Investigating the Foundations of Physical Law
8 Gravity

General relativity or quantum mechanics?

Everyone recognizes that gravity is one of the most difficult problems in physics. To try to make progress here, we will definitely have to go back to fundamental thinking. We are going to have to use the information available from lectures 3 and 4 to make key decisions, and to use the methodology outlined in lecture 2, which rejected compromise and model-dependent thinking in favour of abstractions and symmetries. If we believe what the fundamental symmetries tell us, we will have to accept what they imply, and see if the consequences match up to experiment.

At some point, any physicist thinking about fundamentals is going to have to confront the biggest dilemma facing us at the present time: quantum mechanics and general relativity cannot both be correct. One is going to have to be modified or abandoned. Perhaps both will. Every theory which tries to tackle this problem – for example, string theory, membrane theory, quantum loop gravity – makes such changes. What is changed depends on where you start. It looks like one or other of these theories must be our starting point, because their results will have to be explained in the new theory, and no other viable starting point has ever been proposed. So, is general relativity the basis of fundamental physics or is it quantum mechanics? Interestingly, the bias so far has always been in the direction of starting from general relativity. For some reason, general relativity's appeal to something more tangible or accessible to human experience has prevailed over the abstract calculating engine of quantum mechanics. Nevertheless, I make no secret of the fact that, for me, the only route to a fundamental theory lies in quantum mechanics. There is no doubt that quantum mechanics has been put to far more extensive and rigorous testing than general relativity, and that it explains many more physical effects and in much greater detail and precision. General relativity has explained only a few things and these are at a considerable distance from its core principles.

So, can we accommodate general relativity if we privilege quantum mechanics? My answer is that we can, as long as we realise that it is primarily a *mathematical*, and not a physical theory, and that it is in no way an *explanation* of gravity, but a mathematical description of its effects on space and time, as we *measure* them. This, I believe, is another area where our lack of knowledge of the true history has been damaging to our understanding of the physics. The problem is that the relation between Einstein's and Newton's theories has been determined by the historical verdict that one is a *revolutionary replacement* of the other, the physical description of gravity as a force being superseded by the geometrical description of a Riemannian space-time, in which the time component is indistinguishable from space, although this is in contradiction to its status in quantum mechanics. Such an argument, however, takes no notice of the fact that GR makes no physical sense on its own until it replaces its mathematically-derived curvature with the physically-derived Newtonian potential. Taken on their own, the GR field equations are simply a mathematical description of space-time curvature. It is only when they approach the 'Newtonian limit' that an arbitrary connection with a Newtonian 'physical' force can be established.

The Newtonian explanation uses an inverse-square force, a typical result of 3-dimensional spherical symmetry, analogous to the Coulomb force for charge and responding to the $U(1)$ symmetry group. If we consider this only as an experimental 'limit', we lose a possible fundamental symmetry with a fundamental meaning which we have to reintroduce arbitrarily as an 'approximation' which has no explanation. Since we are convinced that symmetries are more fundamental, in many respects, than mathematically sophisticated theories, we will have to make a decision on whether this symmetry is really broken in nature and, if it is, find some explanation why it appears to be 'approximately' true.

Gravity and quantum mechanics

Going back to quantum mechanics, we have to ask why gravity has never been successfully quantized. Quantum gravity presents us with nonrenormalizable infinities, largely because of the fact that its sources all have the same sign and there is no way it can be 'shielded', unlike any of the other forces. Quantum field theory attributes this to its intrinsically attractive force requiring a spin 2 mediating boson. Simple quantization principles also present us with the worst fit to data ever recorded in physics in deriving a value for the cosmological constant Λ. In fact, it is not only the worst ever fit to data, it is also the worst *possible* fit. A basic quantum-style argument makes Λ about 10^{120} to 10^{123} times higher than experimental data, based on 'dark energy', would suggest. Now, if, as Seth Lloyd has proposed, on the basis of the holographic principle, there are 10^{123} possible 'bit flips of data' in the observable universe over a Hubble lifetime, then the prediction is as wrong as it could conceivably be. Not even 'not even wrong'!

A prediction that is so completely wrong is not so much a problem, more an opportunity. It seems to be telling us that we should be looking for an answer in entirely the opposite direction to the one we have assumed. Gravity is certainly quantum – it is almost inconceivable that anything involving matter could not be – but maybe 'quantum' needs to be taken in a wider context than the one we have assumed. We have assumed that 'quantum' refers only to *local* interactions. However, 'quantum' also demands nonlocality, and, at present, we have no mechanism for nonlocal correlations. Gravity is clearly quite different from all the recognized local forces. Its source, mass-energy, is completely continuous throughout the universe, whereas the other sources come from discrete localised points. In fact the continuity of mass-energy could indicate that it is even more unlike discrete sources than we have already imagined; for example, the Higgs field might well suggest that the amount of mass-energy is the same at every point in space, and that changes occur only in its method of manifestation because of the presence of charges of various kinds.

Gravity is also extremely weak, only equalling the strength of the local forces between particles if summed up on a *universal* scale. Further, the gravitational energy between centres of mass is negative, compared to the positive energy between similar centres of charge, and negative energy, in fermionic states, is associated with vacuum rather than real space. Now a possible route forward at this point, and a way of incorporating gravity into a nilpotent structure, is to suggest that the total vacuum –1($\pm$ $i\boldsymbol{k}E$ $\pm$ $\boldsymbol{i}\mathbf{p}$ + $\boldsymbol{j}m$), which is partitioned by the $\boldsymbol{k}$, $\boldsymbol{i}$, $\boldsymbol{j}$ operators, can be thought of as the negative energy continuous gravitational vacuum, which supplies the mechanism for the instantaneous transmission of quantum correlation. According to this way of thinking, the gravitational vacuum has a special nature as a kind of 'sum' of all the others. Such a possibility is also hinted at by the 3 + 1 nature of charge and mass within the parameter group, and was present in the original theories of particle structures based on it.

