{2T} > a_o. \tag{1.10}$$

There is a characteristic acceleration of

$$a = \frac{v_h}{\Delta T} = \frac{c}{T}, \tag{1.11}$$

assuming a propagation delay following the transition of r/c[5].

The idea of a transition when the acceleration is of the order c/T emerges from an energy threshold condition. This clears many of the objections of MOND, but of course the threshold condition in this case is variable because T is not a constant. In that case, here then does the MOND constant a_o come from?

Closer investigation of the data can reveal this. The RAR best-fit function [4] is:

$$a = \frac{a_N}{\sqrt{1 - e^{-\frac{a_N}{a_o}}}}. \tag{1.12}$$

Consider a modified interpretation where the transition occurs as expected once the acceleration is c/T (because of the energy condition) *but the accompanying change in the gravitational constant takes time to build up to the steady rate given by Equation 7.*

---

[4] The same argument can instead be applied to the total energy of the bound mass, $-Gm_1/2r$, in which case the threshold condition is c/T.

[5] Why include this? For a system in equilibrium the gravitational and electromagnetic forces act on the instantaneous (proper) position of the interacting entities. Thus, for all intents and purposes, the propagation delay is zero and the interaction can be considered instantaneous (though, of course, it is not). However, for systems that are the order of the universe in size, this can no longer hold and there is a genuine propagation delay which needs to be factored in. In this case, if an entity suddenly joins the Hubble flow and acquires a velocity, a matching acceleration is needed.

If the transition point is labelled with the subscript $_e$, then varying T_e and m_1 will result in a range of possible values for r_e. The observed RAR distribution will be a summation of these (which perhaps explain the huge scatter in Fig. 1). An optimised response for one specific case ($T_e = 1.8$ billion years and $m_1 = 2$ billion solar masses) is shown in Fig. 3 and illustrates the fit that can be achieved with the most basic exponential growth function.

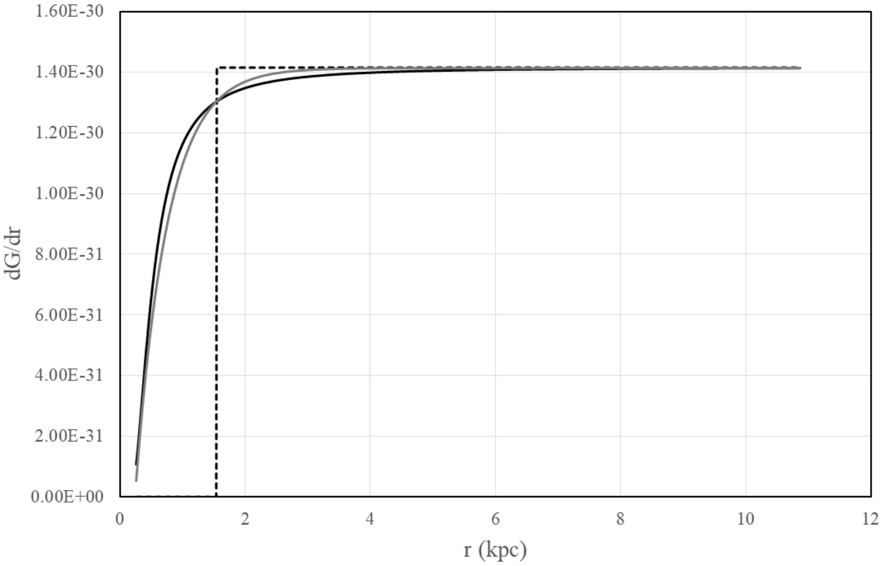


Fig. 2. For a central mass of 2 billion solar masses, a test mass joins the flow after 1.8 billion years (cosmological time), at radial distance 0.23 kpc. The grey line is the development of G with radius according to equation 13. This is compared to the best-fit RAR function (solid black line) and a step transition once the acceleration reaches a_o (dotted line).

Following the modelling of a wide range of mass/time/radius scenarios, the general equation for the response for any scenario is found to be consistent with

$$\frac{dG(r)}{dr} = \sqrt{\frac{G_e a_o}{m_1}} \left(1 - exp\left(\frac{r_e - r}{\frac{1}{3}\sqrt{\frac{G_e m_1}{a_o}}}\right)\right), \tag{1.13}$$

where G_e is the value of G prior to transition. As r tends to ∞,

$$\frac{dG(r)}{dr} \to \sqrt{\frac{G_e a_o}{m_1}} = \frac{G_e}{r_o}, \tag{1.14}$$

where r_o is the radius at the point the acceleration touches a_o. Assuming G is G_e at that point (which is now seen to be only an approximation) the a_o

'threshold condition' now emerges naturally and finds its way into the Tully-Fisher relation:

$$\frac{v_\theta^2}{r} = \frac{Gm_1}{r^2} = \frac{G_e m_1}{r_o r}; \tag{1.15}$$

and rearranging,

$$v_\theta^4 = \frac{G_e^2 m_1^2}{r_o^2} = G_e m_1 a_o. \tag{1.16}$$

Equation 13 is a solution of the differential equation

$$m_1 \frac{d^2G(r)}{dr^2} = 3a_o - 3\sqrt{\frac{m_1 a_o}{G_e}} \frac{dG(r)}{dr}. \tag{1.17}$$

By this interpretation a_o refers not to the transition acceleration as suggested by MOND but instead relates to the rate at which G can increase ('accelerate'). This of course eliminates the main problem with the standard MOND interpretation where the effect was dependent on acceleration, and makes the more logical energy transition interpretation previously presented now largely consistent with RAR data.

Note that the best-fit RAR function, equation 12, can be directly interpreted as a change in G with radial distance:

$$\frac{dG(r)}{dr} = \sqrt{\frac{G_e a_o}{2m_1}} \; \frac{\alpha}{\left(\sinh \frac{\sqrt{\alpha}}{2}\right)^2}, \tag{1.18}$$

where

$$\alpha = \frac{G_e m_1}{a_o r^2}. \tag{1.19}$$

However, in this case there is no transition point and no logical reason for postulating an increase in G with radius.

The equation for the change in G with radial distance, equation 13, is trivially transformed to one of time variation because the expansion velocity is assumed to be constant. The time equation has the same form as the equation governing the motion of a body falling from height h under the influence of gravity and subject to friction (friction constant, b):

$$\frac{d^2h}{dt^2} = -g - b\frac{dh}{dt}. \tag{1.20}$$

By analogy, it can be postulated that $dG(r)/dt$ becomes constant when 'friction' balances the driving 'acceleration'[6].

---

[6] The fact that the value of G might be determined by a second order homogeneous differential equation might help explain the curious coincidence of how the overlaid spiral galaxy rotation curves resemble the response of a second order system to a step function[5].

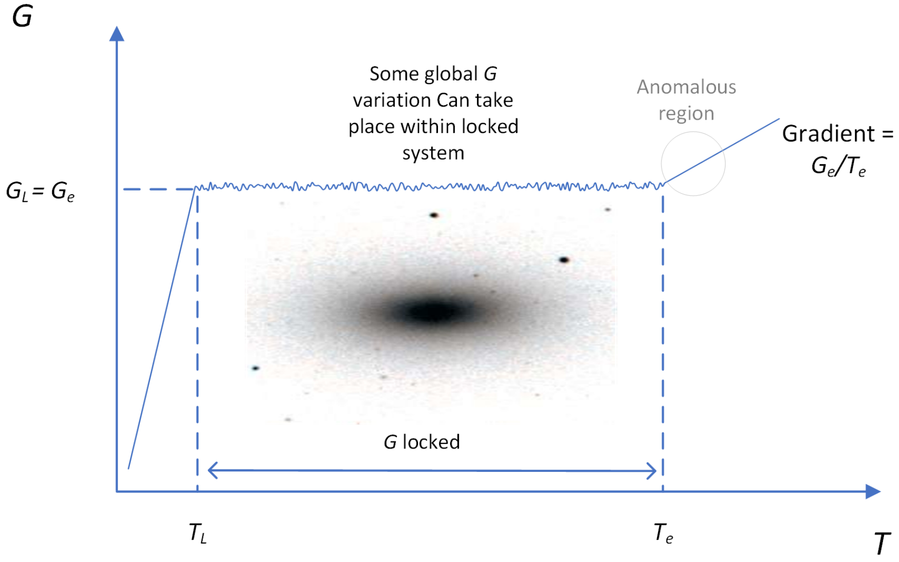


Fig. 3. A mass locked within the galaxy will have more-or-less constant G, and on joining the low the value will increase in proportion as shown. The change in G around the point where the mass joins the flow as determined by equation 13 is not shown. Note that it is more accurate in this interpretation to consider G to be a relative rather than an absolute parameter.

1.4 Continuous Outflow

The picture then is of a stop-start development of G for a particular mass arising from an individual history made up of periods of expansion and time locked within a galaxy. A typical time line might be as shown in Fig. 3.

If G is varies as claimed, then all the atoms that make up the Solar System are likely to have different G values because of the variable time spent in the expansion prior to becoming bound in the Milky Way galaxy. But it is unknown if bodies will retain their accumulated G value if they later become bound and locked (separated from the expansion). However, a system cut off from the expansion may present an average G value throughout. That being the case, all free fall experiment will be consistent with the strong equivalence principle, i.e. all masses will fall at the same rate, regardless of size and material type [3]. But it is possible for the common value to change universally as material is either captured, ejected or joins the flow[7].

The implication from the entire discussion above is that the disk material in spiral galaxies is the result of expansion outflow[8]. The velocity magnitude is very low, of the order of 1% of the rotational velocity, and can easily escape detection when primarily measuring rotational velocity - the outflow contributes to the redshift only through the less significant transverse Doppler effect.

It is evident that in the vast majority of cases the disk is continuous out to large distances. It is therefore necessary that any outflow explanation should involve a process that is self-sustaining once outflow begins. This is the case if the threshold condition is based on the total gravitational energy because the net gravitational force inside the inner edge of a thin ring is non-zero, and will tend to draw out additional material both directly from the unbalanced gravitational forces and indirectly by pulling the acceleration below threshold to allow more material to join the flow. As the disk grows in size, the increasing value of G with radius ensures the distant material makes a greater than expected contribution, which amplifies the pull.

If the transition condition is based only on material not involved in the expansion, the process is more difficult to explain. For a particular disk galaxy, the expansion process can be played backwards in time to the predicted threshold point when the acceleration matched the time-variable threshold condition. Assuming an initially spherically symmetric bulge, the point of origin of material currently at radial position r with accumulated mass $M(r)$ inside that radius and expansion velocity r/T is

$$\frac{G_e(M(r)+\lambda\tau)}{(r-v_h\tau)^2}=\frac{c}{T-\tau}, \tag{1.21}$$

where τ is the time difference between the current cosmological time and the time the material joined the flow. λ is the rate of change of mass over that time

---

[7] Another possibility is that only certain preferred G values such 6.67 x 10^{-11} m^{v} kg^{-1} s^{-2} are allowed.

[8] This does not refer to the visible arms which are thought to have a density wave origin.

(assumed in the simplest model to be a constant); essentially, all the complex effects that take place in the long history of a galaxy have been subsumed into one variable, a steady loss or gain of mass with time.

Rearranging:

$$\tau = \frac{cr^2T - g_e M(r) T^2}{G_e \lambda T^2 + cr^2}. \tag{1.22}$$

Logical space ordering requires that older material originate from greater radius r_e than newer material. However, this is only possible if a galaxy steadily loses mass. For typical galaxies, a loss of 1 - 100 solar masses a year is required, normal for quasar action but far from universal.

1.5 Peculiar Velocity

Up to now, it has been assumed that the material in the expansion has no significant peculiar velocity, but that is rarely the case, and the inclusion of a peculiar velocity introduces significant complication. The combination of peculiar and expansion velocities leads to interference effects once the requirement for energy conservation is applied[9].

With e as the total energy per unit mass of a test mass subject to both the expansion and gravity from central mass M,

$$\frac{de}{dt} = \frac{d}{dt}\left(\frac{1}{2}v^2 - \frac{GM}{r}\right) = v\frac{dv}{dt} + \frac{GM}{r}v_p = 0, \tag{1.23}$$

where $v = v_h + v_p$.

To conserve energy the peculiar velocity develops as follows:

$$\frac{dv_p}{dt} = -\frac{GM}{r^2}\frac{v_p}{v} - \frac{v_p}{T}. \tag{1.24}$$

With no expansion, the equation simplifies to the familiar gravitational acceleration equation. Without a significant gravitational force, the first term disappears and what is left is the familiar 'Hubble drag' effect [6] which converts peculiar motion to expansion velocity whilst keeping the total velocity constant[10].

When both gravity and expansion are present, the equation introduces a barrier effect. If $v_p >$ -v_h, the peculiar velocity is pulled towards zero and stays there. If $v_p <$ -v_h, the body is pulled towards the central mass. This is shown in Fig. 4 for a range of initial peculiar velocities.

---

[9] Care is required here. The relation $dG/dr = G/r$ is the restricted result of the more general $dG/dT = G/T$ in the case no peculiar velocity. Incorrectly using the former relation means a moving body is no longer accelerated by gravity. In addition, it is better to write the general expression as $dG/dT = G/r\ v_h$ to correctly reflect a dependence on the entity being in the expansion.

[10] The Hubble drag here is with reference to the proper frame rather than in comoving coordinates.

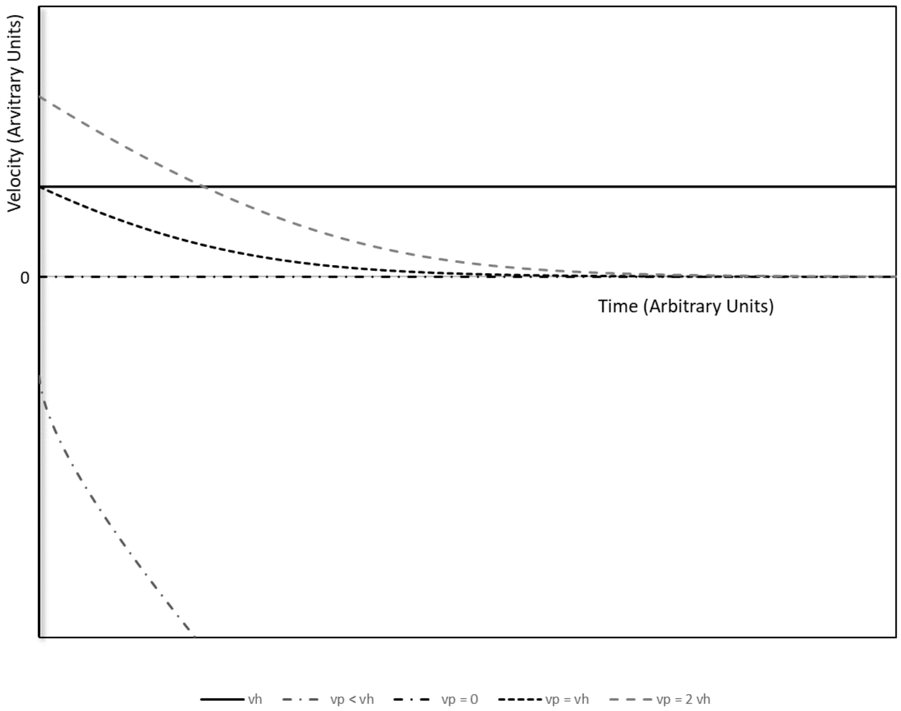


Fig. 4. The solid grey line is the expansion velocity. The effect on various initial v_p values is shown: $2v_h$; v_h; 0; -$v_{h\text{-}}$.

A velocity horizon that is an echo of the black hole event horizon has appeared. In the absence of other forces, all bound masses beyond the threshold velocity will forever be on the outside and locked into the expansion, with the peculiar velocity eventually disappearing. Bound masses with more negative velocity than threshold are trapped within.

1.6 Discussion

MOND is a puzzle. It is very effective in certain domains but the basic premise is hard to reconcile with established physics. One reaction might be to dismiss the theory out of hand as a huge coincidence, but there is the chance that it is revealing something important about gravitation. The hypothesis was first presented in 1983 by Milgrom who has recently speculated on cosmological links suggested by the value of the MOND constant [7]; and though there have been some tantalising similarities he discovered nothing concrete and there is no real evidence of anything beyond coincidence.

The MOND assumption that gravitational acceleration depends on the acceleration is circular referencing that undermines the entire concept and makes one suspicious of the entire idea. The alternative approach here is to try to normalise the concept by making the effect, whatever it may be, dependent on an energy condition in such a way that the apparent acceleration dependence is an emergent effect and not in itself a cause.

The proposal is that the expansion of the universe does not operate down to atomic level but there is a threshold point where, upon reaching it, an entity joins the expansion. The gravitational constant then increases in proportion to elapsed time, removing many of the objections to the MOND hypothesis whilst demonstrating the correct dynamic behaviour. The MOND constant a_o has been significantly demoted in its role, which means that the places where MOND fails such as larger scale systems and unusual galaxy types may not be that damaging in the modified hypothesis [8]. a_o relates to the way the gravitational constant is able to change and is not a significant threshold condition. It is an emergent effect rather than the driver.

This work has not been an attempt to create a new theory but instead a plausibility argument has been presented to justify an alternative approach. Significant problems have emerged that are yet to be resolved, but these are different to and technically more manageable than the problems of MOND.

The approach has been Newtonian, but any plausible modification will need to be general relativistic. Once one accepts that changes in gravity are possible on the large scale, whilst being (relatively) constant locally, there are a number of relativistic models to choose from, but there is insufficient data to make a selection[11]. More data is needed to investigate the concept further and refine

---

[11] There is generally no objection to fundamental constants that vary under certain circumstances, and there is a vast body of relevant research, for example the work of Barrow [17].

the basic idea. Nevertheless, it is tempting even at this stage to look to a ready-made theory such as Brans-Dicke gravity and piggy-back onto it by setting the characteristic coupling parameter ω to infinity at a sub-galaxy level and choosing an appropriate value for the scalar component that returns the expansion behaviour described here, but it does not work - in the Brans-Dicke model a variable G that is decreasing rather than increasing with time is chosen in order to incorporate Mach's principle [16].

References

1. M. Milgrom (1983) "A modification of the Newtonian dynamics - Implications for galaxies". Astrophysical Journal 270 pp 371–389.
2. M. Milgrom (2010) "New Physics at Low Accelerations (MOND): an Alternative to Dark Matter". INVISIBLE UNIVERSE: Proceedings of the Conference. AIP Conference Proceedings, Volume 1241, pp 139-153.
3. C. M. Will (2005) "Was Einstein Right? Testing Relativity at the Centenary". Annalen der Physik. 15: pp 19–33.
4. S. S. McGaugh, F. Lelli and J. M. Schombert (2016) "Radial Acceleration Relation in Rotationally Supported Galaxies". Phys. Rev. Lett. 117, 201101.
5. Y. Sofue, Y. Tutui, M. Honma, A. Tomita, T. Takamiya , J. Koda and Y. Takeda, (1999) "Central Rotation Curves of Spiral Galaxies". The Astrophysical Journal, 523:136-146, Fig. 2. *Compare with a standard control text book such as R. C. Dorf and R. H. Bishop, Modern Control Systems, Pearson Education, 11th edition, 2008.*
6. E. Bertschinger (1995) "Cosmological Dynamics". arXiv:astro-ph/9503125v1.
7. M. Milgrom (2020) "The a_0 – cosmology connection in MOND". arXiv:2001.09729v1.
8. C. Di Paolo, P. Salucci and J. P. Fontaine (2019) "Radial Acceleration (RAR): key cases of dwarf discs and galaxies with low surface brightness". The Astrophysical Journal, Volume 873, Issue 2, article id. 106.
9. C. Brans and R. H. Dicke (1961) "Mach's Principle and a Relativistic Theory of Gravitation". Phys. Rev. 124, 92.
10. J. D. Barrow (2005) "Varying constants". Royal Society of London Transactions Series A, vol. 363, Issue 1834, pp 2139-2153.

2 Cosmology: An Operational Approach

Summary

It is useful to associate a model with the mathematical description of a physical process in order to show the underlying mechanism. This is true of cosmology: cosmological data is consistent with the Friedmann equations, and these are closely linked to a model of a finite universe curved under its own gravity and expanding at a rate constrained by the enclosed mass. The model is effective and accessible, and is widely accepted as a realistic representation of the universe. An alternative is to analyse cosmological data through an operational approach. An operational approach presumes no underlying model and merely involves the rigorous application of the laws of physics relevant to that domain, with modifications made to these if needed in order to obtain good agreement with observational data. By analysing distant supernovae data in this way, a reasonable match with observation is obtained; and the match can potentially be improved by investigating the range of additional influencing factors that are identified.

2.1 Introduction

Finding a pattern in data and expressing it in mathematical form is a significant step towards understanding the underlying process(es), but understanding is often only complete when a model is developed that gives the mathematics a physical form. But a distinction should be made between a model that is primarily an analogy introduced to aid understanding and a model that purports to be an accurate system representation. In the later case, if the model is a faulty representation of the underlying process, the model itself can form a barrier to deeper understanding and hold back or misdirect further theoretical developments. One way round this is to adopt an operational approach[1] where the objective is to analyse the data without any concern for models or even 'meaning'. The system is analysed using the established laws of physics that are applicable in that domain (with the sole focus of matching observational data). This can act as a useful check on an accepted model. The features of the model that are relevant and important should emerge naturally from the analysis.

The operational approach applied to cosmology is essentially an independent and objective analysis that is without preconceptions. Everything is taken to be as it seems: distant cosmological objects are moving away from the observer; this is interpreted as a velocity because there is nothing in the observations that would indicate otherwise (the redshift effect is just as expected). The idea of this motion being associated with the expansion of space is a concept originating

---

[1] This is loosely based on the work of Bridgman who emphasised the importance of measurement, and who considered the act of measurement to be the application of a set of repeatable and clearly defined operations [1]. Similarly a collection of physical laws is effective if, by applying them in a sequence of operations, the result of the measurements is reproduced.

from a model, and is therefore dismissed. There is found to be a linear relation between apparent velocity and distance:

$$v_h(T) = \frac{r(T)}{T} = H_o r(T), \tag{2.1}$$

where r is the proper distance, T is the cosmological time and H_o is the current Hubble constant value. This will be referred to as the 'Hubble velocity' and is exactly what would be expected if entities are largely non-interacting (because of the distances involved) and acquire their velocity at a time around $T = 0$ (for whatever reason and whatever the value, again not a concern). As a consequence of energy and momentum conservation, the velocity will have remained constant over the lifetime of the universe, hence an object starting off with velocity v will be a distance $r = vT$ away now, which is just Equation 1.

It is also evident that gravitation on a large scale can be treated in the Newtonian approximation because the gravitational field strength is almost negligible with the immense separation distances involved.

The data also indicates space-time is flat, hence observations should be subject to special relativity (SR). The brightness of distant objects is required to be consistent with the standard relativistic derivation. This can be tested by comparing the derived distance-luminosity relation, d_L-z, with observational data.

However, there is a serious problem - energy appears not be conserved. This must be addressed before any analysis can begin.

2.2 The Energy Problem

The receding galaxies must do work against the local gravitational field. There is no obvious place the energy for this could come from because the recessional velocity remains unaltered. The energy conservation violation is small, but even a tiny violation is problematic. The only way to recover energy conservation is to make modifications to the laws of physics, i.e. new physics is needed[2].

The simplest possible modification is to loosen the definition of the Cavendish gravitational 'constant' and allow it to vary. A severe constraint on this strategy is that the value is known from measurements here on earth to be absolutely constant. It is proposed that there exists a transition point on the galactic scale where an entity joins the expansion (or in operational terminology, takes on a Hubble velocity). From this point on, G no longer remains constant but changes over time, increasing steadily from its previous constant value in order to conserve energy. Gravitational energy is conserved if $dG/dt = (G/T)$ holds. The

---

[2] In standard cosmology, it is claimed that the gravitational pull slows down the expansion. It is unclear how this happens as it implies a binding between space and matter which is incompatible with normal dynamics - a normal force moves a mass with respect to space; it does not affect space. In addition, where is the evidence for this - should expansion not be found to vary with matter density in order that energy is conserved locally as well as globally? The idea of mass slowing the expansion is a foundational idea in conventional cosmology, but it is poorly justified and has not been sufficiently evaluated.

consequence is that the effect of gravity on a body with a Hubble velocity can be ignored[3].

This is taken as the working hypothesis. The energy problem is resolved and there is no conflict with accepted physics locally[4].

2.3 The Luminosity Distance in Special Relativity

An important test of any cosmology is predicting the apparent brightness of distant standard candles such as supernovae. A standard SR analysis will generate a d_L-z relation that may be compared with observational data. The derivation here of the luminosity distance in flat space-time follows closely the work of Chodorowski [2][5], but is presented in more detail to demonstrate the difficulty in making significant modifications that could address the any gap that might be found to exist between prediction and observational data.

Applying special relativity, the Hubble velocity is the proper distance at emission divided by the time at emission, or, equivalently, the proper distance when a photon is received (i.e. the entity is observed) divided by the time of absorption, all measured or inferred from the rest frame of the observer. This is illustrated in Fig. 1. The proper separation distance is the constant Hubble velocity times the elapsed time in the observer frame. This is consistent with the simple linear Hubble velocity equation presented as Equation 1. The diagram shows the relation between the same events in the emitter and absorber rest frames.

The Lorentz transform equations are:

$$x' = \gamma(x - vt); \tag{2.2}$$

---

[3] It means that G is relative rather than absolute to ensure consistency for all observers and all mass distributions.

[4] Bear in mind that this is not a binary situation - the standard interpretation and what is proposed here are only two on many possible interpretations. For example, it is also possible to claim a mass that is subject to the expansion has an energy that is directly related to the expansion process. This 'Hubble energy' could take the form

$$E_h = -m_o v_h c.$$

A photon from a distant source will be received with less than expected energy resulting redshift,

$$z = -\frac{\Delta E}{E} = \frac{\left(\frac{E_\gamma}{c^2}\right) v_h c}{E_\gamma} = \frac{v_h}{c}.$$

This is consistent with standard cosmology at low redshifts.

[5] The derivations appear very different because Chodorowski is operating in the emission frame and is more verbal in the description, whereas the derivation here is with respect to the observation frame. However, they are completely equivalent and the end result is the same.

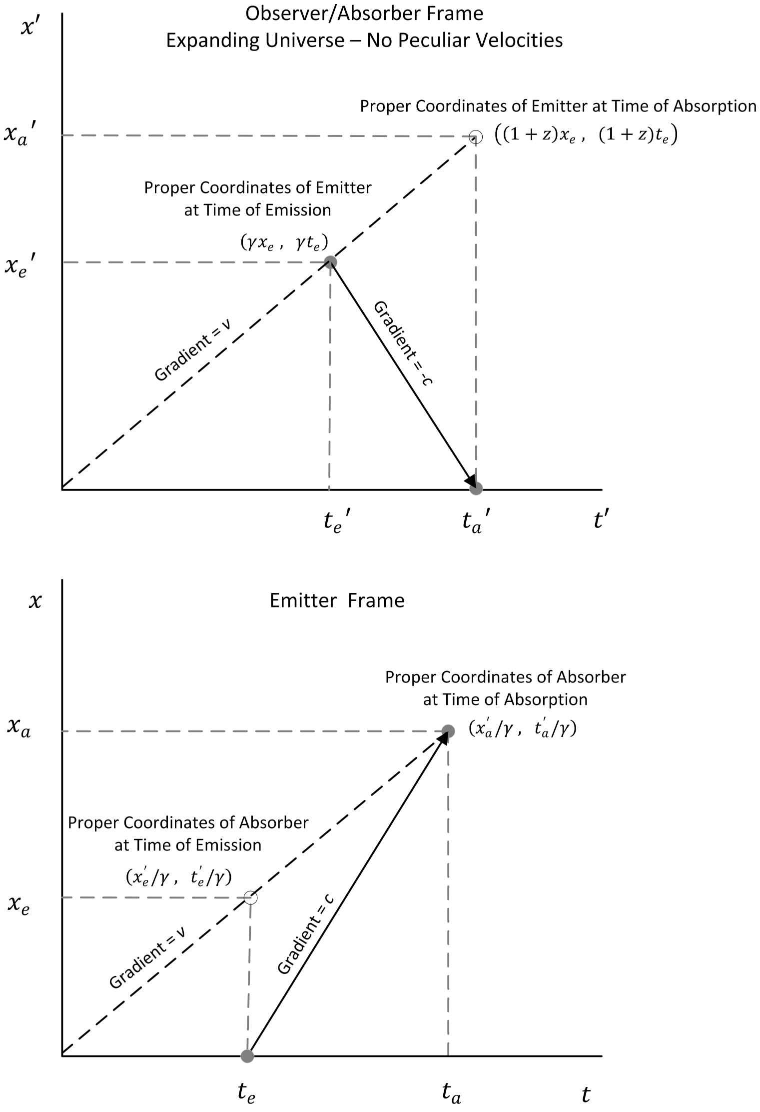


Fig. 1. The diagram shows the emission and absorption of a photon between two entities moving apart at a constant velocity over cosmological time. The coordinates in the emitter frame are shown in the lower graph, and the primed absorber frame is shown above. The emitter and absorber are coincident at $T = 0$.

$$t' = \gamma\left(t - \frac{vx}{c^2}\right). \tag{2.3}$$

The luminosity distance in SR is related to the measured flux (f) and the intrinsic luminosity (L) by the following equation:

$$f = \frac{L}{4\pi d_L^2}. \tag{2.4}$$

This can also be written as

$$f = \frac{L}{4\pi(1+z)^4 d_A^2} \tag{2.5}$$

where the diminution of flux as a result of distance, $(1{+}z)^2 {d_A}^2$, is combined with the energy reduction because of the Doppler Effect and the loss of energy as photons 'fill' the growing volume between emitter and observer (the additional $(1{+}z)^2$ contribution).

Clearly, the flux will reduce with distance, but it is not entirely obvious what the distance d_A is. Do we work with the separation distance between emitter and absorber at the time of emission, or at the time of absorption; should we be in the frame of the emitter or the absorber; do we use proper or retarded distances? A full (and correct) SR analysis must identify invariants and transform these appropriately. This is explained in [3] sections 22.6, 29.4 and exercise 29.5. It is shown that the measured flux from an isotropic luminous body receding at velocity v is

$$f = \frac{L}{4\pi(1+z)^4 r^2}, \tag{2.6}$$

where r is the proper position of the emitter in the rest frame of the observer (the photon absorber) at the time of emission.

Referring to Fig. 1, the Hubble velocity is

$$v_h = \frac{x'_e}{t'_e} = \frac{x'_a}{t'_a}. \tag{2.7}$$

There is the implicit assumption that the observer's local time tracks cosmological time.

The proper distance in the observer frame at the time of emission is ${x_e}'$ (r in Equation 6) and once the event is transformed to the correct coordinates in the observer frame the photon will move at velocity c, covering the distance to the observer in time ${x_e}'/c$. The proper separation at absorption time ${t_a}'$ is just $v{t_a}'$, which is ${x_e}' + v/c\ {x_e}'$. We can therefore write (dropping the subscript $_h$)

$$(1+\beta)x'_e = vt'_a; \tag{2.8}$$

and rearranging gives

$$x'_e = \frac{c\beta}{(1+\beta)}t'_a. \tag{2.9}$$

The luminosity distance is

$$d_L = (1+z)^2 x'_e. \tag{2.10}$$

Noting that

$$\beta = \frac{(1+z)^2 - 1}{(1+z)^2 + 1} \tag{2.11}$$

and

$$1 + \beta = \frac{2(1+z)^2}{(1+z)^2 + 1}, \tag{2.12}$$

the final result is

$$\begin{aligned} d_L^{SR} &= c(1+z)^2 \, \frac{(1+z)^2+1}{2(1+z)^2} \, \frac{(1+z)^2-1}{(1+z)^2+1} t'_a \\ &= ct'_a \left(z + \frac{z^2}{2} \right) \\ &= \frac{c}{H_o} \left(z + \frac{z^2}{2} \right). \end{aligned} \tag{2.13}$$

Equation 13 is the correct luminosity distance expression for flat SR space-time. The equation describes a model-less flat space-time cosmology where the velocity is imparted at $T = 0$ and thereafter remains constant.

This should of course match observational data otherwise the laws of physics that were applied are either incorrect or incomplete (or have been incorrectly applied). A number of d_L-z functions have been derived that match observational data, and it is reasonable to compare Equation 13 with these as an alternative to matching the actual data (recognising the d_L-z function in standard cosmology has evolved over time and is still developing).

Terrell's formula is [4]:

$$d_L^{TR} = \frac{c}{H_o} \left(z + \frac{z^2(1-q_o)}{1 + q_o z + (1 + 2q_o z)^{1/2}} \right), \tag{2.14}$$

where q_o is the deceleration parameter.

There is equality when the deceleration parameter is 0. Of course, there is no deceleration in the SR model where G increases with time in the manner proposed earlier.

There are a number of other d_L-z functions in the literature. For example in [5],

$$d_L^{DL} = \frac{c}{H_o}(1+z)\frac{(1+z)^2-1}{(1+z)^2+1}, \tag{2.15}$$

supposedly also a special relativistic analysis, though not consistent with the argument presented here - the authors seem to be using the proper separation time in the emission frame for r in Equation 6 rather than the proper separation time in the absorber frame (though they could also be using a completely different definition of the Hubble velocity).

Equation 13 can also be compared with the redshift-distance equations in the consensus ΛCDM model (which is known to be a very good fit for supernova data).

Consider first the FLRW equation:

$$d_L^{FLRW} = (1+z)\,\frac{c}{H_o}\int_0^z \frac{dx}{\left(\Omega_m(1+z)^3 + \Omega_L(1+z)^{3(1+\omega)}\right)^{1/2}} \tag{2.16}$$

which is solved for Ω_m=0.266, Ω_L=0.734 and ω= -0.667.

A comparison is also be made with the approximation of the ΛCDM consensus model presented in [6] (using the same density values):

$$\begin{aligned} d_L^{ULP} = \frac{3.876\ c(1+z)}{H_o}&[0.8956 - ((1+z)^4 \\ &- 0.1549\ .\ 1.402^1\ (1+z)^3 +\ 0.4304\ .\ 1.402^2\ (1+z)^2 \\ &+ 0.19097\ .\ 1.402^3\ (1+z)^1 +\ 0.0669415\ .\ 1.402^4)^{-1/8}]. \end{aligned} \tag{2.17}$$

The four Equations 13, 15, 16 and 17, are plotted and compared in Fig. 3.

The d_L-z relation derived simply by directly interpreting the data without reference to a model is a good match for the observational data as represented by the best-fit standard cosmological relation.

But the fit is not perfect. This is problematic because it is very difficult to make any changes to the operational methodology. If we insist on energy and momentum being conserved, there is no valid definition of Hubble velocity except the one presented as Equation 1. All other formulations lead to inconsistencies.

However, some assumptions were made. For example, the effect of peculiar velocity was not considered. In addition, it was assumed that clock time and cosmological time are the same (or almost the same). Another possible source of error is the assumption that entities have been moving apart from $T = 0$. We will consider each of these in turn along with other factors that might be relevant.

2.4 Peculiar Velocity

Looking again at the discrepancy between the predictions of SR cosmology and supernovae brightness data, one possibility is that there are systematic peculiar velocities as a result of the gravitational action of mass accumulating over great distances[6]. In Equation 7, the total velocity (the relativistic addition of peculiar and Hubble velocities) enters into the equation through the $(1+z)^4$ term, but

---

[6] Huge streaming velocities have been observed [7].

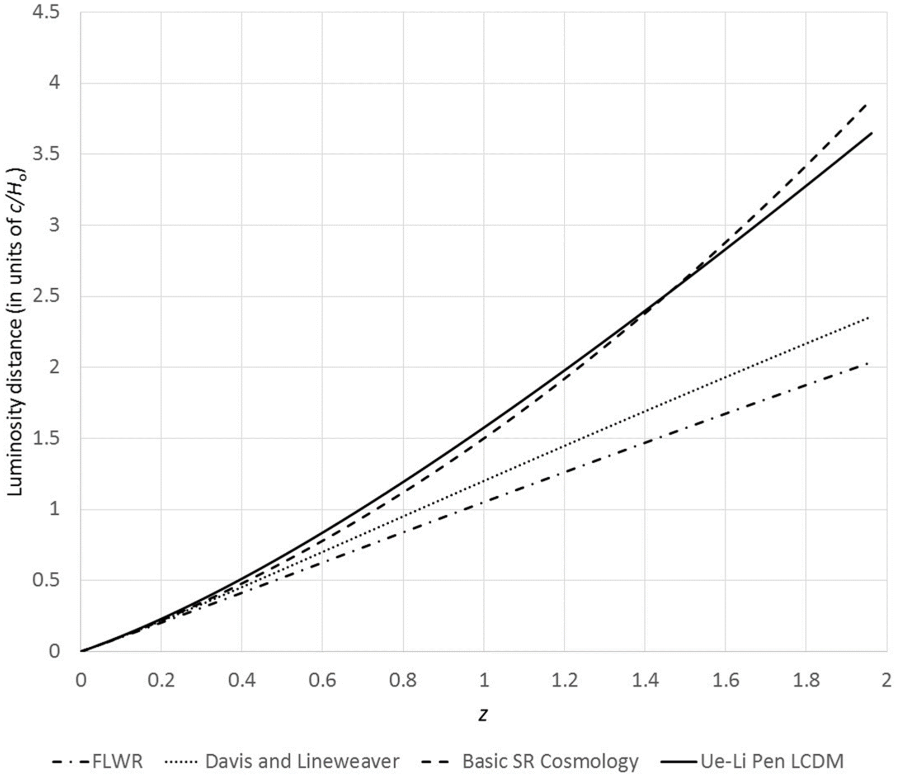


Fig. 2. The Luminosity Distance-Redshift plots for the four expressions in the text. The older FLWR model, though largely discredited, has been included to show how far GR cosmology as moved while following the data. By contrast, the SR cosmology function has not changed at all.

it is the Hubble velocity that establishes, to an extent, the r^2 value. The modified relation depends very much on the time over which the peculiar velocity is present and it is not possible to derive a simple relation. However an idea of the influence of a peculiar velocity can be obtained by assuming that the emitter has a peculiar velocity for a short time starting just at the point of emission[7]. The peculiar velocity 'pulse' needed to bring SR cosmology into line with the data is found by modifying Equation 9:

$$x'_e = \frac{v_h}{(1+\beta)} t'_a = \frac{c\beta \ominus v_p}{(1+\beta)} t'_a, \tag{2.18}$$

where v_p is the peculiar velocity and the $\ominus$ indicates a relativistic subtraction of velocities.

The luminosity distance becomes (assuming $v \ll c$):

$$d_L^{+p} = \frac{c}{H_o} \left\{ \left(z + \frac{z^2}{2} \right) - \left(1 + z + \frac{z^2}{2} \right) \frac{v_p}{c} \right\} \equiv d_L^{ULP}. \tag{2.19}$$

It follows that

$$d_L^{SR} - d_L^{ULP} = \Delta d_L = \frac{c}{H_o} \left(1 + z + \frac{z^2}{2} \right) \frac{v_p}{c}. \tag{2.20}$$

Rearranging,

$$\frac{v_p}{c} = \frac{H_o}{c} \frac{\Delta d_L}{\left(1 + z + \frac{z^2}{2}\right)}. \tag{2.21}$$

The peculiar velocity is calculated and plotted in Fig. 3. For objects nearby, the peculiar velocity is negative (i.e. in the opposite direction to the Hubble flow). The redshift is consequently smaller than that from the Hubble velocity alone, hence an object is dimmer than would be expected if the redshift distance were assumed to correspond to the actual distance. Far away, the peculiar velocity is positive, giving rise to an apparent brightening.