A similar concept of gravity / gauge theory correspondence has now emerged in string theory. Essentially, gravity and gauge theory (strong and electroweak) are dual. This is also evident in the way that the holographic principle privileges gravity to obtain information about the entire system. One can be used to provide information about the other. The fundamental duality is that of the nonlocal (gravity) and local (gauge theory), which tells us why gravity is so weak and why it is not obviously a quantized force. (Later in this lecture, we will approach gravitational quantization through the inertial reaction.) It also appears to tell us why the 'cosmological constant' is at the opposite end of the possible physical scale (in information terms) to the one worked out from quantum gravity.

Gravity-gauge theory correspondence appears to show that gravity acts as a dual to the combined gauge theories of the electric, strong and weak interactions, and duality in our language usually refers to negation and totality zero. The same seems also to apply in the continuous combined vacuum, which we have associated with gravity and the negative value of the fermionic nilpotent, –($\pm$ $i\boldsymbol{k}E$ $\pm$ $\boldsymbol{k}\mathbf{p}$ + $\boldsymbol{j}m$), compared to the discrete specific vacua created by the partitioning introduced by the quaternionic charge operators $\boldsymbol{k}$, $\boldsymbol{i}$, $\boldsymbol{j}$. What we are proposing here is that, rather than applying to a single local quantum state, gravity is a manifestation of the total vacuum, –($\pm$ $i\boldsymbol{k}E$ $\pm$ $\boldsymbol{k}\mathbf{p}$ + $\boldsymbol{j}m$), the sole source of nonlocal correlation, and applies to the whole universe at once. Gravity, it would seem, provides the most directly observable aspect of nonlocality. This is entirely the logic of all the symmetry arguments we have so far produced, and we have not yet found evidence that the fundamental symmetries are anything but absolutely true. Our methodology insists on us following this logic without exception.

According to this reasoning, unless something is communicated instantly, then we are unable to complete quantum mechanics – it isn't any of the known local interactions, and, if it isn't gravity, then it must be outside of what we have defined as physics. Of course, nonlocal correlations must create local effects. So, what are the local effects? It would seem here that the local manifestation associated with gravity must be the positive energy inertial local reaction, which is what we really observe and equate to the gravity of localized clumps of matter, and which, for a fermion, we could represent by the nilpotent structure ($\pm\, i\boldsymbol{k}E \pm \boldsymbol{k}\mathbf{p} + \boldsymbol{j}m$) itself. This force, however, is fictitious and repulsive, which means that it could be 'quantized' in the local sense, as a repulsive force, and so require a spin 1 gauge boson, rather than the spin 2 particle which creates the renormalization problems associated with quantum gravity.

The need for a spin 1 mediator and QED-like theory in 'quantum gravity' has been discussed in many of my publications. There, it has been suggested that the continuity of mass-energy, the filled vacuum, the Higgs field, and the need for instantaneous correlation between Dirac states, together with the fact that energy does not actually move (as opposed to the form of its realisation in connection with a discrete state), require an instantaneous gravitational force, which is undetectable by direct observation, and only ever observed through the c-dependent inertial reaction on discrete fermionic or bosonic states, which, being repulsive, requires a mediator of spin 1.

So, according to this way of thinking, gravity provides a nonlocal dual to the combined interactions, which has inertia as its local manifestation. While we can do the calculations for weak, strong and electric forces from the perspective of the local structure of the fermion state, and work iteratively to determine the effect of the rest of the universe, with gravity we have to do the calculation universally or recursively to work back to the inertial effect on the quantized fermionic state. Gravity is the carrier of all the information separately available from the other three forces, and, in fact, of all their local manifestations, but it is their inverse rather than their summation, exactly as reflected in its negative energy, and we can even regard the attractive nature of the gravitational interaction between like particles as a *result* of its being a vacuum, rather than a local force, and therefore requiring negative energy. If this is a true description, gravity will be able to complete quantum mechanics in a way that the proposed (but so far unsuccessful) 'quantum gravity' cannot, because if localised 'quantum gravity' is valid, there will be no vacuum through which it can act.

This has some very interesting consequences, a number of which have been part of this theory from the beginning, but have now begun to appear in other areas of physics. The holographic principle is a conjecture which states that all the information in a system is contained in the bounding area. In the nilpotent structure, the *E* and **p** terms create the effective 'bounding area' (the mass term being redundant as additional information). This is an area in 'vacuum space', bounded by ***k*** and ***i***, but it is dual to an area in real space, and can be observed directly by reversing the roles of vectors (connected to space) and quaternions (connected to charge) in the nilpotent structure. The holographic information will then determine the nature of the system, including connected information about its inertial mass and charge structure. A related concept is quantum holography, which is like ordinary holography but not degenerate. It has now been unequivocally demonstrated in the case of 'quantum holographic encoding in a two-dimensional electron gas' (C. R. Moon, L. S. Mattos, B. K. Foster, G. Zeltzer and H. C. Manoharan, *Nature Nanotechnology*, **4**, 167-72, 2009), but probably has a much more general application. In the nilpotent formlism, we can accommodate the operation of quantum holography, by defining *E* and *m* as the phase and reference phase.

General relativity and Newtonian theory

Where, then, does this leave general relativity, which assumes that the gravitational force is governed by a Lorentzian metric, with maximum speed at the velocity of light? To answer this we need to recall that general relativity, unlike Newtonian theory, with its 'God's eye view', is an *observer-centred* theory. All observation is local and we still have to accommodate this fact in our mathematical description of gravity as a nonlocal correlation. It may well be that, in resolving this problem, we can find an explanation for the otherwise arbitrary use of the Newtonian potential in general relativity.

Newtonian theory, as used by Newton, is very different from the way it is often portrayed. The key early problem that general relativity had to solve was the anomalous precession of the orbit of the planet Mercury. By the end of the nineteenth century, Leverrier and others had calculated, that, from the perturbing effects of all known bodies in the solar system, the orbit of Mercury should precess at a rate of 531 seconds of arc per century, whereas the measured rate was 574 seconds of arc, leaving 43 seconds of arc unexplained. Now, Newton had first introduced perturbation theory in the *Principia*, but he didn't do it in the way people often think. To solve the orbit of a body perturbed by some external influence, he created a system of coordinates varying in space and time such that, *in that system*, the body would still describe an ellipse, sweeping out equal areas in equal times and obeying the conservation of angular momentum.