One hypothesis is that the peculiar velocity profile is the result of great spheres of matter beginning to contract and coalesce. [8] The inwards peculiar

---

[7] This is sufficient because peculiar velocity acting for a long time previously is gradually converted to expansion velocity, and the velocity after emission is not relevant to the apparent flux.

[8] The shape of the function out to the turning point is consistent with the collapse of a large sphere of hydrogen gas. If modelled as shells of gas of initially equal density, assuming adiabatic compression and expansion with no flow of heat or matter between shells, the shells will develop this velocity profile as growing pressure acts against the contraction. In the model, constant $G)$ and no expansion of the universe can be assumed because increasing G postulated in Section 2cancels the effect of the expansion. However the collapse of a gas cloud has a timescale that is many orders of magnitude shorter than the age of the universe (the timescale evident in Fig. 3). However, it is an unreasonable comparison as large discreet objects do not behave like an ideal gas. There are issues such as radiation pressure and ejecta from su-

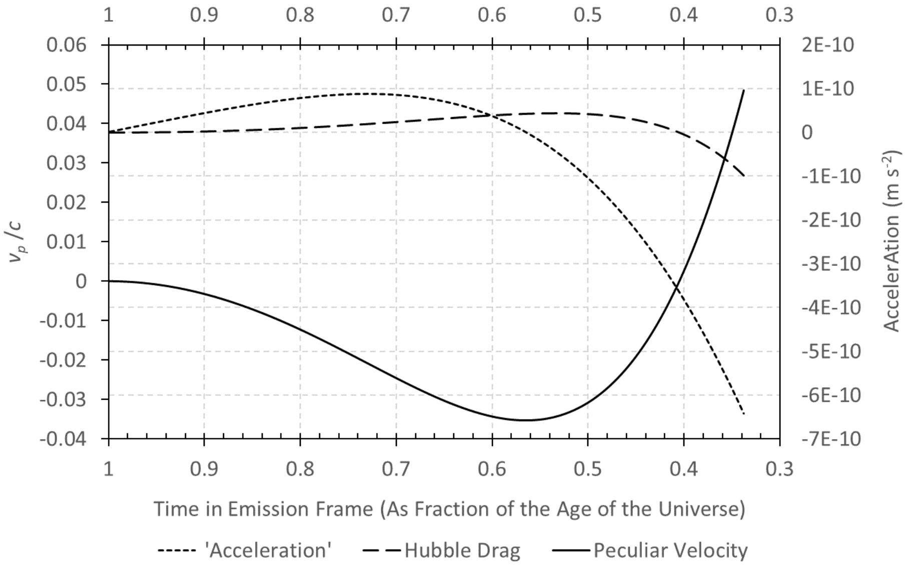


Fig. 3. The peculiar velocity (and the associated acceleration) required to get SR cosmology to align with the data. Note the use of special relativistic time units. The peculiar velocities are large but not unreasonably so - for example, the relative velocity of sub-clusters in the colliding bullet cluster at $z = 0.3$, $T = 0.75\ T_o$ is consistent with 0.018 c.

velocity is of course the effect of gravity. However, the velocity drops to zero only at $z = 1.47$, which is much larger than the radius of the observed superclusters of galaxies. These have a radius of the order $z = 0.05$[9]. But bear in mind the observational data used in the d_L-z analysis involves the reduction of 3-D data to a single dimension and also the loss of directional effects[10] and may not be entirely suitable for this analysis.

The change in peculiar velocity per time step *in the emission frame* is also plotted in Fig. 3. This is tentatively labelled 'acceleration', in inverted commas: the calculation does not give the instantaneous acceleration but instead is an indication of how the peculiar velocity field as a whole is developing over time. The data shows a peculiar velocity field *at any fixed point out to 0.6 T* that becoming more negative with time, as might be expected because of the central gravitational attraction. Note though that the Hubble drag, -v_p/T, must be considered and acts as a constraint - it arises because peculiar velocity is steadily converted to Hubble velocity.

If the disagreement between SR predictions and observational data really is a result of the interfering effect of peculiar velocities, additional work is required to explain the detail of this very significant peculiar velocity field. There are also a number of other approaches available to resolve the discrepancy.

2.5 Additional Factors

Supernovae data is considered reliable and there is no evidence of time evolution. That being the case, bringing the d_L-z function and data completely into line without relying on a very large peculiar velocity field can only be achieved by modifying the operational procedure in some way. This is very difficult because there is so little flexibility in the methodology. Of course, other issues such as subtleties relating to the increase in G with time have not been fully assessed, but these can only be factored in when considerably more information is available.

But a closer examination of the process followed highlights a number of assumptions that could be questioned. One key assumption was that all entities

---

pernovae and galactic jets, transverse rotation and the effect of large-scale magnetic fields, that need be taken into account.

[9] Though there are claims of significant peculiar velocity fields out to the Gpc scale [8].

[10] This is fine for expansion velocity (which is entirely radial), but not for peculiar velocity. If the peculiar velocity for each galaxy is assumed to be the deviation from the SR prediction, plotting the position and and peculiar velocity for each galaxy in 3-D space defines the galaxy distribution in an SR universe. Note that the peculiar velocity of the earth with respect to the expansion frame - about 0.002 c - places us somewhere along the axis rather than the centre, and suggests that observational data should not be isotrophic. After all it would be unlikely the earth lies at the centre of the contracting sphere; this could possibly explain the weak anisotropy found in the supernovae data [16] (which is assumed to arise from an imbalance in the peculiar velocities with respect to the off-centre position of the earth-bound observer) and, in particular, the reason why the effect is found to be more prominent at low z [17].

gain their Hubble velocity at $T = 0$, but that is not necessarily the case. It perfectly conceivable that a random or unknown velocity could be assigned at any time: v_u at T_u when the proper separation distance is r_u; this is consistent with special relativity. That being the case, the Hubble velocity estimated from the proper position, v_h, will deviate from the actual Hubble velocity v_u,

$$v_h = \frac{r_u + v_u(T - T_u)}{T} \simeq v_u \left(1 - \frac{T_u}{T}\right), \tag{2.22}$$

potentially leading to huge errors, but is largely an unknown factor that can only be addressed when there is more knowledge about how an entity takes on a Hubble velocity[11].

Another factor to consider is that local clock time is unlikely to be the same as cosmological time because the observer has been subject to multiple acceleration events over the lifetime of the Universe. However, the associated error is considered minimal because time enters into Equation 1 through the predicted age of the universe rather than local clock time.

There are physical factors that could have an impact as well: these include the effects of residual gravity, relativistic beaming, aberration, increasing angular size with distance, evolution of all physical constants and light absorption.

Though the operational approach has not been completely successful, the predictions are tantalisingly close to the observational data, hence further investigation along these lines is justified.

2.6 Discussion

The agreement between Equation 13 and the formula applied in consensus cosmology is surprisingly good, and the small difference could be explained by a specific peculiar velocity field, though it is not clear if that is a velocity distribution that might naturally arise.

Had the operational approach produced a perfect match, one might argue that there is nothing in the Universe that is inconsistent the laws of physics as

---

[11] But a word of caution, it is doubtful if the Hubble velocity can emerge as anything other than r_u/T_u because otherwise gravity will change according to $dG/dr = (G/r)$. That will mean peculiar velocities are not correctly affected by gravity.

In addition, the SR model has been tested against the standard cosmological model rather than real data, but this could also be a problem. Some evidence is emerging the consensus model may not be the 'gold standard' assumed. The issue is as follows: the value of H_o that needs to be entered into the standard equation, Equation 17, to match observation is 67 ± 1.5 km s^{-1} Mpc^{-1} if the energy density values consistent with the cosmic background radiation are used; as z tends to zero in the universe around us, H_o can be directly measured and the value is found to be 73 ± 1.5 km s^{-1} Mpc^{-1} [18] [19]. In other words, Equation 17 does not converge to the correct luminosity values for small z. A higher-than-expected H_o value for local observations implies that an observed redshift is associated with a shorter distance and higher luminosity than when the smaller value is used (and also that the age of the universe is significantly less than was thought).

we know them, apart from a need to modify G under some conditions, though the same result (i.e. recovering energy conservation) could be achieved in other ways as well, such as introducing dark matter and dark energy. Without strong evidence to support any of the options, the individual's choice is largely a matter of taste.

Every effort has been made to avoid the use of models, but one might now at this stage be tempted to try to construct a model to explain 'special relativistic cosmology', but this is not possible. In SR, every observer has a unique window on the world in retarded time, and any attempt to sew together all the views of all observers into a single overall picture of the Universe will fail because no observer can be conceived for whom this is a proper representation of the Universe. It is not even clear if the SR universe has an edge or if it is isotropic. In reality, 'special relativistic cosmology' is not cosmology at all. It merely ensures every separate view of the universe is consistent.

Is it acceptable to leave it at that and just employ an operational procedure to perform calculations? Probably; it is the standard approach in quantum mechanics, another theory for which no effective model is known to exist. No model is better than a bad model.

References

1. Bridgman., 1927, The Logic of Modern Physics, New York: Macmillan
2. Chodorowski M., 2005, Publications of the Astronomical Society of Australia. 22. 287-291
3. Misner C. W., Thorne K. S. and Wheeler J. A., 1973, Gravitation, W.H. Freeman, New York
4. Terrell J., 1977, Am. J. Phys. 45:869
5. Davis T. M. and Lineweaver C. H., 2004, Publ. Astron. Soc. Austral. 21 97
6. Ue-Li Pen, 1999, Ap.J.Supplement Series, 120, 1
7. Kashlinsky A. et al, 2008, ApJ 686 L49
8. Atrio-Barandela F., Kashlinsky A., Ebeling H., Fixsen D. J. and Kocevski, D., 2015, Astrophys. J. 810 (2): 143
9. Gupta S. and Singh M., 2014, Monthly Notices of the Royal Astronomical Society, 440, 10.1093/mnras/stu416
10. Rong-Gen Cai and Zhong-Liang Tuo, 2012, Direction dependence of the deceleration parameter, Journal of Cosmology and Astroparticle Physics, Volume 2012
11. Riess A. G. et al, 2018, The Astrophysical Journal, 861, 2
12. Riess A. G. et al, 2019, arXiv:1903.07603 [astro-ph.CO]

3 Acceleration and Special Relativity

Summary

A body χ is moving at a constant velocity towards a stationary body, ψ. As they meet, ψ is quickly accelerated to the same velocity as χ and clocks are synchronised. χ and ψ had distinct worldviews before the merger (in accordance with special relativity), and an infinite time afterwards they will share the same worldview; but how long after the acceleration event does full alignment take place? The suggestion is that the worldviews of χ and ψ instantly align, because there is no mechanism in the accelerated frame for a gradual merging that is consistent with conservation principles. The physical consequences of this are investigated.

3.1 Introduction

Special relativity (SR) is a formalism that relates the space-time coordinates of the same event viewed from different inertial frames. Though it is commonly believed that general relativity (GR) is required when frames are accelerated rather than inertial, that is not the case. SR can effectively handle acceleration, though some care is required.

Working with one space dimension only, consider an event $P = (x,t)$ in a particular inertial frame (which will be identified as the stationary frame of observer ψ). Another observer, χ, moving at velocity v, and at the same position as observer ψ at $t = 0$, will allocate coordinates to that event, P'=(x',t'), where:

$$x' = \gamma(x - vt); \tag{3.1}$$

$$t' = \gamma\left(t - \frac{vx}{c^2}\right). \tag{3.2}$$

They will not agree on the coordinates of any distant event. Now assume the stationary observer, ψ, is instantaneously accelerated to velocity v at time $t = 0$ (Fig. 1). Does the set of points P instantly transform to P'? In other words, does the newly-accelerated observer take on the same space-time as observer χ who has always been moving at that velocity?

The question is important because there are measurable differences. In the moving frame following acceleration, a distant light source may be immediately brighter or dimmer compared to before acceleration, depending on whether or not the frame updates instantly, because the luminosity of a source in SR varies with a $(1 + z)^{-4}$ dependence (z is the redshift factor, which changes with relative velocity).

A sensible approach is to check for any energy conservation violation. Assume that prior to acceleration ψ receives a stream of photons of energy $E^p{}_o$ from a distant emitter stationary relative to ψ. Moving observer χ will receive these with reduced energy

$$E^p = \gamma\left(t - \frac{v}{c}\right) E^p_o \tag{3.3}$$

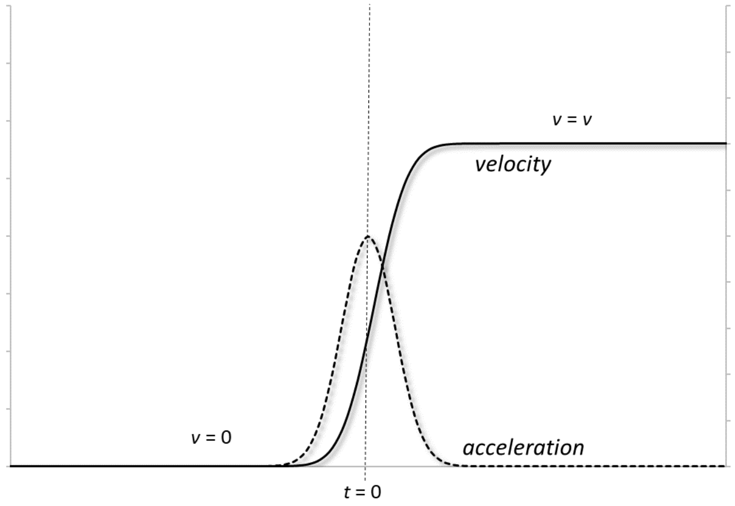


Fig. 1. Observer ψ is accelerated from 0 to velocity v over a short interval $\Delta t \to 0$. The velocity change is generally very small.

because of the relativistic Doppler redshift effect (note that the formula is not quite correct - see Appendix 1).

If, following acceleration, the world view of ψ as it relates to distant objects remains the same (even for a short time) then the received photons will have the same energy as before acceleration, $E^p{}_o$. This energy can be immediately passed from ψ to χ with no loss or gain because of the shared rest frame. Imagine that, following the transaction, ψ is decelerated back to the original velocity. Comparing the system energy just before the initial acceleration to just after the deceleration, there is a discrepancy: χ has more energy than it should have. Energy is not conserved.

It is therefore necessary to conclude that for consistency, the world view of an accelerated body is instantly transformed[1].

Note that an instantaneous transformation across all of space-time does not violate causality; events that are light-like in the stationary frame will also be light-like in the moving frame. At no point is there transfer of information faster than the speed of light. The transformation is linear and will not reverse the causal order of any pair of events[2].

---

[1] In other words, redshift occurs instantly as is commonly assumed. The separation distance is not relevant. For example, the distortion caused by relative velocity to the cosmic microwave background radiation is corrected for the instantaneous velocity of the observer at the time of observation.

[2] For any observer, causally connected entities lie on the surface of the forward and backward light cones centred on the observer (and probably includes every particle in the universe, once and once only). The causal connection in the backward light

This is quite different to how the same acceleration in Fig. 1 affects external observers. If ψ is a charged particle such as an electron, the effect of the velocity change propagates outwards at the speed of light and is not instantaneous to external observers. This asymmetry is significant, as shown in the next section.

3.2 The Twin Paradox

An instant space-time transformation clarifies situations in SR involving acceleration.

Consider an observer a distance x from a stationary emitter. Expressed in space-time coordinates, the event (0,0) marks the reception of a photon originating from emission event (x, -x/c) in the observer's inertial frame.

Assume the observer is instantly accelerated to velocity v at $t = 0$, just as the photon is about to be received. The received photon will then be associated with event (x',t')[3]:

$$x^{'} = \gamma(x - vt) = \gamma(x + \frac{vx}{c}) = \gamma x(1 + \beta) \tag{3.4}$$

$$t^{'} = \gamma(t - \frac{vx}{c^2}) = \gamma(-\frac{x}{c} - \frac{vx}{c^2}) = -\gamma\frac{x}{c}(1 + \beta) = -\frac{x^{'}}{c} \tag{3.5}$$

The emitter will be perceived to have been further away in retarded space when the photon was emitted: retarded space is stretched in the forward velocity direction and compressed in the reverse direction. The emitter, who has not been accelerated, will see no change in her space-time (though eventually, after a delay x/c, the distant absorber will suddenly be seen to move towards her at velocity v).

A benefit of identifying the transformation as global and instantaneous is seen in the way it clarifies one of the contentious predictions made when acceleration is mixed with SR, the twin paradox (also known as the clock paradox). One form of the supposed paradox concerns two hypothetical twins A and B. Twin B is accelerated and travels to a distant star then decelerates and returns to earth. SR predicts that twin B will have aged less than the stay-at-home twin A. The 'paradox' arises because from the stationary or rest frame of twin B, it is the earth that appears to be moving, first away and then towards, with the result that twin B will predict that the twin on earth will have aged less using the same argument as the sibling. This is obviously contradictory: they cannot both have aged less.

Clearly the situation is asymmetrical because of the acceleration of twin B, but the persistence of the paradox is astonishing - Shuler in 2014 stated there

---

cone is absolutely maintained through an acceleration if acceleration is treated as an instantaneous change in velocity, or as a series of instantaneous changes. What this means is that there is no point during an acceleration where either the emitter or absorber 'disappears' from view, hence the natural order of events is conserved.

[3] Or perhaps more accurately, the next photon emitted immediately afterwards at t'=0_+.

is an increasing number of research publications appearing that address the paradox [1]. One reason for its persistence is that, having accepted that the acceleration on one twin is the cause of the asymmetry, it is not made clear by many authors how precisely it affects the calculations, particularly when it is well known that SR can be applied in a piece-wise manner to situations involving acceleration without the need for corrections (even with accelerations as large as 10^{19} g, as demonstrated by Bailey *et al* [2]). A typical example of the lack of detail offered by some sources is the conclusion to Sciama's explanation of the paradox [3]:

> .. the difference .. is that B has accelerated relative to distant matter while our stay-at-home A has not. . . . These considerations dispose of the clock 'paradox'.

Considering without loss of generality a single leg of the journey with twin A a distance x from twin B and twin B accelerated towards A. The transformation of space-time is associated the instant coordinate change:

$$\left(x, -\frac{x}{c}\right) \rightarrow \left(\gamma[1+\beta]x, -\frac{\gamma[1+\beta]x}{c}\right) \tag{3.6}$$

We can see this immediately resolves the clock paradox. Stationary twin A is distance x away from stationary twin B. If twin B is accelerated towards A, the proper distance to twin A apparent to twin B suddenly and instantly changes to x/γ (having converted the retarded position to the proper position[4] by subtracting $v|t'|$), hence the journey back takes time $x/(\gamma v)$. Twin A sees no difference in relative position as an immediate result of the acceleration and registers an elapsed time of x/v till they meet. The measured intervals are not the same. Time seems to go faster for the accelerated twin when a comparison is made. This is summarised in Table 1.

Time	Parameter	Twin A View	Twin B View
Before Acceleration	Separation Distance	x	-x
After Acceleration	Retarded Distance	x	-$\gamma x(1+v/c)$
...	Proper Distance	x	-x/γ
...	Time To Meeting	x/v	$x/(\gamma v)$

Table 1 The effect of acceleration on the twins.

The resolution of the paradox lies in the instantaneous change in the space-time of the accelerated party, and this is what is missing in Sciama's explanation

---

[4] This is not to suggest that it is proper space-time that is transformed. The switch to proper time is only for convenience - working out the time intervals in retarded time and get the same result.

above. There is nothing new in the explanation itself - a simple mechanism has just been introduced to explain how the difference between the space-times arises.

The suggestion that space-time transforms instantly following acceleration would seem to suggest that space-time is neither absolute or substantive but is instead a relational concept, but that is not necessarily the case[5]. Space-time can be substantial (as might be necessary to build sensible cosmological models of an expanding universe) but entities within see it from a particular perspective, thus neither interpretation is more correct than the other. The variation between observers could therefore be a perspective-based illusion. In this context, the idea of an instant transformation is neither challenging nor problematic[6].

3.3 Gravitational Energy

Immediately following a change in velocity, the transformed space-time will be completely consistent with the accepted laws of physics. But what about gravitational energy – assuming the acceleration was not the result of gravity and that the net gravitational force in the vicinity is so weak that space-time distortion is insignificant? It is possible to recalculate the gravitational binding energy with respect to all other bodies in the universe, assuming a symmetric mass distribution.

The effect of an instantaneous space-time reconfiguration on a static mass distribution centred on the accelerated observer of mass m will be calculated. The mass distribution is assumed to have radius R, total mass M and constant density $\rho = 3M/(4\pi R^3)$. This is shown in Fig. 3. The sphere is split into thin disks rather than shells as might be expected for a calculation of this type because the transformation breaks spherical symmetry.

The volume of the sphere is obtained by integrating as follows:

$$V = \int_{-R}^{R} \left[\int_{0}^{\cos^{-1}\left(\frac{x}{R}\right)} 2\pi x^2 \frac{\sin\phi}{\cos^3\phi} d\phi \right] dx = \frac{4}{3}\pi R^3. \tag{3.7}$$

If we assume also an initial small velocity boost v in the $+x$ direction the difference in gravitational energy is:

$$\frac{\Delta E}{m} = -\int_{0}^{R} \left[\int_{0}^{\cos^{-1}\left(\frac{x}{R}\right)} 2\pi G\rho x^2 \frac{\sin\phi}{\cos^3\phi} \left(\left[\frac{\gamma}{r_+} - \frac{1}{r}\right] + \left[\frac{\gamma}{r_-} - \frac{1}{r}\right] \right) d\phi \right] dx, \tag{3.8}$$

where

---

[5] The ontology of space-time as it relates to the argument presented here is discussed in depth by [4], [5], [6] and [7].

[6] Although the implication that each and every particle somehow maintains its own possibly unique view of the entire universe is a little disturbing. How is this sustained?

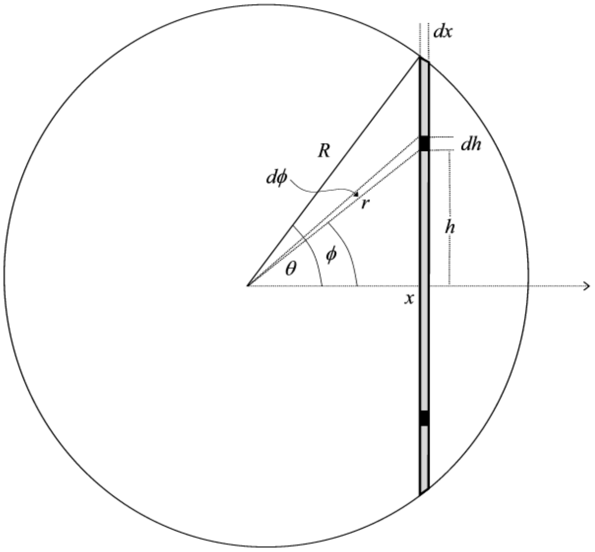


Fig. 2. The spherical mass distribution is split into disks as shown.

$$\begin{aligned} r_+ &= \sqrt{h^2 + x_+^2}; \\ r_- &= \sqrt{h^2 + x_-^2}; \\ x_+ &= \gamma(1+\beta)x; \\ x_- &= \gamma(1-\beta)x. \end{aligned} \tag{3.9}$$

It is assumed that the interaction is retarded. The mass element is not changed except for the relativistic mass increase.

Some preliminary calculations are necessary:

$$\frac{\gamma}{r_+} = \frac{\gamma/x}{\sqrt{\sec^2\phi + \gamma^2(1+\beta)^2 - 1}}; \tag{3.10}$$

$$\frac{\gamma}{r_-} = \frac{\gamma/x}{\sqrt{\sec^2\phi + \gamma^2(1-\beta)^2 - 1}}; \tag{3.11}$$

$$-\frac{2}{r} = -\frac{2\cos\phi}{x}. \tag{3.12}$$

Slotting these into Equation 8:

$$\frac{\Delta E}{m} = -\int_0^R 2\pi G\rho x I dx, \tag{3.13}$$

where

$$I = \int_0^{\cos^{-1}\left(\frac{x}{R}\right)} \frac{\sin\phi}{\cos^3\phi}\left(\frac{\gamma}{\sqrt{\sec^2\phi+\gamma^2(1+\beta)^2-1}}+\frac{\gamma}{\sqrt{\sec^2\phi+\gamma^2(1-\beta)^2-1}}-2\cos\phi\right)d\phi. \tag{3.14}$$

The result is

$$\begin{aligned} I &= \left[\gamma\sqrt{\gamma^2(1+\beta)^2-1+\sec^2\phi}+\gamma\sqrt{\gamma^2(1-\beta)^2-1+\sec^2\phi}-2\sec\phi\right]_0^{\cos^{-1}\left(\frac{x}{R}\right)} \\ &= \gamma\sqrt{\gamma^2(1+\beta)^2-1+\frac{R^2}{x^2}}+\gamma\sqrt{\gamma^2(1-\beta)^2-1+\frac{R^2}{x^2}}-2\frac{R}{x}+2-2\gamma^2. \end{aligned} \tag{3.15}$$

Function I should be approximated before performing the second integration:

$$\begin{aligned} \frac{\Delta E}{m} &\approx -4\pi G\rho\int_0^R\left[\gamma R+\frac{\gamma^3\beta^2x^2}{R}-R+(1-\gamma^2)x\right]dx \\ &\approx -4\pi G\rho R^2\left(\gamma+\frac{\gamma^3\beta^2}{3}-\frac{1}{2}-\frac{\gamma^2}{2}\right). \end{aligned} \tag{3.16}$$

The Lorentz factor will be small ($\gamma \approx 1$) hence this can be simplified further:

$$\Delta E \approx -\frac{1}{2}mv^2\left(\frac{2GM}{Rc^2}\right) \tag{3.17}$$

There is a second-order effect. The accelerated observer will be more strongly tied into the mass distribution than before. The observer will deduce that energy -ΔE has been extracted from the gravitational field during the acceleration[7].

Instead of assuming a retarded force, working with proper distances with no propagation delay gives the result:

$$\begin{aligned} \frac{\Delta E}{m} &\approx -\int_0^R 4\pi G\rho x\left[\int_0^{\cos^{-1}\left(\frac{x}{R}\right)}\frac{\sin\phi}{\cos^2\phi}\left(\gamma-1+\frac{1}{2}\gamma\beta^2\cos^2\phi\right)d\phi\right]dx \\ &\approx -\int_0^R 4\pi G\rho x\left[\frac{\gamma-1}{\cos\phi}-\frac{1}{2}\gamma\beta^2\cos\phi\right]_0^{\cos^{-1}\left(\frac{x}{R}\right)}dx \\ &\approx -4\pi G\rho\left[(\gamma-1)Rx-\frac{\gamma\beta^2x^3}{6R}-\frac{1}{2}\left(\gamma-1-\frac{1}{2}\gamma\beta^2\right)x^2\right]_0^R \\ &\approx -\frac{1}{3}\pi G\rho\beta^2R^2 \approx -\frac{5}{4}\frac{GM}{4Rc^2}v^2 \approx -\frac{1}{2}\left(\frac{GM}{2Rc^2}\right)v^2. \end{aligned} \tag{3.18}$$

---

[7] The procedure followed is rather similar to a (failed) attempt in early part of the last century to attribute the mass of an electron to its electromagnetic field energy. This is described in one of Feynman's lectures [8].

The expansion of the universe has negligible effect on the calculation.

The analysis has been far from rigorous, neither completely Newtonian nor entirely relativistic, with convenient mass aggregations assumed. But the objective was merely to show that there is a possible physical effect associated with the instantaneous frame transformation, and this indeed appears to be the case.

The calculations suggest that the instantaneous frame transformation has an effect on the gravitational energy. This is surprising - where does the energy apparently released by the process go? It is not an artefact of the calculations because the same effect is clearly seen in a toy universe comprising just 3 particles. There is potentially a link here with inertia as there seems to be a hidden energy transformation associated with the reconstruction of space-time that has nothing to do with the applied force.

3.4 Mach's Principle

Over 100 years ago, Ernst Mach proposed that local inertial effects are connected with the large-scale distribution of matter in the universe [16]. The local resistance to a change in motion was considered to be a manifestation of some kind of coupling that exists between all matter in the universe. Distant matter should have the greatest influence by virtue of its overwhelming quantity, but a mechanism by which local motion could be affected by the stars, apparently instantaneously, was not known then, and is still unknown: it cannot be a facet of the normal forces, all of which significantly weaken with growing separation distance and exhibit a response delay. Furthermore, the supposedly dominant effect of distant material makes it impossible to set up a practical experiment to test the principle. Nevertheless, it is an enticing idea that has attracted the attention of many scientists: refer to Bondi and Samuel for the range of modern interpretations of the principle [17]. But in spite of all the work to date the hypothesis has largely remained in the realm of philosophy.

There have been unsuccessful attempts to associate Mach's principle with general relativity, but perhaps the place to look for Mach's principle is in special relativity. If the Lorentz transformation is instantaneous as proposed, the entity being accelerated will instantaneously move all the distant stars and galaxies, with measurable effects in terms of energy measurement. This may be a promising way of linking inertia with gravitational mass. The calculation in the previous section suggests a possible link between the large scale and the small.

There is an interesting link with the so-called large numbers hypotheses that try to explain the coincidence of the similarity of some large dimensionless numbers by proposing (in one variant) that if all the mass (M) in the universe is gathered in a spherically symmetric structure, then the outer surface (whatever that means) at distance R is the event horizon of a black hole. The criterion is:

$$\frac{2GM}{Rc^2} = 1. \tag{3.19}$$

Because the universe expands at a rate that is approximately c, this condition in terms of the age of the universe T is:

$$\frac{2GM}{Tc^3} = 1. \tag{3.20}$$

For this to be always true then G must increase with cosmological time:

$$G = \left(\frac{c^3}{2M}\right) T. \tag{3.21}$$

Slotting this expression for G into equation 17[8]:

$$\Delta E \approx -\frac{1}{2} m v^2. \tag{3.22}$$

This is a curious coincidence. It suggests that energy is required to transform space-time (which we are unaware of) and that this is currently supplied by the change in gravitational energy[9].

The problem with this, of course, is that a change in G has not been detected in local bounded systems. However, this is not necessarily a fatal objection. Some observations can be made:

1. The point at which G increases may be linked with the transition from the apparently instantaneous interaction of local gravitational systems and the the retarded interaction of very distance systems.
2. If the LHS of equation 20 were to be greater than 1, then there is the possibility of an entity moving spontaneously because the gravitational energy harvested then exceeds the kinetic energy. An external force is not required.
3. Though there has been a focus on G varying with time, equation 21 indicates equally an important variation effect with mass. Adding or removing mass to/from the universe can effect inertia[10], and it is possible to conceive of a big bang where the change in G with time and mass makes it energetically favorable for mass to be spontaneously created. The endpoint of creation would be when the inertial and gravitational masses become equal.

3.5 Conclusion

Special relativity is normally applied in situations where velocity differences are already established. It has been proposed here that special relativity can also deal with acceleration, and the results are perfectly consistent so long as

---

[8] The focus is on the retarded interaction because the discovery of gravitational waves from merging neutron stars conclusively demonstrates that gravity is subject to a propagation delay on a large scale that is consistent with that of the electromagnetic interaction [18].

[9] The equivalence between inertial and gravitational mass is then because the gravitational energy supplies the correct amount of energy.

[10] It may well be that only one other particle is required to establish the property of inertia because the observer is anchored to this particle via its presence on the surfaces of both the future and past light cones.

space-time for the accelerated entity is instantly transformed with the acceleration. The overarching picture is of every particle holding a complete space-time picture of the entire universe, a picture that can transform instantly. But the transformation has energy consequences. Looking just at the gravitational interaction (though other forces are by no means excluded), it was established that the gravitational energy changes. This suggests a link between the individual acceleration event and the bulk of the universe, the type of link described as Mach's principle.

References

1. Shuler R. L. 2014. The Twins Clock Paradox History and Perspectives. Journal of Modern Physics 5, pp 1062-1078.
2. Bailey J. *et al.* 1977. Measurement of relativistic time dilatation for positive and negative muons in a circular orbit. Nature 268, pp 301-305.
3. Sciama D. 1986. Time 'paradoxes' in relativity. in 'The Nature of Time', Flood R, Lockwood M, Eds, BASIL BLACKWELL.
4. Pooley, O. 2012. Substantivalist and Relationalist Approaches to Spacetime.
5. Baker, D. 2004. Spacetime Substantivalism and Einstein's Cosmological Constant. Philosophy of Science 72 (5):1299-1311.
6. Belot G. 1999. Rehabilitating Relationalism. International Studies in the Philosophy of Science 13 (1):35 – 52.
7. Allori V. 2016. Space, Time, and (how they) Matter: a Discussion about some Metaphysical Insights Provided by our Best Fundamental Physical Theories. In G. C. Ghirardi and S. Wuppuluri (eds.), Space, Time, and The Limits of Human Understanding. Springer. pp. 95-107.
8. Feynman R., Leighton R., and Sands M. 1964. The Feynman Lectures on Physics. Edition: Basic Books (2010) Volume 2 Ch. 28. Also available at http://www.feynmanlectures.caltech.edu/ [Accessed 30 July 2019].
9. Sciama D. W. 1953. On the Origin of Inertia. MNRAS 113 1, pp 34–42 https://doi.org/10.1093/mnras/113.1.34 .
10. Bondi H. and Samuel J. 1997. The Lense–Thirring Effect and Mach's Principle. Physics Letters A. 228 3, pp 121–126 (1997). Available at arXiv:gr-qc/9607009 [Accessed 30 July 2019].
11. Abbott B. P. *et al.* (LIGO Scientific Collaboration and Virgo Collaboration). 2017. GW170817: Observation of gravitational waves from a binary neutron star merger. Physical Review Letters 119:161101.

Appendix A: The Doppler Redshift Formula

The relativistic redshift formula is:

$$E^{\gamma} = \gamma \left(1 + \frac{v}{c}\right) E^{\gamma}_{o} = \frac{E^{\gamma}_{o}}{1 + z} \tag{3.23}$$

where ${E^{\gamma}}_{\mathrm{o}}$ is the rest frame photon energy and E^{γ} is the observed photon energy. The Lorentz factor refers to velocity v, the velocity of the emitter at the time of emission in the observer frame. This is easily derived using the invariance of the energy-momentum 4-vector, along with the conservation of energy and momentum:

$$EdE = pc^2dp + m_oc^4dm_o \tag{3.24}$$

where E and p refer to properties just prior to photon emission and m_o is the system rest mass. From energy and momentum conservation:

$$\begin{aligned} dE &= -E^\gamma; \\ dp &= \frac{-E^\gamma}{c}; \\ dm_o &= \frac{-E^\gamma_o}{c^2}. \end{aligned} \tag{3.25}$$

Slotting in and rearranging gives the standard equation.

It seems like the derivation includes the recoil velocity (through the conservation of momentum requirement), but that is not entirely true. If initially the source velocity is 0, then the formula states $E^\gamma = E^\gamma{}_\mathrm{o}$. However, the emitter must recoil, but this is not included in the energy balance[11].

The problem is that the differential equation is a continuous equation, whereas the photon emission event is discrete, hence the commonly used equation is only an approximation. In most cases it is an excellent approximation, but the error might be significant, perhaps in the context of cosmology.

The full and complete redshift formula is

$$E^\gamma = \gamma\left(t + \frac{v}{c}\right)\left(1 - \frac{E^\gamma_o}{2m_oc^2}\right)E^\gamma_o = E^\gamma_o\frac{\left(1 - \frac{E^\gamma_o}{2m_oc^2}\right)}{1+z}. \tag{3.26}$$

Take as an example a hydrogen atom (mass 938.8 MeV) emitting a H-α 1.89 eV Balmer-alpha photon. The correction amounts to

$$\frac{E^\gamma_o}{2m_oc^2} = 10^{-9}. \tag{3.27}$$

The error is negligible. It might be expected that there may be cosmological sources where there could be a significant effect, such as photons from electron-positron annihilation, and synchrotron radiation from relativistic electrons, but the cosmological redshift is wavelength stretching caused by space expansion and it is unclear if the recoil calculations above are relevant, or indeed if photon energy is even conserved with the cosmological expansion.

---

[11] It is possible in some cases for there to be no recoil or almost no recoil, this is when the emitter mass includes the crystal lattice, and not just the emitting atom. For this to work of course, the extended lattice must be a single quantum system.

4 Can G Vary?

Summary

A simple cosmological model with a gravitational constant G that varies over space and time is presented. The variation is specifically selected to result in a universe expanding at a constant rate. The effectiveness of the model is evaluated and the possible connection between space-time and gravity when the gravitational constant is variable is considered.

4.1 Introduction

The constants of nature give form to the universe, but why they take on the specific values that give rise to this particular world is completely unknown. A better understanding of the universe can potentially be gained by changing the fixed value of these constants to model what might happen, or even allowing variation over space and time. This is not to suggest that such variation is actually taking place, but rather that in doing so we may learn more about the nature of these constants and the universe itself.

Of all the constants, the Cavendish gravitational constant stands out because it operates on the larger scale and can therefore say something about the structure of the universe as a whole. Can it vary, and what would happen if it did, or could, vary?

The question will be investigated by constructing a simple cosmological model where the key features of the real universe are reproduced by carefully modifying the gravitational constant. Whilst such a model does not exactly match observational data, it does so sufficiently well to bring out some very fundamental problems with varying G, problems that one should be wary of in developing more realistic cosmological models or theories where G is a variable.

The possibility of a varying gravitational constant has been widely explored in a cosmological context, for example, as a way of explaining dark energy and the accelerating expansion [1], [2]; in scalar-tensor theories [3], [4], [5]; and alternative cosmologies [6], [7]. The body of work so far in the area of varying gravity provides a variety of interesting ideas which in many cases are largely consistent with observation, but simple density distributions subject only to the cosmological expansion tend to be adopted. The difference here is that the variation is treated as a tool to investigate the relation between gravity and space-time. The basic simplicity of the model will allow more realistic dynamics to be investigated. It will be seen that more complex systems with discrete masses and significant peculiar velocity fields leads to predictions that are much more difficult to reconcile with observation.

The approach taken will be to develop a natural coasting model using varying gravity. A number of coasting cosmologies have been seriously proposed, both general relativistic and Newtonian (for example, [17], [18], [19], [28], [29],

[15]), but these have been severely criticised, [21], [23][1]. In coasting models, the universe is expanding in size (and has always been expanding in size) at a rate equal to the speed of light. This approximation is largely justified because it appears correct to first order - the current value of the Hubble parameter in the concordance model ranges from 67 - 74 km s^{-1} Mpc^{-1} [8], [16], equivalent to a Hubble time of between 13.2 and 14.6 billion years, which has in the middle of the range the age of the universe (about 13.8 billion years)[2].

In the majority of models, G as a time-varying function is assumed to be a simple power of cosmological time, most frequently a negative power because of the lasting impact of Dirac's observation that to maintain at all times the curious ratio of gravitational to electromagnetic forces between a proton and an electron requires a G value that should be inversely proportional to time [24]. A basic Newtonian model with $G \propto T^{-n}$, where n is a positive number, is covered in detail by Barrow [27], [20]. The focus here will be on negative n and energy conservation mechanisms associated with this type of variation[3].

4.2 The Model

Any serious attempt to incorporate varying G into a realistic cosmology needs to account for the fact that measurements within the solar system reveal no significant variation either over space or over time [21].

For that reason, a very specific type of G variation is proposed. For bound systems with high internal acceleration the intrinsic value of the gravitational constant is a constant 6.67 x 10^{-11} m^3 kg^{-1} s^{-2} (labelled G_o) throughout and it is presumed that the mass within is not subject to the cosmological expansion. The G value will increase with time between systems or sub-components that are part of the expansion, and the increase begins once the space in between begins

---

[1] The Milne Special Relativistic model is of particular interest: it refers to flat Minkowski space-time with unrestrained expansion and is the original coasting universe. However, it has been translated into FRW cosmology as the special case of zero mass density and negative curvature. Though the equations for the expansion in each case are the same, the Milne model is not a general relativistic space-time with negative curvature. The two models are conceptually incompatible. Milne made this clear in his textbook 'Relativity Gravitation and World-Structure' (OXFORD UNIVERSITY PRESS, 1935), sections 493 and 501. Milne cosmology has some unusual features such as a rejection of the expansion of the universe and a trick where an infinite amount of matter is fitted into the universe using the special relativistic contraction effect, and it is considered that general relativity dominates locally with special relativity operating globally (the assumption being that gravity on the cosmic scale is too weak to result in significant space-time curvature): these features are not typical of modern coasting models. Because of an association with the earlier Milne model, more recent coasting models seem to have acquired a bad reputation.

[2] The size of the universe is consistent with growth at the speed of light over the lifetime of the universe

[3] A warning from the outset: a time-varying G does not emerge naturally from any cosmological model, and a model into which this feature is added has a tendency to appear forced and rather contrived.

to expand[4]. The threshold condition upon which an entity joins the expansion is not precisely specified, but implies some connection between gravity and the expansion of the universe. This is illustrated in Fig. 1 where object 1 joins the flow at cosmological time T_2 and object 2 joins the flow at later time T_2.

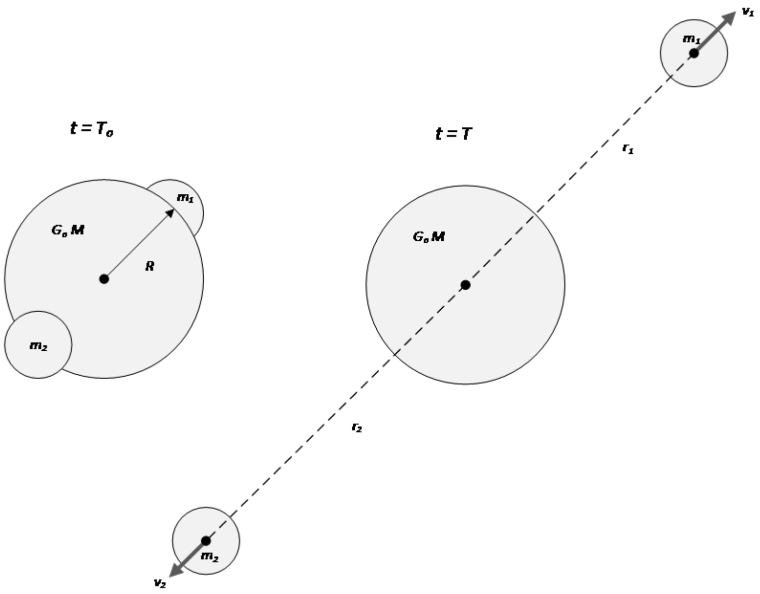


Fig. 1. A system with interacting components initially locked and protected from the expansion changes over time as the two components join the flow at different times.

Assuming the simplest Hubble law and a Hubble velocity that remains constant:

$$v_{h1} = \frac{r_1}{T} = \frac{R + v_{h1}(T - T_1)}{T} = \frac{R}{T_1}, \tag{4.1}$$

$$v_{h2} = \frac{r_2}{T} = \frac{R}{T_2}. \tag{4.2}$$

The Newtonian potential of the central mass is

$$\Phi(r) = -\frac{G(t)M}{r} \tag{4.3}$$

where r is the radial distance[5]. Spherical symmetry is assumed.

---

[4] The inspiration here is rest mass and the effect of velocity on mass, in this case the expansion velocity.

[5] The potential is essentially the energy that can be extracted from the field by moving unit mass from r to infinity. It is unclear if this procedure is even valid when G is

If the body moves only through the expansion, the gravitational potential energy at time T is:

$$U_{10}(T) = -\frac{G_{10}(T)m_1}{r_1} = -\frac{Mm_1}{R}\frac{T_1}{T}G_{10}(T), \tag{4.4}$$

$$U_{20}(T) = -\frac{G_{20}(T)m_2}{r_2} = -\frac{Mm_2}{R}\frac{T_2}{T}G_{20}(T), \tag{4.5}$$

The potential energy for time T in each case is equal to the initial potential energy if the following relation holds:

$$\frac{dG(T)}{dT} = \frac{v_h G(T)}{r} \tag{4.6}$$

If the object does not participate in the expansion, v_h is 0 and G remains constant as expected; if the object is in the expansion the rate of change is G/T (using Equation 1).

It therefore follows that

$$G_{10}(T) = \frac{T}{T_1}G_o; U_{10} = -\frac{Mm_1}{R}G_o; \tag{4.7}$$

$$G_{20}(T) = \frac{T}{T_2}G_o; U_{20} = -\frac{Mm_2}{R}G_o. \tag{4.8}$$

With the increasing G value, the Hubble velocity remains constant as does the gravitational potential energy - no work is done against gravity by bodies that have joined the expansion. The coasting effect has been achieved without introducing anything except the time varying G. Equation 6 in the absence of peculiar velocities can be written as

$$\frac{dG(r)}{dr} = \frac{G}{r} = constant \tag{4.9}$$

for bodies in the flow. Essentially, a very specific G variation function that ensures energy conservation has been applied.

But care is required when looking at complex rather than point systems, when the simple Equations 6 and 9 do not necessarily apply. Referring back to Fig. 1, this is evident when we look at the G value that relates to the gravitational potential energy between m_1 and m_2. Energy must be conserved by the G variation that arises, hence

---

changing because there is no practical way of making such a measurement (also, the universe is not infinite and is actually continually growing in size). In addition, defining the force on a test body as the derivative of the potential results in work done that does not correspond to the change in gravitational potential. The potential function no longer gives rise to a conservative vector field. In any event the expansion is a complication which means the force is no longer conservative when a proper coordinate system is used.

$$U_{12}(T_1) = -\frac{G_0 m_1 m_2}{2R} \tag{4.10}$$

$$U_{12}(T_2) = -\frac{G_{12}(T_2) m_1 m_2}{R\left(1 + \frac{T_2}{T_1}\right)} = -\left[\frac{1}{2}\left(1 + \frac{T_2}{T_1}\right) G_o\right] \frac{m_1 m_2}{R + r_1} \tag{4.11}$$

$$U_{12}(T) = -\frac{G_{12}(T) m_1 m_2}{RT\left(\frac{1}{T_1} + \frac{1}{T_2}\right)} = -\left[\frac{1}{2}\frac{T}{T_2}\left(1 + \frac{T_2}{T_1}\right) G_o\right] \frac{m_1 m_2}{r_1 + r_2} \tag{4.12}$$

It is clear that a piece-wise change in G is necessary because of the way the energy is referenced to the centre of mass of bound objects. However, on a large scale the error associated with this is small.

Varying G is an effective mechanism for maintaining energy conservation both locally and globally through the expansion, and it is actually difficult to think of another way of achieving the same result. The approach in standard ΛCDM cosmology makes an interesting contrast. The universe is assumed to be homogeneous and isotropic and on this basis the rate of expansion is derived (with various tweaks made to ensure both energy conservation and predictions that match observational data). The energy required to oppose gravity is supplied by a damping of the expansion. This approach assumes space-time is substantial, a backdrop, and it is space that is expanding and carrying masses with it; in a sense space is being created. It is a global model but significant problems appear when the local consequences are calculated. Looking again at Fig. 1, applying the global expansion rules at a local level to discrete mass distributions results in energy conservation violations that are evident in the motion of m_1 and m_2 with respect to the host when the expansion rate is exerted from without[6]. How is it possible to reconcile these energy violations (even though they may average out over the universe to zero) with a universe that appears constructed on the fundamental building block of energy conservation?

Alternatively, if the global expansion is the accumulation of local effects all of which conserve energy (a bottom-up approach), then the rate of expansion should vary with local mass density, but there is no evidence for this happening. Worse, the local expansion depends on the local distribution of masses, which implies an insubstantial and relational space-time. Trying to tightly link matter with space to recover a substantive space-time hits a problem because in the presence of normal forces mass happily slides across space-time and there is no evidence of a locking effect that would be necessary for the action on matter to affect the dynamics of space itself. Adding new forces to supply energy (such as dark energy or some cosmological constant) is an interesting measure that appears effective on a global level, but again expect problems when a local interpretation

---

[6] It is the question of how the universe as a whole influences the individual parts - Mach's problem again. What is the balancing mechanism?

is needed - if expanding space releases energy by some process, how is it harvested and targeted to offset the energy consumed fighting gravity[7]?

4.3 A Comparison with MOND

The very simple model that has been developed can be shown to be equivalent to MOND under some circumstances. MOND is a modified gravity proposed by Milgrom in 1983 [22] where gravity becomes much stronger than expected as the normal Newtonian acceleration approaches a lower threshold value. The modification is remarkably effective at explaining the anomalous rotation curves of spiral galaxies, but is less effect on a larger scale.

MOND is described by the relation

$$a_N = \mu\left(\frac{a}{a_o}\right) \tag{4.13}$$

where a is the actual acceleration, a_N is the expected Newtonian acceleration, a_o is the MOND 'constant' approximately equal to 1.2 x 10^{-10} m s^{-2} and $\mu(a/a_o)$ is an interpolating function. The most basic function used is

$$\mu\left(\frac{a}{a_o}\right) = \frac{1}{1+\frac{a}{a_o}}. \tag{4.14}$$

The standard interpolating function is[8]

$$\mu\left(\frac{a}{a_o}\right) = \sqrt{\frac{1}{1+\left(\frac{a}{a_o}\right)^2}}. \tag{4.15}$$

Rearranging the basic interpolating function,

$$a = -\frac{|a_N|}{2}\left(1+\sqrt{1+4\frac{a_o}{|a_N|}}\right). \tag{4.16}$$

When $|a_N| \gg a_o$, $a = a_N$; if $|a_N| \ll a_o$, $|a| = (a_o|a_N|)^{1/2}$.

Identifying R in Fig. 1 as the transition radius with acceleration corresponding in magnitude to a_o at T_1 for m_1,

$$a_o = \frac{G_o M}{R^2}. \tag{4.17}$$

---

[7] The usual escape is 'energy is not conserved in general relativity', but then it is possible to continuously create energy from nothing once peculiar forces are included. In addition, the key general relativistic equation describing the dynamical evolution of the universe is identical to the Newtonian equation of motion of a test particle outside a sphere of radius r and mass M, an equation which essentially expresses energy conservation.

[8] Care is required with signs; the normal convention being the gravitational acceleration with respect to a central source is negative.

Now using equation 9,

$$|a(r_1)|= \frac{G(T)M}{r^2} = \frac{G_oM}{Rr_1} = \sqrt{\frac{G_oMG_oM}{R^2r_1^2}} = \sqrt{|a_N|a_o}. \tag{4.18}$$

Varying G as defined by Equations 6 and 9 is equivalent to MOND when the transition point is identified as the point when the acceleration matches the MOND characteristic acceleration a_o.

The effect can be quantified by looking at a hypothetical galaxy where material on the outer edges joins the flow (without affecting the boundedness of the material) as the acceleration drops below a_o. Let the current cosmological time be τ and assume mass leaves the outer edges of a galaxy (1 kpc) at time $\tau/10$. The flow velocity will be 1 kpc/$(\tau/10)$, which is approximately 750 m s^{-1}, and the current position will now be about 10 kpc from the centre. The apparent G value will therefore be 10 G_o.

Anomalies like this are typical, and are apparent to a lesser extent in the motion on the sun around the Milky Way galaxy centre. The sun lies about 8 kpc from the centre. The current Hubble velocity is therefore 570 m s^{-1}. The kinematics suggest the sun is gravitationally influenced by 1.9 times the mass that is visible, hence the material around the sun will have joined the flow at a distance of about 4.1 kpc from the centre about 7 billion years ago. The total baryonic mass of the Milky way galaxy is about 3 x 10^{11} solar masses but assuming 5 x 10^{10} within the solar orbital radius, the acceleration at the point of joining the flow was approximately a_o as expected.

But varying G is only the only the same as MOND under certain restricted circumstances. The enhancement of G depends on history and the time interacting entities have been expanding relative to one another and significant deviation from the MOND force equation is possible if peculiar velocity is present - this breaks teh the equivalence between Equations 6 and 9. In situations beyond the boundaries of galaxies, significant deviations from the MOND equation can be expected[9]. Another possibility is that the G value can in certain circumstance be reset back to the standard value $G_o = 6.67$ x 10^{-11} m^3 kg^{-1} s^{-2}.

4.4 Peculiar Velocity

The inclusion of peculiar velocities adds significant complication. The assumptions of varying G cosmology can be summarised:

- G scales with the change in cosmological time
- The Hubble velocity is the proper separation distance divided by the absolute cosmological time

---

[9] The highest mass discrepancy of about 100 is in clusters (though values as high as 500 have been suggested) and with a universe age of 13.8 billion years, expansion of the material that has gone on to form the cluster components will therefore have starting at 138 million years. This is consistent with the evidence of galaxies forming about 200 million years after the big bang.

- Energy and momentum are conserved locally and globally
- The value of G can be reset to G_o under certain circumstances

Again working only in one space dimension and including a peculiar velocity in the radial direction, the effect of the central Newtonian gravitational force on the peculiar velocity is:

$$\frac{dv_p}{dt} = -v_p \left(\frac{GM}{r^2 v} + \frac{1}{T} \right). \tag{4.19}$$

However, it is more convenient to work with proper distance and total velocity when looking at infall kinematics. It is not necessary to make a distinction between expansion and peculiar velocities, except as a start condition[10]:

$$\frac{dv}{dt} = \frac{GM}{r^2} \left(\frac{v_h}{v} - 1 \right) = -\frac{GM}{r^2} \frac{v_p}{v}. \tag{4.20}$$

With a large separation distance and therefore a large expansion velocity, the gravitational pull is amplified dramatically if the peculiar velocity is in the opposite direction to the expansion and of slightly greater magnitude.

On the Mpc scale, large infall velocities can be achieved in point mass freefall simulations, and the velocity is retained as the accelerated body passes through the source (because the direction of expansion relative to peculiar velocity changes)[11]. A domino effect can be imagined where a sequence of accelerations builds up a significant mix of peculiar and expansion velocity. The increasing G with time contributes significantly to the growth in peculiar velocity in order to conserve energy.

The change in G over time is modelled using λ as a gravitational constant enhancement that can range from 1 to 500[12]. A comparison can be made with the standard MOND interpolating function and also with ΛCDM following the procedure of Angus and McGaugh [23] using $H_o = 72$ km s^{-1} Mpc^{-1}, $\Omega_\Lambda = 0.73$ and $\Omega_\mathrm{m} = 0.27$ in the equation

$$\frac{da(t)}{dt} = H_o \left[\Omega_m a^{-1} + \Omega_\Lambda a^2 \right]^{1/2} \tag{4.21}$$

---

[10] Equations 19 and 20 follow from

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{r}} \right) = \frac{\partial L}{\partial r}$$

where the Lagrangian is

$$L = \frac{1}{2} m \dot{r}^2 + \frac{G(t)Mm}{r}.$$

[11] This is analogous to a gravitational slingshot and explains the very large peculiar velocities needed for infall even to start.

[12] It may be that there are cosmic particles from close to the beginning of time with a huge gravitational enhancement.

where a is the scale factor. Fig. 2 shows the results starting from a distance of 4.5 Mpc from a gravitational point source of 10^{14} solar mass and an initial proper velocity of -1,000 km s^{-1} 3.5 billion years ago.

By the time the test mass is 1 Mpc away, the peculiar velocity in each case has increased significantly because of the gravitational attraction and the conversion of expansion velocity to peculiar velocity.

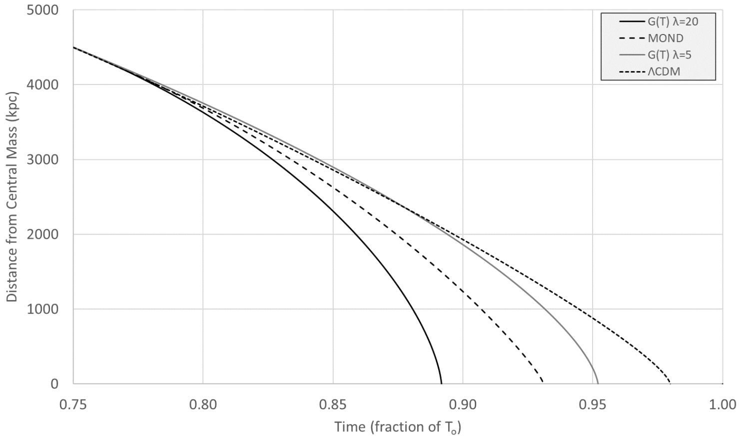


Fig. 2. Infall model with a central mass is 10^{14} solar mass. The initial proper velocity is -1,000 km s^{-1} (the peculiar velocity excess over expansion velocity). The velocities at 1,000 kpc are respectively -4,160 km s^{-1}, -2,440 km s^{-1}, -2,330 km s^{-1}, -1,660 km s^{-1}.

Very large peculiar velocities on the scale of galaxy clusters are observed [24] but some are difficult to explain in the standard cosmological model [25]. However, the flexibility in the choice of λ means the anomalous dynamics at cluster level can potentially be explained by fine-tuning the λ value, which in turn reveals the expansion history of the interacting bodies.

One specific example is the bullet cluster at $z = 0.296$ where one component of the galaxy has passed through a larger component at speed of the order of 4,500 km s^{-1}. Modelling suggesting a relative speed of 3,000 km s^{-1} at 2,500 kpc separation (redshift $z = 0.5$). The cluster as it is now is shown in Fig. 3. Achieving the required speed is not a problem for varying-G cosmology, but one result of the collision is that gas has been stripped and accumulates at the point of collision. The components that have passed though appear to have greater than expected mass (conventionally interpreted as retaining the dark matter), whilst the stripped gas has the normal gravitational pull for the visible mass. In varying G terms, the component cores are mostly stars and pass through with minimal interaction, retaining the enhanced G that has accumulated. On the other hand, the colliding gas will release energy though inelastic collisions and some of this can be used to pay back the gravitational energy needed to reset the G value to G_o. Particle collisions are the only viable mechanism that ensures energy and momentum are conserved.

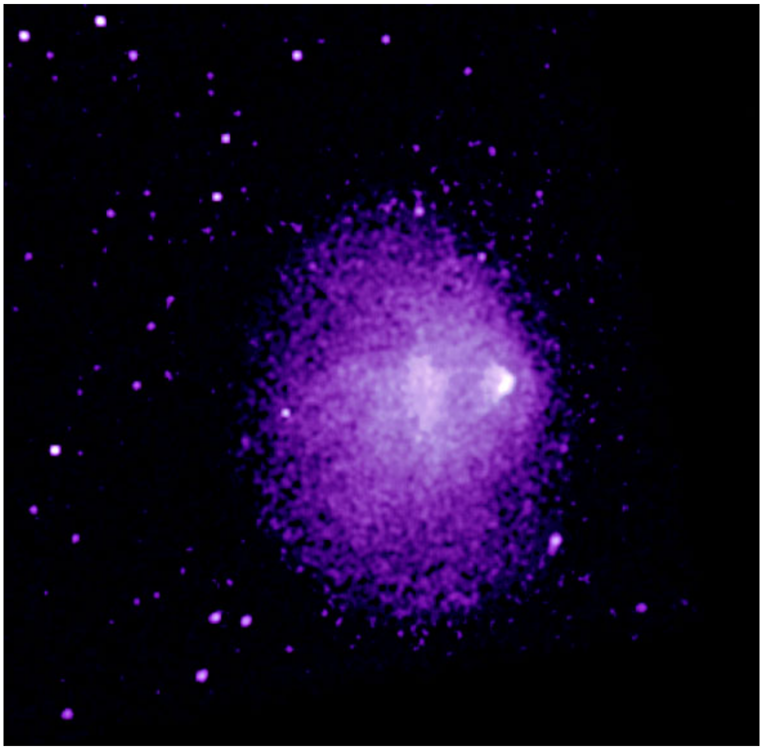


Fig. 3. The shock wave located near the right side of the Chandra X-ray image is the result of the collision of a smaller group or sub-cluster of galaxies with the larger group. Only the hot gas is visible. The target and bullet sub-clusters have passed through and gravitational lensing data indicates the mass of the bullet and target galaxies are higher than indicated by visible matter. [26].

But it is here that the huge problems with a varying G of the manner described are most clearly seen:

1. The G value as described is a relative value between two interacting entities. Whilst colliding gas must have some mechanism for returning the value to normal so that a bound gravitational structure with a common G value both internally and as seen by distant entities can result, how can this be reconciled with distant matter when each of the distinct components will have had different G values prior to collision[13]?

---

[13] It is therefore unclear how this type of reorganisation can take place locally when there are global consequences. An echo of the problems of inertia and Mach's principle.

2. Gravitational lensing indicates that the effective mass of the galaxies in Fig. 3 is much greater than suggested by the visible material. However, if G is relative, how is gravitational lensing possible? Lensing is essentially a field effect (or, equivalently, the bending of light by curved space-time) which depends on gravitational being absolute, either real space-time curvature, a field, or by some local particle-by-particle interaction. In a relational interpretation, it should be based on the relative G value of the photon source and the lensing mass, but this is not consistent with observation. General relativity is certainly correct locally, but it remains to be demonstrated that the expanding universe affects the curvature in a way that conserves energy.

However, the problems might arise because of the way the G is considered to vary with time rather than position. Equation 9 was derived from Equation 6 in relation to galaxy dynamics. It is possible to make G a function of space to rather than time to facilitate lensing and still support a coasting universe, but an inevitable consequence is that bodies are no longer influenced in the normal way by gravity[14]. But G can also be a mix of space and time variation whilst still achieving the free expansion required. The variation needs to satisfy the following equation:

$$\frac{dG}{dr} = \frac{T}{r}\frac{\partial G}{\partial t} + \frac{\partial G}{\partial r} = \frac{G}{r}. \tag{4.22}$$

Rearranging,

$$G = T\frac{\partial G}{\partial t} + r\frac{\partial G}{\partial r}. \tag{4.23}$$

This has solution

$$G(r,T) = c_n r^n T^{1-n}. \tag{4.24}$$

with the variation just with time or just with space as special cases:

$$n = 0 : G(r,T) = c_o T; \tag{4.25}$$

$$n = 1 : G(r,T) = c_o r; \tag{4.26}$$

An n value in between is a mix of both ($n < 0$ or $n > 1$ are nonphysical). The constant, c, is determined by the G_o, T_o and r_o when the entity joins the expansion:

$$G(r,T) = \frac{G_0 R^n T^{1-n}}{r_o^n T_o^{1-n}}. \tag{4.27}$$

Generalising Equation 20[15],

---

[14] As a function of time, G is relative; as a function of space it is absolute. This is reminiscent of the photon – particle or wave, substantial or insubstantial?

[15] Using d/dt instead of the clumsy d/dT, assuming clock time and cosmological time do not significantly differ.

$$\frac{d}{dt}\left(\frac{1}{2}v^2 - \frac{GM}{r}\right) = 0 = v\frac{dv}{dt} - \frac{GM}{r^{1+n}T^{1-n}}\frac{d}{dt}\left(r^n T^{1-n}\right) + v\frac{GM}{r^2}. \tag{4.28}$$

Rearranging,

$$\frac{dv}{dt} = -\frac{GM}{r^2}\left(1 - \frac{v_h + nv_p}{v}\right). \tag{4.29}$$

An appropriate choice of n can be selected to match observational data but the simplicity of G varying with the expansion is lost. It is nevertheless one way of proceeding.

4.5 Discussion

Constant Expansion

A universe expanding at the speed of light has been presented as a plausible cosmology, but without full justification. The concept is worthy of serious consideration because it opens up the possibility of the expansion being explained without the need for an initial explosion. The idea of an initial explosion is natural but objectively is a very weak metaphor – explosions do not create space (and time). The expansion is undoubtedly taking place, but surely there is a more convincing explanation?

In the spirit of special relativity where the contortions of space and time are the result of a requirement that the speed of light should be constant, one can inquire if the constant expansion of the universe arises from a need for some other consistency. The question can be phrased as follows: Why is it necessary for the universe to expand in order that consistency be maintained? Or equivalently, what inconsistency would arise if the expansion did not occur?

The question can be approached from the viewpoint of the observer in the context of what is directly observable. From this perspective, the retarded frame is appropriate. In the retarded frame, the observer is in the privileged position of being at the centre of a finite universe and its oldest occupant. Light that is received has been emitted at earlier times from objects that lie on concentric spherical surfaces tracing the regression of time to an outer surface where $T = 0$, the point when the universe came into existence and time began. This outer surface is the transition from no time (or universe) to existence and the progression of time. Not surprisingly, a major discontinuity can be expected at this point. If it is the case in our universe that the observer should be protected from singularities and discontinuities, it would be expected that the universe will operate in some way to prevent the $T = 0$ discontinuity becoming apparent at an observational or operational level. One way this can happen is if the outer surface recedes from the observer at the speed of light. The way it works can be illustrated with a simple example: Consider the observer (now) at time $T = T_o$. Let the observer be in causal contact with an emitter at $T = 0$ at the edge of the universe, a distance cT_o away. After a time increment δT in the observer's frame, time has moved on to $T_o + \delta T$. The universe has also expanded by $c\delta T$ hence photons received at this later time from the same source will also appear to have originated from time $T = 0$. The observer can remain in causal contact with entities for whom time does not progress, thus the transition point as time began to 'flow' is incorporated into the observable world.

It is tentatively suggested that this may be the reason for the cosmological expansion. In the example given, if there were no recession velocity the observer would perceive photons received at $T_o + \delta T$ to have originated from the source on the boundary at a later time, which of course is impossible if time is not progressing at the boundary - there is no later time. The cosmological expansion is then a direct consequence of the finite age of the universe: mass density and gravitation are irrelevant.

The model is simple and philosophically very appealing because it deals nicely with the beginning of time and incorporates it into observation. It obviously predicts the universe is expanding at a constant rate c and we may presume intermediate locations scale proportionally (although there is no compelling reason why this should be the case). Although there are problems with the coasting model, the possibility of explaining the expansion of the universe is a compelling reason for persisting with it and trying to resolve the problems that have been identified.

Space-Time

A varying gravitational force constant has been presented as a mechanism for decoupling the expansion process from the matter content of the universe[16]. A steady expansion rate guaranteeing energy conservation is the result.

But the effectiveness of a basic cosmology with varying G is questionable given that the observed world is not exactly as predicted[17]. It is therefore not necessarily a viable alternative to the consensus model of a curved universe with inflation, dark matter and dark energy[18] but that it is an effective vehicle

---

[16] The idea that the expansion should conserve energy has been a basic assumption throughout, and the only obvious way of doing this is to vary the gravitational constant - the idea that the expansion can be slowed as work is done against gravity seems fanciful.

[17] The model is not entirely satisfactory and suffers very similar problems to those of consensus cosmology when examined at the local level. However, the same problems are applicable to standard cosmology, but given the huge mathematical separation of the current consensus model and the initial concept of almost a century ago there is little benefit or interest in the scientific community in revisiting foundational principles. In addition, how could you ever distinguish varying G from invisible mass for objects at great distance because any test will only probe the product GM?

[18] The introduction of dark energy to drive the relatively constant expansion that is observed (and which is a natural consequence of varying G) is particularly interesting. There are certain parallels with the frenetic activity between 1890 and 1910 in search for the medium to support the electromagnetic radiation theorised by Maxwell and

for posing and perhaps answering some very difficult questions relating to the nature of space and time and what gravitation really is. Certainly, it is possible in the concordance model to get better and better agreement with observational data by elaborate fine-tuning, but these are not ground-breaking and disruptive changes that delve into the nature of the universe but are merely the addition of more of the type of stuff we know about in order to improve the fit with observational data.

But what can a varying G value tell us about space-time? The relationalist interpretation of space and time considers that all causes and effects can be ascribed to the relative properties of substantive entities and the introduction of space as a backdrop is unnecessary, and that fields therein are abstract constructions of no ontological value. Though this interpretation is fully compatible with special relativity (which relates to inertial systems)[19], it is hard to square with situations involving acceleration: acceleration is absolute; but absolute relative to what? What is the zero reference? On balance, the preferred and prevalent view is that space is a substantive medium in which particles are embedded[20].

An obvious approach is to reorganise material according to relative G value, but the topology of the G-domain is similar to that of an expanding universe and adds little to it[21].

Any connection between gravity and space-time lies on the boundary between physics and metaphysics, and it was common to ponder such things in the past. What might Lord Kelvin, who vexed over the aether (and if it could occupy the same space as ponderable matter) have made of this additional information

---

experimentally produced by Hertz. The aether was considered to be either a perfect fluid or perhaps an elastic solid. Berkson [27] describes the wide range of extremely imaginative ideas all of which were ultimately unsatisfactory. Examples included: pulsating spheres in an incompressible fluid (Bjerknes); eternal vortices in a perfect fluid (Lord Kelvin); the rotationally elastic ether that resisted twisting (Macullagh); a gyroscope and connected rods model (Lord Kelvin); the imaginary fluid ether with an additional spec dimension (Pearson, of statistical test fame). Berkson refers to them in terms of 'remarkable and ingenious failure'. The development of special relativity dispensed with the aether as no mechanical system with the required properties could be conceived.

[19] And suggests there is no absolute space and time and no aether. In addition, the constancy of the speed of light is more fundamental as a principle than space and time. This removes a degree of freedom as it could be said that space is just c times interaction time difference between entities.

[20] The very different viewpoints can be illustrated by considering the series of integers 1, 2, 3, .. and asking what lies between them. A substantivalist response might be to describe the fractional numbers in between that support the entire edifice. The relationalist may say the question is irrelevant and arises from a misunderstanding – the series is simply a list of distinct entities as is demonstrated if the labelling is altered to a, b, c..

[21] Equation 6 can be rewritten

$$\frac{\dot{G}}{G} = H$$

and is essentially the standard cosmology scale factor equation.

had it been available 120 years ago [27]? As a substantivalist he might have reasoned as follows: 'Let a universe of fixed size be filled with a perfect fluid. A unit of matter is created/inserted and in doing so displaces and compresses the fluid into the remaining space in such a way as to create a pressure gradient. Matter moving through the aether will be affected by the pressure gradient in a manner analogous to Bernoulli's fluid equation mimicking the apparent effect of a gravitational force. Repeating the process of mass insertion 10^{80} times, the pressure builds up until the universe is no longer able to oppose it and starts to expand, an expansion that has yet to stop.' Lots of holes there, but the basis for a model. Alternatively, he may have made a comparison with static and dynamic/kinetic friction where no motion takes place until a threshold force is applied.

The relationalist will reason quite differently: 'Varying G is evidence of the universe losing control over one the constants. Under normal circumstances interacting entities exchange information effectively and under these conditions come to an agreement on the value of the constants, in this case, the gravitational constant. However, if the distance is too great, propagation lags mean that agreement is not achieved and the value of G drifts. And because the apparent universe is a construction based on the set of relations, the entities must appear to move steadily apart in space to maintain the illusion of energy conservation.'

I am sure you can think of more imaginative and convincing models yourself and in this way investigate the nature of space (and time).

4.6 Conclusion

It is very difficult to adapt normal physics to include the expansion of the universe; specifically, ensuring energy conservation on the global and local level is challenging. There is no evidence that the recognised ΛCDM model manages this, as it only really works if the detail is ignored and one deals with a hypothetical smoothed out universe which is clearly not the observed universe. Varying G in a particular way is a promising alternative approach[22], but it is likely that a viable theory incorporating this principle can only be developed by gaining a better understanding the link between space-time and gravitation, a deep link which one presumes must exist.

References

1. H. Hoavo, 2020, 'Accelerating universe with decreasing gravitational constant', Journal of King Saud University – Science 32, 1459-1463
2. A. Paliathanasis, 2018, 'Dynamical analysis and cosmological viability of varying G and Λ cosmology', Eur. Phys. J. C 78, 681
3. C. Rubano and P. Scudellaro , 2005, 'Variable G and Λ: scalar-tensor versus RG-improved cosmology', General Relativity and Gravitation 37, 521–539

---

[22] There are other possibilities, such as allocating energy to the expansion.

4. D.-J. Wu, Y.-S. Piao, Y. Cai, 2013, ‘Dynamical analysis of the cosmology of mass-varying massive gravity’
5. C. H. Brans, 2010, ‘Varying Newton’s constant: A personal history of scalar-tensor theories’ in: Einstein Online Band 04, 04-1002
6. S. Chakraborty, U. Debnath, M. Jamil and R. Myrzakulov, 2012, ‘Statefinder Parameters for Different Dark Energy Models with Variable G Correction in Kaluza-Klein Cosmology’, International Journal of Theoretical Physics 51, 2246–2255.
7. J. V. Narlikar, 1983, ‘Cosmologies with variable gravitational constant’, Foundations of Physics 13, 311–323
8. N. Aghanim, T. Ghosh, J.J. Bock, B. Crill, O. Doré, S. Hildebrandt, and G. Rocha, 2018, ’Planck 2018 results. VI. Cosmological parameters’, https://arxiv.org/abs/1807.06209
9. A. Riess, S. Casertano, W. Yuan, L. Macri and D. Scolnic, 2019, ’Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics Beyond LambdaCDM’, https://arxiv.org/abs/1903.07603
10. P. Vickers, 2009, ’Was Newtonian cosmology really inconsistent?’, Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics, 40, 197-208
11. J. P. Vinti, 1977, ‘Newtonian cosmology with a varying gravitational constant’, Celestial mechanics 16, 391–406
12. A. Dev, M. Safonova, D. Lohiya and D. Jain, 2002, ‘Cosmological Tests for a Linear Coasting Cosmology’, Physics Letters B 2, 814-9
13. M. V. John, 2016, ‘Realistic coasting cosmology from the Milne model’, 1610.09885 [astro-ph.CO]
14. Benoit-Lévy, A., Chardin, G.: Introducing the Dirac-Milne universe. Astron. Astrophys. 537, A78 (2012)
15. F. Melia and A. S. H. Shevchuk, 2012, ‘The $\mathrm{R_h}$=ct universe’, Mon. Not. R. Astron. Soc. 419, 2579-2586
16. A. Mitra, 2014, ‘Why the $\mathrm{R_h}$ = ct cosmology is unphysical and in fact a vacuum in disguise like the Milne cosmology’, Mon. Not. R. Astron. Soc. 442, 382-387
17. G. F. Lewis, 2013, ‘Matter matters: unphysical properties of the $\mathrm{R_h}$ = ct universe’, Mon. Not. R. Astron. Soc. 432, 2324–2330
18. P. A. M. Dirac, 1937’, ‘The Cosmological Constants’, Nature 139, 323
19. J. D. Barrow, 1996, ‘Time-varying G’, Mon. Not. R. Astron. Soc. 282, 1397-1406
20. J. D. Barrow, 1998, ‘Varying G and Other Constants’, In: Sánchez N., Zichichi A. (eds) Current Topics in Astrofundamental Physics: Primordial Cosmology. NATO ASI Series (Series C: Mathematical and Physical Sciences), vol 511. Springer, Dordrecht
21. C. M. Will, 2014, ’The Confrontation between General Relativity and Experiment’, Living Reviews in Relativity 17, 4
22. M. Milgrom, 1983, ’A modification of the Newtonian dynamics: implications for galaxy systems’, Astrophys. J. 270,384-389
23. G. W. Angus and S. S. McGaugh, 2008, ‘The collision velocity of the bullet cluster in conventional and modified dynamics’, Mon. Not. R. Astron. Soc. 383, 417–423
24. R. B. Tully, H. M. Courtois, A. E. Dolphin, J. R. Fisher, P. Héraudeau, B. A. Jacobs, I. D. Karachentsev, D. Makarov, L. Makarova, S. Mitronova, L. Rizzi, E. J. Shaya, J. G. Sorce and P.-F. Wu, 2013, ‘Cosmicflows-2: The Data’, The Astronomical Journal 146, 25
25. L. Jounghun and E. Komatsu, 2010, ‘Bullet Cluster: A Challenge to ΛCDM Cosmology’, The Astrophysical Journal 718, 60-65

26. M. Markevitch, A. H. Gonzalez, L. David, A. Vikhlinin, S. Murray, W. Forman, C. Jones and W. Tucker, 2002, 'A Textbook Example of a Bow Shock in the Merging Galaxy Cluster 1E 0657-56' Astrophys. J. 567, L27-L31
27. W. Berkson, 1974, 'Fields of Force: The Development of a World View from Faraday to Einstein', Routledge and Kegan Paul LONDON
28. Right. Hon. Lord Kelvin, 1901, 'I. Nineteenth century clouds over the dynamical theory of heat and light', The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 2:7, 1-40

5 A Universe Stretching at Light Speed

Summary

A very simple cosmology is described where the universe expands at the speed of light. This model has a luminosity distance-redshift relation which differs from that in ΛCDM cosmology, with the functions steadily diverging beyond a redshift of 2. The difference is huge once the redshift exceeds 1,000, and, in theory, an effective comparison of the two can be made from the way the functions converge on the microwave background data if the cosmic background is treated as a mesh of point sources.

5.1 Introduction

The universe is effectively described by the ΛCDM model, and the success of the model has almost completely banished rival cosmologies with the result there is now little talk about alternative descriptions of the universe. But the consensus model did not just appear as a fully formed package – there has been significant evolution over time, with major modifications introduced to fine-tune its predictions to match emerging data. Over the last 50 years, observational techniques have dramatically improved and this has forced a series of changes that have culminated in the consensus model (a model that cannot yet be considered complete [1]). Whilst the research focus is naturally on trying to understand what the model is telling us about the universe, an evaluation of rejected theories from the past against the current data set could offer additional insight.