As an example he calculated the effect of one or two of the larger planets on Mercury's orbit at about 240 seconds of arc per century. He did this specifically to show that, at the level of accuracy possible to observational astronomers in his own time, there would be no observable effect. It was therefore more useful to him to solve the anomalies that could then be observed, like those of the lunar orbit. One thing that would have made even less sense for him to pursue was the fact that the velocity of light was finite, though he knew perfectly well that this would have an effect of its own on astronomical observation, for this was, after all, the precise way in which the finite speed of light was discovered by Roemer in 1676, and Newton used this astronomical anomaly for an optical test of his own in 1691. In addition, he believed that light would be deflected in a gravitational field (*Opticks*, 1704, Query 1). He made no calculation, though others did later in the century.

The point of this is to show that, to extend Newtonian theory to accommodate astronomical results obtained at the precision available at the end of the nineteenth century requires more than basic Newtonian theory, but Newton was fully aware of the technique that had to be used: the creation of a *varying* coordinate system in which those effects which disturb the otherwise perfectly elliptical orbit of a two-body system can be accommodated by making the orbit perfectly elliptical in the new set of coordinates. In principle, this is still the technique which general relativity uses: find a geometry in which the effects which disturb an otherwise perfect path – in this case a straight line – are accommodated into the structure and then assume that the perfect path is followed in the new geometry. The idea of substituting geometry for physical effect is, in no sense, new with general relativity, only the idea that there is *only* geometry, not real physical effect; but even this is only an idealisation of the process, for all practical calculations from general relativity restore the Newtonian potential. In fact, it is impossible to tell from the mathematical structure alone whether there really is a physical effect or whether there is only geometry. Subsequent to the adoption of Einstein's theory, Cartan showed that standard Newtonian gravitational theory, without the velocity of light, could be represented by a geometrical structure very like Einstein's.

The effect of observation on nonlocal gravity

If we assume that gravity is instantaneous and nonlocal, but physical *observation* is local, involving time-delayed luminal or subluminal interaction, we can no longer use the Lorentzian space-time of a local coordinate system in the description of gravity. So, what happens if we do so, as we must? The answer will be that we will create the equivalent of a noninertial frame for the gravitational system, with the resulting appearance of fictitious inertial forces, and a rotation of the coordinate system. We may recall that Newtonian theory already uses inertial forces as a standard way of dealing with mismatches between theory and observation, and that this is needed every time Newtonian theory is used, because, although Newtonian theory is defined only for inertial frames, *all* experimental observations use frames which are noninertial. So, in effect, we avoid the problem by assuming that the noninertial frames are actually inertial, and then add on the purely fictitious centrifugal and Coriolis forces to accommodate the noninertial effects.

We can, in fact, show that a gravitational system involving instantaneous interaction will require dynamic equations in the space-time of measurement of exactly the form required to predict the relativistic effect of planetary perihelion precession. It is easy to see that the main difference from pure Newtonian theory will be the equivalent of the gravitational deflection of light. In effect, the straight lines of observation, found using local forces limited by the speed of light (and mostly by direct electromagnetic signals), will become rotated in the same way as rays of light. Now, in the absence of the gravitational field, a ray of light would travel in a straight line according to the equation:

$$\left(\frac{dr}{dt}\right)^2 + r^2\left(\frac{d\phi}{dt}\right)^2 = 0 \,.$$

This is also the Newtonian 'straight-line' or 'default' position for a body of unit mass in the absence of any deflecting force or source of dynamic energy, whereas the application of a Newtonian potential $-GM / r$ to deflect a body of unit mass from straight line motion, in a system of total mechanical energy E, requires a dynamic equation of the form:

$$\left(\frac{dr}{dt}\right)^2 + r^2\left(\frac{d\phi}{dt}\right)^2 = \frac{GM}{r} - E \,.$$

Now, in a gravitational field, with Newtonian potential $-GM / r$, the equation for the light ray, according to general relativity, becomes

$$\left(\frac{dr}{dt}\right)^2 + \left(1 - \frac{2GM}{rc^2}\right) r^2\left(\frac{d\phi}{dt}\right)^2 = 0 \,,$$

while the relativistically-corrected dynamical equation from straight line motion, in a system of total mechanical energy E, would require a dynamic equation of the form:

$$\left(\frac{dr}{dt}\right)^2 + \left(1 - \frac{2GM}{rc^2}\right) r^2\left(\frac{d\phi}{dt}\right)^2 = \frac{GM}{r} - E \,.$$

This is the equation which leads to relativistic perihelion precession, and we can see that the only difference it displays from the Newtonian equation is the term on the left-hand side equivalent to the light-bending. The effect is as if the gravitational potential had rotated the local coordinate system based on Lorentzian space-time, by an amount which added the term $-(2GM / rc^2)\ (d\phi / dt)^2$ to the equation of motion, and this constitutes the 'inertial force' term (in the form of 'inertial energy') expressing the fact that the ordinary process of measurement provides a noninertial frame for a gravitational system. It can even be seen in terms of a curvature of space and time, though as an effect, rather than source, of the gravitational field. No such ‘rotation’ would occur, of course, if it was possible to use some ‘absolute’ system of coordinates separate from the process of measurement.

Now, general relativity has never constituted an *explanation* of gravity. As it is constructed it is purely a theory which says that space-time is curved, and that curvature constitutes what we call gravity. However, the mathematics of the general relativistic field equations includes no physical content, only an expression of the curvature. We have to put in the physical content when we solve for special cases, when terms which play similar roles to those in Newtonian theory are identified with them. So, within this context, we are perfectly entitled to find any physical description which corresponds to the equations. In effect, what we find is that general relativity is a theory of the *effects* of gravity, and particularly those relating to observation. It brings us no nearer to understanding what gravity is, but neatly packages and codifies the effects that gravity produces using a geometric system. In our terms, it is about epistemology or measurement, rather than ontology or the thing in itself. It doesn’t distinguish between whether the gravitational influence travels at the speed of light or whether we just measure it that way.