5.2 Historical Background

The discovery by Edwin Hubble almost 100 years ago that light from extra-galactic nebulae is red-shifted in linear proportion to apparent distance was quickly associated with general relativity (GR) and explained by the model of a finite universe curved under its own gravity and expanding at a rate constrained by the enclosed mass. The model still prevails. At the time, not everyone agreed with this interpretation. Edward Arthur Milne argued that curved space-time and suchlike were unobservable and a consistent model should be presented only in terms of observables. This so-called operational approach presumed space-time to be flat (because it appeared so) and therefore all observations should be subject to special relativity (SR). The redshift of distant nebulae is then required to be a Doppler effect arising from a proper recessional velocity [2]. By rigorously applying SR, the luminosity distance-redshift d_L-z relationship that emerges from the analysis is in surprising good agreement with the current observation of distant standard candles, and is largely consistent with other cosmological data. Of course, standard GR cosmology also predicts a luminosity distance-redshift relationship for standard candles. By comparing supernova Ia observations, the favored ΛCDM model is found to be a better match for the data than Milne's model. Nevertheless, it is curious that two apparently unrelated approaches to cosmology can lead to very similar predictions.

A closer look at the physical model conceived of by Milne is needed. His initial or boundary conditions are unusual. Essentially, all the matter in the universe is thought to be accumulated at a single spatial point at $\tau = 0$[1]. This is the synchronization point for all clocks. A simple aggregation of non-interacting particles is assumed. Matter explodes outwards into preexisting Minkowski space-time, with each particle allocated a random velocity in the range 0 to c. Because space-time is flat, there is none of the acceleration associated with geometric effects. Therefore, in the absence of forces, each particle maintains its initial velocity as the system develops over time. Thus, the particles move apart and separate into expanding shells with radial velocity proportional to the shell radius. From there, Milne developed his theory [3] working on the principle that one can,

> "*attempt a complete reconstruction of physics from the bottom up, on an axiomatic basis.*"

The full model conceived by Milne is highly imaginative, essentially a finite bounded universe that could be made to hold an infinite quantity of matter by a neat trick. He recognized that greater and greater quantities of matter could be packed into space with increasing recessional velocity by exploiting the Lorentz contraction effect.

It is common now to identify Milne cosmology with a variant of standard cosmology where the total mass is zero and the curvature is negative, but such a mapping is misleading.

5.3 Updating the Milne Model

Milne cosmology is incorrect for a number of reasons, both practical and aesthetic[2]:

- The initial explosion is contrived and without explanation. The distribution of fragment velocities would need to be implausibly fine-tuned to give an isotropic homogeneous universe from any observation point.
- The flat canvas of special relativity on the large scale is incompatible with gravity. Gravity cannot be ignored when objects are moving in the manner described. How then is are the fragments able to retain their initial velocity against the braking effect of gravity? Energy conservation violations are usually fatal to a theory and cannot be ignored. This is especially the case with SR which is built on the sound foundation of energy conservation.

---

[1] To avoid confusion with temperature, τ will be used to indicate cosmological time for explanatory purposes, though it is unclear if a cosmological time exists in SR cosmological models. In SR cosmology, all entities have their own differing universal times, and everything becomes relative, even the expansion of the universe.

[2] This is not to detract from the impressive intellectual achievement of Milne in building an comprehensive world structure from such simple ideas by pure thought alone [4].

- Where did this preexisting infinite space-time come from? In special relativity, each observer is presented with a completely consistent view of the universe. However, it is not possible to blend all these views into an underlying ‘real’ model of the universe. The universe may exist as a ’thing’ but there no practical procedure by which it can be revealed. It is difficult to see how a physical model can be linked to special relativity.
- Emerging from the theory is two time scales, a ’normal’ time and a cosmic time. A separate cosmic time is clearly needed in order to attach to an SR based cosmology a world model, but there is no evidence that cosmic time exists separately from clock time[3].

But there are also positive aspects to the theory that fit well with observation: a flat space-time; an expansion velocity of c (or close to it); a natural explanation for Hubble’s Law. The apparent recession velocity of distant cosmological objects is given by

$$v_h(\tau) = \frac{r(\tau)}{\tau} = H(\tau)r(\tau), \tag{5.1}$$

where r is the proper distance and τ is the time on the observer’s clock running from the beginning of the universe. It evident that the recession velocity remains constant as the system develops over time. Equation 1 is largely consistent with observational data.

If the Milne model is pared back to the basics, including the removal of the model but retaining energy and momentum conservation, a perfectly respectable mathematical description of the developing universe is obtained that can be tested against observable data. This can be labelled ’SR cosmology’ and describes an expanding universe where the recession effect can be treated as a velocity, with all entities causally connected and the furthest point receding from the observer at the speed of light.

Energy conservation is ensured by recognising that the expansion is not a normal physical process and can therefore be exempt from the requirement that it should do work against gravity.

By rejecting Milne’s interpretation but retaining the methodology, the largely discredited model will be separated from the analysis and a d_L-z function is obtained that can be compared to observational data.

5.4 The Luminosity Distance - Redshift Relation

The measured flux from an object of luminosity L moving away from an observer at velocity v is given by

$$f = \frac{L}{4\pi(1+z)^4 x_e^2} = \frac{L}{4\pi d_L^2}. \tag{5.2}$$

---

[3] Though the obvious question then is how cosmic time works its way into equation 1?

This is a standard exercise whose derivation can be found in many special relativistic textbooks [5]. The redshift is

$$1 + z = \sqrt{\frac{1 + \frac{v}{c}}{1 - \frac{v}{c}}}, \tag{5.3}$$

and x_e is the separation distance at the time of emission (τ_e) in the frame of the observer.

The formula applies to any emitter in Minkowski space-time and the distinct features of SR cosmology only enters into the calculation through this characteristic property of the expansion:

$$x_e = v_h \tau_e, \tag{5.4}$$

hence

$$\tau_o = \left(1 + \frac{v_h}{c}\right)\tau_e, \tag{5.5}$$

where τ_{o} is the current age of the universe (referring to distant photons received now). Assuming no peculiar velocity ($v = v_h$), and using equation 3 [6],

$$d_L^{SR} = c\tau_o\left(z + \frac{z^2}{2}\right) = \frac{c}{H_o}\left(z + \frac{z^2}{2}\right). \tag{5.6}$$

This is the luminosity distance - redshift relation d_L-z in SR cosmology.

In the ΛCDM model, the size of the universe is indicated by a scale factor; by default the current scale factor, a_o, is 1. The redshift is not a Doppler effect but instead arises from the stretching of the light wave as the photon moves from emitter to observer. The increased wavelength is equivalent to energy loss giving rise to an observed redshift effect:

$$\frac{a_o}{a} = \frac{\lambda_o}{\lambda_e} = 1 + \frac{\Delta\lambda}{\lambda_e} = (1 + z). \tag{5.7}$$

It is easily shown that this differs from the Doppler redshift if z is large. In the SR model, the same separation distance is registered as a recession velocity of v_h:

$$\frac{a_o - a}{a} = \frac{v_h}{c} = \frac{z}{1 + z}. \tag{5.8}$$

The SR redshift is then (for small z:

$$1 + z' = \sqrt{\frac{1 + \frac{v_h}{c}}{1 - \frac{v_h}{c}}} = \sqrt{1 + 2z} \approx 1 + z. \tag{5.9}$$

The flux - luminosity relationship in ΛCDM cosmology is

$$f = \frac{L}{4\pi(1 + z)^4 X_e^2} = \frac{L}{4\pi D_L^2}. \tag{5.10}$$

This is identical to the SR relation except that the separation distance at emission is different (hence the apparent brightness will differ). Upper case X and D have been used to distinguish the ΛCDM values. X_e is the angular diameter distance, the proper distance at the time the light left the emitting object (at least in flat space-time).

The luminosity distance for a flat universe is [7]

$$D_L^{\Lambda CDM} = (1{+}z)\frac{c}{H_o}\int_0^z \frac{dz'}{\sqrt{\Omega_m(1+z')^3 + (1-\Omega_m)(1+z')^{3(1+w)} + \Omega_R(1+z')^4}} \tag{5.11}$$

where Ω_m and Ω_R are the matter and radiation densities respectively, and w is the equation of state parameter.

The equations for the two cosmologies are overlaid in Fig. 1 to make a comparison for high z values. For the ΛCDM model, values of w = -1, Ω_m = 0.266 and Ω_K = 0 have been chosen. The SR model has no free parameters.

Whilst there are differences at low z it seems that the best place to compare the two is at high z, and there is adequate data available from the cosmic microwave background radiation where the z value is believed to be about 1090.

Of course, the luminosity distance - redshift relation is normally applied to point sources, but it is possible to think of the surface of last scattering as mesh of small emitters, with the size tending towards zero.

5.5 The Luminosity of the Cosmic Microwave Background Radiation

A spherical surface of radius R can be imagined around a portion of the emitting surface to create a virtual source. Let the intrinsic temperature be T. The luminosity is

$$L = 4\pi R^2\sigma T^4 \tag{5.12}$$

where σ is the Stefan-Boltzmann constant. Consider first the SR observer who will perceive an emitter with emission area of πR^2 because coordinates perpendicular to the direction of motion are not affected by the Lorentz transformation from the emitter to the observer frame. The apparent flux is therefore

$$f = \frac{R^2}{x_e^2}\sigma\left(\frac{T}{1+z}\right)^4 . \tag{5.13}$$

But the flux at the position of the observer is the total from the entire sky filled with a mesh of identical emitters. The total energy density is therefore

$$U = \frac{1}{c}\frac{4\pi x_e^2}{\pi R^2}\frac{R^2}{x_e^2}\sigma\left(\frac{T}{1+z}\right)^4 == \frac{4}{c}\sigma\left(\frac{T}{1+z}\right)^4 . \tag{5.14}$$

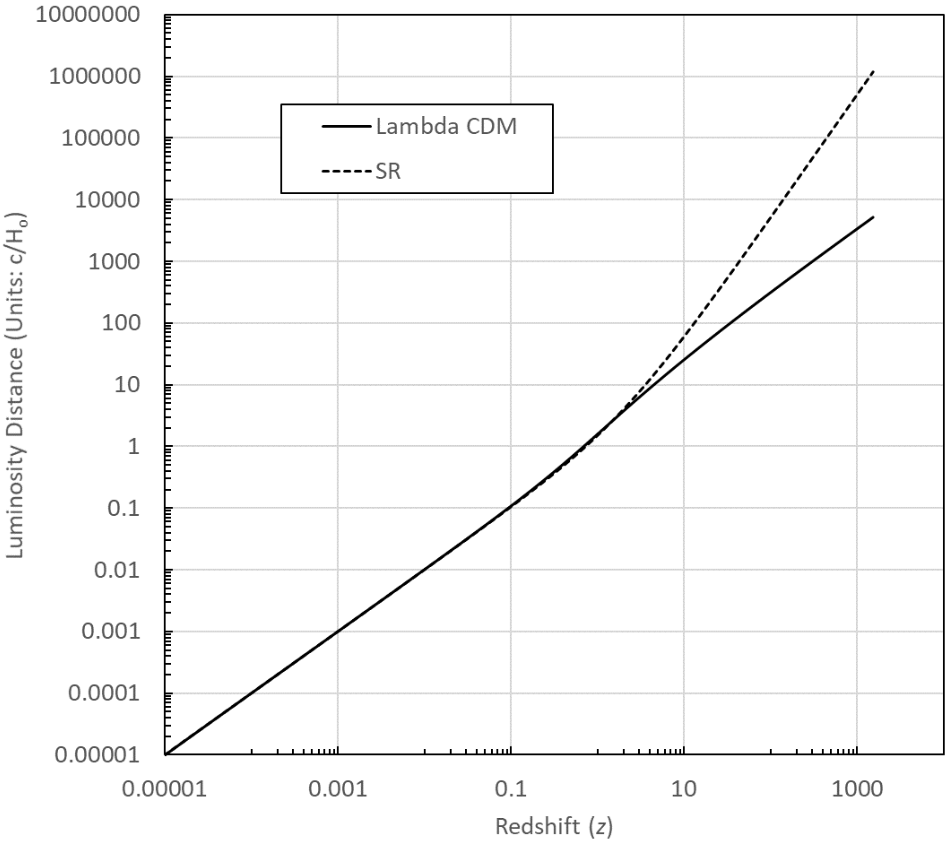


Fig. 1. The luminosity distance redshift relation for both SR and ΛCDM models are shown. For small z values the curves are similar but diverge dramatically at higher z values.

The observer appears immersed in a blackbody of much lower temperature, $T/(1 + z)$. This is consistent with observation satellites detecting a *local* blackbody spectrum corresponding precisely to a temperature 2.725 K.

Note that the blackbody shape is retained over time, which means the source temperature cannot be deduced from the data - the ultimate source could equally well be 5,700 K, the temperature of the surface of the sun (another blackbody). Selecting 1,090 for the redshift value gives a misleading idea of the accuracy with which the emission source is known. Taking the argument to its conclusion, it may not be correct to associate the radiation with a surface - it has depth developing over time to the present day and no obvious source.

In addition, it is not just the total flux that scales - individual wavelengths also scale correctly. The spectral radiance at the point of emission is

$$B(\lambda, T)d\lambda = 2hc^2 \frac{d\lambda}{\lambda^5} \frac{1}{e^{\frac{hc}{\lambda kT}} - 1}. \tag{5.15}$$

Received is

$$Bd\lambda = 2hc^2 \frac{(1+z)d\lambda}{(1+z)^5\lambda^5} \frac{1}{e^{\frac{hc}{(1+z)\lambda kT}} - 1} = 2hc^2 \frac{1}{(1+z)^4} \frac{d\lambda}{\lambda^5} \frac{1}{e^{\frac{hc}{\lambda k(1+z)T}} - 1}. \tag{5.16}$$

The peak has moved to $T/(1 + z)$ and the amplitude has decreased by the correct amount.

Switching now to ΛCDM, the flux from the single source is

$$f = \frac{R^2}{X_e^2} \sigma \left(\frac{T}{1+z} \right)^4 \tag{5.17}$$

and when the entire sky is included, the result is

$$U = \frac{4}{c} \sigma \left(\frac{T}{1+z} \right)^4. \tag{5.18}$$

Both flux equations transform the blackbody radiation in the manner observed. However the luminosity distance is not the same. Using $z = 1090$, the results shown in Table 1 are obtained.

Parameter	SR	ΛCDM ($\Omega_\mathrm{m} = 0.266$)
Luminosity Distance (c/H_o)	594,000	3,658
Angular Diameter Distance (c/H_o)	0.4999996	0.0030797
Emission Time (million years from $\tau = 0$)	12.7	43.6

Table 1 The cosmic microwave source in SR cosmology and ΛCDM model.

The ΛCDM emission time was found by performing a numerical integration. The light travel time, t_γ,

$$t_\gamma^{\Lambda CDM} = \frac{1}{H_o}\int_0^z \frac{dz'}{(1+z')\sqrt{\Omega_m(1+z')^3+(1-\Omega_m)(1+z')}} \tag{5.19}$$

is subtracted from the current cosmological time (13.8 billion years) using $\Omega_{\mathrm{R}} = 0$ and w = -1. The angular diameter distance (42.3 million ly) is usually thought of as the physical distance, and is slightly less than the absolute emission time multiplied by the speed of light.

However, the emission time is not normally calculated in this way but found instead by solving the Friedmann equations for a flat matter dominated universe. The age as a function of redshift is then given by the formula

$$\tau_e^{\Lambda CDM} = \frac{2}{3H_o\sqrt{\Omega_m}}(1+z)^{-3/2} \tag{5.20}$$

which predicts an emission time 0.495 million years from the time of the big bang (using the same parameter values as before). This is clearly inconsistent, but can be rationalised - see for example Cahill [8]. ΛCDM does not have the same limitations as SR cosmology and generates unusual horizons[4].

So which of the two relations is correct? Unfortunately it is not possible to say because the proper position of the cosmic background at the time of photon emission is unknown. But neither theory is excluded either. The fact that both flux relations (equations 2 and 10) are valid is a consequence of Etherington's reciprocity theorem [17] which states the luminosity distance is $(1 + z)^2$ times the proper distance at emission time in the frame of the observer, a relationship true for any cosmology.

What is significant from the analysis is the difference in luminosity distance predicted by the functions. In ΛCDM the luminosity distance can be readily modified to match low-z observational data by adjusting the parameters available, and a match with high-z cosmological data can be obtained by allowing these parameters to vary over cosmological time. It is therefore not possible to invalidate the ΛCDM model; it simply becomes less plausible as more corrections are applied.

5.6 Energy Loss by Cosmological Photons

One important difference between the two approaches is how energy conservation applies to cosmological photons. In 1930, Hubble wrote [18],

> " *...redshifts, by increasing wavelengths, must reduce the energy in the quanta. Any plausible interpretation of redshifts must account for the loss of energy.*"

---

[4] Davis and Lineweaver have attempted to explain the relation between different distance measures in cosmology in an influential paper [16]. However, it should be born in mind that Sections 3.1 and 4 of that publication, referring specifically to special relativity, are inaccurate.

In fact, no plausible interpretation has since emerged and it is generally accepted in conventional cosmology that energy is not conserved.

However, that is not the case in the SR interpretation. The transfer of photons is associated with the normal Doppler effect which does conserve energy. If an atom emits a photon of energy ϵ, the distant absorbing atom will register this as $\epsilon/(1 + z)$ in their rest frame. The change in energy associated with the transaction at the absorber (when viewed from the emission frame) is therefore

$$\Delta E = \frac{pc^2 dp + m_0 c^4 dm_o}{E}, \tag{5.21}$$

where E is the original system energy and recoil has been ignored. Because $dm_{\mathrm{o}} = \epsilon(1 + z)^{-1} c^{-2}$, $dp = \epsilon c^{-1}$, $p = \gamma m_{\mathrm{o}} v$, and $E = \gamma m_{\mathrm{o}} c^2$, it follows that $\Delta \mathrm{E} = \epsilon$. Energy is conserved. The apparent loss of energy is just an artefact of the frame change from emitter to absorber.

The same explanation does not work with the cosmological redshift in ΛCDM because photons are stretched, losing energy as they progress. It is a dynamic effect. In contrast, the Doppler effect is kinematic, essentially a transaction, a function only of the emitter and absorber 4-momenta. In SR, it is unnecessary to describe photons as travelling through the medium; they are completely characterised by the emission and reception events. If we try to replicate the calculation above, the momentum on reception in ΛCDM is $dp = \epsilon(1 + z)^{-1} c^{-1}$ and clearly energy is no longer conserved.

In SR, if we consider a sequence where a photon is absorbed then the received energy is immediately emitted along the same propagation line, the received energy will gradually fall in a consistent way because the distant photons have increasing recession velocity with respect to the original emitter. On the other hand, if two receding particles bat a photon to and fro in the manner of a mirror, the total energy will fall but only because the reaction force converts some of the photon energy into peculiar velocity. This is a standard exercise in special relativity [19][5].

Although energy is conserved in SR, the energy emitted in the emission rest frame cannot be deduced from the energy of the photons in transit. It is therefore very difficult in SR to answer questions relating to the total radiation energy in the universe[6].

---

[5] Interestingly, cosmological photons do not increase the entropy of the universe. The change in entropy associated with a cosmic microwave photon observed now is

$$\Delta S = \frac{\epsilon}{T_e} - \frac{\epsilon/(1+z)}{T_o} = \frac{\epsilon}{T_e} - \frac{\epsilon/(1+z)}{T_e/(1+z)} = 0.$$

[6] The idea of stretching space as something distinct from recession velocity is not accepted by all astronomers. Bunn and Hogg write [28]:

> From the point of view of general relativity, the 'stretching of space' explanation of the redshift is quite problematic. Light is of course governed by the Maxwell equations (or their general relativistic generalization), which contain no 'stretching of space term' and no information on the current size of the Uni-

5.7 Discussion

The analysis of SR cosmology and ΛCDM cosmology using the cosmic microwave background has not enabled the theories to be distinguished, but it did highlight some important differences.

The SR universe at 3,000 K is older than expected, and this is problematic - there is less time for material to organise in order for structure to form. In addition, the total photon energy is very high. Table 2 shows the result of a basic energy audit (starting from Gamow's original idea of a primordial cluster of neutrons [29]). The sums do not add up.

Parameter	Value	Units
Temperature	3,000	K
Number of baryons	10^{79}	
Universe volume	7.2 x 10^{69}	m^3
(-) Photon energy	4.4 x 10^{68}	J
(-) Thermal kinetic energy	6.0 x 10^{58}	J
(-) Neutrino / antineutrino energy (est.)	8.0 x 10^{63}	J
(+) Neutron decay energy	1.1 x 10^{66}	J
(+) Fusion energy (26% He)	4.7 x 10^{67}	J
(+) Electron binding energy in hydrogen atom (max.)	1.6 x 10^{60}	J

Table 2 The energy balance in the universe @ 3,000 K.

It is unclear what the source of all the radiation energy could be. Because the SR approach is so different to that in ΛCDM, none of the work that has been done in establishing the source of the microwave anisotrophy peaks is applicable and a completely new analysis is required. For example, in ΛCDM inflation at an early age is an important feature but this is not a necessity in SR because the universe is by definition flat[7].

> verse. On the contrary, one of the most important ideas of general relativity is that space-time is always locally indistinguishable from the (non-stretching) space-time of special relativity, which means that a photon does not know what the scale factor of the Universe is doing.
>
> The emphasis in many textbooks on the stretching-of-space-time interpretation of the cosmological redshift causes the reader to take far too seriously the stretching-rubber-sheet analogy for the expanding Universe. For example, it is sometimes stated as if it were obvious that 'it follows that all wavelengths of the light ray are doubled' if the scale factor doubles.
>
> While this statement is true, it is certainly not obvious. After all, solutions to the Schrodinger equation, such as the electron orbitals in the hydrogen atom, don't stretch as the Universe expands, so why do solutions to the Maxwell equations?

[7] The SR model could have had the equivalent of an inflationary phase where the growth of the universe accompanied the particle creation process, and following that

Though the focus has been on a high-z comparison, it should not be forgotten that there is a mismatch between the luminosity distance - redshift relations at low z as well. Though the difference is small it is significant. But the ΛCDM model has been continuously modified to match the data, but the SR function has never changed. Though SR has no free parameters that can be adjusted, the predictions can be modified by accounting for possible systematic peculiar velocities. Equations 4 - 6 are invalid when there are peculiar velocities. Decoupling occurs because redshift is dependent on the instantaneous relativistic sum of peculiar and recession velocities, but the proper distance in the frame of the observer at the time of emission depends the integral of the velocities from $\tau = 0$ to the emission time. However, it is unclear if the velocities required to remove discrepancies between function and observation are consistent with the driving large-scale gravitational forces[8]. In contrast, the ΛCDM model neatly incorporates dark matter and dark energy in such a way that normal established physics in not undermined, which is a huge advantage, but it is unclear what the nature of these substances might be.

One of the biggest differences is how the universe is visualised in SR and ΛCDM. With the SR formalism, there is no underlying model because all special relativity does is ensure that space-time is consistent for any particular observer. It is not possible to blend together the viewpoint of each observer to construct and reveal the hidden or 'real' universe behind it all. There is consistency in that each observer is in causal contact with all other entities at an earlier time, but that is all that can be said. In contrast, the ΛCD model is of a real universe expanding through an additional dimension. The model is accessible, easy to understand and intuitive. It presents an understanding of the structure of the universe which the SR model lacks. However, it suggests that the universe should be curved, which it is not. However, this can be explained with modifications to the dynamics of the early universe, inflation for example.

A significant drawback is that the concept of energy conservation is unclear. It is space that is expanding and the relative motion of otherwise stationary entities should not be considered a velocity. However, that is a problem because the change in scale enters into the Friedmann equations as a velocity, which in turn appears in the energy equation as a non-relativistic kinetic energy. How is this possible? And the recession velocity can exceed the speed of light – how is it possible to incorporate this into any equation?

Overall, both approaches have significant problems and it is possible that neither is an effective description of the universe.

---

phase the 'normal' expansion began. The SR approach has no difficulty with homogeneity as there can be reliance on an initial homogeneity in the particle creation process that is sustained by the continuous causal link between all particles.

[8] A consequence of the expansion and peculiar velocities entering into the flux equation in a different way is that correcting the CMB data for observer motion is non-trivial: transient velocities such as the motion of the earth around the sun need to be treated differently to large scale peculiar motion flows with respect to the expansion frame.

References

1. Lusso E., Piedipalumbo E., Risaliti G., Paolillo M, Bisogni S., Nardini E. and Amati L., 2019, A & A 628, L4
2. Milne E. A., 1940, Ap.J. 91, 129-158
3. Milne E. A., 1933, Z.Astrophysik 6: 1-35
4. Milne E. A., 1935, ‘Relativity, Gravitation and World Structure’, Oxford University Press
5. Misner C. W., Thorne K. S. and Wheeler J. A., 1973, Gravitation, W.H. Freeman, New York
6. Chodorowski M., 2005, Publications of the Astronomical Society of Australia. 22. 287-291
7. Peebles P. J. E., 1993, Principles of Physical Cosmology, Princeton University Press, Princeton
8. Cahill K., 2020, European Journal of Physics, 41, 3
9. Davis T. M. and Lineweaver C. H., 2003, Publ. Astron. Soc. Austral. 21, 1
10. Etherington I. M. H., 1933, Phil. Mag., 15, 761
11. Hubble E., 1936, The Realm of the Nebulae, Yale University Press
12. Gjurchinovski A., 2004, European Journal of Physics. 26. 10.1088/0143-0807/26/4/009
13. Bunn E. F. and Hogg D. W., 2009, Am. J. Phys. 77, 688
14. Alpher R. A., Bethe H. and Gamow G., 1948, Physical Review 73, 803–804

6 The Local Expansion in a Coasting Universe

Summary

There has been some debate around the question of whether or not the cosmological expansion acts down to the scale of the solar system or even to atomic level. The consensus seems to be that it does, but that the effect is overwhelmed and masked by the action of the normal forces. The simplest possible cosmological model is used to reduce the complexity of the problem and evaluate in detail the key ideas, in particular conservation of momentum and energy, and the interaction of peculiar and apparent expansion velocities.

6.1 Introduction

Though the expansion of space on the cosmological scale is well established, it is not certain if the expansion operates on all scales down to atomic level. The general thinking seems to be that it does, but that up to galaxy scales it is overwhelmed by the gravitational and electromagnetic forces. This has largely been confirmed by the varied analyses of the problem [1] - [8]. However, there has been some criticism of the approaches taken [16]. The current status is that the question has not been conclusively answered.

It is proposed that the issue could be clarified and perhaps even resolved by simplifying the problem. In much the same way as the Milne model is frequently used as a reference model in standard cosmology, the issue of the local expansion will be assessed in the context of the simplest possible cosmology, a universe that has been expanding and continues to expand at the speed of light, the so-called coasting cosmology. Refer to John [17] for a full description. Though basic, the model still contains the critical time variation in the Hubble parameter which makes the assessment of the impact of a local expansion tricky in the consensus ΛCDM cosmological model, hence we would expect the effect of the expansion in the complete model to deviate only a little from that in the coasting models. The use of the coasting model is also justified (up to a point) by the fact that the reciprocal of the current value of the Hubble parameter is a good approximation for the age of the universe; hence the result of this simplified analysis is broadly transferable to the more complex models. The advantage of this approach is that the effect of the expansion on local dynamics, and how it enters into the equations of motion, is brought to the fore and is clearly seen.

In a coasting universe, gravity is assumed to have no effect on the expansion rate, though in all aspects energy is conserved.

6.2 The Expansion in the Absence of Gravity

The simplest possible scenario is a set of particles so far apart that gravitational effects become insignificant, and whose relative motion arises only from the expansion[1]. The apparent velocity in the coasting model is given by Hubble's law:

$$v_h(T) = Hr \approx \frac{r}{T}, \tag{6.1}$$

where v_h is the expansion velocity, r is the proper radial distance, H is the Hubble factor and T is the absolute cosmic time. Note that the current Hubble factor value is labelled H_o (with a corresponding T_o).

Focusing on one pair of particles, it is clear that the apparent expansion velocity over time is conserved and there are no momentum or energy violations: after time interval ΔT, the proper position changes,

$$v_h(T + \Delta T) = \frac{r + v_h(T)\Delta T}{T + \Delta T} = v_h(T). \tag{6.2}$$

However, the situation is more complex when peculiar velocities are introduced. A peculiar velocity $\boldsymbol{v}_p$ is the velocity relative to the expanding frame, and the most natural interpretation is that it should carried by the expansion without being affected by the expansion. This is shown in Fig. 1(a) for an arbitrary peculiar velocity. The total measured velocity is expected to be the sum of these two velocities (added relativistically, though a direct sum is an acceptable approximation with the small velocities being considered here).

Using polar coordinates in the plane of motion and noting that $v_p(T+dt) = v_p(T)$ it follows that:

$$\frac{dv_{p\theta}}{dt} = \frac{dv_\theta}{dt} = -v_{pr}\frac{d\theta}{dt}\left(= -\frac{v_{pr}v_\theta}{r}\right); \tag{6.3}$$

$$\frac{dv_{pr}}{dt} = v_{p\theta}\frac{d\theta}{dt}\left(= \frac{v_\theta^2}{r}\right); \tag{6.4}$$

$$\frac{dv_h}{dt} = \frac{v_{pr}}{T}; \tag{6.5}$$

$$\frac{dv_r}{dt} = v_{p\theta}\frac{d\theta}{dt} + \frac{v_{pr}}{T}\left(= \frac{v_\theta^2}{r} + \frac{v_{pr}}{T}\right). \tag{6.6}$$

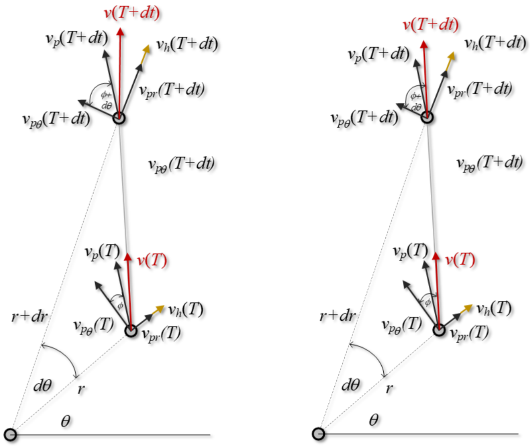
Energy, linear and angular momentum are not conserved unless the peculiar and the apparent expansion velocities are somehow separated in the kinematics.

An alternative approach is to demand that energy and momentum are conserved. The velocity vectors change as shown in Fig. 1(b). The corresponding equations are:

$$\frac{dv_{p\theta}}{dt} = -v_{pr}\frac{d\theta}{dt} - \frac{v_{p\theta}}{T}; \tag{6.7}$$

---

[1] Though the expansion is not necessarily a velocity, it can be treated as such in an operational sense.



(a) $v_p(T)$ is constant (b) $v(T)$ is constant

Fig. 1. A body with a mix of peculiar and expansion velocities observed from a reference point (bottom left). The incremental change over time is exaggerated for clarity and it is assumed from this point on that $dT = dt$, i.e. changes in clock and cosmic time are the same (or almost the same). Vector diagram (a) is based on a constant peculiar velocity, unaffected by the expansion. Diagram (b) on the right presumes the total velocity is constant in the absence of forces.

$$\frac{dv_{pr}}{dt} = v_{p\theta}\frac{d\theta}{dt} - \frac{v_{pr}}{T}. \tag{6.8}$$

Essentially, accompanying peculiar velocity is a virtual force that steadily converts between peculiar velocity and expansion velocity regardless of the direction of motion with respect to the expansion. This is a known effect sometimes referred to as Hubble drag (hd)[2]:

$$\left.\frac{d\boldsymbol{v}_p}{dt}\right|_{hd} = -\frac{\boldsymbol{v}_p}{T}. \tag{6.9}$$

In this scenario, the peculiar velocity is modified to compensate for expansion effects to avoid energy and momentum conservation problems, but the expansion is unaffected. The change in peculiar velocity is potentially detectable locally. An alternative approach is to propose that the expansion is modified to ensure energy is conserved. This has the advantage of no local energy violation and ensures

---

[2] Equation 9 takes the form of a standard damping equation where kinetic energy is converted to another form of energy (such as heat if there is friction). The damping effect is not a specific feature of the coasting model, but is a present in general relativistic cosmological models (refer to [18], [19]), where it may take on a different form if comoving distances are used instead of proper distances- working with proper position and velocity in cosmological time instead gives rise to equation 9.

global energy conservation. Referring back to Fig. 1(a), this essentially requires that v_h should be constant. But a significant consequence is that Hubble's law will no longer hold.

The analysis brings into sharp focus the problem of incorporating the expansion of the universe into standard physics. Whilst it is true that gravitation adds somewhat to the problem (as we will see in the next section), it does not initiate it. The problem is already seen when constant peculiar velocities are introduced into the expanding universe. It is only possible to retain both Hubble's law and energy conservation by the introduction of a mechanism for shifting the balance between peculiar and expansion velocities whilst keeping the total velocity constant. If that is really what is happening, the physical process driving this mechanism is unknown. The result is a gradual change in the peculiar velocity over time, and whilst this can be inferred from multiple measurements by various observers, it is uncertain whether or not it is directly measurable. Certainly it is the best of the possible ways of binding in the expansion; the alternatives that were looked at are less attractive: the idea of the expansion mechanism somehow being changed for balancing purposes would need a complex reworking of space-time *for each observer* in a situation where there are many entities moving at different velocities; and the idea that the expansion velocity should be excluded from the energy balance is not credible as it requires the universe at large to be completely different to the local universe with which we are familiar.

6.3 Central Gravitational Force

A mass subject to a weak gravitational force is drawn towards a central mass, M. Let us first assume the expansion acts independently of the gravitational effect and determine the effect of the expansion on the orbital path. Equations 3 and 6 are modified as follows to include the radial gravitational acceleration:

$$\frac{dv_\theta}{dt} = -v_{pr}\frac{v_\theta}{r}; \tag{6.10}$$

$$\frac{dv_r}{dt} = -\frac{GM}{r^2} + \frac{v_\theta^2}{r} + \frac{v_r}{T} - \frac{r}{T}. \tag{6.11}$$

For stable orbits, angular momentum has to be constant:

$$\frac{d}{dt}\left(rv_\theta\right) = 0 \implies v_r = v_{pr}. \tag{6.12}$$

This is clearly not possible as the equality holds only when the expansion velocity is zero, i.e. no expanding universe. The effect on orbits is not immediately obvious, but can be illustrated on the galactic scale in the present epoch in Fig. 3.

The energy violation per orbital cycle is small but not negligible, and no energy conservation violation can be ignored; it is normally indicative of a misunderstanding of the physics of a process.

Turning now to the scenario where Hubble drag is included, and starting off with a net radial velocity of zero:

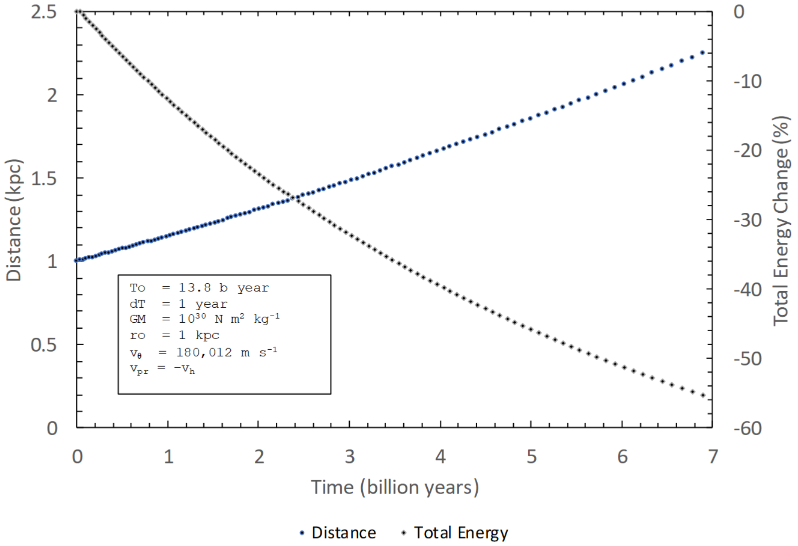


Fig. 2. The change in energy and position for an orbiting body is plotted for each orbital cycle from the current epoch with the initial parameters shown. Data from 112 orbital revolutions are plotted.

$$v_{pr}(0) = -\frac{r}{T}; v_r(0) = 0. \tag{6.13}$$

Slotting this into equation 7,

$$\frac{dv_{p\theta}}{dt} = \frac{r}{T}\frac{d\theta}{dt} - \frac{v_{p\theta}}{T} = 0; \tag{6.14}$$

This defines a circular orbit, which is stable.

We can generalise: differentiating rv_θ with respect to time demonstrates that angular momentum is conserved for any orbit, as is energy:

$$\frac{dE}{dt} = \frac{d}{dt}\left(\frac{1}{2}v_\theta^2 + \frac{1}{2}v_r^2 - \frac{GM}{r}\right) = v_\theta\frac{dv_\theta}{dt} + v_r\frac{dv_r}{dt} + \frac{GM}{r^2} = 0, \tag{6.15}$$

In addition, if $v_\theta = 0$, then the sum of the peculiar and expansion velocities are accelerated by gravity. It is evident that the expansion can be incorporated into standard physics through the mechanism of Hubble drag, but it is also necessary to check if it compatible with cosmological models at the outer distance limits.

6.4 Cosmological Considerations

If the expansion is treated as a velocity then the gravitational force acts on the total velocity in a way that is operationally consistent. The expansion is not affected by the process and has 'priority', i.e. the peculiar velocity is modified to ensure the total velocity correctly follows the acceleration with the expansion velocity always following Hubble's law.

In the cosmological context, the transverse velocity (v_θ with respect to the observer) is of lesser importance and we can see that the key equation is

$$\frac{dv_{pr}}{dt} = \frac{-GM}{r^2} - \frac{v_{pr}}{T}. \tag{6.16}$$

If we set $v_{pr} = 0$ as well, the effect of the acceleration is to pull on a particular mass, and for any single observer, the dynamics are consistent. However, if we look at the same situation from another vantage point, this mass can develop a peculiar velocity with respect to the expansion that is inconsistent with the original viewpoint. In fact, the only way to achieve consistency is to propose that the expansion is physically slowed by the masses moving apart (to negate the effect of the expansion in a direct way).

However, care is required here. The problem only arises because we introduced a privileged 'all-seeing' observer who is outside the system and is able to see the same interaction from multiple viewpoints. No such observer can exist in the real world, hence the requirement for consistency that has been introduced is quite artificial[3].

Nevertheless, a foundational principle of general relativistic cosmology is that the the expansion of the universe is slowed by the action of the mass within. An interesting challenge is to implement in a way that it emerges from a local effect - there is no natural way of incorporating it into the equations discussed above without disrupting Hubble's law in a way that would be observable.

But even more complex ways of handling the expansion are possible. One cosmological curiosity is the flat rotation curves of spiral galaxies [28]. This can be explained by invoking hidden dark matter, but gravity can also be modified in a number of ways with ideas such as MOND being investigated [29]. More radical ideas such as G being a function of time are also possible:

$$\frac{dG}{dt} = \frac{G}{T}, \tag{6.17}$$

then objects in a circular orbit will technically maintain constant energy rotation speed as the orbit grows with the expansion. But it is difficult to see how a constant transverse velocity is maintained in practice for a total radial velocity of $v_r = r/T$ because of the associated transverse acceleration which will tend to reduce the orbital speed over time:

$$\frac{dv_\theta}{dt} = -\frac{v_\theta}{T}. \tag{6.18}$$

One source of the torque needed to compensate is retardation in the gravitational effect. If there is a lag of r/c, the line of force is shifted by an angle of v_θ/c. The associated transverse acceleration is

---

[3] This is analogous to cosmological redshift photons. For any particular observer energy is conserved, but from the 'outside' energy appears to be continuously lost.

$$\frac{dv_\theta}{dt} = -\frac{v_\theta}{T} + \frac{GM}{r^2}\frac{v_\theta}{c}. \tag{6.19}$$

The acceleration is zero when

$$\frac{GM}{r^2} = \frac{c}{T}, \tag{6.20}$$

coincidentally similar to MOND threshold acceleration. Though a lag is expected on a cosmic level, note that there is no apparent lag on a small scale. This is not a violation of causality with influences exceeding the speed of light, but is an indication of forces acting on projected rather than retarded positions. Refer to Carlip for a full explanation [15][4]. Refer also to Appendix A.