It might be objected that the equation of light deflection, on which this discussion is based, can only be derived using the full field equations of general relativity, and that this proves, for example, that gravity must travel at the speed of light or less. In fact, this has long been known to be incorrect. The light deflection (and even the perihelion precession) can be derived purely from special relativity, as long as we include length contraction as well as time dilation. (The presence of both effects is obvious in the Schwarzschild solution for a spherically-symmetric point source.) Light, of course, is special relativistic, and the gravitational part of the calculation is only an application of gravitational energy, treated classically, to the special relativity of light. There is no assumption that gravity is special relativistic. In any case, the special nature of light even allows a *Newtonian* derivation of the effect. Eighteenth Newtonian century calculations were based on the idea that light, because of its great speed, would follow a hyperbolic orbit, with eccentricity $e >> 1$, in being deflected by a large gravitating body. Twentieth century authors have assumed that this would be equivalent to taking the potential energy equation for a body already in orbit, with velocity c and angle 0^o at distance of closest approach R to the central body:

$$c^2 = \frac{GM}{R}(e+1),$$

from which we calculate the half-deflection angle as $\delta \approx 1 / e \approx GM / Rc^2$, and the full-deflection as $2GM / Rc^2$, which is only half the measured value. The doubling can be considered as due to the addition of the length contraction effect to the time dilation caused by the presence of the potential energy. However, light’s ‘velocity’ c

when it is emitted from its source is nothing to do with any dynamical orbit. Dynamical characteristics are only acquired when it is drawn into an orbit, for which we need the kinetic energy equation

$$\frac{1}{2}c^2 = \frac{GM}{R}(e+1).$$

This gives the correct deflection, and was, in fact, the basis of Soldner's calculation of 1801, although Soldner failed to get the correct deflection by only doing the half-deflection integration. In fact, energy considerations should lead to the same results for special relativity and classical physics, because special relativity is deliberately constructed to preserve the classical conservation of energy.

The aberration of space

We have interpreted general relativity in the only way in which we believe it can be compatible with quantum mechanics, and with our understanding of the nature of mass-energy, as derived both by symmetry and from observation. It is totally in keeping with GR as an expression of a purely abstract mathematical structure, realised as geometry. The most important result is that the theory is linearised because the space-time curvature is no longer the *source* of the gravitational field as well as its consequence. Other people might try alternative interpretations involving a real speed of c, but they have no support in the mathematical structure which constitutes the theory. They also lead to anomalous results, due to the assumed nonlinearities, such as wormholes, infinite gravitational collapse, the violation of conservation laws, unrenormalizable infinities, and the clearly unsustainable theory of 'quantum gravity'. Some of the suggestions might entertain the public interest in science fiction, but they don't involve science which can be imagined as verifiable by observation. Other theories might propose complex and usually model-dependent schemes of 'unification', but they will not be based on foundational thinking, and are unlikely to find experimental confirmation any time soon.

Every effect predicted by GR which has been observed, and every effect which has been predicted to be observable within a reasonable time frame, including 'gravitational waves' (here gravitational-inertial waves), will follow from the GR field equations as interpreted here. They will be gravitational in the sense that they are *caused* by gravity, but they will not be gravitational in the sense that they express the *nature* of gravity. A fuller description might require us to use a term like 'gravitation-inertia', since both concepts are involved.

Without the nonlinearities, the GR field equations will show much greater accuracy to many more orders of magnitude than if they are present. J0348 is a neutron star with a mass of two solar masses, only the second discovered after J1614-2230 in 2010. It was previously thought that the maximum mass for a neutron star was 1.5 solar masses, although the result has now been pushed to 2 or 3. J0348 has a white dwarf companion, with short-period orbital period of 2.46 hours. According to results just announced, the system produces gravitational waves, but not the anticipated extra ripples expected from nonlinearity. Clearly, at the level of sensitivity revealed by the current measurements, nonlinear effects are not seen, and some nonlinear models are already ruled out. Further experiments will be needed in this area, but we have here a first indication that the GR equations are preserved to a much greater accuracy than some interpretations would have previously imagined, but that is certainly predicted by the interpretation proposed here (*Zero to Infinity*, p. 481).

All the relativistic effects, in this interpretation, will be due to the action of gravity on the space-time of measurement, and could be described, by analogy, as the 'aberration of space'. We incorporate this aberration by adding on the required inertial force terms, and this becomes equivalent to using the GR field equations as the combined expression of gravitation and inertia(as Einstein actually required when he invoked the principle of equivalence). The inertial force correction terms are, typically, of magnitude $-3GMv^2 / r^2c^2$ or $-3GM\alpha^2 / r^4\mathrm{c}^2$, and they are identical, in all respects, to the terms representing the gravitational effect on energy transmitted at the velocity of light.

So, if, to a light-ray geodesic defined in the absence of a gravitational field, we introduce inertial forces exactly sufficient to cover the effects produced by a gravitational potential $-GM / r$, we will, *by definition*, obtain the standard new geodesic equation expected under these conditions, with the extra terms now interpreted as being equivalent to the rotation of the space-time coordinates. Our calculations will produce exactly what we would expect to obtain from the general relativistic field equations. Under the most general conditions, we would recover the full theory of general relativity, though with the emphasis shifted from ontology (the nature of the gravitational field) to epistemology (the nature of measurement in a gravitational field). Quantum aspects, however, would now be available on the basis that they would be concerned with local inertial repulsion and spin 1 transmission, and so generate a renormalizable theory.

Gravitomagnetic effects

Many of the effects due to the presence of *c* in GR equations can be dealt with by assuming the existence of a 'gravitomagnetic' field analogous to the magnetic field of classical electromagnetic theory. This is, in fact, a relatively standard procedure in GR, although there are some extra consequences here of the fact that *c* is itself gravitationally affected. In fact, many of the results (like most of the standard consequences of GR) can be derived from special relativity, with additional 'GR' modifications arising from the use of both time dilation and length contraction. The assumption of linearity is assumed in standard calculations because the fields are relatively weak, but here we assume that it is fundamentally true in any case.

Kolbenstvedt has, in fact, derived the equations for the gravitomagnetic field from special relativity, using a kinematical argument, but has allowed only the effect of time dilation as special relativistic because it requires only the principle of equivalence and the Doppler effect. He considers that the contraction of measuring rods requires general relativity or the 'curvature of space'. Nevertheless, as we have already seen from the case of gravitational light deflection, the length-contraction, or some other doubling effect, is a necessary component of a full special relativistic or even classical treatment, and this is the key case because it is really an expression of the effect of the gravitational field on space-time geometry.