6.5 The Expansion in Standard Cosmology

It has been shown that by including Hubble drag, the expansion can be incorporated into orbits on the small scale without any observational effects. Circular orbits are particularly easy to analyse and are not affected by the expansion because the change in expansion velocity over time is exactly balanced by a reduction in the magnitude of the balancing radial peculiar velocity. In fact it would be impossible to say if the expansion were acting locally or not.

This can be compared with the equivalent analysis in ΛCDM models. One of the most influential contributions in recent years is the work of Cooperstock, Faraoni and Vollick (CFV) [3]. They considered a two-body gravitational problem with circular motion and showed that the expansion is present but the effect is negligible. The equation of motion developed in that analysis was derived from general relativity (Appendix 1 of [3]);

$$\frac{d^2\boldsymbol{r}}{dt^2} - \frac{\ddot{a}}{a}\boldsymbol{r} = -\frac{GM}{r^2}\hat{\boldsymbol{r}} \tag{6.21}$$

where a is the universe scale factor. The Hubble factor is defined as $\ddot{a}/a$. Recalling from equation 1 that $H = 1/T$ in a coasting cosmology, the following is true:

$$\frac{d}{dt}\left(\frac{\dot{a}}{a}\right) = \frac{\ddot{a}}{a} - \frac{\dot{a}^2}{a^2}. \tag{6.22}$$

Hence,

$$\frac{\ddot{a}}{a} = \frac{dH}{dT} + H^2 = -\frac{1}{T^2} + \frac{1}{T^2} = 0. \tag{6.23}$$

The expansion term disappears from equation 21, and the equation is identical to those discussed above. This implies that the effect of the expansion on dynamics on the small scale is only evident if the expansion rate varies over

---

[4] Locally the gravitational acceleration far exceeds c/T and with a propagation delay there would be no stable orbits.

time, perhaps with phases of positive and negative acceleration. This is also the conclusion of Sereno and Jetzer [7]. One implication is that the way the expansion rate is currently varying from the speed of light is potentially detectable through small-scale dynamic effects (though measuring these very small effects is extremely challenging).

References

1. McVittie G. C.1933. The mass-particle in an expanding Universe. Mon. Not. Royal Astron. Soc. 93 325.
2. Noerdlinger P. D. and Petrosian V.1971. The effect of cosmological expansion on self-gravitating ensembles of particles. Astrophys. J. 168 1.
3. Cooperstock F.I., Faraoni V., and Vollick D.N. 1998. The influence of the cosmological expansion on local systems. Astrophys. J. 503 61.
4. Domınguez A. and Gaite J.2001. Influence of the cosmological expansion on small systems. Europhys. Lett. 55 458.
5. Barnes L. A., Francis M. J., James J. B. and Lewis G. F. 2006. Joining the Hubble flow: implications for expanding space. Monthly Notices of the Royal Astronomical Society, Volume 373, Issue 1, 21 November pp 382–390.
6. Faraoni V. and Jacques A.2007. Cosmological expansion and local physics. Phys. Rev. D 76 063510.
7. Sereno M. and Jetzer P. 2007. Evolution of gravitational orbits in the expanding Universe. Phys. Rev. D 75 064031.
8. Giulini D. 2014. Does cosmological expansion affect local physics? Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics, Volume 46, Part A, pp 24-37.
9. Dumin Y. V. 2016. Local Hubble Expansion: Current State of the Problem. Conference: Cosmology on Small Scales 2016, Prague. Preprint arXiv:1609.01793v1.
10. John M. V. 2019. Rh = ct and the eternal coasting cosmological model. Monthly Notices of the Royal Astronomical Society: Letters, Volume 484, Issue 1,pp L35–L37.
11. Kwan J., Lewis G., and James J. 2010. The Adventures of the Rocketeer: Accelerated Motion Under the Influence of Expanding Space. Publications of the Astronomical Society of Australia, 27(1), pp 15-22.
12. Linder E. V. 2005. Cosmic growth history and expansion history. Phys. Rev. D 72, 043529.
13. Sofue Y. Rubin V. 2001. Rotation Curves of Spiral Galaxies. Annual Review of Astronomy and Astrophysics 2001 39:1, 137-174
14. Bekenstein J. D. 2006. The modified Newtonian dynamics—MOND and its implications for new physics. Contemporary Physics, 47:6, 387-403
15. Carlip S. 2000. Aberration and the speed of gravity. Physics Letters A, 267 (2–3), 81-87
16. Data available from https://spdf.gsfc.nasa.gov/research/visualization/visbard/download.html [Accessed 01/06/2020]

Appendix A: Local Retardation

We can look for a lag in an obviously accelerated system. This might be expected once a critical acceleration is reached. The best data available is from space probes. A number of space probes have reached solar system escape velocity, most notably the Pioneer probes. Contact with Pioneer 10 was lost in 2003 with the spacecraft 10 billion km from earth (and the sun). The approximate gravitational acceleration then was of the order of 10^{-6} m s^{-1}.

If the gravitational force is retarded, the force is felt from where the sun was at time t-r/c; the result will be an apparent increase in acceleration. Using the ViSBARD full trajectory data [21] the expected deviation for Pioneer 10 is plotted in Fig. 3. Whilst there is a suggestion of some sort of anomalous acceleration associated with the spacecraft, the precise signature shown is not present. We can therefore conclude there is no evidence of retardation out to 80 AU.

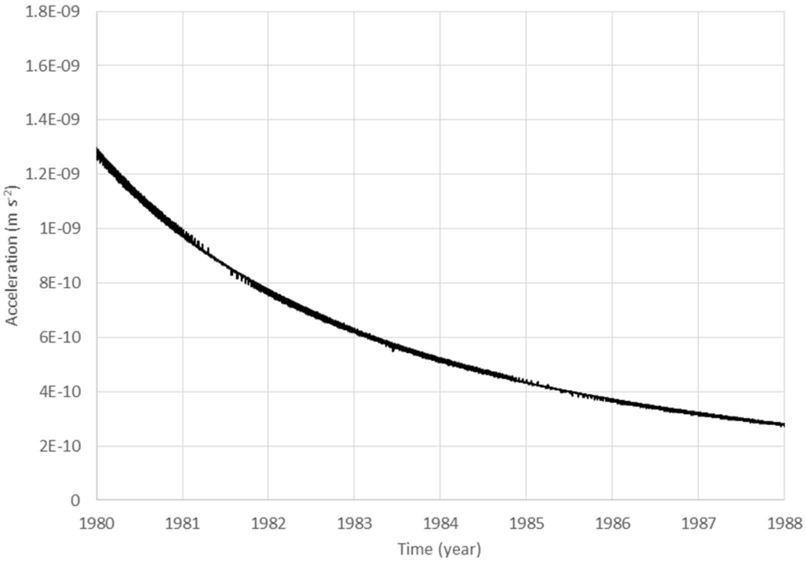


Fig. 3. If the gravitational force acting on Pioneer 10 were retarded, the anomalous acceleration shown would result. The spacecraft would be slowed as a result.

7 The Expanding Balloon Analogy in Cosmology

Summary

The influence of the expanding balloon analogy on the development of the accepted cosmological model is examined. Though many scientists reject this analogy as simplistic, it is nevertheless an effective way of illustrating the way the universe can expand without necessarily having to expand into *something*. The balloon model relates easily to the scale factor that is derived using General Relativity and which expresses the size of the universe as a function of time. It is inferred from the balloon model that space should be curved. However, no global curvature is observed in the real universe, though it is possible to modify the properties of the universe by postulating such things as inflation, dark matter and dark energy to explain the lack of observed curvature. However, is this reasonable: there really is no logical requirement for an expanding universe be curved. Is the analogy being taken too far?

7.1 Introduction

Analogies are commonly used in education to help explain a new idea or phenomenon by referring to a familiar concept which possesses similar attributes. The familiar concept is called the analog and the new concept is the target. A systematic comparison between the key features of the analog and target is called a mapping. Jonāne [1] explains that the use of analogies is a key feature of the learning process and suggests that every learning process includes a search for similarities between what is already known and the new, and that this is part of thinking, everyday speech, and artistic expression.

Dilber [2] assesses the use of analogies in the context of physics education and notes that in some cases features of the analogy that are unlike the target can cause impaired learning, and that the use of analogies in the teaching of science does not always produce the intended effects, especially when students take the analogy too far and are unable to separate the analogy from the content being learned. The message is clear: choose analogies with care because that initial understanding, if faulty, is not easily unlearnt later.

Some analogies in common use can be misleading. For example, electrons in an atom are compared to planets orbiting the sun[1]. This is the Bohr atom, also

---

[1] Around 1600 the German mathematician Johannes Kepler analysed the available data on the position of the planets. He found the planets followed a well-known mathematical curve called the ellipse. A simple expression with just a few numbers to define the width and height of the ellipse replaced the entire mass of data. Condensing the information in this way made it possible to understand what was happening to the planets. However, it was unclear to Kepler why the planets were constrained to follow these orbits. One way to try to solve the puzzle is to look at how an ellipse is drawn on paper by a schoolchild. A closed loop of string is constrained in two separate places, the foci of the ellipse (usually with thumbtacks). This constrains the movement of the pencil to trace the curve. Could the same thing be happening with

referred to as the Rutherford model. This is a nice way of explaining the forces acting on an electron and some sensible calculations can be made. However, further down the line it becomes evident that the electrons are associated with complex wave functions and do not actually follow precise orbital paths. But does the student ever forget the initial orbiting particle model? It is important when using analogies to emphasise the differences as well, and to choose the most appropriate analogy [3][4]. To maximise the effectiveness of an analogy, the mapping should be detailed and comprehensive at the formative stage, essentially providing robust instructional scaffolding.

Though it is entirely possible to express a concept entirely in mathematical terms, and this is often the approach when exploring physics on the very small and the very large scales, it is nevertheless desirable to link an abstract idea to a day-to-day concept in order to experience a sense of understanding. Work by Suleyman [5] backs this up: many more analogies were found in physics textbooks in comparison with biology textbooks where the material is more familiar and less abstract. The use of analogies can be seen as an extension of the process where scientific instruments are used to access information not available to our senses: a bat detector will pull the frequency of bat chirps down into the audible range; a thermal imaging camera will push infrared radiation into the optical range. An analogy will move an abstract idea into the world of our experience.

Analogies are not just used to communicate ideas, but are also used to understand new phenomena and acquire new knowledge. Fisher [6] has looked at reasoning by analogy in the context of discovery and has argued that this is a valid methodology to assess unexplained experimental or observational data and open up new lines of inquiry. There are many examples from the history of science of analogies being used in this way. James Clerk Maxwell developed the equations that describe all electromagnetic phenomena by likening the lines of force proposed by Faraday to thin tubes of fluid [7]. Once the equations were extracted from the model, the underlying model was abandoned. Maxwell's equations are stand-alone and explain all classical electromagnetic phenomena.

Analogies are very useful when trying to understand new information because phenomena in different disciplines and scales tend to be described by the same equations, or the same type of equations. In fact, there are remarkably few named equations in basic science, and many share the same structure, with second order differential equations particularly common. In addition, causality and the conservation laws appear to operate on all scales and this contributes to the effectiveness of analogies that cross scales and domains.

---

planets? Whilst the sun sits on one of the focal points of the ellipse, the other focal point is unoccupied. Though we can think of the sun somehow pulling on the planets like the tension in the string, it cannot be the same. The string model does not really relate that well to what is happening to the planets and the nature of this attractive force remains unclear through this comparison. This should also be a warning of the problems with using common-day analogies to understand the relationships that emerge from the mathematics and assuming they will be applicable.

But, even in the fundamental sciences, by focusing too much on the similarity, there is a risk of incorrect inferences. Consider the familiar case of the gravitational and electric forces between two charged particles, where one is often presented as an analogy of the other (as we saw with the Bohr atom). Both are inverse square laws and possibly the main scientific challenge of today is to unify these two forces based on this apparent similarity. However, we should remember the basic rule with analogies mentioned before - emphasise the differences! The electric force saturates; the gravitational force accumulates; they are completely different things when examined in this way. The inverse square property common to both could simply be a consequence of both forces operating in the same three-dimensional space, and is therefore an incidental link rather than a key to understanding.

Another important analogy is the expanding balloon model that has been used to explain the expansion of the universe. This has had a role both in the development of the current cosmological model and in explaining the general features of the model to the lay person. We will consider the relationship between the model and the universe it purports to explain. Is the model a good representation of the universe, or has the analogy has given us a faulty understanding of the universe which may now be impossible to dislodge?

7.2 The Balloon Analogy

The lambda-cold-dark-matter (ΛCDM) model of the universe has been developed and refined over the last 80 years and has proved extremely effective in explaining observations [8]. This consensus model is widely accepted in the scientific community. There are a few who question the model and propose alternatives, but the dissidents are very much on the fringes of cosmology [16][17].

The basic idea is that, from nothing, the universe began to expand in size dispersing the material within and cooling down in the process. The expansion continues to this day with each entity viewing itself as the centre of the expansion - there is no preferred location. There is a lot of evidence supporting this model including the existence of the cosmic microwave background radiation (CMBR) and the increase in the redshift of cosmological radiation sources with distance.

It is even possible to propose a topology for the expanding universe by drawing on a simple analogy. If we imagine an observer bound to the surface of a vast expanding balloon, distant points will be moving away at a rate proportional to the distance (assuming light moves along surface geodesics). An observer located anywhere on the surface will consider the expansion is centred on them. This is similar to what we are observing in the universe, though of course we need to switch from two dimensions to three dimensions. It is possible to extrapolate from the analogy a universe that is a closed system curved in an unseen fourth dimension. There is some theoretical support for this from the very effective general relativistic interpretation of gravity as curvature[2]. The balloon

---

[2] Certainly, the gravitational effect is elegantly described using the concept of curvature, it is also possible just to associate a slowing of the speed of light with gravity.

analogy was promoted by Misner, Thorne and Wheeler in their influential textbook 'Gravitation' [18][3]. In developing a mathematical model for the expansion, this analogy has had a significant influence. It is a particularly powerful analogy because this vision of a curved universe is very natural given our familiarity with the shape and curvature of the earth, where we inhabit an active more-or-less 2-D surface with locally hidden curvature. Another huge advantage is that the analogy illustrates how the universe can expand without necessarily expanding into something, and how we can have a finite universe without an edge[4].

But an analogy of the universe is not the universe itself, and like all analogies we need to clearly separate the differences from the similarities. And there are a number of very significant differences that should warn us that the expanding balloon model is possibly a convenient and loose analogy, nothing more.

If the real universe is very like the analogy, it should share a number of distinct and measurable properties. For example, triangles drawn over vast distances with lines of light should have a sum of internal angles which exceeds 180 degrees (as is the case with lines along the surface of the earth). And because material is carried by the expansion, work requires to be done against gravity by the expansion. The expansion should therefore slow down as kinetic energy is given up in order to conserve energy.

But these predictions are not correct: there is no evidence for space curvature and the expansion is not being slowed down by the need for energy conservation. What this does is highlight the differences between the analog and the target, and shows the limitations of the analogy. It would appear that the expanding balloon analogy is of limited value. However, this is not the approach taken by cosmologists. It is as if the analogy/model is too good to discard; and the approach is instead to fine-tune the universe to match some key features of the model[5].

After all this is what happens to light in glass, and there is no suggestion of curvature there. In addition, gravitation is space-time curvature, not exactly what is proposed here.

[3] The model is introduced in Section 27.5 of MTW. In the section of the book that follows, the metric for different curvatures is derived and clearly shown to originate from a 3-D geometry embedded in a 4-D Euclidean space. For a positive curvature, the radius is the scale factor. However, the extra dimension is not time but a space dimension. The position of the 3-D space within the 4-D manifold is changing with time. You could say that the 'edge' has been subtly moved into another dimension.

[4] This interpretation of the shape of the universe is so natural (obvious even), that we are predisposed to think of the balloon as more than an analogy, but a description of the universe itself. After all, what alternative geometry could be consistent with the available data? The alternative 'raisin cake' analogy refers to a blob of dough in an oven. As the cake bakes, the raisins move apart. This is not as effective an analogy because it does not solve the problem of the edge.

[5] The problem with fine-tuning an errant model is that, as more becomes known, the corrections become more elaborate and then implausibly contrived. Some have suggested that the consensus big bang model has reached such a state. Is the model too simple? Should we be able to understand it? In contrast, interactions on a fundamental level are based on quantum mechanics where causation is unfathomable.

However, it should be noted that many physicists consider the analogy described to be rather naive and that it has outgrown its usefulness, and should be rejected. Nevertheless, the scale factor is commonly referred to as the radius of curvature of the universe; this is the balloon analogy, still there in the background.

7.3 Fixing the Universe

We have noted that the real universe has properties which deviated significantly from what we would expect from the balloon analogy. Though it may seem an odd approach (given the earlier discussion concerning the limitations of analogies), it is possible to modify the universe so that it is consistent with expectation from the balloon analogy. We can address the problem of lack of curvature first.

If we assume the expansion proceeds at close to the speed of light (c), the greatest possible expansion rate consistent with the established laws of physics, we can estimate the size of the universe now, and from that calculate the expected curvature. The measured curvature is found to be much less than the calculated value, indistinguishable from zero, in fact[6]. The discrepancy can be explained by assuming the expansion at some stage in the lifetime of the universe was much greater than c and caused a dramatic increase the radius of curvature. This explosive growth is known as 'inflation' would need to have happened sometime in the early history of the universe as there is no evidence for it anywhere in the observational record that stretches all the way back to the cosmic background radiation (refer to [19] for an overview and [28] for some plausible alternatives).

Inflation is a contrived solution, certainly, but not completely implausible. At the early stages of the universe, there had to be a mechanism to rapidly produce between 10^{78} and 10^{80} fermions; perhaps the fermion creation process was accompanied by the creation of space.

The second problem is the expansion is not in reality being restrained by the gravitational force. The expansion rate, almost 14 billion years after the big bang, is still close to c. This is explained by adding in a new force feeding the expansion, labelled 'dark energy' (more about this later). This in itself is insufficient to give the correct expansion, and additional matter referred to as 'dark matter' is required to get the balance right. Making dramatic predictions of this sort is the sign of a good theory, but the problem is that there is no indication of what dark matter or dark energy might be, and they could feasibly be fudge factors added merely to shore up the model.

It would be wonderful if it were correct, but if it is found that just too much fixing is needed, we need to decide if persisting with the model is just a 'suspension of disbelief'? There is a huge reluctance on anyone's part to abandon the basic model of course, given humanity's millions of person-hours of investment, and especially so when all alternatives seem considerably worse.

[6] An expansion rate smaller than c would not improve the situation but would result in an even greater error.

There are some other problems as well but these are generally overstated. The horizon problem is the apparent mystery that the universe looks the same at opposite sides of the sky. The issue is described something like this:

> When we look at the CMBR from one side of the sky, we see the same temperature as the opposite side of the sky, though they are almost 28 billion light years apart and could not be in causal contact.

This is quite spurious because we should be referring to the state of the CMBR at the time of emission when we the absorbers are just a blob of unformed atoms somewhere in the middle of it all, a blob from which the galaxy, stars and planets and people will develop as the photons emitted proceed on their long journey from emitter to absorber (we are moving away from the emitters at close to the speed of light). Our view of the CMBR is that of a surface of equal cosmological time when the universe was about 300,000 years, and of course no point on that surface is in causal contact with any other point (more about this later). Notwithstanding the fact that the surface of last absorption has significant depth, every point on that surface will be in causal contact with every other point at an earlier time, perhaps close to the start of the universe; and this is true so long as the growth rate of the universe never exceeds c.

The homogeneity of the universe at that time is merely a reflection of the process that created all the matter. The initial density could have been perfectly uniform, and the inhomogeneity we see now grew later, largely as a result of gravity. The very small variation in the CMBR is a sign of the process starting; very limited because of the local expansion rate greatly exceeded the clustering action of gravity. There no real need to call on inflation to solve the horizon problem, because it is not necessarily a problem at all. We instead need to look at the process where matter is created to understand the initial homogeneity.[7]

A second minor problem (which is usually ignored) is that energy is not conserved in general relativistic cosmology[8]. Sean Carroll in a blog[9] stated:

> ‘It’s clear that cosmologists have not done a very good job of spreading the word about something that’s been well-understood since at least the 1920’s: energy is not conserved in general relativity.’

He focuses specifically on cosmological photons that are redshifted and lose energy as a result.

---

[7] It will not be anything like the way particles are created (and destroyed) in the universe today where baryon number violation seems to be forbidden.

[8] This is an amazing statement really, because energy conservation is embedded so deeply into physics. An analogy (as we are talking about analogies!) is probability. The sum of probabilities relating to an event is always one. We cannot see how it could not be one! In addition, the basis of the whole big bang model is energy conservation – work is done against gravity, which slows down the expansion. Either energy is conserved or it is not – which is it?

[9] http://www.preposterousuniverse.com/blog/2010/02/22/energy-is-not-conserved/ Accessed 25th Nov. 2018.

However, if we construct a system of particles moving apart through space and apply special relativity, then we get the same redshift effect, but there is no loss of energy. Energy is conserved from the frame of the emitter, absorber, or indeed any external inertial observer. The issue of energy non-conservation only arises in general relativistic cosmology because the frame that is used is unusual, as we will see further on.

We can obtain a good match between balloon analogy and observation by proposing inflation, dark energy and dark matter. If we were to discard the balloon analogy as irrelevant, could we then lose inflation, dark energy and dark matter as well?

7.4 Cosmology without a Model

In a flat cosmology consistent with special relativity, entities can move apart at an apparent velocity proportional with distance because of the elementary properties of Minkowski spacetime. If we consider a set of emitters centred on the surface of a sphere a proper distance r from the absorber at the time of photon emission, and all moving away at velocity v, this surface is projected onto a greater surface of radius $(1+z)r$ at the time of absorption (z is the redshift). The scaling ensures angular distances on the observed surface are the same as the angular separation of emitters at the time of emission. Thus, there is no curvature implicit in the observations.

If the distance separation at emission is allowed to vary, the surfaces of interaction can be combined to obtain a 3-D space for each observer that is centred on the observer, and with the property that photons propagation has essentially been eliminated. The act of emission and absorption is just the loss of energy at one point and the simultaneous gain of energy at another point (re-purposing the word ‘simultaneous’ to this particular frame), and energy is conserved when the relative velocity is considered.

This retarded frame can be ‘projected’ to consider where distant emitters might be at the proper time of the observer (the normal meaning of the word ’simultaneous’), and because of the time dilation effect, the position of the most distant emitters tends to infinity because the recession velocity tends towards the speed of light. The result is an infinite universe, echoes of the Milne cosmological model [29][15].

In the retarded frame, the observer horizon is interesting, in particular for the case of a universe whose expansion has never exceeded the speed of light. In this case, all particles remain in causal contact. If, in addition, the expansion has never been never less than the speed of light, there exists a static horizon representing the start point of the universe.

How are we to understand this horizon? A fascinating video piece by artist Mark Wallinger, *Construction Site*, 2011, seems to reflect on the difficulty. Workers laboriously construct scaffolding on a pebble beach to a height where the top platform exactly aligns with the distant horizon from the viewpoint of the observer. It is evident that this mechanism for understanding the horizon is completely inappropriate because the scaffolding is then dismantled. The process is repeated[10]. It is clear that local activity has revealed nothing useful concerning the nature of the distant apparently static horizon. We are perhaps in a similar position - the observable horizon of the universe could be the big bang itself, but how could we ever investigate this?

Returning to the retarded model, an attempt can be made to piece together the views of different observers to construct a model that represents the position of every particle at universal time T(just like the expanding balloon model). A significant obstacle is that the absolute time cannot be the local clock time because the local clock is affected peculiar acceleration and velocity[11]. If we do build this composite picture of the position of each particle at absolute time T, then there is no interaction between particles, and it is a dysfunctional model. In addition, the same model cannot be valid for all observers, and even if an agreed model could be constructed it could be neither isotropic nor homogeneous. Refer to Appendix B for further discussion.

The concept of an observer somewhere outside the universe viewing from this privileged position is something that the balloon analogy encourages, and is the basis for the concordance model, but no such position exists[12]. On an even more basic level there is the question of how the cosmological time that is assumed to drive the expansion, if such a thing exists, should be incorporated into normal physics?

Certainly we can make our universe work by treating it as flat expanding space with no thought for geometry[13], but by abandoning models and analogies, we lose all understanding: what is this expansion, and why is it happening? There are benefits with just sticking with the maths and doing without a visual model. First of all, the idea of 3-D space embedded in 4-D is problematic. Though we have talked of the 2-D surface of the earth, there are actually no 2-D objects on this planet – everything is 3-D. The entire balloon concept is flawed unless we think of everything in this universe as 4-D objects trapped in a 3-D world.

If exotic tricks such as invoking higher dimensions are allowed, there are other cosmologies that can equally well be invented. Refer to Appendix A for a discussion.

7.5 The Energy Problem

In flat space-time, as in standard cosmology, there is an apparent energy problem. The expansion does work against gravity, but the problem can be solved

---

[10] Presumably echoing the incoming and outgoing tides.

[11] For example, if an entity is moving in a circular orbit, the clock will run slow compared to the centre, hence the local time will drift from the cosmological time.

[12] It is possible to go even further. Philosopher van Fraassen argued that the experimental data we have at our disposal is nothing more or less than a representation of an observable fragment of a fundamentally unobservable universe [21].

[13] The focus on flat space-time is not a denial of general relativity - it is just that general relativity does not need to be applied globally because the net gravitational force in deep space is so small.

in standard cosmology by introducing dark energy[14]. We will show how it is introduced in standard cosmology and check if the same substance present in a flat space-time model would ensure energy conservation in a similar way.

To explain dark energy, we again return to the balloon analogy, but this time looking at the work done blowing it up. We can work towards this by giving a simplified derivation of the expansion starting with the Friedmann Equation. The general relativistic form the Friedmann Equation gives the radius of the universe $R(t)$:[15]

$$\left(\frac{\dot{R}}{R(t)}\right)^2 = \frac{8\pi G}{3}\frac{\epsilon(t)}{c^2} - kc^2, \tag{7.1}$$

where $\epsilon(t)$ is the total energy density (all sorts of energy) and k is the curvature.

As the curvature is observed to be zero, the equation can immediately be simplified:

$$\left(\frac{\dot{R}}{R(t)}\right)^2 = \frac{8\pi G}{3}\frac{\epsilon(t)}{c^2}. \tag{7.2}$$

Because the energy density is proportional to $R(t)^{-3}$, we get the following differential equation:

$$\dot{R} = \frac{\sqrt{2GM}}{R(t)^{1/2}}, \tag{7.3}$$

where M is the total gravitating mass.

The solution is

$$R(t) = \left(\frac{9GM}{2}\right)^{1/3} t^{2/3}. \tag{7.4}$$

Energy is conserved as kinetic energy is used up making the potential energy less negative. We deduce that the expansion rate must inexorably slow down[16]. However, this is not what we see. The expansion rate in the current epoch is still

---

[14] The trick in the concordance model is that the action of gravity on matter can slow down the expansion, which implies a connection between matter and space that is not evident in the normal action of the gravitational force. The normal force moves entities through space and does not create or destroy space. The proposal that gravity can slow the expansion really needs justification.

[15] As mentioned previously, there is a limited acceptance that energy is not conserved in general relativity, yet the Friedmann Equation that describes the expansion is very much an expression of energy conservation. This radius $R(t)$ is the size at cosmological time t, and it is completely unclear how this can be measured or related to clock times.

[16] You could suggest that the universe is rotating in order to generate a pseudo force such as a centrifugal force, but this is extremely contrived.

close to the speed of light. And the solution gives an age for the universe that is too small.

But we have not considered all the effects of expansion. We have only looked at the gravitational forces at work. What if we consider matter in the universe to be in the form of a gas? There is work associated with the expansion of a gas, and should this not be considered in an expanding universe? Let us return to the balloon. When we blow up a balloon, we need to do work to expand the skin against the tension forces in the material[17]. The internal pressure needs to exceed the external pressure, and a quantity of mechanical work done equal to $\Delta p \Delta V$ should be supplied (where p and V are the pressure and volume respectively). The equation applies equally well to either a small region or the entire volume.

If we have a sphere of radius *R(t)* expanding at velocity $\dot{R}$,

$$dE + \Delta p(t) dV = 0, \tag{7.5}$$

and

$$\dot{V} = 3V \frac{\dot{R}}{R}. \tag{7.6}$$

The total energy is

$$E = V\epsilon \tag{7.7}$$

hence,

$$\dot{E} = \dot{V}\epsilon + V\dot{\epsilon}. \tag{7.8}$$

Combining equations 5, 6 and 8:

$$3\frac{\dot{R}}{R}\left(\epsilon + \Delta p(t)\right) + \dot{\epsilon} = 0. \tag{7.9}$$

If we presume pressure to be a function of energy density, postulating $\Delta p(t) = w(t)\, \epsilon$, then

$$\frac{\dot{\epsilon}}{\epsilon} = -3(1+w)\frac{\dot{R}}{R}. \tag{7.10}$$

Back to the Friedmann equation, let us differentiate it with respect to time:

$$\frac{\ddot{R}}{R} = \frac{8\pi G}{3c^2}\left(\frac{\dot{\epsilon}R}{2\dot{R}} + \epsilon\right). \tag{7.11}$$

Substitute in the equation 10 into equation 11 (with the energy loss associated with the work done converted to the equivalent gravitational mass):

$$\frac{\ddot{R}}{R} = \frac{4\pi G}{3c^2}\epsilon\left(3w + 1\right). \tag{7.12}$$

---

[17] In the case of free expansion, for example a balloon burst in a vacuum.

If w = -1/3 there is no acceleration (more or less what is observed).

This is a very interesting fix to the energy problem. The concept was derived from the analogy of an expanding gas, but it is now quite different. As the universe expands (for whatever reason) the expansion releases energy because the pressure within is negative. We can hardly interpret this in terms of the pressure outside the universe affecting the expansion, but it will have to be some unknown type of internal fluid, which we can call dark energy or it can be Einstein's cosmological constant, or a combination of both. For this simple analysis, it is not necessary to make a distinction. Regardless, it certainly fixes the problem of an expansion which is not slowing down as expected. The gravitational energy that has to be supplied is provided by this fluid. If $w <$ -1/3 the expansion accelerates.

You would be forgiven for finding this extremely fortuitous (or contrived), but we do have a viable cosmology. We fixed the energy problem by assuming our balloon is filled with a fluid that supplies energy as it expands.

Though dark energy is not necessarily exclusive to consensus model, it cannot in this form be the energy fix that will work in flat space-time cosmologies. The reason is that dark energy by its nature is closely coupled with the expanding balloon model, as has been shown. Any other model will need its own solution to the energy problem.

7.6 Conclusion

The accepted model of the universe is closely linked with the expanding balloon analogy[18]. It makes it possible to develop simple equations for the expansion. The model is therefore largely effective. However, there are significant differences between what is naturally predicted and what is observed in terms of energy conservation and geometry. The measures taken to make the model correspond to reality can be interpreted as revealing hitherto unknown properties of universe, or alternatively as an unnecessary contrivance.

But we come back to the derivation of Maxwell's equations. Once the equations were derived from the crude mechanical model, the model was discarded. The equations took on a life of their own and immediately started making amazing predictions. Can we say the same about the balloon analogy in cosmology? Certainly, but only after dark matter and dark energy are found.

And there is not a better alternative theory sitting in the wings ready to step in. Flat space-time cosmologies, as examples, do not come with a convenient understandable model. But it may be that neither approach is correct - something completely new may be needed.

---

[18] The link is that the Friedmann equations which describe the local expansion based on energy density are joined with the cosmological principle (the universe is spatially homogeneous and isotropic) to give the apparent uniform expansion at cosmological time T.

References

1. Jonāne, L. (2015). Using Analogies in Teaching Physics: A Study on Latvian Teachers' Views and Experience. Journal of Teacher Education for Sustainability. 17. 10.1515/jtes-2015-0011.
2. Dilber, R. (2012). The effects of analogy on students' understanding of direct current circuits and attitudes towards physics lessons. European Journal of Educational Research. 1. 211-223. 10.12973/eu-jer.1.3.211.
3. Glynn, S. M., and Takahashi, T. (1998). Learning from analogy-enhanced science text. Journal of Research in Science Teaching, 35, 1129-1149.
4. Wiley, J., Jaeger, A., Taylor, A. R. and Griffin, T. D. (2017). When analogies harm: The effects of analogies on metacomprehension. Learning and Instruction. 10.1016/j.learninstruc.2017.10.001.
5. Suleyman, A. (2016). Analysis of analogy use in secondary education science textbooks in Turkey. Educational Research and Reviews. 11. 1841-1851. 10.5897/ERR2016.2984.
6. Fisher, A. A. (2018). Inductive reasoning in the context of discovery: Analogy as an experimental stratagem in the history and philosophy of science. Studies in History and Philosophy of Science Part A, Volume 69, Pages 23-33.
7. Bokulich, A. (2014). Maxwell, Helmholtz, and the unreasonable effectiveness of the method of physical analogy. Studies in History and Philosophy of Science Part A. 50. 10.1016/j.shpsa.2014.09.012
8. Hamilton, J.-Ch. (2014). What have we learned from observational cosmology? Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics, Volume 46, Part A, Pages 70-85.
9. Merritt, D. (2017). Cosmology and convention. Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics, Volume 57, Pages 41-52.
10. López-Corredoira, M. (2017). Tests and problems of the standard model in Cosmology. Found. Phys. Volume 47, 6, Pages 711-768.
11. Misner, Thorne and Wheeler. (1973). Gravitation. Freeman and Company, San Francisco
12. Uzan,J.-P. (2015). Inflation in the standard cosmological model. Comptes Rendus Physique, Volume 16, Issue 10, Pages 875-890.
13. Brandenberger, R. (2014). Do we have a theory of early universe cosmology? Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics, Volume 46, Part A, Pages 109-121.
14. Torretti, R. (2000). Spacetime Models for the World. Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics, Volume 31, Issue 2, Pages 171-186.
15. Milne, E. A. (1948). Kinematic relativity; a sequel to Relativity, gravitation and world structure. Oxford: Clarendon Press.
16. van Fraassen, B. C. (2008). Scientific Representation: Paradoxes of Perspective. Oxford: Oxford University Press.
17. Nature, doi:10.1038/nature.2013.13743 commenting on Pourhasan, R., Afshordi, N. and Mann, R. B. (2013). Preprint available at http://arxiv.org/abs/1309.1487.

Appendix A: Alternative Cosmologies

If it is the case that invoking higher dimensions is acceptable, many cosmologies become possible[19]. Consider the following conversation between an an engineer (E) who is not a subject specialist but is very familiar with mechanical systems, and a cosmologist (C):

E: 'I have a new model for the universe which I consider better than the established model. Space is infinite and there is no such thing as the expansion: flat space is rotating with respect to the observer (i.e. us) and the angular frequency is ω. A cosmological object at radial distance r has transverse velocity $v = r\omega$. The observed Doppler redshift, $1 + z$, is therefore γ (because of the transverse Doppler effect). In special relativity, the luminosity distance of a moving source is

$$d_L = (1+z)^2 r_e, \tag{7.13}$$

where r_e is the proper distance in the observer frame at the point of emission[20]. With transverse velocity v, and actual photon velocity c, the proper distance at emission is not the apparent direct distance but the photon travel time. Because the photon path is curved, the change in radial distance with time is[21]

$$\frac{dr}{dt} = c\left(1 - \frac{r^2\omega^2}{c^2}\right)^{1/2}, \tag{7.14}$$

hence the photon travel distance is

$$r_e = \left| \int_r^0 \frac{1}{\left(1 - \frac{r^2\omega^2}{c^2}\right)^{1/2}} dr \right| = \frac{c}{\omega}\sin^{-1}\left(\frac{v}{c}\right). \tag{7.15}$$

Slotting r_e into equation 13 and noting that the redshift is γ, the result in terms of observables is

$$d_L = \frac{c}{\omega}(1+z)^2 \sin^{-1}\left(1 - \frac{1}{(1+z)^2}\right)^{1/2}. \tag{7.16}$$

Choosing a value for ω of 2.5 times the age of the universe in conventional cosmology, an excellent match for the available supernovae redshift - luminosity distance data is obtained.'

C: 'But, if space is rotating, objects the same radial distance away towards the direction of the rotation axis have smaller rotation velocity and the relation $v = r\omega$ no longer holds.'

---

[19] And many new concepts as well. This is seen in modern cosmology where there may be multiple universes occupying multiple dimensions, some of which interact with our universe.

[20] This follows because the specific intensity divided by the frequency cubed is Lorentz invariant.

[21] A relativistic Coriolis effect.

E: 'The rotation is in a 4th dimension.'

C: 'Even if this were meaningful, we can refer to the lower dimension analogy of a 3-D observer looking out onto a rotating 2-D surface infinite in extent, and there is still a problem: how is possible for all observers to experience the same rotation effect - the rotation still is centred on the observer?!'

E: 'In the manner of special relativity, each observer constructs their own consistent world view, though it is even possible that it is the observers who are rotating.'

C: 'This postulates a model with a cutoff radius when the transverse velocity reaches c beyond which nothing is visible, but this needs justification. And what about the cosmic microwave background which is evidence of a developing rather than a static universe?'

E: 'Without expansion diluting the energy, repeated cycles are possible giving rise to completely different particle aggregations at different cosmic times. Secondly, the cutoff problem also exist in conventional cosmology and is ignored. The apparent recession velocity in conventional cosmology can exceed c, hence there are apparently galaxies that are not visible, but the cosmic background predates all galaxies and is visible in all directions - how then can galaxies that formed later be out of view?'

C: 'But surely, this rotation, if it exists would be observable.'

E: 'Not necessarily. Near and far objects do not move relative to one another and the absolute change is about 10^{-9} radians of arc a year. And how would you establish your reference point?'

C: 'Your proposal seems to suggest that all redshift is Doppler, but what about gravitational redshift?'

E: 'No, the claim is that the cosmological redshift is actually a Doppler effect. However, gravitational redshift could be reproduced by a transverse oscillation.'

From this exchange (which could go on forever) it is evident that the biggest leap, the assumption that rotation (or indeed any motion) in a higher dimension is possible, passes without sufficient scrutiny. Perhaps the same is true of the big bang analogy. Once a model gains traction and the focus moves to the detail, there may be a reluctance to revisit or even do a sense-check on underpinning ideas.

And if one is to deny the need for a model of any sort, as would be the case with any special relativistic interpretation, it is just anarchy - anything goes. Consider the following conversation with a philosopher (P).

P: 'Gravity is the curvature of space, but the influence is greater than that. I believe gravity also binds space and can prevent it expanding. In the way the slight curvature of a fingernail imparts significant strength, sufficient curvature can halt the expansion. If curvature as normally expressed by the gravitational acceleration drops below a critical value, the natural tendency for expansion can no longer be halted and the expansion proceeds. This critical value is the a_o of alternative theories like MOND, with a value of about 1.2 x 10^{-10} m s^{-2}.'

C: 'That is completely inconsistent with reality. How can a region that is not expanding be seamlessly joined into an expanding space-time manifold? The only consistent approach is to presume the expansion proceeds unhindered everywhere, down to the atomic level. Stable circular orbits therefore have an unseen peculiar radial velocity that cancels the expansion velocity, rendering the expansion undetectable in our neighbourhood. But it is still taking place everywhere.'

P: 'You are completely locked into the naive idea of the universe being something that makes topological sense when somehow viewed from the outside. In addition, that assumption of unhindered expansion goes against standard cosmology where the expansion is slowed by having to do work against gravity (though the mechanism for that is never explained).'

C: 'Turning to the a_o you mentioned, I am aware it has been related to the current age of the universe by this formula:

$$a_o \approx cH_o \approx \frac{c}{T_o}. \tag{7.17}$$

But if this were the case, we would expect the value to change over time. However, there is no evidence of a difference in the value of a_o when comparing older and younger galaxies, or indeed any evolutionary effect. Furthermore, if your transition value does have a cosmological origin, and it is the ratio of two constants as suggested by equation 17, it is unclear what the time constant could be. Perhaps surprisingly, there are no such constants in nature. There are constants for mass, charge, velocity etc, but not for time. The decay rates of the neutron or other particles cannot be considered fundamental constants.'

P: 'That is not true. One time constant that does exist is the lifespan of the universe from the big bang to destruction. This could be the time that enters into equation 17 instead of the current age of the universe. Slotting in the numbers:

$$T_U = \frac{c}{a_o} = \frac{3\ .\ 10^8}{1.2\ .\ 10^{-10}}\ \mathrm{s} =\ 2.5\ .\ 10^{18}\mathrm{s} = 79 \text{ billion years.} \tag{7.18}$$

If the universe now is 13.8 billion years old, it has about another 65 billion years to go.

It can be suggested that once the characteristic acceleration of the entire universe drops to a_o then a rapid process of collapse is suddenly initiated. In the language of particle physics (and with reference to the Higgs field) current space is a false vacuum and a spontaneous drop to a true vacuum may occur 65 billion years from now triggered by this acceleration (or curvature). The movement of mass in the expanding frame which we see now when the gravitational acceleration drops below threshold is just an echo of the process that will ultimately be triggered to end the universe.'

C: 'But how do all the particles fit into this picture? They seem irrelevant, when we know gravity acts not on space but on matter.'

P: 'The origin of matter can be imagined as a box tightly packed with little springs opened at the moment time began. They have since been flying around, pulled along by the shifting space-time background.'

And again this discussion can go on. Note here the philosopher's heavy dependence on a variety of very loose analogies rather than a concrete model or any solid physics, and these are analogies that do not try to represent reality but instead to illustrate and explain processes. Whilst we may criticise models as possibly distorting reality, they do in some cases sensibly constrain ideas[22].

Appendix B: Models in Special Relativistic Cosmologies

It is perhaps useful to explore in more detail the claim made in the text that physical models are incompatible with special relativistic cosmologies.

A special relativistic observer is causally linked to all other entities in retarded time. The time lag is a function of distance hence it is possible to infer the current position of these distant objects by projecting forward using the apparent velocity of the object. However, there is no guarantee that the objects will reach these actual positions at that observer time because the velocity can vary over the propagation interval. Nevertheless, it is technically possible to build an exact picture of the object configuration by collating observations from future times. Unfortunately the complete picture cannot easily be closed at the edges because the furthest objects have a recession velocity approaching the speed of light, hence infinite time is needed to access the data required. And to ensure picture is the same for all observers, normal clock time cannot be used (again because of acceleration and velocity effects[23]) but instead some velocity-independent cosmological time is needed (which does not have a natural place in special relativity).

And even if we were to succeed in constructing this simultaneous time slice of the universe, the only significant feature of this foliation (spatial hypersurfaces with constant cosmic time) is that no entity is causally linked to any other[24].

---

[22] Perhaps, surprisingly, ideas of this type find their way into the academic literature. For example, the gravitational energy of the entire universe (GM/r) is almost equal to c^2, which suggests we inhabit the interior of a black hole, or almost a black hole. The cosmic microwave background radiation is then energy leaking from the mother universe past the event horizon. Presumably the Hubble redshift is then some sort of gravitational redshift. The universe is not expanding but collapsing into a singularity. Also throw in an extra dimension. But this is very flimsy and does not seem to fit the most basic facts, but it is seriously described in *Nature News* [23].

[23] For example, if we have a star and a planet rotating round the star, the clocks will differ. How is this long-term clock drift incorporated into the model?

[24] The simultaneous frame is the default in physics because it is not normally necessary to include retardation in typical problem situations.

A particular problem with flat space-time cosmologies is the presence of an edge and a matter distribution is neither homogeneous nor isotropic[25] The matter creation process is also difficult to comprehend when there is a 'edge'. It is inconceivable that all the matter appeared instantly, but more likely by a process that was energetically favourable in the very early universe. But we do not know the rules that applied at the time. Though matter was created then, it is not now. How have the rules changed?

But without rules anything goes. We can make up any idea, no matter how ridiculous and it is equally valid. For example, we can look at the uncertainty principle equation:

$$\Delta E \Delta t \geq \hbar. \tag{7.19}$$

This usually refers to the amount of time energy conservation can be violated, the duration for which energy can be borrowed. What if instead we interpreted it as the energy that must be supplied to buy time? We start off with one particle. The energy, mc^2, is borrowed and it moves backwards in time by time step $\hbar/mc^2$. The universe after this is empty and therefore ceases to exist, the earlier universe then proceeds along a different line with two particles and no temporal paradoxes. This doubling proceeds until it is no longer energetically possible. The particles repel in an act to eject the invader from that space.

But if the speed of light were infinite, the problems disappear. This implies instant communication of some form; but is instantaneous communication possible? In the realm of quantum mechanics, instant communication may even be essential to explain the process of quantum entanglement. Quantum entanglement means that many particles are linked together in a way such that the measurement of the quantum state of one particle determines the possible quantum states of the other particles. The connection between in no way depends on the location of the particles in space.

But it may go deeper than that. The wave functions that are solutions to Schrödinger's equation are not localised and are infinite in extent, or perhaps more accurately delocalise a particle to a point anywhere in the universe. Consider a heavy atom with all electrons in their ground state. Now imagine this atom undergoes beta decay and emits an electron. The wave functions must all change as a result because the central charge has increased by +1e. We may presume the new functions are generated instantly such that the old electron wave functions are now a mixed state of these new wave functions. There is no evidence of a propagation setup time, which would be problematic as there is a need for all wave functions to adjust to the expanding universe in some way (though wave functions may be a convenience in the same way field lines are). Overall, there is no evidence of interaction delays once an electron becomes bound into an atom.

Furthermore, the wave functions of stable bound states of a hydrogen atom when calculated using Schrodinger's equation do not include retardation. In addition, the wave functions solutions extend to infinity – how is the instantaneous transition between states possible (particularly with the emission or absorption of a photon)? And of course, the universe is not infinite as far as we know – how do wave functions adapt to expanding space?

[25] However, the requirement for isotropy and homogeneity are assumptions and there is no compelling reason why they should be correct.

Ridiculous, perhaps, but with such limited knowledge, almost nothing can be excluded.

The bottom line is that we think there should exist a description of the universe that is comprehensible to observers within who are an intrinsic part of it, but is this really reasonable?

8 Is Potential Energy Real?

Summary

It is generally considered that matter is 'real', but, in contrast, the nature of energy is indeterminate, and this particularly applies to potential energy. The view is taken that although total energy is conserved, potential energy is merely a bookkeeping mechanism introduced to hide a degree of ignorance concerning energy flows and is therefore not real. Ways in which energy can be conserved without having to introduce potential energy are proposed and evaluated.

8.1 Introduction

Energy can neither be created nor destroyed and the total energy in any closed system is constant[1], students are told from an early age. This is considered important underpinning knowledge in any science class. It is not even necessary for the student to understand what energy actually is to find the idea starts to rapidly pay dividend: problems are quickly and easily solved and systems become ordered. But is energy really indestructible? If so, how can a concept have this property? Surely it must be a 'thing' rather than an abstract mathematical quantity[2]?

There is no clear answer to be found in physics textbooks, and we need to pay attention when an authority such as Feynman claims that, ' .. in physics today, we have no knowledge of what energy is' [2], [3]. Furthermore, when we consider cosmology, doubts about energy conservation begin to creep in, as it is casually stated that energy is not conserved in general relativity [4], [5], and this is the theory on which the accepted model of the universe is based. Of course, the pragmatic approach is to accept energy as a useful calculation device and leave it at that, but there is the possibility it is telling us something important about the structure and nature of the universe.

If there are questions about the nature of energy in general, the role of potential energy is particularly difficult. It is commonly believed that potential energy is an artefact invented as a simplistic way of ensuring energy conservation; it is no more than a fix. The shortcoming of the concept becomes apparent when we look at the sophisticated representation of the gravitational force in general relativity - the simple adjustment labelled potential energy is found to be inadequate and indeed has no part in the geometric interpretation of the gravitational effect.

Overall, the nature of energy has been subject to much investigation. In general terms, philosophical investigations into the topic have a rich history.

---

[1] And already we have a problem - is it the energy **in** a system or **of** a system? A key question, but there is no definitive answer.

[2] Coelho argues, 'If energy cannot be destroyed, it must be a real existing thing. If its form changes, it must be something real as well [1]. The commonly accepted view is that energy is the property of a system and that it needs to be associated with something real, but this interpretation is open to argument.

Hecht provides a good review of the subject [6], and refers to the definition of potential energy attributed to Maxwell as the energy a system has 'the power to acquire', which in any relativistic system is infinite[3].

By referring to a number of situations, we will try to establish if potential energy is anything other than a useful tool. In its simplest form, a falling object gains kinetic energy by harvesting gravitational potential energy. But if we dispense with potential energy, is it still possible to claim that energy is conserved?

8.2 Quantifying Potential Energy

Potential energy is labelled U, but it is more convenient to work with potential (label, V). This is potential energy in a form that is independent of the properties of the test body. If the active force is gravity, the potential is the potential energy per unit mass.

Gravitational potential is defined as the work that must be done to move the body to infinity. Whether or not there is a practical procedure by which this might be accomplished is immaterial. But the definition itself is problematic in the context of cosmology. Consider the integral below[4]:

$$V(R) = -\int_R^\infty \frac{GM}{r^2} dr. \tag{8.1}$$

The universe is not infinite but is of approximate radius cT where T is the current age of the universe, and the size is increasing at rate c (more or less). The binding energy therefore becomes more negative over time:

$$\frac{dV(R)}{dt} = -\frac{GM}{cT^2}. \tag{8.2}$$

The effect seems small, but it is not. Consider a sphere of matter of constant density ρ and radius R. The total potential is

$$V(T) = 4\pi\rho\; G\int_0^R \left(\frac{M(r)r^2}{r_U} - M(r)r\right) dr = \frac{16}{3}\pi^2\rho^2 G\int_0^R \left(\frac{r^5}{r_U} - r^4\right) dr, \tag{8.3}$$

and evaluating,

$$V(T) = -\frac{3GM^2}{5R} + \frac{GM^2}{2R_U}. \tag{8.4}$$

If all the matter in the universe had somehow resisted the expansion (and stayed over time as this sphere), the anomalous energy (second term in RHS of

---

[3] Hecht suggests that potential energy is measurable because of the mass reduction associated with it, but it is really the lost energy that is tracked. The potential energy refers by default to the prior state, but it can really refer to any of the possible states of the system.

[4] Reduced mass effects have no impact on the argument and are ignored throughout.

equation 3) would then of the same order of magnitude as the rest-mass energy of the universe.

An alternative definition of potential, which avoids the problem of referencing infinity, is to refer only to the local change in energy per unit test mass as work is done against the field:

$$dV(r) = -\frac{V}{r}dr. \tag{8.5}$$

This definition breaks the direct link with the expansion, but a disadvantage is that the identification of potential energy as a global property relevant to cosmological modelling is lost. By the definition in equation 5, potential is clearly a balancing mechanism. The change in potential is the exact opposite of the kinetic energy change. But is it more than that? Consider two bound stable particles. The effective mass is less than the total rest mass, but only because the kinetic energy that was gained during the binding process has been lost. It can certainly be interpreted as potential energy contributing negative mass, but there is no need to do so.

However the situation is more complex in general relativity. Einstein's field equations are (without the cosmological constant):

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu}, \tag{8.6}$$

where $R_{\mu\nu}$ is the Ricci curvature tensor, R is the scalar curvature, $g_{\mu\nu}$ is the metric tensor and $T_{\mu\nu}$ is the stress–energy tensor. The stress-energy tensor determines the curvature and is of particular interest here. It takes the form

$$T = \begin{bmatrix} T_{00} & T_{01} & T_{02} & T_{03} \\ T_{10} & T_{11} & T_{12} & T_{13} \\ T_{20} & T_{21} & T_{22} & T_{23} \\ T_{30} & T_{31} & T_{32} & T_{33} \end{bmatrix} \tag{8.7}$$

Referring to a volume element at a particular time, T_{00} is the energy density within that volume, T_{01} - T_{03} are the energy flows in each of the coordinate directions, T_{10} - T_{30} are the corresponding momentum flows, and the remaining matrix elements represent pressures and shear stresses[5].

Now consider the case of a static mass distribution with zero pressure. Only the element T_{00} is required and the density takes into account the mass reduction associated with the gravitational binding energy. However, if we were instead to express the distribution as individual particles, the density derived from the rest masses should be entered and there is no natural place in the matrix for the inclusion of the binding energy between them (though the tensor could be modified to include binding energy because the gravitational effect can be interpreted as a stress effect).

---

[5] The first index, $\mu = 1 .. 3$, specifies the direction in which the stress (force over area) component acts, and the second index, $\nu = 1 .. 3$, identifies the orientation of the surface upon which it is acting.

This is a tricky problem if we were to refer to the missing energy as gravitational potential energy, because it can lead to circular arguments (e.g., 'does gravitational potential energy gravitate?); and if we simply refer to the binding energy as missing energy[6], it is unclear to which rest masses and in what proportion the loss of rest mass should be allocated[7].

8.3 At the Limits

The concept of potential energy can be evaluated further at the limit where the mass concentration is so great that matter and energy can no longer escape the gravitational force and a black hole is formed.

The possibility of a black hole first arose in the context of Newtonian gravity (though as we will see, it is only a genuine feature of general relativity). The basic idea is that the gravitational escape velocity at radial distance R from the centre of a compact spherical mass concentration is

$$v = \left(\frac{2GM}{R}\right)^{1/2}. \tag{8.8}$$

If the combination of M and R is such that v must be greater than c, a test mass cannot escape. Another approach is to consider a photon of energy E with equivalent gravitational mass E/c^2. If the magnitude of the potential energy is greater than the photon energy then light cannot escape. The condition in this case is:

$$\frac{GM}{Rc^2} > 1. \tag{8.9}$$

However, both approaches are seriously flawed and should be considered as conceptual arguments only, not least because the standard expression for the potential energy is assumed to be correct and just dropped in. Following on from the preferred approach in the previous section, it is preferable to calculate the work done by the escaping mass or photon over a restricted path. In a quasi-Newtonian model, the potential energy associated with moving under the influence of a central force from radial distance r_1 to r_2 is related to the work done *by* mass m $(= \gamma m_o)$:

$$\int_{E(r_1)}^{E(r_2)} \frac{1}{E} dE = -\int_{r_1}^{r_2} \frac{GM}{r^2c^2} dr. \tag{8.10}$$

This gives:

$$U(r_2) - U(r_1) = mc^2 \left(1 - \exp\left(\frac{GM}{rc^2}\left(\frac{1}{r_2} - \frac{1}{r_1}\right)\right)\right). \tag{8.11}$$

---

[6] Kinetic energy has weight [7], hence missing kinetic energy has negative weight.

[7] However, one would expect a correction to exist for consistency with Noether's theorem which states that the laws of physics not altering with time, as seems to be the case, implies energy conservation.

If r_1 is set to infinity then in the weak field approximation the potential $V(r_2$ tends to -GM/r_2 as expected. Dividing by relativistic mass at infinity converts from potential energy to potential in this case. There is no circumstance where the a black hole can form with this calculation.

Other approaches are possible, but non-general relativistic approaches all fail. It is possible to explain in general terms why this is the case: Imagine a photon does not have the energy to escape to infinity if the emission radius is R. An observer at position $R + \Delta R$ will be able to receive redshifted photons from behind what is ostensibly the threshold radius and for this observer the object is not a black hole. In addition, the apparent mass of the black hole after external mass has been captured is increased by the relativistic mass, but at the same time decreased by the mass equivalent of the gravitational binding energy. The net mass change is zero at the point of capture: a Newtonian black hole cannot grow. As a result, any discussion of black holes should only be in the context of general relativity where there exists an event horizon which is observer independent and past which no material can escape.

In general relativity, the potential at any point (again with respect to infinity) can be deduced from the gravitational redshift effect. From the Schwarzschild metric, a photon with energy E_o emitted from radial position r with respect to a central mass M has an energy at infinity

$$E_\infty(r) = E_o\left(1 - \frac{2GM}{rc^2}\right)^{1/2}, \tag{8.12}$$

hence

$$V(r) = c^2\left[\left(1 - \frac{2GM}{rc^2}\right)^{1/2} - 1\right]. \tag{8.13}$$

If $2GM/rc^2$ is small, the result converges on the Newtonian value but it is clear that though potential cannot be entered into Einstein's field equations, potential does emerge from general relativity (certainly in the case of the simplest discrete mass configurations).

In addition, it is possible to calculate the redshift of a photon going from point r_1 to r_2:

$$1 + z = \frac{\left(1 - \frac{2GM}{r_2c^2}\right)^{1/2}}{\left(1 - \frac{2GM}{r_1c^2}\right)^{1/2}}. \tag{8.14}$$

If r_1 is the Schwarzschild radius, then it is clear that an event horizon emerges. The denominator of equation 10 is zero and no external observer, no matter how close to the event horizon, can interact across the barrier.

It is clear that that although GM/r recurs, there is no unique formula for potential energy comparable with that of kinetic energy, and this is most evident when dealing with black holes where weak field approximations are no longer valid. The concept of potential energy is useful but it is just that - there is no indication that potential energy is real, and we now see the justification for not including gravitational potential energy in the field equation - it is simply not real energy so why should it be included?

But if we were to do without potential energy, surely it means that energy is no longer conserved, or is there a way round this? We certainly need to review statements such as 'potential energy is stored energy' [8] which are easy to understand but are simply incorrect, or at the very least inaccurate. There is no indication of energy being stored anywhere; for example in the case of gravity the field is not changed.

An exceptional case is the propagation of a transverse wave, which is usually associated with the transfer of energy between kinetic an potential and specifically the effect this has on velocity, but applying this to electromagnetic radiation is problematic. The wave equation is

$$c^2 \nabla A_\alpha - \frac{\partial^2 A_\alpha}{\partial t} = 0 \tag{8.15}$$

where A_α with $\alpha = 0$ is the electromagnetic scalar potential $\alpha = 1$ .. 3 are the components of the vector potential. When the equations are expressed in terms of observables, **E** and**B**, it would seem that energy is stored in electric and magnetic fields which are created and destroyed through the oscillation cycle to balance the dynamic aspect of the wave. However, it is unclear what this physically means. Feynman in Chapters 27 and 28 of his lectures [16] wrestles with the problem of the energy associated with fields without supplying any real clarity or solution. He comments, 'that there seems to be no unique way to resolve the indefiniteness in the location of the field energy' and concludes 'Anyway, everyone always accepts the simple expressions we have found for the location of electromagnetic energy and its flow. And although sometimes the results obtained from using them seem strange, nobody has ever found anything wrong with them—that is, no disagreement with experiment. So we will follow the rest of the world—besides, we believe that it is probably perfectly right.' This is another illustration of the success of the methods employed in physics though there is limited understanding of underlying mechanisms. It is unknown where energy resides, or even if it is valid to ask such things.

What is certain is that energy transfer that changes the binding energy or, equivalently, the potential energy of a system is associated with a change in mass. A number of experiments have verified that the mass of an atom changes after either the emission or absorption of a photon by an amount equal to E/c^2 to an accuracy better than 1 ppm [17].

8.4 The Role of Forces

If it is accepted that energy is always conserved, the rules of interaction can be reformulated to ensure this is the case and in this way completely removes the need for potential energy, even as a concept.

Consider first the case of a single atom or molecule. It will absorb and emit photons as electrons change state. It is possible to track individual photons to a destination point from where they are emitted. From the viewpoint of any observer, energy is conserved through the interaction. The role of the perturbing forces involved in the process is very interesting. The force facilitates the energy transfer process, but it is essentially a secondary and indirect role. The perturbing action does not provide energy, the perturbing force is not necessarily the direct source of energy.

This energy transfer mechanism is seen by an inertial observer as a transaction, and energy conservation is ensured no matter how the transaction is observed. For example, the observer can slice the space-time manifold into three-dimensional hypersurfaces of simultaneity. With this foliation, there is the need for a propagating photon to ensure energy is conserved. However, working in retarded time (the lightlike frame of relativity) there is no need for photons that physically propagation; energy disappears at one point with the same amount (after relative velocity adjustments) appearing elsewhere. Because the probability of photon emission can be altered by applying an external force (a field), it is clear that the role of the force is to mediate the process in some way.

What if this were true of all forces, that a force is essentially a mechanism for accessing energy, a conduit only? For interactions other than the case of electrons is atoms and molecules, we do not know what the energy transfer mechanism could be, or at least we are unaware that such a process might be taking place[8]. Nevertheless, we can investigate the possibility of the existence of such processes associated with other forces by focusing on energy conservation; because we do not know the energy transfer mechanism, an energy reservoir can be introduced which becomes a source and sink for the energy change associated with the force[9].

This is illustrated in Fig. 1 for a mass subject to the gravitational force. The 1-D representation is of a mass oscillating in a gravitational field[10]. As the test mass (on the left) is pulled towards the centre the velocity increases and it gains energy. Having passed through the centre, energy is given up as the mass is slowed down.

The advantage of representing the process in this way is that there is no longer a question of how forces create or destroy energy, and the concept of potential energy becomes entirely irrelevant. But more importantly, it explain where energy comes from and why energy is conserved. Instead of forces somehow creating and destroying energy, the energy is merely transferred from another

---

[8] Presumably because there is such a volume of transactions that individual events cannot be isolated.

[9] These reservoirs represent ignorance of the process and do not necessarily need to exist!

[10] Though the same concept is equally applicable to events ranging from chemical reactions producing heat - kinetic energy - or the compression of a spring (essentially a quantum mechanical change in the lattice structure).

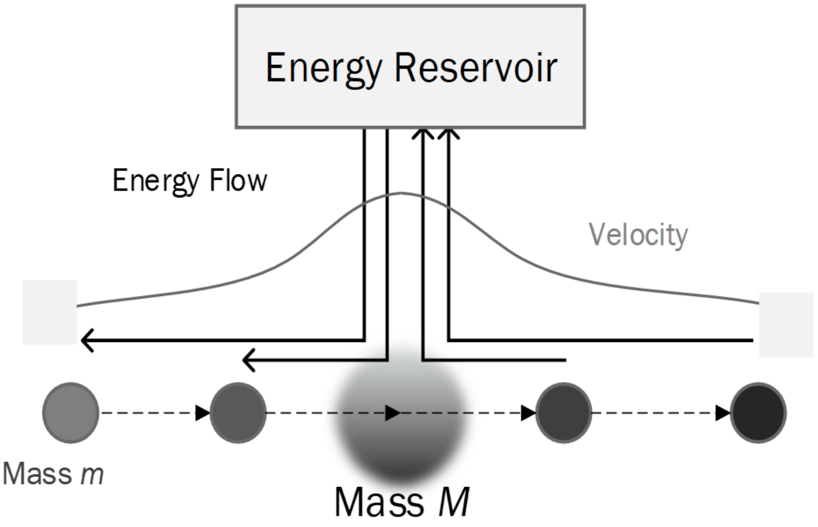


Fig. 1. Mass m is drawn towards larger mass M. At it does so, it gains kinetic energy until it passes through the central mas, after which it is slowed down. The diagram shows the energy flows to and from a hypothetical energy reservoir.

location[11]. The quantity of energy transferred depends on the masses and the position, but also on the the gravitational constant, G, and it is conceivable that G could vary over space and time depending on the availability of energy. For example, the expansion of space will put energy into the reservoir as all mass does moves against the gravitational force and is slowed down, whilst the condensation of matter into stars and galaxies will take energy out of the reservoir (energy that is later converted predominantly into electromagnetic radiation)[12].

But remember the reservoir is only a convenience - ultimately energy events need to associated with transactions; it may be that energy is gained only as energy was lost from an earlier time and that there is a direct transaction mechanism analogous to the propagation of the photon (or, equivalently, the graviton); but we know nothing about such processes - they may even run backward in time rather than forward.

Though we are no closer to understanding what energy is, we can at least say that potential energy is not real by this interpretation. However, it is potentially still useful as a concept because the energy acquired through the gravitational

---

[11] This is particularly relevant to general relativity. The Einstein field equations are essentially a machine for determining curvature. Because the notion of a gravitational force has been eliminated, the gravitational field has also gone and there is no indication where the energy a test mass gains originates from, or where lost kinetic energy goes.

[12] With the implication that without the gravitational energy supplied by the expansion, stars and galaxies could not form.

force can be dissipated in a number of ways, with the result that the kinetic energy can then be less than is necessary to escape the force. By comparing the kinetic energy with the potential energy calculated in the normal way, an overall negative value indicates the test body is bound. Again referencing the electromagnetic force, it once systems become bound that quantum mechanics come into play, or at least present effects that are observable to us.

Of course the entire reformulation may be no more than a very contrived and artificial explanation for the action of forces, but it is impossible to prove the idea wrong[13]. All energy gain can be claimed to a energy transfer from somewhere else, and we know that any theory that cannot be disproved is of limited value.

8.5 Conclusion

Energy is conserved, and there is no obvious explanation why. With a full understanding of energy transfer processes, there is no need for a concept such as potential energy, but as it stands, potential energy though undoubtedly unreal, is a useful mechanism for keeping tack of energy. However, potential energy is not necessary to ensure energy conservation.

The real problem is our total ignorance of what energy is. The ancient Greeks had an understanding of energy as something originating from fire, and though we may now look on this as naive, arguably no progress has been made. We have merely pushed our ignorance of what energy is to a deeper level.

References

1. R. L. Coelho, On the concept of energy: History and philosophy for science teaching, Procedia Social and Behavioral Sciences 1 2648–2652 (2009)
2. R. P. Feynman, R. B. Leighton, and M. Sands, The Feynman Lectures on Physics, Addison-Wesley, (1963)
3. J. Coopersmith, Energy the Subtle Concept: The discovery of Feynman's blocks from Leibniz to Einstein, OUP Oxford, (2015)
4. H. Bondi, Conservation and non-conservation in general relativity, Proc. R. Soc. Lond. A427 249–258, (1997)
5. C. Hoefer, Energy Conservation in GTR, Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics, 31, 2, 187-199, (2000)
6. E. Hecht, Understanding energy as a subtle concept: A model for teaching and learning energy, American Journal of Physics, 87, 495 (2019)
7. S. Carlip, Kinetic Energy and the Equivalence Principle, American Journal of Physics, 66, 409-413, (1998)
8. Article 'Potential energy', Encyclopædia Britannica (2020) https://www.britannica.com/science/potential-energy [Accessed 20/06/2020]
9. R. P. Feynman, R. B. Leighton and M. Sands, Feynman Lectures on Physics Vol II, Basic Books; New Millennium ed. (2011)
10. S. Rainville, J. K. Thompson, E. G. Myers, J. M. Brown, M. S. Dewey, E. G. Kessler Jr, R. D. Deslattes, H. G. Börner, M. Jentschel, P. Mutti and D. E. Pritchard, A direct test of $E=mc^2$, Nature 438, 1096–1097 (2005)

---

[13] Photons subject to the gravitational redshift effect will be affected.

9 The Influence of Future Absorbers on Photon Emission

Summary

Absorber theories of radiation are a type of interaction model used to explain some of the apparently non-local aspects of quantum theory. The basic idea is that interactions are somehow influenced by potential absorbers from the future. Though the theories are widely acknowledged as interesting and imaginative, they are not in any way considered representative of reality because processes that originate in the future do not conform to established physical theories based on causality, and the mechanism of influence is completely unknown. As a result of the obvious conflict with our accepted world model, absorber theories have had little impact over the years. Here we will consider the possibility that both absorber theories and standard physics are correct rather than in conflict, and suggest alternative models of the universe where both could be true. The analysis reveals an unusual picture of a universe that is holistic and deterministic; in addition, a novel interpretation of the purpose and nature of forces emerges.

9.1 Introduction

When a charged particle is accelerated, electromagnetic radiation is emitted in the form of a stream of photons. The process is described effectively using only the concept of a local force acting on the particle and causing the acceleration. There is no requirement that the entities that will eventually absorb these photons, the 'future absorbers', should play any part in the process. Nevertheless the suggestion that future absorbers influence the photon emission process has been repeatedly made.

A mathematically consistent absorber theory was proposed by Wheeler and Feynman in 1945 [1]. The basic idea exploits the time symmetry of Maxwell's equations which allows electromagnetic waves to travel both backwards and forwards in time. The full electromagnetic interaction is presumed to be a combination of the influence of forward traveling waves and waves working backwards in time from the future.

Though the mathematics works, the concept is not considered completely sound because causality is violated. Nevertheless it has inspired a number of related ideas that try to address some of the conceptual difficulties of quantum theory. Rather than make a distinction here, we will classify all variants of the basic theory, including transactional interpretations, as essentially absorber theories because they share a common feature, the direct influence of future events on the present. In conventional physics, the future only comes into existence once it has been created by the advancing present; in contrast, in absorber theories the future is already there.

Quantum mechanics has an intrinsic randomness that is not thought to be the result of a lack of information or understanding (i.e. not the result of simple 'hidden variables'). Though quantum randomness cannot be explained by causal effects, it may be possible to explain it as acausal interactions involving future absorbers of the type proposed in the Wheeler-Feynman absorber theory. A notable example of the acausal hidden-variable descriptions of quantum theory is the de Broglie-Bohm interpretation that proposes a guiding equation that spans the entire universe [2], but the theory has never been incorporated into a unified world model and does not convincingly explain why quantum mechanics works so well as a local theory. Other examples include the work of John Cramer [3], James Woodward [4], Heidi Fearn [5], John Gribbin [6] and to an extent Julian Barbour [7]. These theories sometimes reference Mach's principle, a suggestion that inertial effects are an acausal effect. Hoyle and Narlikar [8] have extended the basic absorber theory to take into account cosmological models.

It is clear that there is at least some philosophical value in investigating absorber theories, but can they be reconciled with quantum theory? We will consider ways this could be done, though the consequence is likely to be a need to modify our existing world model. We will first look in more detail at the non-local aspects of quantum theory.

9.2 Cosmic Experiments

Paul Davies in his book *About Time* [2] relates the story of American astrophysicist Bruce Partridge who took a microwave transmitter to the top of a hill and pointed it to the heavens. The transmitter orientation was slowly altered to test for a change in output power, but there was no discernible variation. Partridge was checking the Wheeler-Feynman absorber theory which proposed that photons are only emitted when a signal traveling backwards in time (from the future absorber) is received by the emitter. In the directions where there is a deficit of potential absorbers, so the theory goes, the transmission power should be lowered. The suggestion may very well be valid, but in any line of sight there are potential absorbers out to enormous distances limited only by the end of the universe or the diminishing particle density that is the inevitable consequence of the expansion of the universe (not forgetting the reduction in energy to an unusable level by an eventually massive expansion redshift).

Partridge's microwave transmitter almost certainly did not approach the energy density needed to probe the future, but we can surely conceive of a source of power that exceeds the integrated absorption density along its travel path. If the emission power of such a source were found to depend on the availability of absorbers, it would become necessary to reappraise the notion of photons as entities that freely propagate through space. It should be noted in passing that Partridge's experiment in itself could not have provided conclusive evidence for the correctness of the Wheeler-Feynman absorber theory because once one admits the possibility of collusion between absorbers and emitters, there can be a multitude of competing theories on offer.

It is unlikely that a powerful enough beam could ever be generated here on earth to properly test the conjecture, but cosmological objects can produce narrow beams of incredible energy density. Examples of these are quasar jets and gamma ray bursts (GRB). Indeed, a possible correlation between the direction of gamma bursts and availability of absorbers (in the form of galaxies) may well be an example of the successful cosmic scaling of Bruce Partridge's experiment: Prochter and Prochaska and their team [3] examined the spectra of GRBs and quasars, focusing in particular on the Mg II absorption lines. Correcting for distance variation and other weighting factors they found that Mg II absorbers were four times more likely to be located in the line of sight of GRBs than in the line of sight of quasars. They considered dust, intrinsic effects and gravitational lensing to be unlikely explanations for this strange discrepancy.

A typical quasar has a power output of around 10^{39} J s^{-1} spread over the entire spectrum [4] - although BL Lac objects are more luminous. In contrast, GRBs can have a peak luminosity of 10^{46} J s^{-1} [5], though the difference in photon fluence (the integrated photon count) is less pronounced because GBRs tend to emit at higher frequencies than quasars. It could be that the photon density of GRBs exceeds the typical density of matching absorbers for some emission cones, therefore a jetting event is suppressed in these directions. Although GRB activity may be intrinsically isotropic, symmetry is broken by the need to select emission directions with at least the minimum number of absorbers.

The GBR evidence, though admittedly tenuous, may hint at a flaw in the established view. It is currently accepted that photons are emitted in response to local causal influences (superimposed with quantum-mechanical randomness); photons then propagate freely until a near encounter with a potential absorber having a matching energy requirement. Although there are some difficulties with the model, not least the fact that no photon has ever been an observed in transit, it is still considered the best available interpretation of the electromagnetic energy transfer process. But what if this interpretation is wrong and emission only occurs if and when a particular future absorber is singled out? In that case it is superfluous to invoke waves traveling backwards in time to mediate the effect. One may instead consider the process merely as a transaction - energy 'disappears' from one location and 'appears' at another (the quantity adjusted of course for various redshift and recoil effects). Certainly we can talk about some particle, a photon, propagating from source to destination as a convenient mechanism to maintain energy conservation between the emission and absorption events, but even this is not an absolute necessity because energy is automatically conserved by applying alternative foliations to space-time (working in the retarded frame, for example). And at the very least the difficult problem of the nature of the propagation medium is resolved (because no propagation medium is required).

The interpretation of the electromagnetic interaction as a transaction also clarifies some deeply puzzling aspects of quantum mechanics. Particularly strange is the apparently non-local behavior of two entangled photons [6] where a measurement to determine the polarization of one of the photons results in the balancing polarization becoming associated with the other photon, even though the detection events are causally disconnected. However, if the photons are understood to be emitted as an agreed transaction only when the state of the detectors is consistent at the projected absorption times (regardless how they reach that state), the experimental results are trivial to interpret and are as expected. The actions of the experimenters between the times of emission and absorption is often emphasized in this class of experiment and linked with the collapse of the wave function, but with this alternative methodology which speaks of an acausal arrangement, the actions of the experimenters are clearly irrelevant. One might say the results of this and many related experiments at the very least show the current understanding of the electromagnetic interaction is incomplete and in need of revision (rather than accepting that quantum processes are incomprehensible).

Interaction as a two-, three- or many-body arrangement presents us with a universe that is completely deterministic. The universe is required to be a composite entity, a completely interlocked machine that progresses through its life cycle rearranging its parts in a coherent way in the process. Change is driven from the top, not the bottom as seems natural from our own lowish position as observers - and one should avoid the anthropomorphic tendency to look at individual particles, atoms and molecules (which are merely changing as part of the greater whole) and attempting to attribute to them decision-making capabilities. For instance, one cannot legitimately inquire how an atom knows at what point of time to emit a photon and in what direction - this is only possible with knowledge of the entire system, knowledge not available to its parts. It is interesting to note the concept of local decision-making is the presumption of our established physical theories and intrinsic to them: this shows how radically different an holistic model would of necessity be.

If the universe is truly of this form, the search for understanding is clearly going to be extremely difficult. All that can be said from the outset, and even then with little conviction, is that the universe develops in an unknown way from a point of beginning presumably towards a point of termination. It is possible to investigate global action by extrapolating from the apparently local effects, but there is no certainty that the process can ever be understood because of the complexity accompanying the extremely high degree of connectivity, though there may be simpler overriding principles controlling the development which may be accessible. We also seem to lose the natural concept of the universe developing in and with time, driven by actions from earlier time.

This is speculative of course, and there is not even a hint of a mechanism by which actions are coordinated across all of time. In addition, if the idea of photon exchange as a transaction is inconsistent with quantum mechanics it cannot be correct.

9.3 Non-Local Effects in Quantum Theory

A general one-dimensional quantum mechanical system is characterized by eigenfunctions $u_n(r)$ of energy E_n where $n = \{0,\ 1,\ 2,\ ...\ \}$, the total wave function develops over time t in the following manner [14]:

$$\Psi\left(r,\ t\right) = \sum_n c_n\left(t\right) u_n\left(t\right) e^{-iE_n t/\hbar}, \tag{9.1}$$

where *$c_n(t)$* are the expansion coefficients. This the general solution of the Schrodinger equation

$$-\frac{\hbar^2}{2m}\frac{\partial^2\Psi(r,t)}{\partial r^2}+U(x)\Psi(r,t)=i\hbar\frac{\partial\Psi(r,t)}{\partial t}. \tag{9.2}$$

Note that U is the potential energy derived from the internal forces acting on mass m. External influences do not enter into the equation.

A system initially in a particular eigenstate is generally driven into a mixed state by a perturbation after which there is the probability of a transition to a new eigenstate with the absorption or emission of a photon. The probability depends on the value of the coefficients, and in the case of a periodic perturbation, the frequency and amplitude. If the perturbation is expressed as a small modification to the Hamiltonian operator, $H^{'}(r)\, cos\ \omega t$, a first order approximation of the transition probability for a system moving from a lower energy state n to a higher state m, is

$$|c_{mn}|^2=\ \frac{|H^{'}_{mn}|^2}{4\hbar^2}\frac{sin^2\left[(\omega_{mn}-\omega)\,t/2\right]}{\left[(\omega_{mn}-\omega)\,/2\right]^2}, \tag{9.3}$$

where

$$\omega_{mn}=\left(E_m\ -E_n\right)/\hbar \tag{9.4}$$

and

$$H^{'}_{mn}=\ \int u^*_m H^{'}\,(r)\ \ u_n\ d\tau. \tag{9.5}$$

Integration is over the entire space volume. The probability is a maximum when the driving frequency exactly matches ω_{mn} with a peak value of $\frac{|H^{'}_{mn}|^2 t^2}{4\hbar^2}$ where t is the time elapsed from the onset of the perturbation[1]. These equations describe stimulated absorption, though the probability of a reverse transition from level $n{\rightarrow}m$ is exactly the same (stimulated emission). Both types of transitions are of course also dependent on the occupancy state of energy levels involved. A little more complicated is laser emission which is the result of emitted photons themselves stimulating the emission of additional photons leading to an avalanche effect.