We will allow for this doubling effect, but otherwise follow Kolbentsvedt's procedure, and consider an object of mass M, moving with velocity $\mathbf{u}$ in the positive x-direction in the frame of the laboratory, and a particle of mass m moving with velocity $\mathbf{v}$ under its gravitational influence. In the rest frame of mass M, the Lagrangian $\mathcal{L}_0$ of the mass m particle can be found from the variational principle:

$$\delta\int(-mds) = \delta\int \mathcal{L}_0 dt_0 = 0 \, ,$$

where

$$ds_0^{\,2} = \gamma^{-2}dt_0^{\,2} - \gamma^2 dr_0^{\,2}$$

is the line element in the rest frame, and

$$\gamma^2 = 1 - \frac{2GM}{rc^2} = 1 - \frac{2\phi}{c^2} \, .$$

Integrating for the rest frame,

$$\mathcal{L}_0 dt_0 = -\, m\,(\gamma^{-2}c^2 dt_0^{\,2} - \gamma^2 dr_0^{\,2})^{1/2}.$$

Then, transforming to the laboratory frame, and neglecting higher order terms, we obtain

$$\mathcal{L}dt = -\, m\,[(c^2 + 2\phi)\,(dt - udx)^2 - (1 - 2\phi)\,(dx - udt)^2 - dy^2 - dz^2]^{1/2}.$$

Once again neglecting the higher order terms, and dividing by dt, we obtain the Lagrangian

$$\mathcal{L} = -\left(c^2 - v^2 + 2\phi - 8\phi\frac{\mathbf{u.v}}{c^2}\right)^{1/2} .$$

When $u \ll v \ll c$, and the rest energy term mc^2 and higher order corrections are neglected, the series expansion approximates to

$$\mathcal{L} = ½mv^2 - m\phi + 4m\phi \frac{\mathbf{u.v}}{c^2}.$$

This expression may be compared with the standard Lagrangian for a particle of mass m and charge q moving with speed $v \ll c$ in an electromagnetic field determined by scalar potential ϕ and vector potential **A**:

$$\mathcal{L} = ½mv^2 - q\phi + q\frac{\mathbf{A.v}}{c}.$$

The equations are clearly analogous, with $4\phi\mathbf{u} / c$ being the gravitational equivalent of the electromagnetic vector potential **A**.

Maxwell's equations for gravitomagnetism

We can now extend the argument. The rotational analogue of the magnetic field term $\mathbf{B} = \nabla \times \mathbf{A}$ becomes the 'gravitomagnetic' field

$$\boldsymbol{\omega}c = \nabla \times \left(4\phi\frac{\mathbf{u}}{c}\right) = \left(4\phi\frac{\mathbf{u}}{c}\right) \times \left(\nabla\phi\right) = 4\frac{\mathbf{u}}{c} \times \mathbf{g}$$

We can then set out a series of equations of the Maxwell type:

$$\nabla.\mathbf{g} = 4\pi G\rho$$
$$\nabla.\boldsymbol{\omega}c = 0$$
$$\nabla \times \mathbf{g} = -c\frac{\partial\boldsymbol{\omega}}{\partial t}$$
$$\nabla \times \boldsymbol{\omega}c = 4\pi G\rho\mathbf{v} + c\frac{\partial\mathbf{g}}{\partial t}.$$

Here, $G\rho\mathbf{v}$ takes the place of **j**; and $4\pi G\rho$ that of ρ / ε_0. In empty space **g** and $\boldsymbol{\omega}c$ will satisfy the equations for $\nabla \times \nabla \times \mathbf{g}$ and $\nabla \times \nabla \times \boldsymbol{\omega}c$:

$$\Box\mathbf{g} = 0$$

and

$$\Box\boldsymbol{\omega}c = 0,$$

with the analogous mass-density terms being added where sources are present. Other standard results follow immediately, as they do in electromagnetic theory: the Poynting vector (energy flux in an element of solid angle), energy density of fields, Lorentz force, Larmor precession frequency, equation of motion of a particle, quadrupole radiation, etc. Kolbentsvedt points out that such effects as 'geodesic deviation of spinning particles, precession of gyroscopes orbiting the Earth, and 'dragging' of inertial frames by rotating masses, by leaning on well-known effects from classical electromagnetism and atomic physics involving spin-orbit and spin-spin coupling' would follow automatically from a gravitomagnetic theory.

The gravitomagnetic equations here are concerned with inertial effects and so repulsive forces, rather than attractive, which they would be if they were directly gravitational; this makes the rotational term $\boldsymbol{\omega} c$ positive, unlike the gravitational field, and so, for comparison with Maxwell's equations, in which **E** and **B** are both positive, we take **g**, a 'static' component of inertial repulsion, in place of the gravitational field –**g** in the equations from which the wave solutions are obtained. The form of the equation $\nabla.\mathbf{g} = 4\pi G\rho$ is, of course, insensitive to the velocity at which the interaction is transmitted, so a real gravitational attraction *of* the mass density ρ transmitted at infinite velocity would be formally indistinguishable from a fictitious static inertial repulsion *by* the mass density ρ transmitted at the velocity of light. We will show later that the static component of inertial repulsion, which is assumed by the principle of equivalence to be numerically equal to the attractive force of gravity, has a very important cosmological significance.

The factor 4 in the vector potential is the main other difference from the equivalent term in electromagnetic theory, and, as we have seen, it comes from a combination of the effects of length contraction and time dilation, which means that the pure term c is no longer the defining one as it is in electromagnetic theory. It emerges also in the theory of 'gravitational waves', as produced by other authors from general relativity, but it is not an indication of the need for a spin 2 boson in quantum gravity, as has been claimed, for its origin in special relativity is clear, and it is possible to assign its origin to a gravitational effect on space-time which has no reciprocal effect on the gravitational force itself.