Let us break this down as little further and examine what is happening in relation to the possible influence of distant absorbers. An point in favour of distant influences is that the wave functions are infinite in extent and there is no consideration retardation either in the potential function or the wave equation itself, something that is suggestive of instant communication.

In contrast to stimulated emission arising from clear external influence, it is known that a completely isolated atom in an excited state will also spontaneously drop to a lower level in the absence of a measurable electromagnetic field. It may

---

[1] Only time-varying field has an effect.

be argued that electromagnetic effects are long range and there will be a residual field everywhere in the universe that cannot be eliminated - these transitions events are simply atoms responding to this tiny field; but the evidence is that the spontaneous emission rate is constant and it unlikely that the residual field would be the same everywhere.

This anomaly with spontaneous emission can be effectively addressed by assigning a new local property to the vacuum. Following second quantization (treating space in the manner of an harmonic oscillator and defining creation and annihilation operators that raise and lower the number of photons occupying each quantized level), the ground energy state of any oscillation mode is found to be non-zero. The energy of the ground state, the zero-point energy (ZPE), is presumed to be real and becomes the trigger for 'spontaneous' emission[2]. Even empty space has this field and it must be included in equation 5 where it will trigger emission. Whilst the field can trigger emission, it is not energy that is available to the atom for absorption.

By this device, stimulated and spontaneous emission are neatly unified. The idea that space is a medium able to oscillate in a range of frequencies, with the amplitude of oscillation determined by the number of photons present seems surprising, but it is correct. Even more surprising, the behavior of a complex system can be derived without requiring any knowledge of where the the photons physically are. Consider a closed system such as a three-dimensional box, filled with atoms that can emit and absorb photons. The number of photons per unit volume in the range of allowed frequencies (ν) is

$$dN = \frac{8\pi\nu^2}{c^3} d\nu, \tag{9.6}$$

and the energy of each mode is

$$E(\nu) = \frac{h\nu}{e^{h\nu/kT} - 1}, \tag{9.7}$$

where k is Boltzmann's constant and T is the absolute temperature. Combining equations 6 and 7 gives the energy density Planck distribution function:

$$E(\nu)d\nu = \frac{8\pi h\nu^3}{c^3} \frac{1}{e^{\nu/kT} - 1} d\nu. \tag{9.8}$$

The photon distribution is sustained by the mirror distribution of elevated energy states of electrons in atoms within the cavity (the Fermi-Dirac distribution). Once equilibrium is reached, the probability of absorbing a photon is balanced by the probability of emitting a photon (as determined by the Einstein coefficients - essentially derived from equations 4 and 5). If more energy were added to this closed system, the result would be an increase the temperature with the photon and electron distributions reorganising to reflect the new temperature. Though the system can be generally understood in classical terms

---

[2] The Casimir effect is a strong indication the ZPE is a property of the vacuum, and there is further evidence from the behaviour of quantum dots.

as the interaction of electromagnetic waves within the cavity with dipoles, the photon interpretation is needed to obtain results that match observation[3].

Within this closed system, the normal interpretation (as described in the technical description above) is to presume internal forces as currently understood define allowed energy levels, and that energy changes are the result of the same forces applied externally and operating by random quantum rules. By this interpretation, individual energy transactions become blurred in the sense that the location and propagation of the photons that participate in each event becomes largely irrelevant, and it is not even necessary to consider the detail of what happens during the photon emission and absorption processes. However, it is also possible within this landscape to state that the transitions are instead initiated by future absorbers, in which case the sea of photons presumed to occupy the cavity do not actually exist, but are an illusion created by the interpretation of the transitions as spontaneous and random.

There is no photon absorption and emission process to be concerned with. It can be in instantaneous transaction. However, it is unclear what conditions must be satisfied for the transaction to take place. Clearly there needs to be an element of energy matching, but there may also be phase or oscillation to account for the wave nature of the process. It is certainly an interesting exercise to investigate the possible rules.

However, it would be misleading to claim that solving the problem in this restricted domain is sufficient. The problem is far more challenging. Consider the blackbody cavity with a small hole so that some of the radiation can escape. Photons can possibly travel billions of years into the future before they are absorbed. Understanding this and spontaneous emission remain huge challenges.

9.4 The Zero-Point Energy

In the case of spontaneous emission, there appears to be two distinct and separable factors that influence the transition rate: the initial and final wave functions (which generally only have a significant amplitude close to the source); and a constant perturbation ZPE field of energy $\frac{1}{2}\hbar\omega$. In effect, the probability of interaction depends on an internal process and an independent external factor, both of which are essential: Without the zero-point field, there can be no photon emission.

These equations predict the probability of photon emission and absorption to great accuracy, though it is not possible to say what specific event will take place or when. The events are largely determined by local effects, and the randomness, the attribute which could be ascribed to the distant absorbers, is only a small factor that is of little relevance to the calculations. In addition, the probability of a photon being absorbed is dependent on paths in space - we see this clearly in a twin-slits experiment. It is difficult to see how the varied and variable distribution

---

[3] One might expect the modes to be coherent with photons in phase, and this is indeed the case - light of blackbody origin will show an interference pattern if a single frequency is isolated.

of absorbers throughout the universe combined together can effect emission in the simple manner of the zero-point energy, or how the path between emitter and absorber could be incorporated into the calculations, or even how the laser effect could have anything to do with distant absorbers.

The idea of the future affecting the past is attractive, but a facilitating mechanism is unknown. It is tempting to compare the outcome of quantum state collapse with the tossing of a coin where the possible outcomes and their relative frequency depend only on local factors, and complex 'hidden' physics determines the precise outcome of a particular event. However, such a comparison is not valid; in the case of the coin the forces (though complex and very much dependent on minuscule initial conditions) can in principle be calculated. That is not the case with quantum mechanics - all hidden variable explanations based on causal influences contradict the result of experiments. The influence of future absorbers is acausal and is therefore a way of explaining away the randomness, but, has been demonstrated, quantum mechanics works perfectly well using only local forces. How would you slot this new 'force' or, if you prefer, 'influence', into existing equations?

Though quantum theory is not completely satisfactory, not least because of the renormalisation process required to subtract away infinities in the calculation made to quantify the electromagnetic force, there is no getting away from the fact that the drivers are essentially local not global (as we have just seen), and there is no hint whatsoever of an influence from distant absorbers. But before discarding the possibility of future absorbers influencing photon emission, we need to recheck some of our assumptions. For example, the wave function solutions seem to emerge instantly in complete form from Schroedinger's equation in spite of spanning the whole universe (and integration over the entire universe is necessary to determine the transition probability). How is this possible? Are these wave functions interacting with future absorbers? This idea does not work for a number of reasons. First, the wave functions as they have been generated are all in current or proper time not future time. The second problem is that the interaction probabilities are already incorporated in the waves and do not depend on any point interaction with future entities. The wave functions are products of the process and cannot in themselves be the link with future absorbers.

It is the potential, the forces, that mediate and control the process. Any connection must be there. We assumed forces were local in this entire analysis. Though an natural interpretation, what if this is not correct? If we wished to modify the quantum calculation process to give both the correct overall transition probability (as is already the case) and also predict specific transition events (which is not currently the case), we would need to modify the forces (potentials) so that they are not merely local but are also channeling the influence of future absorbers. In other words, when we state the electric force between two charged particles is:

$$F = \frac{qq'}{4\epsilon_o \pi r^2}, \tag{9.9}$$

we are saying that the interaction constant is a function of future absorber properties, perhaps the combination of complex influences spanning space and time lumped into a single scalar value. If we try and modify the equation to include these future absorbers it is not necessary to add to the force in any way; the influence of distant absorbers is already there else the transition probability would not be correct. Instead the force must in some way be split (or quantized) to reveal its composition.

If this is true, should the interaction constants (including Planck's constant) not change with time? Not necessarily - the number of absorbers is effectively the number of fundamental particles in the universe (electrons and quarks, minus the source particle(s)) and this is number has been more-or-less constant since the big bang.

We began with the challenge of reconciling absorber theories with quantum theory and have found that this is indeed possible, but the unavoidable conclusion (if absorber theories are correct) is that we occupy a very organised and highly connected universe, and forces are one aspect of this organisation that exist with a clear function[4]. We need to reconsider what forces might be. In the next section we will start to reexamine the possible role of forces in a fully connected universe.

9.5 Forces

What is a force and why must there be forces in the universe? The question in the way it is worded already makes a strong assumption concerning the universe - things exist for a purpose and these 'things' include forces. What therefore is the essential role that forces play in a purposeful and ordered universe?

We can of course sidestep the question and describe a force only by its effect, which is to cause acceleration when applied to entities that participate in the force, but this does not tells us what is the function of a force and what its purpose is. We may propose that the purpose of a force is to move or return a system to an alternative preferred state. This is quite obvious and natural: if we stretch a spring it pulls back; a force comes into existence that was not there before.

This explanation is certainly applicable to the electromagnetic force between particles with opposite charge. As two opposite-charged particles are drawn together the residual force (or field, if you like) detected by a distant observer becomes smaller and almost disappears. In terms of purpose, we can say that the act of creating the opposite charges gave rise to a force with the function of returning the system to its equilibrium chargeless state. The force is local and is not able to distinguish between the same charge on different charge carriers. In other words, it does not just apply between the specific pair of opposite charges that presumably were created together, but on the charge attribute when possessed by any charged particle.

---

[4] This is the logical consequence of forcing absorber theory into agreement with quantum theory.

What has been described is a force that saturates, a force with a clear role (though complicated by the existence of other forces and both space and time). The weak, strong and electromagnetic forces are of this type and can be described as local because they have no obvious global influence.

Gravity is different and it may be why there has been so little success in integrating gravity with the other three forces. As gravity draws together mass, the influence grows rather than diminishes.

An additional point to note is that opposite charged particles are drawn together to the point they can eventually annihilate, but only if all other attributes of particles are similarly matched. This is the case with an electron and positron for example. The mass, the property which is subject to gravity, does not seem to be important to the process.

We have already offered an 'explanation' for saturating forces as a restorative action that in effect tries to reverse time, but we might also ask what the purpose of the gravitational force might be. It is considered to operate on the scale of the universe, though it is still very much a local force with an influence that falls of rapidly with distance. The magnitude of the force will reach a peak only when all the matter in the universe is in close proximity, which suggests an origin from around the time of the big bang when all matter actually was in close proximity.

But as suggested in the previous section, we may ask if it is really legitimate to claim that forces are entities separate from mass, energy, space and time. There is an alternative. By demanding that the universe must be consistent we can create a logical structure in which forces find a natural place. We can postulate an holistic deterministic universe free of anomalies and infinities. A universe of this type would be made up of entities that always connected in some way and able to interact. We may then suggest that the system acts to maintain this consistency on a global level through direct control at a local level, the manifestation of this control being the local forces we have described. By this interpretation, forces though apparently local have a global origin, as was proposed at the end of the previous section. We are also able to widen the definition of a force as any process that acts locally to maintain global order. For example, the 'rule' that two identical fermions cannot occupy the same quantum state can be identified as a force by this definition, the degeneracy force.

9.6 Discussion

A way that both absorber theory and standard physics to co-exist has been suggested. The 'cost' is a complex but coherent universe in which the forces in conventional physics are reinterpreted as an approximation arising from treating as a local effect what is really a global control mechanism. The loss of information makes the forces fuzzy and results in the apparent incompleteness of quantum theory. This is very different to the usual understanding of forces in fundamental physics where the objective and direction of research is to try to reduce all forces to a single one. This approach has led to the unification of the weak, strong and electromagnetic forces which are identified as a local effect involving participating entities that can essentially be treated in isolation. However, attempts at grand unification are faltering and progress in terms of a deeper understanding of the universe has been rather disappointing. There is nothing in the standard particle model to explain where the particles come from and why they are the way they are; in fact there has been no real advance in this direction since the mid-1970s. We have some very nice mathematical structures that describe parts of the universe, but they are not tied down to anything. The model leaves us no wiser about what forces are or their purpose.

The alternative as presented here is perhaps at this stage no clearer, but it has the potential to answer some tough philosophical questions. A coherent deterministic universe is proposed where the future already exists and cannot be altered; this follows because the future is seen to be contained within the present. Although we as observers are locked into a particular time we can, in theory, explore the future through a detailed examination of the specific moments and directions of photon emission. This is Partridge's experiment again. With sufficient data we can construct a picture of the universe extended in both space and time, and fully connected - each present-day energy absorber pulling on the past in the same way they are pulled by future absorbers.

One major objection to this model is that causality is clearly violated because of the influences from the future, but this has no consequence because the future cannot be altered. Causality violation is difficult to detect because there is no energy flow back from the future, but can be inferred from the acausal nature of quantum interaction.

Another problem for some is that action-at-a-distance is reintroduced, mysterious influences acting over space and time. However, the idea that some sort of contact is needed to transfer energy is based on day-to-day experience and is not a necessary component of a coherent universe. By working in advanced and retarded time, energy can be moved without violating any conservation laws and action-at-a-distance is possibly then irrelevant.

The most serious objection may be that the proximity of near absorbers does not have a direct and measurable effect on the emission time and direction of photons. But this might be unjustifiable bias: expecting near absorbers to be more influential than distant absorbers could just be another facet of our (incorrect) assumption that forces are local and not global.

A distinction must be made between systems internally bound by the electromagnetic force exchanging photons and the same forces acting in unbound systems. In the latter case a classical analysis where the forces can be treated as local is entirely effective. But the force in each case is the same and it is unclear how an event such as the collision and subsequent deflection of charged particles on free trajectories can also be influenced by distant absorbers, if, as is claimed, that is the explanation for the local force. Of course, it is possible to convert unbound interactions into quantum form through the introduction of virtual photons (and this is extremely effective); the challenge then is to identify virtual photons with the distant charged particles responsible for stimulated and spontaneous emission in such way that non-local influences become hidden or masked.

A way forward to address these very complex issues could be the use of artificial intelligence (AI). A basic approach would be to create a model of the universe of N particles in 4 dimensions (3 space + 1 time), imposing the constraints of energy and momentum transfer, and allowing energy to be transferred at the speed of light. The deep learning algorithm will look for rules for emission and absorption, as how the rule varies with increasing N, and specifically how this rule will look like to a local observer locked into a particular timeline (i.e. individual particles)[5]. This approach is closely related to the Wolfram Physics Project where computation is used to construct universes from simple rules [19]. Many interesting universe are created. However, it would preferable to derive the rules for the data that is available rather than the other way round[6].

But regardless of the next step, it is important to investigate all alternatives to the accepted pragmatic approach, which is that quantum effects are intrinsically random. This is because randomness is essentially the failure of causality, and if that is the case, we should consider all possible modes of failure of causality (with future absorbers as only one example).

9.7 Conclusion

A deterministic model of the universe emerges if we postulate that absorber theories are correct and then fully consider the implications.

---

[5] Information about the nature of future absorbers could also come from the operation of quantum computing, which likely to replace existing technology over the next generation.

[6] Also worth investigating is the quantum interpretation of a force as a sea of virtual particles. Could these virtual particles be the connection to future absorbers? Another approach would be to start removing particles from the universe in a thought experiment - would the removal of one change inertia in any way, or alter the value of the fundamental constants? What if we keep subtracting until there is only a handful left? In addition, quantum theory tells us that energy is connected with time, and we may have to start looking at stranger options, for example the existence of more than one time dimension, perhaps even one for each particle. Ultimately we would expect this holistic model of the universe to do what the current models do not, that is to explain why particles exist and why they have the properties they do have. Though we have no useful information yet, we can already ask questions such as why all electrons are identical. There are a number of possible explanations: it may be that the universe allows only one set of electron-type properties (for as yet unknown global reasons), or the creation of electrons may be a cloning process. We can take it one step further. They may all be the same particle appearing as many in the curving mirror that is space-time, and are created by time anomalies in the beginning. Is this really impossible - think about a person going back in time; two of will exist, at least for a moment. And the process can repeat with the exponential growth of copies. Michael Dummett discusses the issue of time travel in a book chapter, and discusses conditional subjunctives, counterfactual conditionals, backwards causation and causal chains, all relevant to the discussion here [28]. And if you destroy a single electron (and a single proton), you then destroy the entire universe. The universe may be more fragile than we think.

In this coherent universe, future absorbers will influence the photon emission process through nuances in the apparent force at local level (which, remember, is of global origin, channeling the effect of the potential absorbers into the local frame). The convention of a photon being emitted by an atom at a time for no particular reason (though under some sort of overall control) then wandering through all of space until it meets up with an entity able to accept the energy is incompatible with a deterministic universe and is rejected. We may compare the opposing local and global viewpoints to road traffic control: traffic lights control local flow but global coordination is necessary to ensure the maximum vehicle throughput. A motorist coming up to a junction can not always tell if these lights are on a sequence determined just by local traffic rules or if they are part of a wider coordinated traffic management system. Similarly we may not be able to easily identify global influences in the operation of physics locally, perhaps because our mechanisms of analysis are rooted in causality.

This is a very much more complex model of the universe than the existing model. With the accepted model we can calculate the effect of one electron on another using equations that draw only on information concerning the two particles (such as charge, mass, position and velocity); in the holistic model, as many as 10^{79} particles are involved in the process.

We humans simply may not have the machinery to understand the operation of the universe in the manner proposed. Our brains are optimized for causality on a local scale and it is no surprise the development of science has relied heavily on day-to-day analogies. It may be that a description of the world in a way that can be at least partially understood, though not necessarily correct, may be the only description that is acceptable[7].

And if we choose to investigate this further, how might we proceed? We can conceive of an emitting body in the lab with the property that there is no intrinsic preferred direction of emission. Over an extended period of time, we could monitor precisely the emission direction of each photon that leaves and build up a map[8] that may reveal tiny anisotropies which reflect the influence of the future absorbers on the emission process. This would be analogous to the cosmic microwave background radiation map, but looking forwards in time.

---

[7] In other words, continue with the process of standard physics rather than speculate about ideas that are inherently beyond our comprehension. The weakness in our machinery for thought is evident from the astonishment some scientists and philosophers express that anything should exist at all (rather than nothing). This highlights a specific flaw in our fundamental thought processes because things actually do exist. The philosopher Hegel convincingly argued that 'Being' and 'Nothing' are the same and he poses the question of why should nothing exist, or more precisely, why the existence of nothing is our preference[29]. This may be connected with conservation processes, but does the conservation of energy law and the likelihood that the total energy in the universe is likely equal to zero really directly imply that nothing should exist; it only implies that mass-energy should not exist, not other forms of existence with which we can never be familiar.

[8] Following a correction for peculiar motion.

References

1. J. A. Wheeler, and R. P. Feynman, 1945,Reviews of Modern Physics. 17 (2–3), pp.157–161
2. A. Fine, 1996, 'On the interpretation of Bohmian mechanics', in: J. T. Cushing, A. Fine, S. Goldstein (Eds.): Bohmian mechanics and quantum theory: an appraisal, Springer, pp. 231-250
3. J. G. Cramer, 1986, Reviews of Modern Physics. 58 (3), pp. 647–688
4. J. F. Woodward, 1996, Foundations of Physics Letters, 9, 1, pp. 1-23
5. H. Fearn, 2015, Journal of Modern Physics, 6, pp. 260-272
6. J. R. Gribbin, 2003, Schroedinger's Kittens, Phoenix
7. J. B. Barbour and H. Pfister (Eds.), 1995, Mach's Principle: From Newton's Bucket to Quantum Gravity, Springer Science and Business Media
8. F. Hoyle and J. V. Narlikar, 1964, Proc. Roy. Soc. A 282, 191
9. P. Davies, 1995, About Time: Einstein's Unfinished Revolution, Viking. He is referring to this work: B. Partridge, 1973, Absorber Theory of Radiation and the Future of the Universe. Nature. 244. 263-265. 10.1038/244263a0
10. G.E. Prochter, J.X. Prochaska, H.-W. Chen, J.S. Bloom, M. Dessauges-Zavadsky, R.J. Foley, S. Lopez, M. Pettini, A.K. Dupree and P. Guhathakurta, 2006, arXiv:astro-ph/0605075
11. I. Robson, 1996, Active Galactic Nuclei, Praxis Publishing, CHICHESTER
12. C. Guidorzi, F. Frontera, E. Montanari, F. Rossi, L. Amati, A. Gomboc., K. Hurley and C.G. Mundell, 2005, MNRAS 363 pp. 315-325
13. J. S. Bell, 1987, Speakable and Unspeakable in Quantum Mechanics, Cambridge University Press
14. M. Mathews and K. Venkatesan, 1976, A Textbook of Quantum Mechanics, McGraw-Hill Higher Education
15. S. Wolfram, 2020, A Class of Models with the Potential to Represent Fundamental Physics', On-line Publication https://www.wolframphysics.org/technical-introduction/ [Accessed 19/06/2020]
16. M. Dummett, 1986, 'Causal loops', Chapter in *The Nature of Time*, Eds. R. Flood and M. Lockwood, BASIL BLACKWELL
17. G. W. F. Hegel, 2015, The Science of Logic (Cambridge Hegel Translations), Cambridge University Press

10 A DIY Theory of Everything

Summary

Good progress has been made constructing a viable model of the universe, and the discoveries of the Higgs boson and gravitational waves have demonstrated the correctness of the underlying theory. Though the mathematical models are effective at explaining the dynamics of the universe, they say little about the nature of the universe itself. Explanations for the existence of the universe and the particular form it takes have been proposed, but are neither convincing nor compelling. These explanations are usually presented by experts in this field, and the main ideas are briefly reviewed here. Given the lack of progress towards understanding, we ask if expert knowledge of the theoretical models is really a prerequisite - could not the interested non-expert contribute to the process of developing a better understanding[1]?

10.1 Introduction

Most people like puzzles and the origin of the universe is biggest puzzle of all. Humanity has accumulated a vast quantity of information about the world through observation and experiment, and by carefully examining the data it seems that (to an extent) it is now possible to identify the mechanisms that govern change from the smallest to the very largest scales, and we can even make some plausible guesses as to how the universe came into existence.

All our theories are based on the common-sense notion that each effect has a prior cause, and following centuries of steady scientific progress, there are now a number of mathematical models that together describe the world very effectively. The cosmological model is particularly effective, though it needs unproven ideas, perhaps even speculative ideas, such as inflation and the existence of dark matter and dark energy to hold it together. Though there are a few discrepancies still to be ironed out in all the models, it is felt by many scientists that a single all-encompassing mathematical theory, a theory of everything (ToE) will eventually emerge over time[2]. It is unlikely to be dramatically different from what is currently known, and it is a natural endpoint as we continue to make this steady progress.

Would the emergence of a ToE consistent with all data be the end for physics? Probably, but it would not necessarily be the end of the story. We need to make a careful distinction between describing the universe and understanding the universe. Our existing theories are very good at describing but not so good at explaining. Attempts to understand current mathematical models have not

---

[1] DIY stands for 'Do-It-Yourself'

[2] However, one should bear in mind that there are still gaps in our knowledge, which means there is not just one definitive explanation that could fit the data currently available. Other theories really are possible. The recognised models, the so-call consensus models, are the most likely to be correct, but one should still keep an open mind.

been particularly successful, and present a bizarre world on the small scale and an inexplicable world on the cosmic scale. It is likely that once the ToE emerges the problem of understanding what the theory means will remain, and that the model may then need to be passed over to philosophers to try to make sense of it.

Currently, this search for understanding is rather a closed shop with only those with the appropriate highly specialised technical training able to contribute to developments, and it is the work of these same people in attempting to interpret the models that is recognised; though in truth, they may have no greater philosophical legitimacy than anyone else. There is no compelling evidence that the ideas of the experts are better than those of the enthusiastic amateur - the act of classifying and reducing data to a mathematical form is a different discipline to the task of understanding why these rules should exist at all. Edward Arthur Milne, an English physicist active from the 1930s, strongly advocated that the investigation of the universe should be the accessible to the 'common person' and not be the private domain of the specialist[1]. He reasoned that it is important for a person to understand his or her own position in the world, and this of necessity involves gaining as deep as possible an understanding of the world. To this end, the information from the science community is a start point, a convenient distillation of the existing pool of data, and it is from this position the individual should start his or her independent thinking. The point Milne was making is that scientific data is not, as is sometimes perceived, specialised knowledge belonging to the scientific community to be locked behind a wall of mathematics.

We will consider explanations being proposed for the existence of the universe being put forward by practitioners, and consider if there is a role for the 'common person'[3] not just in using this knowledge, but also participating in the development of the deeper philosophical ideas.

10.2 The Problem

People have been observing the world round about for thousands of years looking for pattern and order. It is apparent that the occurrence of natural catastrophic acts such as floods, lightning, earthquakes, fire and disease do not follow a clear pattern, consequently these events are unpredictable. In contrast, things on a cosmic scale appear completely predictable; the turning of the days, the pattern of the seasons, the years, the positions of the moon and stars in the sky. This is rather odd and not what we might expect. Our experience is that if you want to build something big and highly ordered, like a house for example, careful design, good organisation and the supply of a great deal of energy for construction are essential requirements. The house will not happen by itself. And there is a theoretical basis for the feeling that things are not quite right – order should not increase with scale. Our science is built on a 'bottom up' approach where the big

---

[3] This will be an interested reader who will not have an advanced level of mathematical and scientific training.

things are made by joining together the little things, a concept dating back to the early Greek philosophers. Would we really expect this apparent order on the large scale to emerge from the chaos of the small scale all by itself[4]? But we are witnessing the astonishing averaging effect of gravity with order emerging from the concentration of matter and the accumulation of the gravitational effect. It is apparent that we cannot explain the universe merely by appealing to ’common sense’.

The Rules

There are many explanations proposed for the origin and form of the universe, presented both by experts and non-experts[5]. For an explanation to have scientific merit, it should not contradict established scientific knowledge. However, we need to be careful to define what is meant by that. The baseline may not be the same for anyone, but a reasonable threshold is that any theory adheres to the following:

1. It does not violate energy conservation;
2. It will be consistent with causality;
3. It recognises the constancy of the speed of light;
4. It acknowledges that the universe is expanding, but is not expanding into anything;
5. An aether is not required.

In addition, we would expect the proposed explanation to address at least one of the following questions:

1. Why do the elementary particles exist, and only these?
2. Why are particles of the same type identical?
3. What determines the rest mass of particles?
4. What is the origin of inertia?
5. Why is energy conserved?
6. Why do forces exist?
7. Why are there four types of force?
8. Why are quantum interactions apparently random?
9. What is space, and why is it apparently three-dimensional?

---

[4] The ancients, though not perhaps directly aware of an inconsistency, resolved this issue very simply by establishing a hierarchy that had the heavens as the source of order, capable of influencing the chaos of the natural world below. Thus began the search for an understanding of the universe by looking upwards and outwards. The picture has changed dramatically over the centuries to accommodate data from instruments developed to observe the sky in more detail than is possible by the human senses alone.

[5] Everyone seems to be proposing a theory for how the universe came about. To paraphrase de Montaigne, ‘Man is quite insane. He wouldn’t know how to create a maggot, and he creates universes by the dozen.’

10. What is time?
11. Why is the universe expanding and will it continue to expand?
12. What was there before the big bang?
13. Why does anything exist at all?

The list is indicative only and is my no means complete, but at least there is something against which the merit of a particular proposal can be objectively tested. One huge caveat that applies to all attempts to create a model of the universe is that understanding may be impossible because we are part of the universe. Though it is entirely possible such things are unknowable, there is no evidence that knowledge become inaccessible to us as we delve deeper; in fact the opposite is the case. This is discussed in more detail in Appendix 1.

What the Experts Say

A great number of scientific theories continue to be proposed that are broadly consistent with the listed criteria[6]. However, the accepted standard, indeed the reference, is the current set of consensus models that have undergone extensive scientific scrutiny. New scientific ideas that are likely to survive will tend to address or explain facet of the model rather than challenge it, and in this way the model evolves and gains further traction.

There are also attempts made at a level above this to explain the universe, and this is usually with reference to the accepted model rather than the dubious alternatives. The proposals will be conceptual rather than mathematical, and whilst the objective is to gain deeper understanding, this is rarely achieved. For example, one idea that has been proposed is that the universe is a computer simulation [2]. This is possibly an extrapolation of our ability to create simple worlds using program code. It does in a sense answer the big questions - the reason for everything is that they are simply programmer choices. It is hard to disprove such a theory, but it is equally hard to assign any value to it.

There are expert theories based on higher dimensions, the world of string theory where 1-D filaments and sheets (branes) oscillate and interact (branes can be understood as equivalent to 2-D rubber sheets moving in 3-D space). It is possible to conceive of branes being able to move about and collide, thus giving rise to a normal universe [3] [4] [5]. A recurring motif is 'bubble universes' that appear, grow and perhaps eventually 'pop' within a higher dimensional manifold.

Scientists seem particularly drawn to the idea of colliding universes. The reasoning is that if one universe can exist, why not many? But it is based on the hard-to-shift belief that the universe is somewhere or exists in something. But where is the evidence for this – the universe is not expanding into anything? Collisions are therefore impossible. You may argue that universes exist against the backdrop of higher dimensions, but again the onus is on proponents to explain exactly how this is physically rather than mathematically possible. We can

---

[6] This includes steady-state universe theories with observations that are more naturally explained by expansion attributed to tired light, plasmas, etc.

certainly conceive of 2-D objects as existing in our 3-D world, but they physically do not. There is no such thing as a physical 2-D or 1-D entity. Everything in the 3-D universe is 3-D.

But any idea based on string theory is dubious - there is no supporting evidence for the validity of string theory. Objectively, it is not even satisfactory as a concept. It is a response to the failure of the standard particle model which hit a brick wall about 50 years ago. The model presents fundamental entities such as leptons and quarks as simple building blocks without structure, and to introduce more complex strings as an explanation seems to be a move away from a reductionist direction. Strings and branes are structures with complex vibrational and interaction properties; even if particles were described as strings, it raises the more difficult questions concerning strings and the environment in which they exist/operate. Just because we are all familiar with strings and completely unfamiliar with point particles[7], does not means strings are more fundamental - perhaps they are, but the advocates for string theory have certainly not made the case. Furthermore, the standard particle model is at the point where it possibly needs interpretation, not more structure. Even scientists working with the models do not find they provide the understanding they need. Prof. Dan Tovey of the University of Sheffield made the following statement to the BBC in 2015 just before the Large Hadron Collider started a new run on much greater power [6]:

> 'And the best thing that could possibly happen is that we find something that nobody has predicted at all. Something completely new and unexpected, which would set off a fresh programme of research for years to come.'

He seems to be suggesting the problem is getting a bit boring and tired and needs to be freshened up.

Other ideas being proposed include the concept of multiverses (perhaps 'parallel' universes created from the multiple ways quantum decision-making can determine the future). There are a lot of different multiverse proposals, but they do not really advance understanding in the sense of providing compelling answers to any of the questions above [7]. A popular variant invokes the anthropic principle to explain the specific properties of this universe (out of the many that have existed) as being those needed to produce and sustain life [8]. Again, it is extremely hard to identify with this as any sort of science, and it is a very superficial way of addressing the questions listed above.

The suggestion has been made that the universe is not spherical but in the shape of a horn [9]. This topology has the advantage of explaining a specific anomaly in the cosmic microwave background map, why the smallest distinct shapes are elliptical rather than circular. Barbour has claimed that we completely misunderstand the concept of time and that time does not really exist [10] - it is our brain that creates the illusion of time.

---

[7] Point particles are zero-dimensional, essentially distinct from space and not part of it, whereas strings are one dimensional.

You are probably getting the idea now. Just take any common day object or effect and check if it can be an analogy for the creation of a universe. It does not matter if it is contrived; the predictions are not really testable anyway. It is clearly a free-for-all but only those within the system can join in. One gets the impression that the successful work being done to build effective mathematical models does not really solve the key problems but instead moves the problems to an even deeper level further away from observation, and perhaps this is why the proposals that emerge appear so outlandish.

We are starting to see that it is not enough to establish a sound mathematical theory that describes the universe. Even for scientists, fixing the math is not enough; like everyone else they want the substance, they want to know what it all means and the implications, form and substance. The limited constraint means that the model can reflect the way people want the universe to be rather than what is actually is - it is clear that scientists put their own personal imprint on a theory, and develop an interpretation which rapidly becomes a belief of what is really happening. Mathematical structures make the manikin we can dress any way we like.

We should be aware that it is possible that we cannot ever understand the universe and that a simple model, though not absolutely correct is the best we can do, and that we have to be satisfied with that.

10.3 A New Approach?

In spite of all the suggestions and ideas being presented by the experts, it is fair to say the big and important questions are still there, largely untouched. It is not even clear if we are close to a breakthrough. Shouldn't new approaches be encouraged and promoted?

The Future

Imagine the following scenario. It is the year 2070. It has been discovered that the neutrino is not a single particle but also comes in three different 'colours'. The extension to the standard model to incorporate these facets of the neutrino reveals a multitude of new, heavy and stable particles that have been found to make up the missing mass in the universe (the mass formerly referred to as dark matter). In addition, it is discovered that to incorporate gravity into the standard model, a small extension to general relativity is required, which on a cosmic level acts like dark energy. All observational data, both on the small and the large scales, are found to be consistent with these modified theories. The loose ends are tied up. However, the conceptual questions listed in the previous section are the same questions one might have asked half a century previously, and in 2070 we are no closer to answering them.

Given that the answers are not certain to emerge from the models being developed, perhaps we should not wait for a complete mathematical model to appear before starting to address the big questions; in addition, we could be hundreds of years away from a ToE (based on the current rate of progress), and probably the only chance of a ToE in our lifetime is a visit to this planet by aliens who already possess this knowledge and are able to pass it on.

One of the reason progress is so slow is that the origin and fate of the universe is of surprisingly little significance to people's lives. The drivers are therefore relatively weak - there is little money to be made. The origin of the universe a puzzle, no more. And it is not even certain that the general public want the facts about science if it is complex and inaccessible – an understandable picture may be preferable. Many people are very interested in developments in fundamental science, but it is just that, an interest. The discovery of the Higgs boson in 2014 did not greatly affect society[8]. Consider the huge effort in sending a person to the moon, a big leap certainly, but people have arguably benefited more from more mundane innovations such as putting wheels on luggage.

Another problem when it comes to engaging with the public is there are no charismatic giants driving the process, as was the case a century ago. Significant advances are being made by teams of scientists rather than individuals because there is so much data. We should not be in awe of this kind of hard science - the work is not necessarily difficult. The scientific method is fundamentally simple. Scientists work with real information which is static and generally nothing is hidden, perhaps from a giant table of the redshifts of astronomical bodies. Problem-solving and decision-making is technically easy and conveniently processed by computers[9].

Cranks and Mavericks

Progress made by carefully following the scientific method is slow but for most scientists acceptably slow. People trying to short-circuit the process by jumping to the final answer without reference to the accumulated data are generally ignored, or labelled cranks. Many people have an inbuilt view of the world and their place in it and are perfectly happy with this. Sustaining their worldview sometimes involves constructing an acceptable explanation that encompasses

---

[8] Though there will be benefits in the form of Nobel prizes for some - scientists are not immune from desiring fame and fortune. The scientific community operates on a strict hierarchy with its own celebrities. Nobel prizes play the role of Oscars. The most famous was of course Albert Einstein with people like Richard Feynman as minor celebrities.

[9] Not so when the scientific focus is on people and how they act. The situation is much more complex; we do not really know what people are thinking and what they will do. Problem-solving and decision making are then difficult. Whilst the common perception is that science is hard, in processing terms that is not the case: science is easy but in comparison to day-to-day life. However, the human brain is optimised for these more complex interactions, and most people over a lifetime establish a set of ideas or beliefs for complex reasons, be they social, moral, religious, or political, and will actively pursue strategies that will support and sustain these views. To adapt for the less natural scientific work requires extensive training and may involve studying the same subject at university for a number of years.

their own understanding and draws on knowledge and experience of the world acquired over a lifetime and will include analogies from other spheres of interest. For example, it is not unusual for an engineer who has spent a lifetime working with electrical machines to try to explain everything in terms of magnetic fields. Experiences from childhood or school (such as science demonstrations or a compelling teacher) can also have profound influences in later life, and these early experiences can have an effect that lingers into adulthood. Many of the ideas are just concepts, or based on analogies that perhaps neatly explaining one thing whilst ignoring others, and whilst they would never be accepted by the scientific community, they do show imagination, and it would be wrong to say these undeveloped ideas have no value.

But there is a class of competent scientist familiar with the scientific process and actively working in the field who do not accept that current theories are relevant or even correct. These are the mavericks[10]. Mavericks are often scientists with solid reputations who are able to present their ideas in the appropriate form, and whose objections cannot be casually dismissed.

An example is Herbert Dingle who strongly opposed the special theory of relativity. He was highly critical of Eddington and Milne. Dingle's book on the subject [11] is a very interesting read and demonstrates the huge and honest effort made by an intelligent person to oppose a theory whose structure is not consistent with his own world view, or how he expects the world to be. Whilst dissenting views usually draw the scorn of the scientific community, it is a perfectly acceptable viewpoint.

However, criticism of accepted physics even from respected scientists is a danger to career and reputation[11]. Many of the critics target cosmology, in particular the big bang theory, presumably because the basic assumptions are unproven and rather weak. A need to eliminate the idea of a point of creation (with its religious connotations) drive some towards an eternal steady-state or oscillating universe. The driver might be a personal disbelief in a universe springing out of nothing, or perhaps the concept of infinity is unacceptable. Ideas proposed include chronometric cosmology and plasma cosmology; the cosmological redshift is attributed to tired light losing energy with distance or that the dynamics of the universe are affected by ionised particles. These ideas were developed by Irving Segal, Hannes Alfven and Eric Lerner. Astronomer Halton Arp has suggested that matter is created in the centre of galaxies (and ejected). William Tifft has proposed that cosmological redshifts are periodic or quantized [12].

It is not suggested that these alternative theories are correct or better than big bang cosmology, but it does indicate that big bang cosmology is not to everyone's taste, and more attention might need to be paid to the views of the

---

[10] It is not enough to criticize, a replacement theory has to be proposed. This puts the maverick as a distinct disadvantage because a single person cannot possibly assemble a complete theory that addresses all the data. There are bound to be gaps and new ideas are judged on their points of failure rather than their successes.

[11] For safety reasons, criticism is often disguised behind glowing praise of the theory being criticised.

mavericks. And sometimes the mavericks after initial scorn are proved right, or at least promoted into the mainstream. Subrahmanyan Chandrasekhar was scorned for his idea of black holes in 1930. The Doppler effect was not accepted in the lifetime of Christian Doppler. Fritz Zwicky suggested dark matter back in 1933. All ideas before their time [13].

Religion

The connection between science and religion should not be discounted in this context. Belief systems can be extremely influential as people search for understanding. Many of the significant figures in the development of cosmology had no difficulty maintaining a strong faith without any conflict with the task of understanding Creation. A notable example was Sir William McCrea. McCrea believed that astronomy, and especially cosmology, could never be separated from deeper meaning. At his Royal Astronomical Society 160th anniversary talk, he stated [14]:

> 'After 160 years, the Universe around us that we contemplate as astronomers and geophysicists, is as mysterious as when the Society was founded. We know a great deal more than our founders did about the structure of the Universe. Who would say that we have learned any more about its meaning and purpose? In raising our glasses to the health of our sciences let us hope that in the coming years they will bring us not only still further knowledge of the structure and operation of the Universe, but also some spark of light upon these profounder mysteries.'

The opposing viewpoint is more prevalent. Carroll argues that cosmology and theism are incompatible (though the argument presented appears to be missing the point) [15]. The point though is not to be dogmatic - ideas driven by religious belief are not automatically incorrect.

An Interdisciplinary Approach

Sometimes the cross-fertilisation of expertise across disciplines works. Take the case of Ernst Mach who was born in 1838 in Brno, the second largest city in the Czech Republic. As a lifelong academic he made a significant contribution to the theory of supersonic sound waves. The Mach scale of speed is named after him. He was also very interested in how the mind worked and supported and assisted in the development of the gestalt theory. This theory proposes that the brain makes sense of a complex and chaotic world by establishing a consistent global whole into which data is fitted. The internal world is then not a collection of things but a single complete entity which develops over time as information is acquired. This is an attractive idea consistent with our perception of identity, consciousness, awareness of self and the fact our thoughts run as a single thread rather than a chaos of mixed thought streams. He then extrapolated these ideas to the whole universe by proposing it was a consistent whole rather than a collection of independent parts. He went on to state that a body 'knows' it should continue to move at the same velocity (conservation of momentum) because the other masses in the universe, the distant galaxies, act as a reference. Motion only makes sense with respect to these other bodies. When we try to accelerate a body, we are fighting the pull of the stars which are trying to keep the motion constant. The mysterious property of inertia then has a natural explanation.

This is a great idea of course, but pure speculation. To incorporate it into existing theories, someone would need to propose the force that conveys the information, gravitation presumably, then perhaps set up a mass distribution representing the universe on a small scale and demonstrate how the mechanism works. You can immediately see there are going to be huge problems and no one has yet found a plausible mechanism. Nevertheless the idea has persisted and is still referred to as Mach's principle. It is an imaginative attempt to explain one very puzzling facet of the universe. Note that Mach developed it as analogy of the mechanism by which the brain works, a subject he was very comfortable with.

This is an example of the transference of an idea from a person's unrelated area of expertise into cosmology and finding that it adds value. The worldview of the Greeks and Romans are now known to be incorrect, but it worked for them and included some nuggets of truth that survive to this day. If it were essential that all theories should be absolutely correct, we would not have any theories.

The Contribution of the Machine

But we should be aware of the possibility be aware the possibility that significant advances in knowledge might not come from any human. Advances in AI, and specifically deep learning, over the last few years have been remarkable, accelerating the technology after a half century of much promise but very little achievement. The older neural network approach was to work with configurations that mimicked the organisation of neurons in a biological brain. Deep learning is not particularly concerned with the structure of the artificial neural network used to analyse the problem but instead presents sufficient neurons and the layering of neurons to enable pattern recognition to occur when significant computing power is applied to large data sets.

The ability to find patterns and relationships without it being necessary to entirely understand the problem suggests that deep learning could profitably be applied to solving difficult problems in science. Machine algorithms are typically tested using the game of chess where the performance is easily quantified. Like physics, chess had an early classical phase with clear rules that guided strategy, then a hypermodern phase from the 1920s where it was discovered that the rules should be broken, and now there are the chess playing machines. Google's ApHaZero program learned the game by playing against itself for 4 hours only. In 2017, it played the best chess program of the time, Stockfish, and won convincingly [16] [17]. The surprise was that the machine had a completely new approach to the game and adopted strategies that were unfamiliar to human players[12]. The next major advance in cosmology of particle physics might be just to present a program of this type with the data available and see if anything new emerges, and this is starting to happen [18] [19].

10.4 The Role of the Amateur

We have outlined how the scientific community and machines might advance knowledge, but what can the amateur do? There is already many roles for interested members of the public as 'citizen scientists'. This involves either manual data processing or the donation of computer time, or perhaps participating in rudimentary data collection, and the collaboration often has huge training and dissemination value. However, participants individually tend to have a minor role. It is unusual for a citizen scientist to initiate and lead a project, and with the way science funding is structured, it is hard to see how it could be otherwise. The philosopher Feyerabend has criticised the way science is controlled by the state and institutions and has called for a greater role for the public [20][13]. But is this really feasible - can the non-expert really make a meaningful contribution? It is argued here that the amateur can make a contribution to cosmology because the problem is clearly defined and very little expert knowledge is needed in order to attack the deeper problems. The amateur will usually accept that standard models are correct (or largely correct) without really being concerned with the detail, and start investigating the big questions identified earlier. In some ways the value of the development of current theories has been the removal of clutter to clearly expose these really difficult questions. For this exercise the enthusiast need not be too concerned with data, the realm of the professional scientist, but instead focus on meaning and interpretation.

The idea of advancing knowledge by thought alone rather than by extrapolating from data is referred to as the rationalistic approach, and though common in the past when scientists were data poor, the approach is now out of fashion (though still valid). It is observational data presently driving theories, and certainly not aesthetic or philosophical considerations. The rationalist epistemology has become progressively less effective over time. As our knowledge of the world improves, an increasingly effective mathematical description is created, but the equivalent physical picture is becoming quite incomprehensible. A good example

---

[12] These three stages of development are evident in many areas including music, art, electrical power systems, and even how we shop.In the latter case there was the small shop when an assistant acted as an intermediary, then there was large-scale self-service; the next phase is online ordering and delivery powered by personal computing.

[13] The referenced work is a very interesting read but is an extremely brutal critique of science. He writes: 'There is hardly any difference between the members of a 'primitive' tribe who defend their laws because they are the laws of their gods, or of their ancestors and who spread these laws in the name of the tribe and a rationalist who appeals to 'objective' standards, except that the former know what they are doing while the latter does not.'

is of this process is quantum theory. If we cannot even understand the concepts that emerge from the data, how could we ever generate them in a rational fashion? It is because of this stark fact the emphasis today is on the ‘bottom up’ empiricist epistemology: Logic triumphing over conjecture.

In the first half of the 20th century there was a limited amount of data available and scientists such as Paul Dirac, Arthur Eddington and Edward Arthur Milne happily speculated on elementary particles and the structure of the universe. In addition, they explored the relationship between the very large and the very small. These philosopher/physicists had no problem with idle speculation or even trying to force or squeeze the universe into their own belief systems.

How does a person in the first quarter of 21st century go about building a universe, explaining how approximately 10^{78} particles were made during a frantic period around the point of creation?

Though it is impossible to present a specific strategy for developing a TOE, some guidance and suggestions on some key topics to focus on, and alternative ways to approach the problems, can be suggested[14] that will possibly enable the amateur to contribute to the problem in a meaningful way. Listed in the subsections below are some possible brainstorming topics and ideas.

The inspiration here might be the Augustinian priest Gregor Mendel [21]. From 1856 onwards, he conducted some 29,000 experiments to find out how characteristics of plants such as flower colour are determined. He discovered that the colour was not inherited by blending of the parent flower colours as was believed at the time, but in a discrete manner that ultimately revealed the operation of genes. His breakthrough results were ignored by the scientific community for the next 30 years.

The tool of the present-day amateur will be creativity and imagination, and perhaps a computer running a model of the universe.

Conservation of Energy

Understanding the universe requires at the very least an understanding the nature of energy.

Energy is often defined as ’the ability to do work’ or ’the capacity to work’, but whilst this is acceptable for a day-to-day understanding of what energy is, it not a particularly useful description on a fundamental level given that everything that is substantial in the universe seems to be energy or a facet of it. It is not really the ability to do work in this context. Energy is certainly a measurable property of all systems, and in a closed system energy is always

---

[14] Feel free to consider imaginary (though not impossible) scenarios, for example a universe with only 3 atoms (which was presumably the case at some stage in its history). With 3 atoms, each has a dimension of space to itself – perhaps forces are only necessary when 10^{78} particles have to share these dimensions.. How would it work out? Thought experiments like this come naturally to us – one interpretation of our dreams is that we play out scenarios that help categorise and store information and solve problems.

found to be constant[15]. As the only completely closed system known is the universe itself, we can infer that the total energy in the universe should also be constant for all time. But how much is this? The exact quantity is unknown. A rough calculation suggests the figure is pretty close to zero because the energy associated with the mass of all particles in the universe (and adding in all the radiation) happens to be almost completely balanced by the negative binding energy of the gravitational forces holding everything together. Certainly many scientists would like the total to be zero, if only to exclude a possible violation of energy conservation in the early history of the universe. A bonus is that zero total energy would conveniently dispose of questions regarding where energy first came from.

As to what energy it, it is as fundamental as space and time and equally puzzling. The models we have of the universe provide no explanation. Energy is treated as a fictional entity that enters into the laws and is manipulated by them. We find there is no more efficient way of explaining causes and effects than recourse to the concept of energy. The laws accurately predict how energy is transformed, but we have to enter in the rest mass of fundamental particles using measured values – the models do not explain why particles have these specific intrinsic energy values. We can however predict how energy varies from that start position.

And we come back to the question of what energy is and why it is conserved. You may care to speculate on this yourself, given there are no answers. Note that conservation of energy is not something we demand in day-to-day life and when there are clear violations, it causes us no problems. Just have a look at TV and movies. Superheroes routinely violate the conservation of energy law with no complaints from the viewing public[16].

An Energy Audit of the Universe

As you might do in your home, conduct an energy audit, this time of the whole universe. Add all the mass, including photons and neutrinos, and make the appropriate adjustments for binding energy. In trying to do this, the problems immediately become apparent. For example, how do we deal with time? Do we mean the energy as it appears to the observer, or do we consider a 'proper' universe where everything is the age of the observer? Does there need to be energy input at the start to get things going? This type of calculation will bring to the fore the very important principle of the conservation of energy.

For your audit you may choose to look at the time of the cosmic microwave background radiation[17], when the universe was about 380,000 years old and the temperature was around 3,000 K. Ignoring inflation and assuming the universe

---

[15] Or perhaps more accurately, is defined in such a way that it is found to be constant - it is actually surprising that this can be done.

[16] A number of people believe in the idea of psychic activity without wondering about the energy source.

[17] Noting that Olber's paradox is not a paradox in the microwave spectrum.

is expanding at the speed of light, the total photon energy is about 10^{64} J. Given that 26% of the universe was helium at that time (by mass), we calculate the energy released by fusion assuming 10^{78} baryons to be 2.7 x 10^{65} J, if we start with protons and electrons, or 3.9 x 10^{65} J beginning with neutrons. You can investigate from there where the missing energy might be (kinetic energy, gravitational energy, neutrinos etc), and decide if inflation is needed to account for the difference.

Free Energy

Can we find a new easy-to-extract energy source somewhere in physics to address the impending energy crisis? The answer is, unfortunately, no. Of course, many claims have been and are being made, but none can be substantiated.

Elaborate machines have been designed that supposedly create energy from nothing. Some of these invoke complex magnetic effects. Magnets can stick together indefinitely, but that does not imply the continuous availability of energy. No continuous supply of energy is required to keep them together - they remain stuck because the total energy in all the magnetic field is less than when the magnets are separated; hence work has to be put into the system to separate the magnets. This is rather counter-intuitive and it is not surprising that people have tried to invent free energy generators based on the imaginary flow of energy between the two bound magnets (or coils[18]).

Designs associated with changing the magnetic field by chopping are characterised by elaborate arrangement of magnets and magnetic circuits, as well as the extensive use of mu-metal. It is believed that the effort to distort the magnetic field is negligible in comparison to the electric power that can be generated over the disturbed magnetic field, but that is not the case.

Free energy designs are really perpetual motion machines which violate energy conservation laws. A typical design uses a large flywheel which is turned by a pulsed motor. Apparently the impulses associated with the pulses add more energy to the flywheel than a constant force. There is no reason why this should be the case.

Attempts to access the possibly infinite zero-point quantum field around are sounder, but no one really knows how to do this. There is no way in existing theory of releasing this energy.

Time Travel

Is time travel possible and is there a place for it in a consistent and unified world model? Though we would normally dismiss the idea as impossible, there

---

[18] Nikola Tesla straddled the creative boundary between the conventional science and fringe science. He is responsible for many inventions and his ultimate aim was to extract energy from the air, something we do now with air source heat pumps. Many free energy ideas are based around Tesla coils, a loosely coupled coreless transformer.

are some precedents. For example, in quantum mechanics, a particle trapped behind a barrier can spontaneously appear outside the barrier with exactly the same total energy as it had in the contained state[19]. This is interpreted as borrowing energy to overcome the barrier[20], but it can also be thought of as a spontaneous space transition. If a space transition is possible, could a time transition be possible, either backwards or forwards in time?

It would explain one of the curious properties of electrons listed earlier – the fact they are all identical. If a particle is able to move back in time, there will at the earlier time be two identical particles and so on[21]. This would be a mechanism for creating matter with the energy required for the time shift equal to the mass created. The two electrons could then not occupy the same space at the same time because they are the same particle – the origin of the repulsive electric force and within an atom, the exclusion principle (or degeneracy forces). There are probably far-fetched ways of making this work.

Quantum Theory

A free electron will radiate energy when accelerated, and this is accurately described by classical theory. This rule does not seem to apply to a bound electron - the electron appears to lose its identity once bound into the atom and is no longer able to radiate as it would outside the atom. Experiments also indicate that free electrons have strange wave-like properties, generating an interference pattern when passing through a narrow grating. We are forced to conclude that classical physics is not completely correct, no matter how unwelcome this conclusion might be. This realisation dawned on scientist between 1911 and 1930. The crucial step made in understanding what is happening inside the atom was to postulate that an electron should be treated as a wave once bound inside the atom. The result is quantum mechanics, but it gives only probabilities that something will happen, not a direct causal result as we would obtain by applying Newton's laws of motion. Such an interpretation would normally be completely unacceptable, except that it makes phenomenally accurate predictions.

The mathematical foundation of quantum theory is formidable, but the interpretation is less satisfactory. If the amateur can contribute anywhere to the theory, the place is in the interpretation. It is just not possible that any completely

---

[19] This is referred to as tunnelling. And this is not some obscure theoretical concept. There are devices based on this principle such as the semiconductor tunnel diode (parts 1N3712 – 1N3721 for example) and the scanning tunnelling microscope. Have a look at http://phys.org/news/2015-05-physicists-quantum-tunneling-mystery.html [Accessed 16-07-2020] which discusses the tunnelling time.

[20] You might ask where the energy is borrowed from? In any case, this is quantum land - the particles really exist as waves. The confusion of the barrier might only exists because we are incorrectly applying the classical model of a particle in a box.

[21] It is not the only way to make identical particles; for example, the creation process may have only certain viable 'solutions' - hydrogen will combine with oxygen in only one way, making identical water molecules.

new theory could have the prediction success of quantum electrodynamics. Any alternative theory of interaction you might provide must therefore reduce to standard quantum theory if you expect anyone to take it seriously. If your theory does this and dispenses with the randomness, perhaps by including ‘hidden variables’, factors which are not evident and control the interaction, then you have solved one of the most difficult problems of physics.

Two scientific giants in the development of quantum theory are Schrödinger and Feynman. Schrödinger was a complex thinker who routinely mixed science and philosophy. He was influenced at an early age by the writings of Schopenhauer. In the debate concerning the existence of ’chance’, or if everything is determined, he stated in 1922 that chance was not necessarily a lack of knowledge but could genuinely be considered to exist. This view certainly aided and influenced the development of his quantum formalism (culminating in the Schrödinger equation in 1926). However, he felt dissatisfied with this interpretation and from then on advocated an underlying determinism which remained to be discovered. From 1930 onwards, his writing on quantum mechanics becomes increasingly philosophical, including consideration of the relationship between indeterminism and free will, but his works are full of interesting points and well worth reading. For example, he explains that energy is in some cases about quantity, but on the interaction scale, it is about frequency (i.e. time) [22].

Feynman developed an interest in mathematics and taught himself advanced methods in school[22]. His talent was understanding clearly what the equations meant and conveying this meaning to others through analogies. Feynman was extremely creative and allowed imagination and maths to work together to bring forth novel and compelling ideas. For example, he proposed that a particle samples all possible paths to a destination and takes the path of least action. He showed that this is reasonable interpretation of quantum theory. One piece of advice to a class was, ’The first principle is that you must not fool yourself - and you are the easiest person to fool’.

Randomness

That randomness is a key aspect of quantum processes is perplexing. How can randomness exist? For example, a nucleus undergoing beta decay will emit an electron (or positron); but why at that particular point of time? How can a nucleus have any decision-making capabilities? [23]. If the decay was a result of internal factors, we might expect to be able to alter the decay rate, for example

---

[22] In 1979 Omni magazine labelled him the world’s smartest man to which his mother commented: ‘If that is the world’s smartest man, God help us’.

[23] Beta decay is completely random and nothing we do can change the decay rate. The chain reaction associated with energy generation by fission is an example of neutron absorption affecting decay rates and is not the same, nor the fact that a neutron is stable in the nucleus but unstable on its own. The stimulated emission effect associated with the electromagnetic force does not seem to have an analogue in the weak and strong forces.

by increasing the temperature. And at what point does an atom in an excited state decide to emit an electron? What triggers that discrete event? Quantum theory gives no insight.

A way round this is to propose that actions are not local but influenced by the future – a photon is emitted by an atom because an absorber in the future has been identified. The advantage is that we no longer have to ask how particles can have decision-making capabilities – they do not, they are part of a coherent connected whole. This view of a completely deterministic universe is disliked by many because of its incompatible with our concept of free will, but that is not a strong objection. The idea is particularly attractive because the laws of physics are reversible in time. If we run the universe backwards, then every emitted photon will go to a specific absorber.

One way to investigate this further is to construct a model of a small number of simple identical atoms with hydrogen-type electron energy levels. Construct a snapshot of the universe over all time. An electron can make a transition to a lower level if an electron in future time can make an identical up transition, the velocity of each changing in the process. The search process (probably using genetic or evolutionary algorithms) will find a solution where the interactions are consistent. In other words, there is not example of a possible transition that does not happen. Note how the emission probability changes by increasing the number of atoms and decreasing the iteration time step. This describes spontaneous emission – what rules would need to be added to explain stimulated emission?

In proposing that there is some unity in the universe, it follows that an explanation for the properties of elementary entities is contained somewhere in overall structure and within the development history of the universe. In turn the particles in their fundamental properties will reflect the universe as a whole; 'a whole world in a grain of sand', borrowing the words of William Blake.

The idea of a deep link between the large and small is well established in subjects such as biology: the largest land mammal on earth is made up of about 10^{15} cells, but each elephant cell in its DNA contains the entire knowledge of the creature. In contrast, the link between the small and large in physics and cosmology is not known (and perhaps does not exist), and we can at this stage merely speculate. The holistic model would imply that decision making is not local but global and driven by an overriding flow and order, but what keeps track? What sets the pulse of the universe keeping everything synchronised?

The holistic interpretation has the potential to provide an understanding for rules that have been observed to determine the dynamics of the universe when observed locally. However, causality is violated.

Do we Really Understand Forces?

Consider the case of two hydrogen atoms. They will share the two electrons through covalent bonding to form a hydrogen molecule. But why does this happen? There is an advantage in closing shells by filling each with the maximum number of electrons allowed, even though this unbalances the charge. The reason is not simply that the unbalanced total electron charge creates a dipole as this would not results in the strength of binding that is observed. Also, it is not sufficient just to claim it is just a quantum effect.

Regardless of the explanation, it is certainly a force and not necessarily a direct derivative of the electromagnetic force. Similarly, the exclusion principle prevents two electrons occupying the same quantum state. Should this not be associated with a force?

But the problems go deeper. If the quantum states of an atom are calculated using Schrödinger's equation, the states (in the form of wave functions) span the entire universe. How are these states established instantly across the universe? It is possible to get around this by assuming each particle is 'born' with its own universe-wide wave function whereupon the calculated states are merely superposition effects of extant wave functions, then what then is the effect of the expansion of the universe on these states? Should this be included in quantum calculations? A bound electron moving between states by emitting or absorbing a photon is also very difficult to understand in terms of wave function transitions.

It is easy to get exasperated by such difficult questions and ask if we really understand anything at all.

Action at a Distance

Because of our experience, it is natural to think of energy to be transferred by physical contact, for example collisions, or photons that leave an emitter and head towards a target. But we need to be wary of mapping our brain world onto the real world, as mentioned earlier. In reality, all interaction is action-at-a-distance (with the dashes because it is a 'thing'). There is no such thing as physical contact as there is no solidity at the particle level[24].

The concept of a force as a mysterious entity that acts across empty space has been largely eliminated in the Standard Particle Model of Physics by proposing instead that forces arise because of the exchange of particles called bosons. These carry energy and momentum. It is a nice way of explaining repulsion but the analogy is again flawed – it does not work with attraction. In addition, the photon that acts as the force particle for charge carries and transfers momentum but, in some cases, no energy. A real photon cannot carry momentum but no energy. This problem is circumvented by identifying the exchange particle as a virtual photon. The model does not currently incorporate gravity, but it unifies the other three forces. Quantum field theory is used to calculate the probability of interaction, and the strength of interaction. The model is self-consistent and complete, but in order to work it needs about 20 numerical constants with very precise values that have to be chosen and entered by hand. There is no way of

---

[24] Everything is action at a distance. Steven Wright tells a joke: 'In my house there is a switch that does nothing. Every so often I flick it on and off. Yesterday I got a call from a woman in Germany. She said 'Cut it out''.

deriving the values from fundamental principles. However, it does not explain how these bosons are able to travel to carry the forces[25].

Most of the information used to validate the model comes by examining the fragments from the high-speed collision of elementary particles in a particle accelerator. The Large Hadron Collider (LHC) is the main instrument used. The standard model does not predict any new particles, and indeed any new particles that may be discovered by the LHC would just mess up the model.

Infinity

Do infinities exist in nature? The mathematics that describes the universe permits infinities, and mathematicians routinely deal with infinitesimals when applying integral and differential calculus. However, there is no strong evidence that infinities exist in the real world[26]. It is difficult to have effective conservation laws in the real world if there are infinities lurking about.

But look at how the universe deals with the *possibility* of infinity - when an electron approaches a proton, the accelerating electron does not rush towards it gaining infinite kinetic energy and negative infinite binding energy; instead quantum effects come into play to prevent the total collapse. Of the other end of the scale, the universe had a start point, but this is receding away from the observer at least the speed of light because of the expansion of the universe, preventing us 'seeing' beyond the start point.

What if it is the case that the universe will always introduce a new rule to avoid a singularity or discontinuity? It certainly seems to be the case, though a notable exception seems to be black holes. There is little doubt that huge aggregations of mass can form, but there could very well be a new effect that comes into play to prevent an event horizon being created, perhaps the ejection of mass as jets[27].

---

[25] Many scientist will claim this does not need to be explained, as without an aether there is no answer. Physics spent a lot of time getting rid of the aether, a pre-existing background for the universe but it keeps coming back in so many guises..

[26] Wigner famously spoke about how mathematics, essentially a human construct, is so effective at describing the universe [23]. However, mathematics is no more than an approximation of the universe. The concept of infinity in mathematics causes huge problems with the result that mathematics in its entirety lacks consistency. Gödel's incompleteness theorems show that it is impossible to derive a complete and consistent set of axioms that underpin the whole of mathematics (As proposed by Hilbert[24]).

[27] If you wish to investigate black holes in greater detail, you need to study general relativity. Though black holes are often described in Newtonian terms, that is misleading. In classical physics the total binding energy is measured with respect to infinity. Though this may be greater than the photon energy, the binding energy with respect to an absorber, say, 1 km away from the surface is allowed. A Newtonian black hole will therefore not be black close up!

Even a quasi-Newtonian approach is misleading: In Newtonian terms, it is hard to see how a genuine black hole can have a very large mass – any new material has

The Idea

The lack of progress to date in understanding the universe is perhaps surprising, but bear in mind there has not been a concerted effort made by active scientists, and much of the activity has been incidental. The feeling though is that a fresh look from the outside could result in a breakthrough. Halton Arp in his controversial book 'Seeing Red' [25] writes:

> 'At this point, I believe we must look for salvation from the non-specialists, amateurs and interdisciplinary thinkers - those who form judgments on the general thrust of the evidence, those who are skeptical about any explanation, particularly official ones, and above all are tolerant of other people's theories.'

Of course, comments like this can never make it to peer-reviewed journals. But in popular literature, the theme recurs. Lee Smolin in a critical assessment of progress in physics states [26] (paraphrased):

> 'The one thing everyone who cares about fundamental physics seems to agree on is that new theories are needed... We are missing something big... How do we find the missing idea? Clearly, someone has to either recognise a wrong assumption we have all been making or ask a new question, so that's the sort of person we need in order to ensure the future of fundamental physics... It goes without saying that people who are good at asking genuinely novel but relevant questions are rare, and that the ability to look at the state of a technical field and see a hidden assumption or a new avenue of research is a skill quite distinct from the workaday skills that are a prerequisite for joining the physics community. It is one thing to be a craftsperson, highly skilled in the practice of one's craft. It is quite another to be a seer... The seer must know the subject thoroughly, be able to work with the tools of the trade, and communicate convincingly in its language. Yet the seer does not need to be the most technically proficient of physicists.'

A new idea will cut through the existing model and start the process of understanding the nature of the universe.

10.5 Sharing the New Idea

A distinction should be made between an idea that addresses the current limitations of the accepted physical models and an idea of personal origin and value.

---

binding energy equal to the mass energy hence the net equivalent mass addition is zero. However, general relativity applies in this situation, and in general relativity gravitational energy does not gravitate, which makes the issue of energy conservation in general relativity problematic. In general relativity, an event horizon does form. Entities inside the threshold radius are disconnected from the rest of the universe and unable to interact through the electromagnetic force.

If the non-expert constructs their own model of the universe that is satisfactory to them, and they believe it has the creative edge that the practitioners appear to lack, then it is natural to try to share this idea. But for anyone trying to present their ideas as some sort of universal truth, the rules change. That is science, and most people have to be trained into the process. Science works by making factual steps based on the data available. If what is presented does not also match the body of accumulated data, it will attract little wider interest[28]. To help establish the fundamental soundness of your idea, you can do some basic checks. Referring back to the list in Section 2, you will at the very least need to show that energy is conserved; in other words, there is no way of executing a closed cycle of extracting energy (or losing energy). Of course, your basic premise may be that energy is not conserved, but if that is the case, the burden of proof becomes massive! You will also have to check all the other items on the list[29].

The chance of acceptance is particularly small if the idea has to displace an existing idea that has by definition already passed through the process of proof. The scientific community is very protective of its efforts and an idea that chips away at the base of the edifice is particularly unwelcome. Current theories are the end-result of perhaps as much as a billion hours of human thought. The edifice has incredible momentum because of those who participate in the process and who have invested their life work in it. For a single individual to be able to make changes that involve backtracking and discarding some of the accepted content in order to correct errors is almost impossible. To contribute to the models, a person has to buy into the whole construct and can only add to it (and only after a long period of support and verification at that). If you have a radical theory, perhaps the best that can be achieved as a first step towards acceptance is the approval of a like-minded group of individuals who may appreciate your idea. For this, the internet will be invaluable.

10.6 Conclusion

Geniuses wanted to solve the mystery of the universe[30]! We have already seen that none of the big questions have been answered. The puzzle is interesting

---

[28] Established practitioners are incredibly unforgiving and intolerant – all new ideas are treated with distrust. It is just the way science works. Only well-proven ideas are incorporated into the body of knowledge. You will not have the resources to check your idea against the enormous volume of data available. The current models have the advantage of having been fine-tuned to match all these (without necessarily having predicted too much of it). In addition, a specialist has good knowledge of their own field which may be extremely narrow, and can completely fail to understand a general idea of the type non-experts usually present. They will understand a set of equations which makes predictions, because this is the common language of all scientists.

[29] Check for infinities in your theory and if present how you intend to deal with them. Infinities are not always bad, but they usually are.

[30] And it is evident there are plenty of geniuses working in sales, advertising, design, packaging and marketing.

and there is sufficient data available to enable anyone to have a go at solving the problem[31], particularly as experts in the field of cosmology have made little useful progress in explaining why what we observe is as it is[32].

A deep mathematical understanding of the models is not a prerequisite and the task of explaining the origin of the universe can be thrown open to the non-expert. Anything proposed by the amateur, no matter how inventive or clever, is likely to be ignored by the scientific expert[33], but that should not be a deterrent. The input from many more people can speed up the process of discovery.

Many people are interested in how the universe works and eagerly follow latest developments. However, though the story and the characters develop nicely with some drama here and there (for example the discovery of the Higgs boson in 2013 and gravitational waves in 2015). But it has to be admitted that it is the slowest of all soap operas. The story moves forward at glacial speed and the moments of excitement can be decades or centuries apart. Though the story progresses through the diligent work of thousand of scientists around the world building on existing knowledge with the most extreme caution, it is just too slow for the modern world. We live in the communication and information age. Ideas move across the world in seconds. The plodding telling of the story of the universe can leave the reader on tenterhooks for a lifetime without really learning anything significant about the world in that time. We even suggested earlier in our scenario of 2050 that the story-line is getting very dull and needs to be spiced up – new characters and plots are needed.

---

[31] In the manner of amateur Michael Ventris who decoded the Mycenaean Greek script Linear B.

[32] The experts have had the opportunity to try and answer important question concerning the universe and the results have been mediocre, perhaps because the training that prepares one to work effectively in this mathematical world imposes constraints that hinders genuine creativity, or even cogent philosophical thought. In addition, the elevation of scientific experts to metaphysicians could be an example of the Peter Principle, given the ineffectiveness of what is presented. For example, Max Tegmark, whose contribution to this field is the idea that anything which exists mathematically exists physically as well, writes in a popular science book [27]: "In the United States, recent polls show that 39% consider astrology scientific . . . If everyone understood the concept of 'scientific concept', this percentage would be zero. Moreover, the world would be a better place, since people with a scientific lifestyle, basing their decisions on correct information, maximise their chances of success. By making rational buying and voting decisions, they also strengthen the scientific approach to decision making in companies, organisations and governments." This is of course both pompous and incredibly naive. In addition, astrology in the sense of birth month affecting personality is not necessarily unscientific. Human beings can be born at any time in the year but climate, school-start dates etc are seasonal.

[33] Unlike the expert, the amateur has no vested interest and can re-examine the basic ideas. It is very hard for practitioners to go all the way back to the beginning and question 'obvious' underpinning assumptions. For example, Aristotle understood the idea of force and mass and believed that F = m x v where v represents the velocity. This seemed obvious and was not explicitly tested for 2,000 years as a result.

There are seven billion people on the planet. The 2011 UK census revealed that 27% of the population had a higher degree. Extrapolating, we can suggest that there are perhaps a billion people might be interested in thinking about the questions of existence. We are talking about open source science and philosophy. Of course, there are many online forums and physics and cosmology news groups, but what is proposed here is that people are taught to refine their thoughts and put them in a format that encourages positive discussion and development with reference to their own ideas not current scientific dogma. This might mean presenting the ideas in a simple mathematical form in some cases. Mathematics does not need to exclude but can illuminate as well.

There are ground rules, not least that the current level of recognised scientific knowledge though possibly flawed in some ways, is the reference point. Our existing models represent one of the impressive products of humanity, more impressive than any technological feat such as travelling to the moon or the greatest constructions on earth.

So let us open up the problems from the closed science community to allow non-specialists to approach the problems afresh, using the internet as the dissemination vehicle. The important questions about the universe are not the property of the professional scientist, and the amateur scientist should be able to investigate these with the support and approval of only their own peers without the need for journal publication.

Over to you.

References

1. H. Kragh, 'Cosmology and Controversy: The Historical Development of Two Theories of the Universe', Princeton University Press, p 70 (1999)
2. N. Bostrom, 'Are You Living In A Computer Simulation?', Philosophical Quarterly, 53, 211 pp. 243 (2003)
3. A. Vilenkin, 'Many Worlds in One: The Search for Other Universes', Hill and Wang (2006)
4. L. Susskind, 'The Cosmic Landscape: String Theory and the Illusion of Intelligent Design', Back Bay Books (2006)
5. B. Greene, ' The Hidden Reality: Parallel Universes and the Hidden Laws of the Cosmos', Knopf (2011)
6. http://www.bbc.co.uk/news/science-environment-32809636
7. H. Kragh, 'Testability and epistemic shifts in modern cosmology', Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics, 46, Part A pp. 48-56 (2014)
8. K. Rudnicki, 'The anthropic principle as a cosmological principle', Astronomy Quarterly, 7 (2) pp. 117-127 (1990)
9. R. Aurich, S. Lustig, F. Steiner and H. Then, 'Hyperbolic Universes with a Horned Topology and the CMB Anisotropy', Classical and Quantum Gravity, 21 (21) pp 4901–4926 (2004)
10. J. Barbour, 'The End of Time: The Next Revolution in Physics', Oxford University Press (1999). *Refer also to arXiv:0903.3489 which summarises the ideas.*
11. H. Dingle, 'Science at a Crossroads', Martin Brian and O'Keeffe (1972)

12. M. López-Corredoira, 'Tests and Problems of the Standard Model in Cosmology' Found Phys 47, pp 711–768 (2017)
13. W. I. Beveridge, 'The Art of Scientific Investigation', The Blackburn Press (2004)
14. W. H. McCrea, 'Address by McCrea, W.H./PROFESSOR, F.R.S. at R.A.S. 160TH Anniversary Dinner', Quarterly Journal of the Royal Astronomical Society, 21, 03 (1980)
15. S. Carroll, 'Why (Almost All) Cosmologists Are Atheists' Faith and Philosophy, 22, 10.5840/faithphil200522523 (2003).
16. D. Silver *et al*, 'A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play', Science 07, 362, 6419, pp. 1140-1144 (2018)
17. M. Tavora, 'Deep Learning Explainability: Hints from Physics - Deep Neural Networks from a Physics Viewpoint', URL: https://towardsdatascience.com/deep-learning-explainability-hints-from-physics-2f316dc07727 [Accessed 31-07-2019]
18. L.-G. Pang, K. Zhou, N. Su, H. Petersen, H. Stöcker and X.-N. Wang, 'An equation-of-state-meter of quantum chromodynamics transition from deep learning', Nature Communications, 9 (1) DOI: 10.1038/s41467-017-02726-3 (2018)
19. S. He *et al*, 'Learning to predict the cosmological structure formation', Proceedings of the National Academy of Sciences, DOI: 10.1073/pnas.1821458116 (2019)
20. P. Feyerabend, 'Science in a free society', New Left Books (1982)
21. G. Mendel, 'Experiments in Plant Hybridisation', CreateSpace Independent Publishing Platform (2017)
22. E. Schrodinger, 'My View of the World', Cambridge University Press (2009)
23. E. P. Wigner, 'The unreasonable effectiveness of mathematics in the natural sciences', Richard Courant lecture in mathematical sciences delivered at New York University, May 11, 1959. Communications on Pure and Applied Mathematics. 13: 1–14.(1960)
24. R. Zach, 'Hilbert's Program', The Stanford Encyclopedia of Philosophy, (Summer 2015 Edition), (2015) URL: http://plato.stanford.edu/archives/sum2015/entries/hilbert-program/ [Accessed 31-07-2019]
25. H. Arp, 'Seeing Red: Quasares, Cosmology and Academic Science', Aperion (1998)
26. L. Smolin, 'The Trouble with Physics', Penguin (2006)
27. 'This Will Make you Smarter' Ed. J. Brockman, BLACK SWAN (2012)
28. It is worth listening to the entire lecture. Transcript available at URL: http://www.bbc.co.uk/radio4/reith2003/lecture2.shtml [accessed 31-07-2019]
29. B. Russell, 'Mysticism and logic, and other essays' Longmans, Green and Co (1918)

Appendix A: Limits of the Mind - Are we Capable of Understanding the Universe?

One argument is that we can understand the universe because we understand mathematics, and mathematics explains the universe. However, that does not necessarily follow. Noted neuroscientist Vilayanur S. Ramachandran in the 2013 Reith Lecture [28] described the case of a patient paralysed on the left side. The patient was presented with a mirror on the right side which was directed in such a way that the left side of the world was reflected back and showed an assistant on her left side holding a pen. She was asked to grab the pen. A person with an undamaged brain would turn to the left and reach for the pen. Instead, she started clawing the surface of the mirror, or tried to go behind the mirror. This is in spite of the patient understanding full well how a mirror works. The patient inferred the object must be on the left, but for her left did not exist hence the object had to be inside the mirror.

What this tells us is that the world is how the brain has organised it in order to function effectively, not necessarily as it really is. That being the case, can be trust our brain[34]?

Our brains are optimized for causality on a local scale and it is no surprise the development of science has relied heavily on day-to-day analogies. We persist with images that are simply wrong: atoms are not balls, but chemists still use ball and stick models to build complex molecules; electrons do not execute planetary-type orbits around the atomic nucleus, yet the logo for the International Atomic Energy Agency[35] (... and just about every national atomic organisation) features just that. However, it may be that a description of the world in a way that can be understood, though not necessarily correct, is the only description that is acceptable.

However, we do desire an explanation that is more than a set of equations that describe the dynamics of the universe. The situation is further complicated by the fact that many humans expect the universe and existence to have a ’point’. The philosopher Bertrand Russell in his 1903 essay "A Free Man’s Worship" wrote [29]:

> ... all the labours of the ages, all the devotion, all the inspiration, all the noonday brightness of human genius, are destined to extinction in the vast death of the solar system, and that the whole temple of Man’s achievement must inevitably be buried beneath the debris of a universe in ruins - all these things, if not quite beyond dispute, are yet so nearly certain, that no philosophy which rejects them can hope to stand.

This is pretty extravagant of course, but scientists should be aware that merely describing how the universe works is not enough for anyone.

But it is not just the limitation of the brain we need to be concerned with – though less likely, we might not have access to all the data required. We have already mentioned the game of chess. Consider a scientist who has not heard about the game being presented with a set, board and the rules. The scientist is intrigued and tries to understand the origin of the game. She notes that pawns move forward one or two squares, the bishop moves any distance along diagonals, the knight has a curious octopus tentacle jump pattern, the rook moves along straight lines, the queen has the power of a rook and bishop combined. The king moves one square in any direction. The adversarial nature of the game leads the scientist to conclude that it is a stylised war game based on medieval principles with the objective of gaining territory and capturing (but not killing) the opposing king. In this context, the move of the pawns as infantry makes

---

[34] The way a scene is coloured by the brain and sound represented internally is remarkable and inexplicable, but it is not the ‘real’ world of waves and particles.

[35] https://www.iaea.org/ [Accessed 19-06-2020]

sense as does the ability of the knight on horseback to jump over pieces. Whilst accepting the moves of the bishop, rook (or castle) and queen are symbolic of their power at the time the rule took on that for. The scientist has deduced a significant amount but can go no further.

The same may be true of our attempt to understand matter, forces, space and time. Whilst we believe they are connected and their properties and even existence can be deduced from a simply stated higher principle, it may not actually be possible from the available information to gain a complete understanding.