Mach's principle

One of the things that general relativity never succeeded in resolving was the relation between gravitation and inertia. In particular, it was never able to accommodate Mach's principle, or the idea that the inertial mass of any object was due to its interactions with the rest of the matter in the universe. An argument that sought to overcome this problem, on the basis of gravitomagnetism, was put forward by Sciama in 1953, though he later abandoned it. However, if we extend this argument on the basis of our new understanding of how gravitomagnetism and inertia are related, we find that it leads to a remarkable prediction which now appears to have a close relation to some significant experimental results discovered subsequently.

According to our understanding, relativistic and 'gravitomagnetic' effects emerge as properties of the local coordinate system and not as intrinsic properties of the gravitating source. If we are able to construct a set of Maxwell equations for the gravitomagnetic field that are exactly analogous to those used in electromagnetic theory, then we would expect gravitomagnetic analogues to all the effects that are known from electromagnetic theory. One such analogue would be an accelerationdependent inductive force with the same structure as that appearing in electromagnetic theory. That is,

$$F = \frac{G}{c^2 r} m_1 m_2 \sin\theta \frac{dv}{dt}$$

would be the analogue of

$$F = \frac{q_1 q_2 \sin\theta}{4\pi\varepsilon_0 c^2 r} \frac{dv}{dt}.$$

This was exactly the force that Sciama proposed for explaining inertia. According to Sciama, the inertia of a body of mass $m = m_1$ might be derived from the action of the total mass $m_u = m_2$ within a Lorentzian event horizon of radius r_u, thus making the force F equal to Kma, with K a constant. Sciama required

$$F = \frac{Gm_1 m_2}{c^2 r} \frac{dv}{dt} \propto m_1 \frac{dv}{dt} = Kma.$$

So, if $m_2 = m_u$, the Hubble mass, is responsible for the inertia of a body of mass $m = m_1$, then $F = Kma$ if $Gm_u \propto c^2 r$, and $K = ½$, if the zero total energy means that $Gm_u = ½c^2 r$.

However, in our interpretation, the continuous and zero-gradient mass-field or vacuum which provides gravitational nonlocality (say, the Higgs field) must produce a local consequence, analogous to unit charge, and define the standard for a unit inertial mass for the entire universe, in exactly the same way as the almost constant gravitational field **g** allows us to define a unit mass at the Earth's surface. As a local effect, inertia, unlike gravity, is time-delayed, and limited by the velocity of light, and therefore also by the Hubble radius which defines its event horizon. So, the Hubble mass m_u will define a radial inertial field of constant magnitude from the centre of any given local coordinate system to the event horizon defined by r_u. At the same time, the gravitational field (Gm_u / r_u^2), independently of the local coordinate system, will define a unit of gravitational mass within the same radius. Supposing that isotropy removes the angular dependence, we apply the principle of equivalence and obtain:

$$\frac{Gm_u}{c^2 r} \frac{dv}{dt} = \frac{Gm_u}{r_u^2}.$$

We are saying that the object defined by mass m requires an inertial force, with an acceleration

$$a = \frac{dv}{dt} = v\frac{dv}{dr} = \frac{c^2 r}{r_u^2}$$

and, by integration with respect to r, a velocity related to the Hubble constant, H_0:

$$v = \frac{cr}{r_u} = H_0 r \; .$$

This is Hubble's law for cosmological redshift, which we have derived without assuming a cosmology to explain it, but we have also found something new, an acceleration term, which, in terms of Hubble's constant, can be written:

$$a = \frac{v^2}{r_u} = H_0^2 r \; .$$

If gravity is nonlocal, this acceleration will be 'fictitious', in the sense that it arises from using a local Lorentzian coordinate system to model an instantaneous interaction. It is a manifestation of a fundamental physical process, analogous to a centrifugal force, which does not depend on a prior assumption of cosmology of any kind, though it may *determine* it. Ultimately, fundamental physical effects require fundamental physical explanations, and one purely based on a historical cosmology – of things that just happened to happen at particular times – is not an adequate explanation in these terms. Cosmology, wherever possible, should follow physics, not the other way round.

If we now combine this repulsive force with the attractive gravitational force due to total mass *m* at any distance, we can calculate an equivalent vacuum density, ρ, from

$$F = \frac{Gm}{r^2} - H_0^2 r = \left(\frac{4}{3} \pi G \rho - H_0^2 \right) r \; .$$

We can re-express this as a Poisson-Laplace equation in which

$$\nabla^2 \phi = 4\pi G \left(\rho - \frac{3H_0^2}{4\pi G} \right) = 4\pi G \left(\rho + \frac{3P}{c^2} \right) = 4\pi G (\rho - 3\rho_{vac}).$$

We can identify here a vacuum density

$$\rho_{vac} = \frac{H_0^2}{4\pi G} \, ,$$

which is equivalent to a 'dark' energy density or negative pressure

$$-P = \frac{H_0^2 c^2}{4\pi G} \, ,$$

which can also be expressed through the cosmological constant

$$\lambda = 8\pi G \rho_{vac} = 2H_0^2 \; .$$

If we define the critical density for a ‘flat’ universe as

$$\rho_{crit} = \frac{3H_0^2}{8\pi G}$$

we obtain

$$\frac{\rho_{vac}}{\rho_{crit}} = \frac{2}{3}.$$

What this means is that we have derived a vacuum density which is 67 % of the density of the universe, assuming that this is at the critical value, and, of course, assuming uniform density and isotropy. It manifests itself through an outward inertial acceleration to the velocity generated in Hubble’s law. The force law obeys Hooke’s law, exactly as would be expected for a term equivalent to the cosmological constant.

This would be interesting just as explanation, but in fact it was originally a *prediction*. Many years after this prediction was first made, and published on at least four separate occasions (the last and most accessible of which was in a book called *A Revolution Too Far*, 1994), an effect of exactly this kind was discovered. The latest data from the Planck satellite say that it is in the region of 68 % of the energy of a universe at critical density. It is described as the *dark energy*, and, according to all accounts, the effect was totally unexpected when it was first found and is still unexplained.

We have now seen that, if we equate the inertial reaction numerically with the undetectable gravitational attraction (so defining an equivalence principle), we justify a form of Mach’s principle, and obtain gravitomagnetic effects, redshift, and acceleration of the redshift. In the simplest case of ‘curvature’, provided by a point source, we generate the Schwarzschild metric and a factor 4 in the gravitomagnetic equations by comparison with those for QED, incorporating a factor 2 for space ‘contraction’ and another 2 for time ‘dilation’.

Can we quantize gravitational inertia?

The theories concerning the gravitomagnetic field and dark energy are interesting as showing a new aspect of fundamental reasoning – that it can cast theories created for quite different purposes into new forms, incorporating fundamental ideas not present in the original theories, and seemingly emerge with quite new results. The basic ideas of gravitomagnetism, the aberration of space, and the inertial origin of redshift, with a concept of acceleration, were known to me from fundamental reasoning before I knew of the theories of Kolbentsvedt and Sciama, but these theories created a much more powerful basis on which they could be constructed, leading to a much more exact and predictive treatment of the phenomena. Another example was a theory of discrete gravity based on 'extended causality' with a specific use of the proper time, which had the exact structure required to be developed using the nilpotent method of quantization. This is rather more technical than some of the other arguments in this lecture, and much of this development is due to the original authors, but, because it is such an important subject, it might just be mentioned as a possible means to bring to fruition that the realization of the local aspect of gravity in an inertial theory might be capable of quantization, as predicted.

According to our earlier arguments, the local theory which we have described as inertia should be susceptible to quantization. The interpretation of inertia as the result of the interaction between discrete matter and the continuous gravitational vacuum suggests that it is gravitational inertia rather than gravity which is subject to quantization. Quantizing a pure 4-dimensional structure has proved to be a problem, because time is not an observable in quantum mechanics as it is in classical relativity theory, nor is quantum space-time truly 4-dimensional, as the two parameters are added together only with the mediation of another structure, based on the γ matrices, or $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$. If we assume that 3-dimensionality is the sole source for discreteness in physics, the mathematical object called a 4-vector will have no physical realisation at the quantum level. This means that, for a fully quantized theory, we need a metric other than the 4×4 representation using x, y, z, t, with added curvature, which is used in classical general relativity.

The obvious one that suggests itself is a 3×3 representation, with diagonal terms $i\boldsymbol{k}t$, $\boldsymbol{i}\mathbf{r}$, $\boldsymbol{j}\tau$, in the absence of the curvature resulting from a gravitational field. This formalism would have the distinct advantage of being a natural 2 + 1 theory of gravitational effects (the 2 representing the 'real' terms $\mathbf{r}$ and τ, and the 1 the imaginary term it, in interesting analogy to the 2 + 1 structure of branes in M-theory), and such theories are already known to be renormalizable, unlike those with a higher number of dimensions. Instead of using a 4-dimensional structure, we use the *3-dimensional* nilpotent structure, represented by the terms $\pm i\boldsymbol{k}t \pm \boldsymbol{i}\mathbf{r} + \boldsymbol{j}\tau$ and $\pm i\boldsymbol{k}E \pm \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m$. The first term may be described as the 'quantum metric'; the second is the realisation of the Dirac state, and may be regarded as the phase space version of the first. No other fundamental structure is both fully quantum and fully relativistic, and the 3-dimensionality of the structure is essential to both of these conditions.

Significantly, the discrete theory also dispenses with the transverse directions, to create a 1 + 1 space-time, paralleling the fact that a quantum Dirac particle, with conserved charge (the kind of object to which quantum gravity or quantum gravitational inertia will apply), requires only $\mathbf{r}$, and a single well-defined direction of spin, rather than the classical x, y, z. The 3-dimensionality comes from the fact that, for gravitational inertial interactions involving individual fermionic states at the quantum level, there is an effective reduction or compactification of the spatial dimensions, to a single well-defined parameter ($\mathbf{r}$). In this case, we can construct a 2 +

1 theory of gravitational inertia, based on a 3 × 3 'quantum metric' (with **r** and τ representing the 'real' parts and *it* the imaginary), which would be both quantizable and renormalizable. It would provide the full Dirac 'atom' solution and *U*(1) QED-type behaviour, with a corresponding photon-like mediating boson, merely on the assumption of spherical symmetry and the multivariate vector nature of the spin term **p** (or, equivalently, conserved charge).

The standard mathematical representation of the gravitational force incorporates no information relating to speed, but the description of gravity as an undetectable property of the vacuum would *require* it to be instantaneous. The *c*-dependence of the inertial reaction, however, determines that, though linear and renormalizable, this force will itself be affected by gravity, giving rise to the 'curvature' terms in the metric tensor, as in general relativity. It is, however, 'curvature' of the metric for inertia, not for gravity, which has no metric.

Extended causality and quantized inertia

A discrete theory of 'extended causality' has been proposed by de Souza and Silveira, based on the fact that a single object (particle or field) at two points in Minkowski space-time (represented by the 4-vector *x*) must satisfy the causality constraint

$$\Delta\tau^2 + \Delta x^2 = 0$$

which defines a hypercone for the object. As soon as we convert this to the nilpotent form

$$\Delta(i\mathbf{k}t)^2 + \Delta(\mathbf{i}\mathbf{r})^2 + \Delta(\mathbf{j}\tau)^2 = 0$$

we are on the way to quantization, and through the multivariate nature of **r** it will automatically incorporate spin. Extended causality then applies when we shift τ and x by infinitesimal steps $d\tau$ and dx. We then obtain

$$(\Delta\tau + d\tau)^2 + (\Delta x + dx)^2 = (i\mathbf{k}\Delta t + i\mathbf{k}dt)^2 + (\mathbf{i}\Delta\mathbf{r} + \mathbf{i}d\mathbf{r})^2 + (\mathbf{j}\Delta\tau + \mathbf{j}d\tau)^2 = 0\ .$$

Defining f as a fibre in the 'spacetime' $x \equiv (\mathbf{i}\mathbf{r},\ i\mathbf{k}t)$, where $f_\mu = dx_\mu / d\tau = \mathbf{i}d\mathbf{r} / \mathbf{j}d\tau + i\mathbf{k}dt / \mathbf{j}d\tau$ and $f^\mu = -\,d\tau / dx^\mu = -\,\mathbf{k}d\tau / d\mathbf{r} + i\mathbf{i}d\tau / dt$, we may combine these two equations to obtain:

$$\Delta\tau + f\,.\,\Delta x = \mathbf{j}\Delta\tau + (\mathbf{i}d\mathbf{r} / \mathbf{j}d\tau + i\mathbf{k}dt / \mathbf{j}d\tau)(\mathbf{i}\Delta\mathbf{r} + i\mathbf{k}\Delta t) = 0,$$

while extended causality now requires

$$(\tau - \tau_0) + f_\mu (x - x_0)^\mu = \mathbf{j}(\tau - \tau_0) + (\mathbf{i}d\mathbf{r} / \mathbf{j}d\tau + i\mathbf{k}dt / \mathbf{j}d\tau)(\mathbf{i}\mathbf{r} + i\mathbf{k}t - \mathbf{i}\mathbf{r}_0 - i\mathbf{k}t_0) = 0$$

Suppose now that we introduce a massless scalar field $\phi_f(x, \tau) \equiv \phi_f(\mathbf{ir}, i\mathbf{k}t, \mathbf{j}\tau)$, the extended causality will constrain it to the hypercone generator. The previous equation will also induce a direction to the field derivatives, so that

$$\partial_\mu \phi_f = (\partial_\mu - f_\mu \partial_\tau) \equiv \nabla_\mu \phi_f.$$

With this expression we can now derive a discrete field equation. If χ is the coupling constant and $\rho(x, \tau) \equiv \rho(\mathbf{ir}, i\mathbf{k}t, \mathbf{j}\tau)$ the discrete field source, then, using standard methods, the action of the field is given by

$$S_f = \int i\, d\mathbf{r}\, dt\, d\tau \{ \tfrac{1}{2}\, \eta^{\mu\nu} \nabla_\mu \phi_f \nabla_\nu \phi_f - \chi\, \phi_f\, \rho(\mathbf{ir}, i\mathbf{k}t, \mathbf{j}\tau) \}.$$

The field equation then becomes

$$\eta^{\mu\nu} \nabla_\mu \nabla_\nu \phi_f(\mathbf{ir}, i\mathbf{k}t, \mathbf{j}\tau) = \rho(\mathbf{ir}, i\mathbf{k}t, \mathbf{j}\tau),$$

with energy tensor

$$T_f^{\mu\nu} = \nabla^\mu \phi_f \nabla^\nu \phi_f - \tfrac{1}{2}\, \eta^{\mu\nu} \nabla^\alpha \phi_f\, \nabla_\nu \phi_f.$$

The solution can be conveniently expressed in terms of a Green's function. Here we write:

$$\phi_f(\mathbf{ir}, i\mathbf{k}t, \mathbf{j}\tau) = \int i d\mathbf{r}\, dt\, d\tau\, (\mathbf{ir}' + i\mathbf{k}t')\, G_f(\mathbf{ir} + i\mathbf{k}t + \mathbf{j}\tau - \mathbf{ir}' - i\mathbf{k}t' - \mathbf{j}\tau') \times \rho(\mathbf{ir}', i\mathbf{k}t', \mathbf{j}\tau')$$

and

$$\eta^{\mu\nu} \nabla^\mu \phi_f \nabla^\nu \phi_f\, G_f(\mathbf{ir} + i\mathbf{k}t + \mathbf{j}\tau) = i\delta\mathbf{r}\,\delta t\,\delta\tau.$$

Using the Heaviside step function, Θ, with $b = \pm 1$, the Green's function is then

$$G_f(x, \tau) = \tfrac{1}{2}\, \Theta(bf^4 t)\, \Theta(b\tau)\, \delta(\tau + f \cdot x)$$

or

$$G_f(\mathbf{ir} + i\mathbf{k}t + \mathbf{j}\tau)$$
$$= \tfrac{1}{2}\, \Theta(b(i d\mathbf{r} / \mathbf{j} d\tau + i\mathbf{k} dt / \mathbf{j} d\tau) i\mathbf{k}t)\, \Theta(b\, \mathbf{j}\tau)\, \delta(\mathbf{j}\tau + (i d\mathbf{r} / \mathbf{j} d\tau + i\mathbf{k} dt / \mathbf{j} d\tau)(\mathbf{ir} + i\mathbf{k}t)).$$

Even in classical, discrete form, this equation is independent of the transverse components of the field, paralleling the quantum reduction to a single well-defined direction of spin.

If we now take a single scalar charge $q(\tau)$, with world line $z(\tau)$, as a field source, then:

$$\rho(x, t_x = t_z) = q(\tau_z)\, \delta^{(3)}(x - z(\tau_z))\, \delta(\tau_x - \tau_z),$$

and the solution for the emitted field becomes:

$$\phi_f(x, t_x) = \chi \int d\tau_y\, \Theta(t_x - t_y)\, \Theta(\tau_x - \tau_y)\, \delta[t_x - t_y + f \cdot (x - y)]\, q(\tau_z),$$

which reduces to

$$\phi_f(\mathbf{ir}, i\mathbf{k}t, \mathbf{j}\tau) = \chi\, q(\tau)\, \Theta(t)\, \Theta(\tau) \big|_f,$$

or $$\phi_f\,(\,\mathbf{ir},\, i\mathbf{k}t,\, \mathbf{j}\tau) = \chi\, q(\tau)\,\big|_f,$$

when $\tau \geq 0$ and $t > 0$. We have found that, applying a massless scalar gives us a discrete field equation, and a field source represented by a scalar charge to generate a 'graviton'-like object and a metric for a discrete field related to gravity, though now the field is a repulsive inertial one, fully quantized via the Dirac nilpotent, and the 'graviton'-like object is correspondingly identified as a spin 1 boson or photon-like pseudo-boson. The field can only be quantized in this way because it is inertial, with positive energy, not gravitational, with negative energy, as negative energy, in our understanding, represents nonlocal vacuum rather than the local quantized state.

In the discrete model, the emission or absorption of a field causes a discrete change in q, and we can apply the standard techniques appropriate to quantum field theory, and, in particular, QED, to develop the formalism for interactions at higher orders, the 2 + 1 nature of the theory ensuring its renormalizability. In effect, though with more difficulty, we have reversed the argument used for the other forces, working from vacuum to the quantized local state.

Peter Rowlands
Physics Department, University of Liverpool
26 April 2013